

# The Carbohydrate-Binding Site in Galectin-3 Is Preorganized To Recognize a Sugarlike Framework of Oxygens: Ultra-High-Resolution Structures and Water Dynamics

Kadhirvel Saraboji,<sup>†</sup> Maria Håkansson,<sup>‡</sup> Samuel Genheden,<sup>§</sup> Carl Diehl,<sup>||</sup> Johan Qvist,<sup>||</sup> Ulrich Weininger,<sup>||</sup> Ulf J. Nilsson,<sup>⊥</sup> Hakon Leffler,<sup>@</sup> Ulf Ryde,<sup>§</sup> Mikael Akke,<sup>\*,||</sup> and Derek T. Logan<sup>\*,†,‡</sup>

<sup>†</sup>Department of Biochemistry and Structural Biology, Center for Molecular Protein Science, Box 124, Lund University, SE-221 00 Lund, Sweden

<sup>‡</sup>SARomics Biostructures AB, Scheelevägen 22, SE-220 07 Lund, Sweden

<sup>§</sup>Department of Theoretical Chemistry, Lund University, SE-221 00 Lund, Sweden

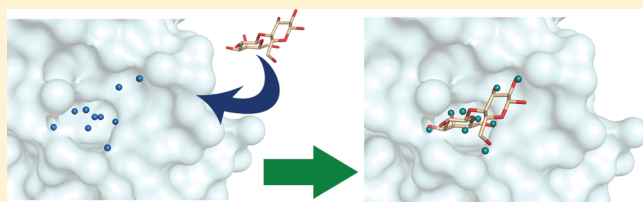
<sup>||</sup>Department of Biophysical Chemistry, Center for Molecular Protein Science, Box 124, Lund University, SE-221 00 Lund, Sweden

<sup>⊥</sup>Department of Organic Chemistry, Lund University, SE-221 00 Lund, Sweden

<sup>@</sup>Section MIG, Department of Laboratory Medicine, Sölvegatan 23, Lund University, SE-223 62 Lund, Sweden

## Supporting Information

**ABSTRACT:** The recognition of carbohydrates by proteins is a fundamental aspect of communication within and between living cells. Understanding the molecular basis of carbohydrate–protein interactions is a prerequisite for the rational design of synthetic ligands. Here we report the high- to ultra-high-resolution crystal structures of the carbohydrate recognition domain of galectin-3 (Gal3C) in the ligand-free state (1.08 Å at 100 K, 1.25 Å at 298 K) and in complex with lactose (0.86 Å) or glycerol (0.9 Å). These structures reveal striking similarities in the positions of water and carbohydrate oxygen atoms in all three states, indicating that the binding site of Gal3C is preorganized to coordinate oxygen atoms in an arrangement that is nearly optimal for the recognition of  $\beta$ -galactosides. Deuterium nuclear magnetic resonance (NMR) relaxation dispersion experiments and molecular dynamics simulations demonstrate that all water molecules in the lactose-binding site exchange with bulk water on a time scale of nanoseconds or shorter. Nevertheless, molecular dynamics simulations identify transient water binding at sites that agree well with those observed by crystallography, indicating that the energy landscape of the binding site is maintained in solution. All heavy atoms of glycerol are positioned like the corresponding atoms of lactose in the Gal3C complexes. However, binding of glycerol to Gal3C is insignificant in solution at room temperature, as monitored by NMR spectroscopy or isothermal titration calorimetry under conditions where lactose binding is readily detected. These observations make a case for protein cryo-crystallography as a valuable screening method in fragment-based drug discovery and further suggest that identification of water sites might inform inhibitor design.



Interactions between carbohydrates and proteins mediate numerous important biological functions, such as signal transduction, cell adhesion, host–pathogen recognition, and the immune response.<sup>1–4</sup> Carbohydrate-recognizing proteins are involved in a number of human disease states, including inflammation and cancer. These key functional properties have stimulated significant efforts in drug design targeting carbohydrate-binding proteins.<sup>5–8</sup> Carbohydrate–protein interactions are typically relatively weak, with dissociation constants of  $\sim 1$  mM. Binding affinity is typically driven by a favorable enthalpic component<sup>9–12</sup> that is partly offset by negative entropy. The low affinity is related to the inherent properties of saccharides, such as their lack of charges and lack of extended hydrophobic surfaces, which both reduce the likelihood of forming strong interactions with proteins. Instead, the formation of carbohydrate–protein complexes involves rela-

tively weak van der Waals interactions and hydrogen bonds to the carbohydrate hydroxyl groups, acetamides, and ring and glycosidic oxygens. As a result of these specific properties of carbohydrate–protein interactions, the design of high-affinity inhibitors has proven to be challenging,<sup>6–8</sup> and the relative importance of the various types of interactions in driving carbohydrate recognition is a matter of intense research.<sup>13–18</sup> During the ligand binding process, significant solvent reorganization takes place across the contact surface.<sup>19–22</sup> Thus, considerable attention has been paid to the structure and dynamics of water molecules in carbohydrate-binding sites and to the role these play in mediating carbohydrate recognition by

**Received:** September 19, 2011

**Revised:** November 19, 2011

**Published:** November 23, 2011

proteins.<sup>20,23–26</sup> Nonetheless, our current understanding of the thermodynamics and kinetics of the solvent reorganization process at the microscopic level remains incomplete.<sup>27</sup> A full understanding of carbohydrate recognition benefits strongly from atomic-resolution descriptions of both the liganded and unliganded states.

Galectins are small soluble proteins that constitute a family of mammalian lectins, defined by a carbohydrate recognition domain (CRD) with a conserved sequence motif that confers affinity for  $\beta$ -galactoside-containing glycans.<sup>28</sup> Galectins have several important functions in both carbohydrate-dependent extracellular and carbohydrate-independent intracellular activities.<sup>28–33</sup> Even though galectins are synthesized in and primarily remain in the cytosol, they reach the extracellular space or lumen of vesicles<sup>9,34</sup> by a nonclassical secretory pathway<sup>35,36</sup> and can then take part in regulation of cellular trafficking of glycoproteins, signaling, and cell adhesion.<sup>31,37</sup> A growing body of evidence links galectins to important roles in cell growth, cell differentiation, cell cycle regulation, and apoptosis, making them potential pharmaceutical targets in inflammation, immunity, and cancer.<sup>29,38–41</sup> Thus, it is critical to understand the molecular driving forces for the ligand binding specificity of galectins.

Structures of several galectins in complex with natural and designed ligands are known, and it has been seen that galectins bind oligosaccharides in a conserved recognition mode involving a network of hydrogen bonds and several bound waters, which form bridges between hydrogen-bonding partners in the protein–ligand complex.<sup>16,25,42,43</sup> As is typical for lectins, galectins bind the monosaccharide galactose with dissociation constants in the millimolar range, in their conserved binding site. However, sugars adjacent to the galactose may interact with neighboring sites to provide stepwise boosts in affinity.<sup>28,44</sup> Thus, for galectin-3 addition of glucose, as in lactose, enhances affinity 50-fold ( $K_d \sim 0.2$  mM), and additional saccharides at position 3 of the galactose can give affinities in the low micromolar range. Moreover, non-natural derivatives at this position may enhance the affinity to the nanomolar range, e.g., by exploiting cation– $\pi$  interactions with the surface residue Arg144.<sup>16,45</sup> Furthermore, recent work shows that triazole derivatization at C3 of galactose results in a high affinity, similar to that observed for aromatic amido compounds.<sup>46,47</sup> Several studies have highlighted the significance of high-affinity and selective galectin inhibitors that act intracellularly, with potential use in modulating inflammatory processes and cancer growth.<sup>30,48–51</sup> To develop effective approaches for the structure-based design of potent galectin-3 inhibitors, it is important to understand the detailed molecular basis for carbohydrate recognition, based on the three-dimensional structure and physicochemical properties of the conserved binding motif. High-resolution structural information greatly aids in this respect. Here we report a study that combines X-ray crystallography, NMR spectroscopy, molecular dynamics (MD) simulations, and isothermal titration calorimetry (ITC) to probe the role of water molecules in the binding site and the details of the hydrogen-bonding networks at the protein–ligand interface of Gal3C. Our results highlight the role of an oxygen coordination framework in the extended carbohydrate-binding site of galectin-3 that potentially can be further exploited in future drug design efforts.

## MATERIALS AND METHODS

**Protein Expression and Purification.** The galectin-3 carbohydrate recognition domain (Gal3C; amino acid residues 113–250) was expressed and purified either as a thioredoxin fusion construct<sup>52</sup> or as isolated Gal3C.<sup>43</sup> In the first case, DNA encoding amino acids 113–250 was amplified via polymerase chain reaction from galectin-3 in pET3C<sup>53</sup> and cloned into the pET-32 Ek/LIC vector (Novagen, Madison, WI) according to the manufacturer's instructions, and as described for galectin-8 previously.<sup>44</sup> In the second case, DNA encoding amino acids 113–250 was inserted into pET 3C without additional tags. The expression protocol for the two constructs is identical and has been reported previously.<sup>52</sup> The purification protocol for the isolated Gal3C domain is highly similar to that reported previously,<sup>52,53</sup> except that the final separation step after cleavage of the expressed product using a lactosyl-Sepharose column is not needed. Typical yields were 150 mg/L of culture of isolated Gal3C. The absence of lactose from the apo-Gal3C samples (used to determine the crystal structures of apo-Gal3C and glycerol-bound Gal3C) was verified using NMR spectroscopy.

**Crystallization and Structure Determination.** All Gal3C crystals were obtained using the hanging drop vapor diffusion method in NeXtal plates (Qiagen). The Gal3C–lactose complex was obtained by mixing 9.5  $\mu$ L of protein solution [20 mg/mL Gal3C, 10 mM  $\beta$ -mercaptoethanol, 100 mM lactose, 10 mM sodium phosphate buffer (pH 7.5), 100 mM NaCl, and 0.02%  $\text{NaN}_3$ ] with 0.5  $\mu$ L of 100 mM 3'-benzamido-*N*-acetyllactosamine on ice. After being incubated for 1 h, the solution was centrifuged at 7000 rpm for 10 min at 4 °C, and hanging drops were set up using 4  $\mu$ L drops containing equal volumes of a protein solution and a reservoir solution [30% (w/v) PEG 4000, 0.1 M Tris-HCl (pH 7.5), 0.1 M  $\text{MgCl}_2$ , 0.4 M NaSCN, and 8 mM  $\beta$ -mercaptoethanol]. Immediately after setup, 0.3  $\mu$ L of a seed bead solution (Hampton Research protocol) was added to the drop. The seed bead solution was made from lactose-containing Gal3C crystals crushed in stabilizing solution [1 mM lactose, 29% PEG 4000, 0.1 M Tris-HCl (pH 7.5), 0.1 M  $\text{MgCl}_2$ , 0.4 M NaSCN, and 8 mM  $\beta$ -mercaptoethanol]. The largest crystals grew to dimensions of 0.4 mm  $\times$  0.4 mm  $\times$  0.5 mm within 1 month. This was an attempt to replace lactose with 3'-benzamido-*N*-acetyllactosamine, which failed because even if the affinity of lactose is lower ( $K_d = 231 \mu\text{M}$ ) than that of 3'-benzamido-*N*-acetyllactosamine ( $K_d = 18.2 \mu\text{M}$ ) for Gal3C,<sup>45</sup> the lactose concentration was 19-fold higher (95 mM vs 5 mM). The complex of 3'-benzamido-*N*-acetyllactosamine and Gal3C has been determined using the same crystallization conditions but without lactose present.<sup>45</sup>

Apo-Gal3C crystals were obtained from 4  $\mu$ L drops containing equal volumes of a protein solution [19 mg/mL Gal3C, 10 mM sodium phosphate buffer (pH 7.5), 100 mM NaCl, 10 mM  $\beta$ -mercaptoethanol, and 0.02%  $\text{NaN}_3$ ] and a reservoir solution [30% (w/v) PEG 4000, 0.1 M Tris-HCl (pH 7.5 or 8.0), 0.1 M  $\text{MgCl}_2$ , 0.4 M NaSCN, and 8 mM  $\beta$ -mercaptoethanol]. Apo crystals appeared overnight and grew within a few days to dimensions of 0.1 mm  $\times$  0.1 mm  $\times$  0.2 mm.

The Gal3C–lactose crystal was flash-cooled to 100 K in the cold  $\text{N}_2$  gas stream of a Cryojet (Oxford Diffraction) using a cryo solution consisting of 15% (v/v) glycerol, 25.5% (w/v) PEG 4000, 0.25 M NaSCN, 85 mM Tris-HCl (pH 7.5), 85 mM

Table 1. Crystallographic Data Collection and Refinement Statistics<sup>a</sup>

	lactose	glycerol	apo (100 K)	apo (room temperature)
Data Collection				
wavelength (Å)	0.9078	0.9078	0.9078	0.9080
resolution (Å)	30.0–0.86 (0.88–0.86)	30.0–0.90 (0.92–0.90)	30.0–1.08 (1.11–1.08)	30.0–1.25 (1.28–1.25)
space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
unit cell parameters ( <i>a</i> , <i>b</i> , <i>c</i> ) (Å)	37.8, 58.3, 63.1	35.8, 58.2, 62.5	35.7, 58.3, 62.8	36.5, 58.2, 63.7
no. of measured reflections	680438	549501	319133	294647
no. of unique reflections	111079	93342	56940	37798
completeness (%)	98.9 (95.4)	95.8 (87.3)	99.9 (99.9)	98.6 (97.7)
multiplicity	6.1 (4.0)	5.9 (4.6)	5.6 (4.7)	7.8 (5.7)
<i>R</i> <sub>merge</sub> (%) <sup>b</sup>	5.0 (85.3)	3.8 (79.2)	7.0 (68.2)	5.3 (85.2)
<i>I</i> / <i>σ</i> ( <i>I</i> )	19.3 (1.8)	20.4 (2.1)	11.6 (2.4)	18.7 (2.5)
Refinement				
resolution limits (Å)	20.0–0.86	20.0–0.90	20.0–1.08	20.0–1.25
<i>R</i> <sub>model</sub> (%) <sup>c</sup>	12.7	13.2	14.9	11.9
<i>R</i> <sub>free</sub> (%) <sup>d</sup>	14.2	15.0	18.2	16.6
no. of waters	275	262	320	168
rmsd from ideal values				
bond lengths (Å) <sup>e</sup>	0.017	0.015	0.13	0.015
angle distances (Å) <sup>f</sup>	0.042	0.032	0.032	0.033
average <i>B</i> factor (Å <sup>2</sup> ) (protein/solvent/ ligand)	11.5/25.8/18.0	12.7/24.8/15.1	15.3/31.9/na	16.6/42.8/na
average anisotropy for all atoms (protein/ solvent/ligand)	0.35/0.30/0.32	0.36/0.34/0.35	0.40/0.33/na	0.38/0.41/na
data/parameter ratio	7.4	6.6	3.6	2.9
Ramachandran plot <sup>g</sup>				
residues in most favored regions	98.5% (134/136)	98.5% (134/136)	97.1% (132/136)	97.8% (133/136)

<sup>a</sup>Values in parentheses correspond to the highest-resolution shell. The resolution limit was taken to be the point at which the (*I*/*σ*(*I*)) ratio was approximately equal to 2. na means not available. <sup>b</sup> $R_{\text{merge}} = \sum_{hkl} \sum_i |I(hkl)_i - \bar{I}(hkl)| / \sum_{hkl} \sum_i I(hkl)_i$ . <sup>c</sup> $R_{\text{model}} = \sum_{hkl} |F_o(hkl) - F_c(hkl)| / \sum_{hkl} F_o(hkl)$ , where *F*<sub>o</sub> and *F*<sub>c</sub> are the observed and calculated structure factors, respectively. <sup>d</sup>A 5% random test set. <sup>e</sup>Calculated from DFIX restraints in SHELXL. <sup>f</sup>Calculated from DANG restraints in SHELXL. <sup>g</sup>Calculated using Molprobitry.<sup>79</sup>

MgCl<sub>2</sub>, 7 mM β-mercaptoethanol, and 4 mM 3'-benzamido-*N*-acetylactosamine. The same cryoprotectant but without 3'-benzamido-*N*-acetylactosamine was used to cryoprotect the Gal3C–glycerol complex crystals, whereas 15% (v/v) PEG 400 [instead of 15% (v/v) glycerol] was used for apo-Gal3C. For data collection at room temperature, the crystal was mounted using the MicroRT kit (MiTeGen). The X-ray diffraction data for all the crystals were collected on a 165 mm marResearch CCD detector on beamline I911-5 of the MAX-II synchrotron in Lund, Sweden. Diffraction data were integrated and scaled using XDS and XSCALE.<sup>54</sup> Data statistics are listed in Table 1.

The structure of the lactose–Gal3C complex was determined by rigid-body refinement using the structure of another Gal3C–ligand complex as the initial model<sup>45</sup> and Refmac5,<sup>55</sup> as implemented in the CCP4 suite.<sup>56</sup> The apo and glycerol–Gal3C structures were determined in a similar manner, with the lactose–Gal3C structure minus the ligand serving as a starting model. In the initial stages, all refinement was conducted to a resolution of 1.4 Å using Refmac5, with 5% of the total reflections randomly set aside for cross validation. Subsequently, the resolution was gradually extended to the full resolution range, and refinement was conducted using SHELXL-97.<sup>57</sup> Manual model building was conducted using Coot.<sup>58</sup> The structure of lactose and the accompanying refinement restraints were generated using the CCP4 program Monomer Library Sketcher. In the final stages of refinement, many of the hydrogen atoms were visible in difference electron density maps and were added at calculated positions using the SHELXL riding model. Alternating steps of anisotropic refinement and minor structure adjustment were performed

until convergence. Refinement statistics are listed in Table 1. Molecular images were generated using PyMOL.<sup>59</sup>

**High-Resolution Protein NMR Experiments.** Lactose-bound Gal3C and apo-Gal3C samples were prepared as reported previously,<sup>52</sup> using proteins expressed in [<sup>15</sup>N,<sup>1</sup>-<sup>13</sup>C<sub>1</sub>]-glucose-containing minimal medium.<sup>60,61</sup> <sup>1</sup>H–<sup>15</sup>N HSQC spectra were acquired with increasing concentrations of glycerol on a sample of Gal3C that was initially in the apo state at 0.46 mM. Nine additions of 10 μL of 25 mM glycerol in NMR buffer were made, yielding a final glycerol concentration of 5.9 mM, corresponding to 16 equiv with respect to Gal3C. The <sup>1</sup>H and <sup>15</sup>N spectral widths were 8000 and 1825 Hz, respectively. The static magnetic field strength was 14.1 T, and the temperature was 301 K.

The tautomeric state of histidines was monitored by <sup>1</sup>H–<sup>15</sup>N HMQC and <sup>1</sup>H–<sup>13</sup>C HSQC spectra, centered at 200 and 128 ppm in the <sup>15</sup>N and <sup>13</sup>C dimensions, respectively, and covering spectral widths of 180 and 28 ppm, respectively. The magnetization transfer delay for <sup>1</sup>H–<sup>15</sup>N HMQC was set to 22.2 ms, to refocus magnetization arising from <sup>1</sup>J<sub>NH</sub> couplings. The static magnetic field strength was 11.7 T, and the temperature was 298 K.

**Low-field <sup>2</sup>H Relaxation Dispersion Experiments.** Lactose-bound Gal3C and apo-Gal3C samples were prepared as reported previously,<sup>52</sup> except that the solvent contained 50% <sup>2</sup>H<sub>2</sub>O and the lactose concentration in the former sample was 20 mM. The longitudinal relaxation rate *R*<sub>1</sub> of the water <sup>2</sup>H magnetization was measured with an accuracy of ~0.5–1.0% at seven different frequencies, using conventional cryomagnets for the higher frequencies and an iron core magnet (Drusch EAR-



35N) interfaced with a Tecmag Discovery console for the lower frequencies. The sample temperature was regulated to  $299.8 \pm 0.1$  K. The relaxation rate was measured at each frequency for both apo-Gal3C and lactose-bound Gal3C, as well as for two matching reference solutions that did not contain Gal3C.

The relaxation dispersion is given by

$$R_1(\omega_0) = R_{\text{bulk}} + 0.2J(\omega_0) + 0.8J(2\omega_0)$$

where the spectral density is modeled as

$$J(\omega) = \alpha + \beta \frac{\tau}{1 + (\omega\tau)^2}$$

in which  $\alpha$  describes a frequency-independent contribution from water molecules that have rotational correlation times significantly longer than those in bulk water, but shorter than 1 ns, due to interactions with the external protein surface.  $\beta$  describes the frequency-dependent contribution from water molecules with longer correlation times due to interactions with internal sites in the protein and is given by

$$\beta = \frac{\omega_Q^2}{N_T} N_\beta S_\beta^2$$

where  $\omega_Q$  is the  $^2\text{H}$  quadrupole coupling frequency,  $N_T$  is the known ratio of water and protein molecules in the sample,  $N_\beta$  is the number of water molecules with a  $\tau$  of  $>1$  ns, and  $S_\beta$  is the orientational order parameter of bound waters.

**Isothermal Titration Calorimetry.** Experiments were performed as described previously,<sup>45</sup> in two series of 30 injections of 10  $\mu\text{L}$  of either 58 or 117 mM glycerol (first injection of 5  $\mu\text{L}$ ) in 5 mM HEPES buffer (pH 7.4). The concentration of Gal3C was 0.1 mM.

**Molecular Dynamics Simulations and Analysis.** MD simulations were performed as described previously,<sup>62</sup> starting from the 100 K apo-Gal3C crystal structure reported here. Ten independent simulations were used, each 5 ns long. Snapshots were saved every picosecond for analysis.

Water occupancy was investigated by the clustering approach of Friesner and co-workers.<sup>63,64</sup> In each snapshot, water oxygen atoms within the active site were stored for clustering. The active site was defined by fitting each MD snapshot to the lactose–Gal3C structure by superimposing the backbone atoms of all residues within 10 Å of lactose. The active site was then defined as the maximal extent of the lactose molecule plus 1 Å in each direction. Subsequently, the water oxygen atoms within the active site were clustered by an iterative approach. In each iteration, the number of water oxygen atoms in all snapshots within 1 Å of a particular water oxygen atom in a given snapshot was counted and the water molecule with the largest number was marked as a water site. That water molecule and all of its neighbors within 1 Å were removed from the list, and the process was repeated until the number of neighbors of a water site was lower than in bulk water.

The clustering procedure was applied to the 10 independent simulations, and the water sites identified in each individual simulation were clustered using a single-linkage hierarchical approach.<sup>65</sup> Constraints were imposed such that two water sites identified from the same simulation could not be in the same cluster. When the minimal distance between two clusters was larger than 1 Å, the clustering was stopped. Euclidian distances between the water sites were used to determine their closeness. For each cluster of water sites, the population number and the maximal spatial extent were calculated. The occupancy is the

fraction of the 10 independent simulations in which the water site was found. The maximal spatial extent is the maximal distance between two water molecules in the same cluster and indicates the precision of the position of the water site. The geometric centers of the water site clusters are reported.

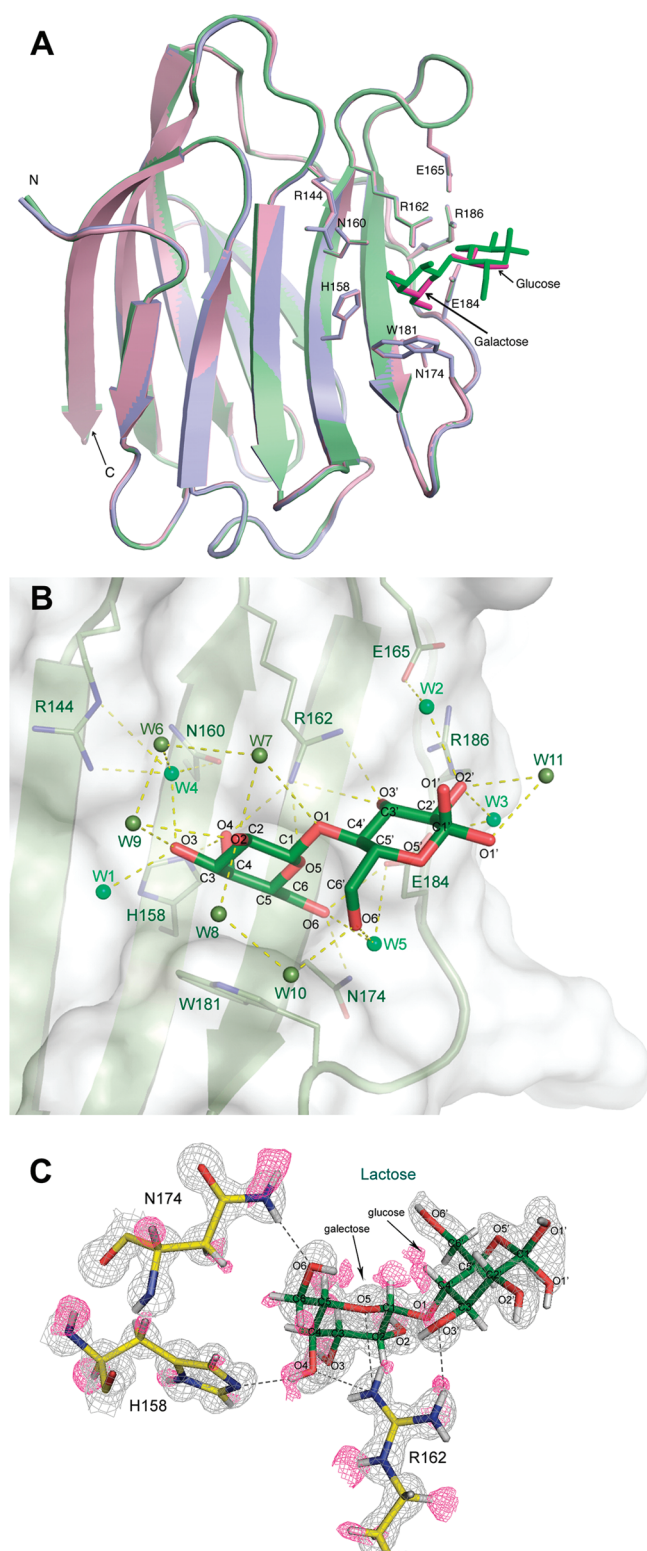
## RESULTS AND DISCUSSION

We have determined the crystal structures of the lactose-bound [Protein Data Bank (PDB) entry 3ZSJ], glycerol-bound (PDB entry 3ZSK), and ligand-free (apo; PDB entry 3ZSL at 100 K and PDB entry 3ZSM at room temperature) states of Gal3C at atomic or near-atomic resolution (Table 1). Overall, these structures are highly isomorphous to each other (Figure 1A and Table S1 of the Supporting Information) and served as a basis for further investigations of oxygen recognition within the binding site of Gal3C. To this end, we performed MD simulations of apo-Gal3C to determine the population density and residence times of water molecules in the binding site. Furthermore, we conducted nuclear magnetic relaxation dispersion measurements to experimentally verify the residence times of water molecules bound to Gal3C. Finally, ITC and chemical shift mapping by NMR spectroscopy were used to study whether glycerol binds to Gal3C in solution under ambient conditions.

Crystal structures similar to those reported here have been published previously,<sup>66</sup> but these were determined at lower resolution (1.35–2.45 Å) from samples with mixed populations of the glycerol-bound, lactose-bound, and apo states. Both of these factors may compromise the interpretation of the observed electron density. Our data mitigate such complications, because the diffraction experiments were conducted with crystals of well-defined samples with only a single state present. Thus, these structures form a solid basis for investigations of water and ligand coordination within the carbohydrate-binding site and for detailed comparisons with results from MD simulations and NMR spectroscopy.

**The Structure of Lactose-Bound Gal3C Reveals Water Molecules and Hydrogen Atoms.** The 0.86 Å resolution structure of Gal3C in complex with lactose yields a highly detailed description of the lactose-binding site (Figure 1B,C), including the positions of many hydrogen atoms (see further below). The conformation of lactose is identical to that in the previously reported crystal structures of lactose- and *N*-acetyllactosamine (LacNAc)-bound Gal3C at lower resolution.<sup>16,43,66</sup> As observed previously,<sup>66</sup> the electron density clearly reveals the presence of both  $\alpha$  and  $\beta$  anomers of lactose with equal occupancy (Figure S1A of the Supporting Information). However, the quality and the detail of the electron density map were markedly improved at this resolution compared to those of previous structures. In particular, we can identify a significantly increased number of water molecules, 289 (of which 14 have partial occupancy), compared to the 197 seen in the previously reported structure at 1.35 Å.<sup>66</sup> In our structures, we have maintained a consistent numbering system for the water molecules, such that wherever possible an experimentally observed water molecule at a given position has the same number in all structures. Details are provided in Table S2 of the Supporting Information.

The binding site includes 11 water molecules. Five of these (W1–W5) are conserved in all available lactose- and LacNAc-bound Gal3C structures.<sup>16,43,66</sup> W2–W4 make important bridging hydrogen bonds between lactose and Gal3C residues Arg144, Asn160, Glu165, and Glu184 (Figure 1B). In addition,



**Figure 1.** Atomic-resolution structures of Gal3C, in the presence of lactose (green) or glycerol (pink) and in the apo form (blue). (A) Gal3C displays an identical conformation in all complexes, and the binding sites are highly comparable. Lactose and glycerol molecules are shown as green and pink sticks, respectively. Key side chains coordinating the ligands are indicated. (B) The carbohydrate-binding site of the lactose–Gal3C complex at 0.86 Å resolution reveals water molecules W1–W5 (dark green spheres) that mediate hydrogen bonding between lactose and Gal3C. Additional water molecules that coordinate lactose, but not the protein, are shown as light green

**Figure 1.** continued

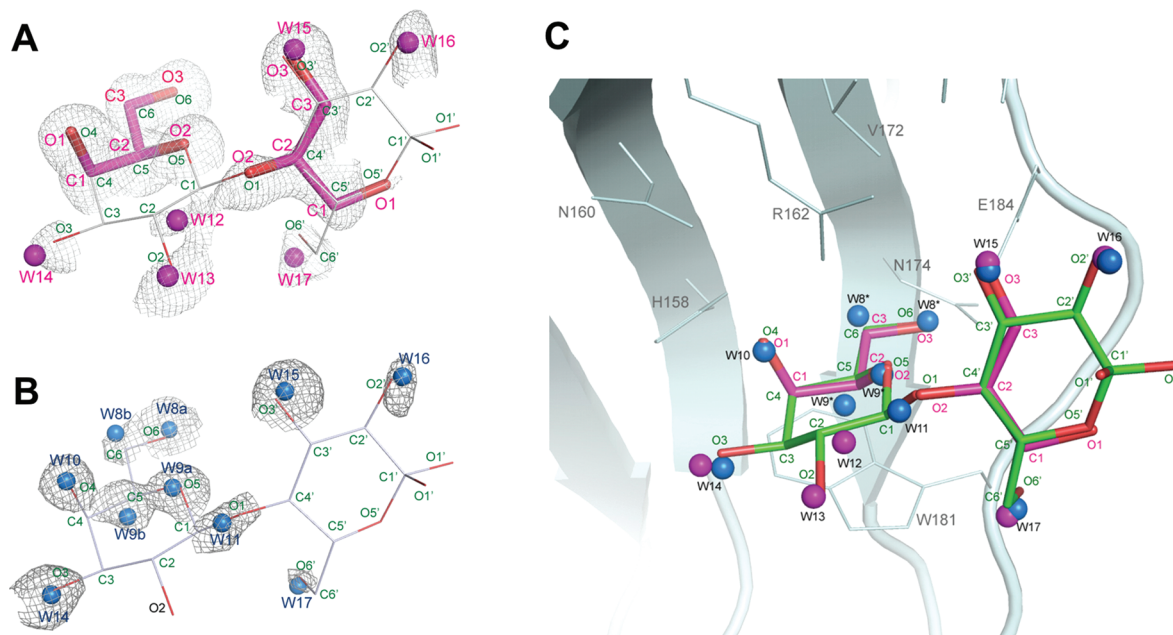
spheres. Hydrogen bonds are represented as dotted lines. (C) Electron density in the lactose-binding site, identifying important hydrogen atoms. The  $2F_o - F_c$  electron density map contoured at  $1.5\sigma$  (gray mesh) and the  $F_o - F_c$  map contoured at  $2.0\sigma$  (pink) were obtained after removal of H atoms followed by 10 refinement cycles to remove bias. The  $F_o - F_c$  map was also contoured at a lower level of  $1.6\sigma$  to visualize the H density around C6 and O6 of lactose.

six water molecules (W6–W11) that have not previously been seen were identified. These coordinate lactose through hydrogen bonds involving O1–O3 of the galactose moiety and O1', O2', and O6' of the glucose moiety (Figure 1B) but are loosely bound and do not contact any protein atoms.

The ultra-high-resolution structure clearly reveals the positions of many hydrogen atoms (Figure 1C and Figure S1B of the Supporting Information). A total of 497 positive peaks above  $2\sigma$  were experimentally observed at ideal hydrogen positions using the  $F_o - F_c$  difference omit map, in which all hydrogen atoms were removed from the model. These peaks represent 45% of the theoretical number and are mostly located close to the backbone N and C $\alpha$  atoms (corresponding to 70 and 85% of all H<sup>N</sup> and H $\alpha$  atoms, respectively). One hundred twenty intramolecular N–H $\cdots$ O hydrogen bonds were predicted (excluding those to water molecules) based on the positions of riding hydrogen atoms, using HBPLUS with a maximal D–H $\cdots$ A distance of 3.5 Å and a minimal D–H $\cdots$ A angle of 90° as criteria.<sup>67</sup> The H atom omit map experimentally confirms the presence of 80 hydrogen bonds, comprising 63 of 80 possible main chain–main chain, nine of 18 possible side chain–main chain, and eight of 18 side chain–side chain interactions. Within the binding site, 49% of the possible H atoms were experimentally observed. These primarily belong to the main chain (80% visible), whereas the side chain hydrogens are less well determined (40% visible).

Because the ligand binding environment is a crucial factor to consider in the design of synthetic ligands, a detailed analysis of the hydrogen bonding pattern in the binding site is of particular interest. Direct identification of the hydrogen atoms improves the description of the hydrogen bonding patterns for key residues. Hydrogens are visible on several conserved amino acid residues of the binding site (Figure 1C). In addition, several hydrogens of lactose are observed, which allow the direct assignment of individual hydroxyl groups as either hydrogen bond donors or acceptors. For example, the electron density clearly defines the tautomeric state of His158, showing that N<sup>δ1</sup> carries the hydrogen (Figure 1C). Further, N<sup>ε2</sup> is seen to accept a hydrogen bond from the galactose O4 hydroxyl group, whose H atom is observed. The tautomer assignment of His158 is verified by the <sup>13</sup>C and <sup>15</sup>N chemical shifts and cross-peak patterns in the <sup>1</sup>H–<sup>13</sup>C and <sup>1</sup>H–<sup>15</sup>N correlation spectra of the imidazole ring (Figure S2A of the Supporting Information), which provide unequivocal evidence that N<sup>δ1</sup> is protonated.<sup>68</sup> Two H<sup>γ</sup> atoms of Arg162 are observed to coordinate O3' of the glucose moiety, which is the only glucose atom involved in a direct interaction with the protein (Figure 1C).

Hydrogens were also identified on the galactose hydroxyl groups O2 and O3, which interact with surface water molecules (W1, W4, and W7–W9). Aliphatic H atoms in four of the lactose CH groups (C1, C3, C5, and C4') show good electron density (Figure 1C), whereas those on C3 and C5 of galactose interact with the  $\pi$ -electron system of Trp181.

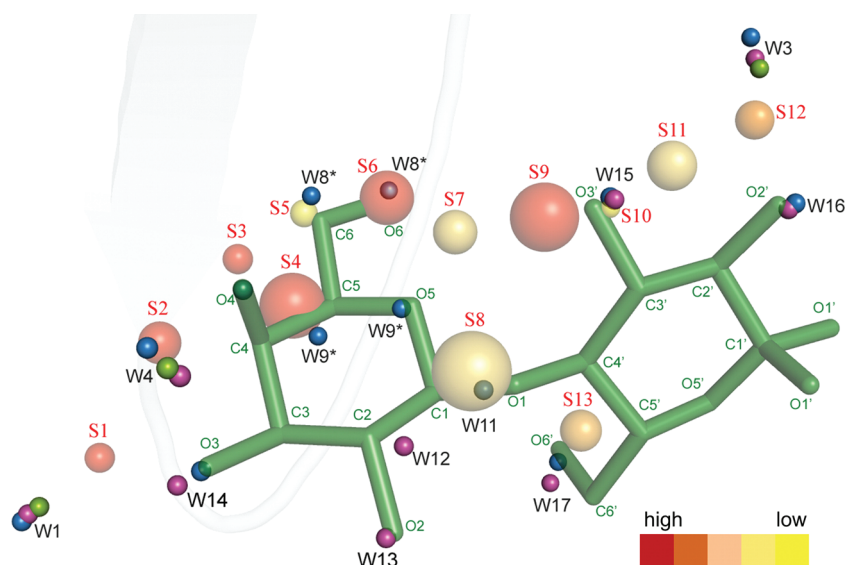


**Figure 2.** Carbohydrate-binding site of Gal3C. The  $2F_o - F_c$  electron density maps, contoured at  $1.0\sigma$ , are shown for the binding sites of glycerol-bound Gal3C and apo-Gal3C, superimposed on the structure of lactose (stick model) in lactose-bound Gal3C. (A) Electron density for glycerol-bound Gal3C. Glycerol and water molecules in the glycerol complex are shown as magenta sticks and spheres, respectively. To illustrate the molecular mimicry of lactose by glycerol, lactose is shown as thin sticks. The lactose and glycerol atom annotations are colored green and magenta, respectively. (B) Electron density for water molecules in apo-Gal3C. Water molecules are shown as blue spheres. Labels a and b are used to denote alternate positions of water molecules W8 and W9. As in panel A, lactose is shown as thin sticks as an aid to interpretation. (C) Superposition of the lactose-bound Gal3C, glycerol-bound Gal3C, and apo-Gal3C structures, demonstrating the common oxygen recognition motif of the binding site. The oxygens of lactose (green) and glycerol (magenta) occupy the same positions as water molecules found in apo (blue spheres) and glycerol-bound (magenta spheres) Gal3C.

**Structure of Glycerol-Bound Gal3C.** It is of interest to compare the binding modes of lactose and glycerol in a complex with Gal3C, because glycerol can be viewed as a fragment of lactose. As a result of a quick soak of apo-Gal3C crystals with glycerol, the carbohydrate-binding site of Gal3C reveals electron density at  $0.9 \text{ \AA}$  resolution that can be fitted with two glycerol molecules. In one of these, the oxygen and carbon atoms of glycerol are positioned identically to galactose atoms O4–O6 and C4–C6. The second glycerol molecule, modeled with 50% occupancy, was found in the same position as the O1, O3', O5', and C3'–C5' atoms of glucose. This second glycerol molecule was not seen at  $1.35 \text{ \AA}$ .<sup>66</sup> In addition to glycerol, the electron density can be fitted with six water molecules (W12–W17), five of which have partial occupancy (Figure 2A). W1–W5, present in the lactose- and LacNAc-bound structures, are also present in glycerol-bound Gal3C. The water- and glycerol-binding sites partially overlap, indicating that the electron density reports on the ensemble average of different glycerol binding modes throughout the crystal. A third partially occupied glycerol molecule might be invoked (not shown) to explain the electron density (partly assigned to W12) that appears near the sites where the galactose C1 and C2 atoms are located in the lactose-bound state. Together, the oxygen atoms of glycerol in these overlapping binding modes perfectly map out the binding site observed for lactose, except for the most peripheral part of the glucose moiety (Figure 2A,C). To this extent, glycerol and lactose constitute a naturally occurring illustration of the principle of fragment-based drug discovery in that the atoms of the initial fragments (glycerol) maintain their original orientations in the elaborated, higher-affinity ligand (lactose).

**Binding of Glycerol to Gal3C Is Insignificant at Room Temperature.** We used ITC and NMR to monitor the binding of glycerol to Gal3C at room temperature and under conditions (protein and ligand concentrations) similar to those used to study lactose binding. ITC experiments show no significant heat evolution upon addition of up to 130 equiv of glycerol to apo-Gal3C, beyond the heat of dilution (Figure S3A of the Supporting Information). Similarly,  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectra of Gal3C acquired with increasing concentrations of glycerol show that the protein chemical shifts are unperturbed by the addition of 16 equiv of glycerol (Figure S3B of the Supporting Information). Thus, both ITC and NMR show that binding of glycerol to Gal3C is very weak ( $K_d > 100 \text{ mM}$ ) at room temperature. By contrast, lactose binds to Gal3C with a dissociation constant ( $K_d$ ) of  $230 \text{ }\mu\text{M}$ , governed by a favorable enthalpy, but an unfavorable entropy,<sup>45,69</sup> as is typical for interactions of galectin with oligosaccharides.<sup>70</sup> The lack of glycerol binding at ambient temperature is consistent with the thermodynamics of binding of a carbohydrate to Gal3C. Evidently, the enthalpy of formation of the glycerol–Gal3C complex cannot overcome the entropic penalty at room temperature. This observation is explained by the smaller size of glycerol, which is essentially one-quarter of a lactose molecule and therefore engages in fewer oxygen-binding sites on Gal3C at a time, relative to lactose. In contrast, at the low temperature ( $100 \text{ K}$ ) of the cryocooled X-ray diffraction experiments, the thermodynamic balance is likely reversed so that the entropic penalty is now overcome by the favorable enthalpy of binding. In addition, the glycerol concentration in the cryoprotectant solution is very high ( $2 \text{ M}$ ), which further serves to drive glycerol into the binding site of Gal3C in the





**Figure 3.** Water sites identified in the MD simulations. The sites are labeled S1–S12 from left to right. The lactose molecule (green) is shown in the active site together with the water sites. The color scheme ranges from high occupancy (red, present in all 10 simulations) to low occupancy (yellow, present in only two of 10 simulations). The size of a sphere is related to the maximal extent (see the text); i.e., water sites defined with high precision are depicted as smaller spheres and vice versa. For reference, the water molecules in the crystal structures of lactose-bound Gal3C, glycerol-bound Gal3C, and apo-Gal3C are shown as green, magenta, and blue spheres, respectively.

crystal. The fact that binding of glycerol to Gal3C could not be detected in solution at room temperature yet occurs perfectly as a fragment lead that can be elaborated into lactose emphasizes the important role of protein crystallography in fragment-based drug discovery.<sup>71,72</sup>

**Structure of Apo-Gal3C.** The crystal structure of apo-Gal3C was determined at 1.08 Å using PEG 400 as a cryoprotectant, to avoid glycerol binding. There are no significant differences in the conformation of the binding site from that of either the lactose- or glycerol-bound forms (Figure 2). Despite the slightly lower resolution of the apo structure, the  $2F_o - F_c$  electron density map contoured at  $2\sigma$  shows electron density at N<sup>δ1</sup> of His158. As for lactose-bound Gal3C, the tautomeric state is corroborated by the <sup>1</sup>H–<sup>13</sup>C and <sup>1</sup>H–<sup>15</sup>N correlation spectra, which are virtually identical to those of the lactose-bound state (Figure S2B of the Supporting Information). Thus, the tautomeric state is the same in both forms, and the hydrogen bonding interaction remains in the apo state in solution.

The electron density at the binding site reveals six water molecules (W8–W11, W14, and W15) that align very well with oxygen positions of lactose (O6, O5, O4, O1, O3, and O3', respectively) (see Figure 2B). The water molecules located at the O5 (W9a) and O6 (W8a) atoms were modeled with alternate conformations (W9b and W8b) that coincide with the C5 and C6 atoms of lactose, and they were assigned occupancies of 0.4 and 0.3, respectively. Two other water molecules with half-occupancy were found at positions equivalent to those of O2' (W16) and O6' (W17) in lactose. For comparison, the previously determined apo-Gal3C structure included only three water molecules in the binding site (corresponding to O3, O4', and O6').<sup>66</sup>

To eliminate the possibility that the observed water molecules could be explained by PEG 400 or fragments thereof, the structure of apo-Gal3C was also determined at room temperature (298 K), without the addition of either glycerol or PEG 400. The resulting structure at 1.25 Å

resolution confirms the position of the water molecules observed in the data at 100 K, except for the two partially occupied water molecules at the glucose O2' and O6' positions (W16 and W17 in Figure 2B), which have the fewest contacts with the protein. Conversely, the room-temperature data actually reveal another fully occupied water molecule at the position equivalent to O5' of glucose. The close agreement between the two apo structures in the number and positions of bound waters indicates that the low temperature of the cryocooled experiment does not alter the water binding properties of Gal3C in any major way.

#### The Binding Site Is Preorganized To Recognize a Sugarlike Framework of Oxygens.

The crystal structures presented here demonstrate that Gal3C coordinates water or hydroxyl oxygens (e.g., of lactose or glycerol) at identical positions in its binding site, as summarized in Figure 2C. The structures also emphasize the significance of conserved water-mediated hydrogen bonding in stabilizing the Gal3C ligands. Further, side chain conformations within the binding site and the loop regions on either side of the binding groove are indistinguishable among the apo, glycerol-bound, and lactose-bound forms; i.e., the site is fully preorganized in the crystallized apo form to accommodate a sugarlike framework. This view contrasts with the solution structure of apo-Gal3C, in which the loops surrounding the binding site have a different conformation in the absence of ligand.<sup>73</sup> Also, the NMR order parameter for the Trp181 side chain of apo-Gal3C indicates that it is highly flexible in solution.<sup>45</sup> These differences in conformation between the solution and crystalline states reflect crystal contacts near the binding site as well as different flexibilities of the protein in solution and in the crystal. Taken together, the NMR and crystal structures suggest that the crystal traps a conformation of the protein that is not the dominant one in solution but is optimized for ligand binding. Consistent with this conclusion, the previous observation that lactose diffuses very slowly out of Gal3C crystals<sup>66</sup> suggests that

**Table 2. Details of the Water Site Clusters within 1.6 Å of Lactose Obtained from the MD Simulations of Apo-Gal3C**

water site	occupancy <sup>a</sup>	maximal extent (Å) <sup>b</sup>	<i>r</i> (Lac) (Å) <sup>c</sup>	<i>r</i> (Gal3C) (Å) <sup>d</sup>	<i>r</i> (H <sub>2</sub> O) (Å) <sup>e</sup>
S1	1.0	0.6	1.0 (O3)	3.3 (H158 CE1)	1.6 (W1/W14)
S2	1.0	0.8	1.2 (O4)	3.0 (H158 NE2/R144 NE1)	0.4 (W4)
S3	1.0	0.6	0.5 (O4)	2.5 (H158 NE2)	0.6 (W10)
S4	1.0	1.7	0.8 (C5)	3.4 (W181 CE2)	0.6 (W9)
S5	0.4	0.6	0.4 (C6)	3.5 (H158 CD2)	0.5 (W8a)
S6	1.0	1.3	0.2 (O6)	3.1 (N174 ND2)	0.2 (W8b)
S7	0.7	1.1	1.0 (O6)	3.0 (N174 ND2)	1.4 (W8b)
S8	0.7	2.5	1.0 (O1)	3.0 (R162 NH2)	0.4 (W11)
S9	1.0	1.7	0.9 (O3')	2.1 (R162 NH2)	1.1 (W15)
S10	0.2	0.5	0.3 (O3')	2.2 (R162 NH1)	0.2 (W15)
S11	0.7	1.0	1.0 (O3')	2.6 (R162 NH1)	1.0 (W15)
S12	0.9	0.8	1.6 (O2')	3.1 (E184 OE2)	1.0 (W16)
S13	0.8	1.1	1.6 (O6')	4.1 (W181 CB)	1.3 (W17)

<sup>a</sup>Occupancy is the fraction of the 10 independent simulations in which a given water site was identified. <sup>b</sup>Maximal extent is the maximal distance between two members in a water site. <sup>c</sup>*r*(Lac) is the shortest distance to a lactose atom, identified within parentheses. <sup>d</sup>*r*(Gal3C) is the shortest distance to a protein atom, identified within parentheses. <sup>e</sup>*r*(H<sub>2</sub>O) is the shortest distance to a water oxygen in the apo crystal structure, identified within parentheses.

the crystalline environment restricts the protein flexibility required for ligand recognition and release.

**Enhanced Occupancy of Water Molecules at Specific Sites in Apo-Gal3C in Solution.** Molecular dynamics simulations of apo-Gal3C in water were conducted to study the water structure in the binding site (Figure 3 and Table 2), as a further probe of whether the observations made for the crystal structures also hold in solution. We identified 13 sites (S1–S13) with enhanced water density within the binding site. In general, the centers of the identified water sites are localized at or close to the water molecules observed in the apo-Gal3C structure. Sites S3, S6, S7, and S9–S12 are very close to the water sites identified in a previous MD simulation initiated from the crystal structure of LacNac-bound Gal3C.<sup>74</sup> However, that study was limited to identifying water sites based on radial and angular distribution functions around expected H-binding atoms in the protein, while our identification was unbiased by prior expectations and identified all sites with high water density regardless of their position relative to any hydrogen bonding partners in the protein.

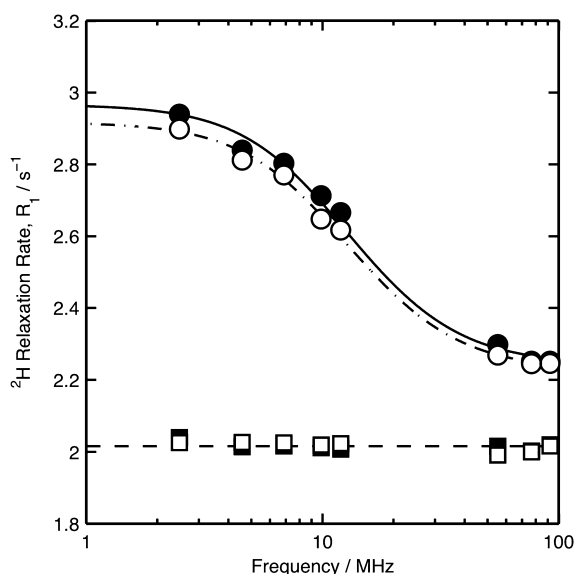
Water sites S2–S6, S8, and S10 are located very close ( $\leq 0.6$  Å) to the water molecules observed in apo-Gal3C at the sites of lactose atoms O4, C6, O6, O1, and O3', respectively [corresponding to W4, W10, W9, W8a, W8b, W11, and W15 in the apo structures, respectively (see Figure 2B)]. Moreover, S1, S7, S9, and S11–S13 are only slightly offset ( $\leq 1.6$  Å) from the observed positions of water or lactose oxygens. Notably, S12 matches well with the position of the acetyl group of LacNac in the LacNac–Gal3C complex.<sup>43</sup> S9 and S11 can be seen as alternative positions of the low-occupancy S10 site, 0.9–1.0 Å from O3'. S5 has a low occupancy of 0.4, in agreement with the 30% occupancy estimated for the water molecule (W8b) close to C6 in the apo-Gal3C structure. Similarly, site S12 is close to W16, which is only half-occupied. Thus, essentially all water molecules in the apo crystal structures are close to water sites in the MD simulation of the apo state, which further supports our conclusion that the apo structure actually is devoid of ligands, such as traces of lactose or PEG, and that the water sites persist in solution.

Comparing the MD-derived water sites in apo-Gal3C with the electron densities of the glycerol-bound Gal3C structure further aids in the interpretation of the latter. As already stated

above, we find water sites on and close to the glycerol molecules. Most of the glycerol molecule that overlaps with the O4–O6 atoms of lactose maps to water sites (S3–S6) in the MD simulation of apo-Gal3C, while the other, partially occupied glycerol molecule overlaps with only two water sites (S10 and S13). Notably, the relative populations of the two glycerol molecules match the relative number and occupancy of water sites. However, two water molecules in the glycerol-bound Gal3C structure [W12 and W13, located near the C2 and O2 sites of lactose, respectively (Figure 2A)] do not correspond to any identified water site in the MD simulations, nor do they correspond to any waters in the apo-Gal3C structure. This observation further supports the hypothesis that the electron density of these tentative waters might in fact be explained by a third glycerol molecule with low occupancy.

**The Bound Water Molecules in the Lactose-Binding Site Exchange Rapidly with Bulk Water.** To investigate the residence times of water molecules bound to Gal3C, we used nuclear magnetic relaxation dispersion (MRD) experiments, which sensitively detect long-lived waters bound to proteins.<sup>75,76</sup> We performed <sup>2</sup>H relaxation dispersion experiments covering resonance frequencies from 2.5 to 92.1 MHz (Figure 4). The MRD profiles of apo-Gal3C and lactose-bound Gal3C are highly similar, indicating that lactose does not displace any long-lived water molecules in the binding site. If this had been the case, the dispersion step for the apo sample would be greatly enhanced compared to that for the lactose-bound sample. Using a rotational diffusion correlation time ( $\tau$ ) of 7.3 ns,<sup>45</sup> the fitted relaxation dispersions yield  $N_{\beta}S_{\beta}^2$  values of  $4.5 \pm 0.1$  and  $4.3 \pm 0.1$  for the lactose-bound and apo states, respectively. The slightly enhanced dispersion observed for lactose-bound Gal3C relative to apo,  $0.2 \pm 0.1$  (mean  $\pm$  one standard deviation), can readily be explained by exchange of <sup>2</sup>H between bulk solvent and the hydroxyl groups of lactose bound to Gal3C, which is estimated to give an  $N_{\beta}S_{\beta}^2$  contribution of 0.20, based on the exchange rate constants for glucose hydroxyl protons.<sup>77</sup> Both apo-Gal3C and lactose-bound Gal3C show a significant relaxation dispersion that is due to long-lived waters and labile deuterons (e.g., in hydroxyl groups) in the protein, the latter of which are estimated to make an  $N_{\beta}S_{\beta}^2$  contribution of 0.5 at the present pH. The remaining dispersion step ( $N_{\beta}S_{\beta}^2 \approx 3.8$ ) is due to long-lived waters. Expecting that these waters





**Figure 4.** Nuclear magnetic relaxation dispersion of apo-Gal3C (○) and lactose-bound Gal3C (●). Empty and filled squares represent the corresponding data from the reference samples without Gal3C. The relaxation rates for the lactose samples are downshifted by a small ( $0.05\text{ s}^{-1}$ ) and frequency-independent difference between the reference samples to facilitate comparison. The errors are of the same size as the symbols.

are relatively well ordered ( $\langle S_{\beta}^2 \rangle = 0.75\text{--}0.95$ ), we obtain an  $N_{\beta}$  of 4–5. This value is in agreement with the apo-Gal3C and lactose-bound Gal3C crystal structures, which each have three to five completely buried internal water molecules, all in regions outside of the lactose-binding site.

Importantly, the MRD data reveal that all bound waters in the binding site of apo-Gal3C exchange with bulk water on a time scale of nanoseconds or faster at room temperature, in keeping with the MD simulations, which show that the mean residence time of water molecules in the binding site is only 14 ps, with individual values ranging between a few picoseconds to 1.3 ns.

**Concluding Remarks: Implications for Drug Discovery and Design.** The water molecules observed in the crystal structures of Gal3C are weakly bound at room temperature and have only slightly prolonged residence times relative to water molecules in the hydration layer surrounding the protein. Nonetheless, the enhanced water population translates into a potential of mean force that attracts oxygen atoms to these sites, as exemplified previously.<sup>25,26,74,78</sup> Thus, the weak interactions of water molecules with the binding site of Gal3C essentially map out a free energy surface for oxygen/hydroxyl binding. Ligands such as lactose, which have several oxygens arranged to fit this surface, will bind to the protein, partly as a consequence of the concerted forces driving each ligand oxygen atom into a given minimum of the energy surface. In contrast, binding of smaller ligands such as glycerol is penalized by the fewer favorable interactions, which cannot overcome the unfavorable entropy of binding (loss of translational and rotational freedom) at room temperature. Indeed, binding of glycerol to Gal3C was not detected at room temperature by ITC or NMR, despite clear evidence of two glycerol molecules in the cryo-crystal structure. This observation has a bearing on fragment-based drug discovery, emphasizing the role of protein cryo-crystallography in

screening for suitable lead fragments that may otherwise escape detection.<sup>71,72</sup>

## ■ ASSOCIATED CONTENT

### Supporting Information

Rms deviations between all four Gal3C structures (Table S1) and the consistent numbering of water molecules in their active site areas (Table S2), electron density of the  $\alpha$  and  $\beta$  anomers of lactose (Figure S1A) and hydrogen atoms (Figure S1B), NMR spectra of histidine side chains (Figure S2), and ITC data (Figure S3A), and NMR spectra (Figure S3B) of glycerol titrations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

### Accession Codes

The atomic coordinates and structure factors have been deposited in the Protein Data Bank as entries 3ZSJ (lactose-bound Gal3C), 3ZSK (glycerol-bound Gal3C), 3ZSL (apo-Gal3C at 100 K), and 3ZSM (apo-Gal3C at room temperature).

## ■ AUTHOR INFORMATION

### Corresponding Author

\*D.T.L.: e-mail, [derek.logan@biochemistry.lu.se](mailto:derek.logan@biochemistry.lu.se); telephone, +46 46 222 1443. M.A.: e-mail, [mikael.akke@bpc.lu.se](mailto:mikael.akke@bpc.lu.se); telephone, +46 46 222 8247.

### Funding

This work was supported by the Swedish Research Council (Grant 2007-5450/2010-4912 to M.A., Grant 2010-5025 to U.R., Grant 2006-3985 to U.J.N., Grant 2009-5656 to H.L., and Grant 2006-4387 to D.T.L.), the FLÅK Research School for Pharmaceutical Sciences at Lund University (M.A. and U.R.), the Carl Trygger Foundation (D.T.L.), the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine (M.A.), and the Knut and Alice Wallenberg Foundation (M.A.).

### Notes

D.T.L. and M.A. are founders and shareholders of the company SARomics Biostructures, which was involved in determining three of the crystal structures reported in this paper.

## ■ ACKNOWLEDGMENTS

Computer resources were provided by LUNARC at Lund University and HPC2N at Umeå University (Umeå, Sweden). We thank Matilda Larsson and Dennis Ivarsson for assistance with NMR experiments and the staff at beamline I911 of MAX-lab for maintaining the facility and for help with data collection.

## ■ ABBREVIATIONS

Gal3C, galectin-3 carbohydrate recognition domain; CRD, carbohydrate recognition domain; MD, molecular dynamics; ITC, isothermal titration calorimetry; PEG, poly(ethylene glycol); HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; rmsd, root-mean-square deviation; LacNAc, *N*-acetylglucosamine.

## ■ REFERENCES

- Geijtenbeek, T. B.; Torensma, R.; van Vliet, S. J.; van Duijnhoven, G. C.; Adema, G. J.; van Kooyk, Y.; and Figdor, C. G. (2000) Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* 100, 575–585.
- Sacchetti, J. C.; Baum, L. G.; and Brewer, C. F. (2001) Multivalent protein-carbohydrate interactions. A new paradigm for

- supermolecular assembly and signal transduction. *Biochemistry* 40, 3009–3015.
- (3) Karlsson, A., Feuk-Lagerstedt, E., Almkvist, J., Leffler, H., and Dahlgren, C. (1999) Priming is required for galectin-3-induced neutrophil NADPH-oxidase activity. *J. Leukocyte Biol.*, 25.
- (4) Henderson, N. C., and Sethi, T. (2009) The regulation of inflammation by galectin-3. *Immunol. Rev.* 230, 160–171.
- (5) von Itzstein, M. (2007) The war against influenza: Discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discovery* 6, 967–974.
- (6) Bernardi, A., and Cheshev, P. (2008) Interfering with the Sugar Code: Design and Synthesis of Oligosaccharide Mimics. *Chem.—Eur. J.* 14, 7434–7441.
- (7) Ernst, B., and Magnani, J. L. (2009) From carbohydrate leads to glycomimetic drugs. *Nat. Rev. Drug Discovery* 8, 661–677.
- (8) Lepenies, B., Yin, J., and Seeberger, P. H. (2010) Applications of synthetic carbohydrates to chemical biology. *Curr. Opin. Chem. Biol.*, 1–8.
- (9) Dam, T. K., and Brewer, C. F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* 102, 387–429.
- (10) Toone, E. J. (1994) Structure and Energetics of Protein Carbohydrate Complexes. *Curr. Opin. Struct. Biol.* 4, 719–728.
- (11) Garcia Hernandez, E., Zubillaga, R. A., Rojo Dominguez, A., Rodriguez Romero, A., and Hernandez Arana, A. (1997) New insights into the molecular basis of lectin-carbohydrate interactions: A calorimetric and structural study of the association of hevein to oligomers of N-acetylglucosamine. *Proteins: Struct., Funct., Genet.* 29, 467–477.
- (12) Garcia-Hernandez, E., and Hernandez-Arana, A. (1999) Structural bases of lectin-carbohydrate affinities: Comparison with protein-folding energetics. *Protein Sci.* 8, 1075–1086.
- (13) Lemieux, R. U. (1999) How water provides the impetus for molecular recognition in aqueous solution (vol 29, pg 373, 1996). *Acc. Chem. Res.* 32, 631.
- (14) Lis, H., and Sharon, N. (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* 98, 637–674.
- (15) Davis, A. P., and Wareham, R. S. (1999) Carbohydrate recognition through noncovalent interactions: A challenge for biomimetic and supramolecular chemistry. *Angew. Chem., Int. Ed.* 38, 2978–2996.
- (16) Sörme, P., Arnoux, P., Kahl-Knutsson, B., Leffler, H., Rini, J. M., and Nilsson, U. J. (2005) Structural and thermodynamic studies on cation- $\pi$  interactions in lectin-ligand complexes: High-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J. Am. Chem. Soc.* 127, 1737–1743.
- (17) del Carmen Fernandez-Alonso, M., Canada, F. J., Jimenez-Barbero, J., and Cuevas, G. (2005) Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc.* 127, 7379–7386.
- (18) Laughrey, Z. R., Kiehna, S. E., Riemen, A. J., and Waters, M. L. (2008) Carbohydrate- $\pi$  interactions: What are they worth? *J. Am. Chem. Soc.* 130, 14625–14633.
- (19) Ladbury, J. E. (1996) Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design. *Chem. Biol.* 3, 973–980.
- (20) Clarke, C., Woods, R. J., Gluska, J., Cooper, A., Nutley, M. A., and Boons, G. J. (2001) Involvement of water in carbohydrate-protein binding. *J. Am. Chem. Soc.* 123, 12238–12247.
- (21) Chervenak, M. C., and Toone, E. J. (1995) Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities. *Biochemistry* 34, 5685–5695.
- (22) Dunitz, J. D. (1994) The entropic cost of bound water in crystals and biomolecules. *Science* 264, 670.
- (23) Kadirvelraj, R., Foley, B. L., Dyekjaer, J. D., and Woods, R. J. (2008) Involvement of Water in Carbohydrate-Protein Binding: Concanavalin A Revisited. *J. Am. Chem. Soc.* 130, 16933–16942.
- (24) Lemieux, R. U. (1996) How water provides the impetus for molecular recognition in aqueous solution. *Acc. Chem. Res.* 29, 373–380.
- (25) Di Lella, S., Ma, L., Ricci, J. C., Rabinovich, G. A., Asher, S. A., and Alvarez, R. M. (2009) Critical role of the solvent environment in galectin-1 binding to the disaccharide lactose. *Biochemistry* 48, 786–791.
- (26) Li, Z., and Lazaridis, T. (2005) The effect of water displacement on binding thermodynamics: concanavalin A. *J. Phys. Chem. B* 109, 662–670.
- (27) Asensio, J. L., Siebert, H. C., von der Lieth, C. W., Laynez, J., Bruix, M., Soedjanaamadja, U. M., Beintema, J. J., Canada, F. J., Gabius, H. J., and Jimenez-Barbero, J. (2000) NMR investigations of protein-carbohydrate interactions: Studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N"-triacetylchitotriose. *Proteins: Struct., Funct., Genet.* 40, 218–236.
- (28) Leffler, H., Carlsson, S., Hedlund, M., Qian, Y., and Poirier, F. (2004) Introduction to galectins. *Glycoconjugate J.* 19, 433–440.
- (29) Rabinovich, G. A., Liu, F. T., Hirashima, M., and Anderson, A. (2007) An emerging role for galectins in tuning the immune response: Lessons from experimental models of inflammatory disease, autoimmunity and cancer. *Scand. J. Immunol.* 66, 143–158.
- (30) MacKinnon, A. C., Farnworth, S. L., Hodkinson, P. S., Henderson, N. C., Atkinson, K. M., Leffler, H., Nilsson, U. J., Haslett, C., Forbes, S. J., and Sethi, T. (2008) Regulation of alternative macrophage activation by galectin-3. *J. Immunol.* 180, 2650–2658.
- (31) Delacour, D., Koch, A., and Jacob, R. (2009) The role of galectins in protein trafficking. *Traffic* 10, 1405–1413.
- (32) Liu, F. T., and Rabinovich, G. A. (2010) Galectins: Regulators of acute and chronic inflammation. *Ann. N.Y. Acad. Sci.* 1183, 158–182.
- (33) Grigorian, A., and Demetriou, M. (2010) Manipulating cell surface glycoproteins by targeting N-glycan-galectin interactions. *Methods Enzymol.* 480, 245–266.
- (34) Nieminen, J., Kuno, A., Hirabayashi, J., and Sato, S. (2007) Visualization of galectin-3 oligomerization on the surface of neutrophils and endothelial cells using fluorescence resonance energy transfer. *J. Biol. Chem.* 282, 1374–1383.
- (35) Seelenmeyer, C., Wegehingel, S., Tews, I., Kunzler, M., Aebi, M., and Nickel, W. (2005) Cell surface counter receptors are essential components of the unconventional export machinery of galectin-1. *J. Cell Biol.* 171, 373–381.
- (36) Baptiste, T. A., James, A., Saria, M., and Ochieng, J. (2007) Mechano-transduction mediated secretion and uptake of galectin-3 in breast carcinoma cells: Implications in the extracellular functions of the lectin. *Exp. Cell Res.* 313, 652–664.
- (37) Dennis, J. W., Lau, K. S., Demetriou, M., and Nabi, I. R. (2009) Adaptive regulation at the cell surface by N-glycosylation. *Traffic* 10, 1569–1578.
- (38) Almkvist, J., and Karlsson, A. (2004) Galectins as inflammatory mediators. *Glycoconjugate J.* 19, 575–581.
- (39) Dumić, J., Dabelic, S., and Flogel, M. (2006) Galectin-3: An open-ended story. *Biochim. Biophys. Acta* 1760, 616–635.
- (40) Liu, F. T. (2005) Regulatory roles of galectins in the immune response. *Int. Arch. Allergy Immunol.* 136, 385–400.
- (41) Liu, F. T., and Rabinovich, G. A. (2005) Galectins as modulators of tumour progression. *Nat. Rev. Cancer* 5, 29–41.
- (42) Walti, M. A., Walser, P. J., Thore, S., Grunler, A., Bednar, M., Kunzler, M., and Aebi, M. (2008) Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from *Coprinopsis cinerea*. *J. Mol. Biol.* 379, 146–159.
- (43) Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998) X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution. *J. Biol. Chem.* 273, 13047–13052.
- (44) Carlsson, S., Öberg, C. T., Carlsson, M. C., Sundin, A., Niisson, U. J., Smith, D., Cummings, R. D., Almkvist, J., Karlsson, A., and

- Leffler, H. (2007) Affinity of galectin-8 and its carbohydrate recognition domains for ligands in solution and at the cell surface. *Glycobiology* 17, 663–676.
- (45) Diehl, C., Genheden, S., Engström, O., Delaine, T., Håkansson, M., Leffler, H., Nilsson, U. J., Ryde, U., and Akke, M. (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J. Am. Chem. Soc.* 132, 14577–14589.
- (46) Salameh, B. A., Leffler, H., and Nilsson, U. J. (2005) 3-(1,2,3-Triazol-1-yl)-1-thio-galactosides as small, efficient, and hydrolytically stable inhibitors of galectin-3. *Bioorg. Med. Chem. Lett.* 15, 3344–3346.
- (47) Salameh, B. A., Cumpstey, I., Sundin, A., Leffler, H., and Nilsson, U. J. (2010) 1H-1,2,3-Triazol-1-yl thiodigalactoside derivatives as high affinity galectin-3 inhibitors. *Bioorg. Med. Chem.* 18, 5367–5378.
- (48) Yang, R. Y., Rabinovich, G. A., and Liu, F. T. (2008) Galectins: Structure, function and therapeutic potential. *Expert Rev. Mol. Med.* 10, e17.
- (49) Wu, M. H., Hong, T. M., Cheng, H. W., Pan, S. H., Liang, Y. R., Hong, H. C., Chiang, W. F., Wong, T. Y., Shieh, D. B., Shiau, A. L., Jin, Y. T., and Chen, Y. L. (2009) Galectin-1-mediated tumor invasion and metastasis, up-regulated matrix metalloproteinase expression, and reorganized actin cytoskeletons. *Mol. Cancer Res.* 7, 311–318.
- (50) Delaine, T., Cumpstey, I., Ingrassia, L., Le Mercier, M., Okechukwu, P., Leffler, H., Kiss, R., and Nilsson, U. J. (2008) Galectin-inhibitory thiodigalactoside ester derivatives have antimigratory effects in cultured lung and prostate cancer cells. *J. Med. Chem.* 51, 8109–8114.
- (51) Lin, C. I., Whang, E. E., Donner, D. B., Jiang, X., Price, B. D., Carothers, A. M., Delaine, T., Leffler, H., Nilsson, U. J., Nose, V., Moore, F. D., and Ruan, D. T. (2009) Galectin-3 targeted therapy with a small molecule inhibitor activates apoptosis and enhances both chemosensitivity and radiosensitivity in papillary thyroid cancer. *Mol. Cancer Res.* 7, 1655–1662.
- (52) Diehl, C., Genheden, S., Modig, K., Ryde, U., and Akke, M. (2009) Conformational entropy changes upon lactose binding to the carbohydrate recognition domain of galectin-3. *J. Biomol. NMR* 45, 157–169.
- (53) Massa, S. M., Cooper, D. N., Leffler, H., and Barondes, S. H. (1993) L-29, an endogenous lectin, binds to glycoconjugate ligands with positive cooperativity. *Biochemistry* 32, 260–267.
- (54) Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* 66, 133–144.
- (55) Murshudov, G. N. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* 53, 240–255.
- (56) Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003) A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* 59, 1131–1137.
- (57) Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* 64, 112–122.
- (58) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D* 66, 486–501.
- (59) DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, Palo Alto, CA.
- (60) Teilum, K., Brath, U., Lundström, P., and Akke, M. (2006) Biosynthetic <sup>13</sup>C labeling of aromatic side chains in proteins for NMR relaxation measurements. *J. Am. Chem. Soc.* 128, 2506–2507.
- (61) Lundström, P., Teilum, K., Carstensen, T., Bezsonova, I., Wiesner, S., Hansen, D. F., Religa, T. L., Akke, M., and Kay, L. E. (2007) Fractional <sup>13</sup>C enrichment of isolated carbons using [1-<sup>13</sup>C]- or [2-<sup>13</sup>C]-glucose facilitates the accurate measurement of dynamics at backbone C<sup>α</sup> and side-chain methyl positions in proteins. *J. Biomol. NMR* 38, 199–212.
- (62) Genheden, S., Luchko, T., Gusarov, S., Kovalenko, A., and Ryde, U. (2010) An MM/3D-RISM approach for ligand binding affinities. *J. Phys. Chem. B* 114, 8505–8516.
- (63) Young, T., Abel, R., Kim, B., Berne, B. J., and Friesner, R. A. (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *Proc. Natl. Acad. Sci. U.S.A.* 104, 808–813.
- (64) Abel, R., Young, T., Farid, R., Berne, B. J., and Friesner, R. A. (2008) Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* 130, 2817–2831.
- (65) Shao, J. Y., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007) Clustering molecular dynamics trajectories: I. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* 3, 2312–2334.
- (66) Collins, P. M., Hidari, K. I., and Blanchard, H. (2007) Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: Possible implications for ligand-exchange protocols. *Acta Crystallogr. D* 63, 415–419.
- (67) McDonald, I. K., and Thornton, J. M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238, 777–793.
- (68) Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated III(Glc), a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. *Protein Sci.* 2, 543–558.
- (69) Bachhawat, K., Thomas, C. J., Amutha, B., Krishnaswamy, M. V., Khan, M. I., and Suroliya, A. (2001) On the stringent requirement of mannosyl substitution in mannooligosaccharides for the recognition by garlic (*Allium sativum*) lectin. A surface plasmon resonance study. *J. Biol. Chem.* 276, 5541–5546.
- (70) Ahmad, N., Gabius, H. J., Sabesan, S., Oscarson, S., and Brewer, C. F. (2004) Thermodynamic binding studies of bivalent oligosaccharides to galectin-1, galectin-3, and the carbohydrate recognition domain of galectin-3. *Glycobiology* 14, 817–825.
- (71) Murray, C. W., and Blundell, T. L. (2010) Structural biology in fragment-based drug design. *Curr. Opin. Struct. Biol.* 20, 497–507.
- (72) Burley, S. K., Hirst, G., Sprengeler, P., and Reich, S. (2010) Fragment-based structure-guided drug discovery: Strategy, process, and lessons from human protein kinases. In *Drug Design: Structure- and ligand-based approaches* (Merz, K. M., Ringe, D., and Reynolds, C. H., Eds.) pp 30–40, Cambridge University Press, New York.
- (73) Umamoto, K., Leffler, H., Venot, A., Valafar, H., and Prestegard, J. H. (2003) Conformational differences in liganded and unliganded states of galectin-3. *Biochemistry* 42, 3688–3695.
- (74) Gauto, D. F., Di Lella, S., Guardia, C. M., Estrin, D. A., and Marti, M. A. (2009) Carbohydrate-binding proteins: Dissecting ligand structures through solvent environment occupancy. *J. Phys. Chem. B* 113, 8717–8724.
- (75) Halle, B., and Denisov, V. P. (2001) Magnetic relaxation dispersion studies of biomolecular solutions. *Methods Enzymol.* 338, 178–201.
- (76) Halle, B. (2004) Protein hydration dynamics in solution: A critical survey. *Philos. Trans. R. Soc. London, Ser. B* 359, 1207–1223.
- (77) Hills, B. P. (1991) Multinuclear NMR-studies of water in solutions of simple carbohydrates. I. Proton and deuterium relaxation. *Mol. Phys.* 72, 1099–1121.
- (78) Di Lella, S., Marti, M. A., Alvarez, R. M. S., Estrin, D. A., and Ricci, J. C. D. (2007) Characterization of the galectin-1 carbohydrate recognition domain in terms of solvent occupancy. *J. Phys. Chem. B* 111, 7360–7366.
- (79) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B. III, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383.