Elucidation of Hydrogen Bonding Patterns in Ligand-Free, Lactose- and Glycerol-Bound Galectin-3C by Neutron Crystallography to Guide Drug Design

Francesco Manzoni,†,‡,○ Johan Wallerstein,† Tobias E. Schrader,*,§ Andreas Ostermann,∥ Leighton Coates,*,‡ Mikael Akke,*,‡ Matthew P. Blakeley,*,§ Esko Oksanen,†,∨ and Derek T. Logan*,†

†Department of Biochemistry & Structural Biology, Lund University, S-221 00 Lund, Sweden
‡Department of Biophysical Chemistry, Lund University, S-221 00 Lund, Sweden
§Jülich Centre for Neutron Science (JüCNS) at Heinz Maier-Leibnitz Zentrum (MLZ), Forschungszentrum Jülich GmbH, Lichtenbergstrasse 1, 85747 Garching, Germany
∥Heinz Maier-Leibnitz Zentrum (MLZ), Technische Universität München, Lichtenbergstrasse 1, 85748 Garching, Germany
→Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, Tennessee 37831, United States
§Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France
∨Instrument Division, European Spallation Source ERIC, Box 176, S-221 00 Lund, Sweden

ABSTRACT: The medically important drug target galectin-3 binds galactose-containing moieties on glycoproteins through an intricate pattern of hydrogen bonds to a largely polar surface-exposed binding site. All successful inhibitors of galectin-3 to date have been based on mono- or disaccharide cores closely resembling natural ligands. A detailed understanding of the H-bonding networks in these natural ligands will provide an improved foundation for the design of novel inhibitors. Neutron crystallography is an ideal technique to reveal the geometry of hydrogen bonds because the positions of hydrogen atoms are directly detected rather than being inferred from the positions of heavier atoms as in X-ray crystallography. We present three neutron crystal structures of the C-terminal carbohydrate recognition domain of galectin-3: the ligand-free form and the complexes with the natural substrate lactose and with glycerol, which mimics important interactions made by lactose. The neutron crystal structures reveal unambiguously the exquisite fine-tuning of the hydrogen bonding pattern in the binding site to the natural disaccharide ligand. The ligand-free structure shows that most of these hydrogen bonds are preserved even when the polar groups of the ligand are replaced by water molecules. The protonation states of all histidine residues in the protein are also revealed and correlate well with NMR observations. The structures give a solid starting point for molecular dynamics simulations and computational estimates of ligand binding affinity that will inform future drug design.

INTRODUCTION

The specific binding of carbohydrate binding proteins (lectins) to glycoproteins or glycolipids directs a large variety of cellular processes, such as adhesion to other cells and trafficking of intracellular components. Hence, there is great interest in understanding the detailed molecular mechanisms governing the binding specificity between lectins and their cognate carbohydrate chains. This binding generally involves many hydrogen bonds, whose directionality can be difficult to characterize, as well as many indirect and dynamic interactions via water molecules.

Galectin-3 is a mammalian protein belonging to the galectin family, which has 15 members in mammals and is defined by affinity for glycans containing β-D-galactoside moieties. The protein is found both in the nucleus and the cytoplasm of cells, and it is also secreted extracellularly, where it interacts with β-D-galactoside-containing glycoproteins and glycolipids. Galectin-3 consists of two domains. The carbohydrate-recognition domain (CRD) is C-terminal, while the N-terminal domain is involved in oligomerization, probably through formation of a coiled-coil structure, a characteristic that is unique among the galectin family proteins.

Recently, human galectin-3 has emerged as an interesting pharmaceutical target, owing to its involvement in various diseases, including inflammation, cancer proliferation and metastasis, and diabetes. Many X-ray crystal structures of the C-terminal domain of galectin-3 (galectin-3C) in complex

Received: January 17, 2018
Published: April 19, 2018

DOI: 10.1021/acs.jmedchem.8b00081
J. Med. Chem. 2018, 61, 4412−4420
with synthetic ligands have been determined as a key tool in the structure-based design of new inhibitors.8–16 The most successful candidates are derived from a core based on the natural substrate lactose, or the very similar dithiogalactoside core, and their sugar moieties conserve the same interactions with the protein residues as lactose. Enhanced affinity has instead been gained by the exploration of additional binding pockets without disturbing the disaccharide core. The structural basis for this inability to vary the core interactions is of fundamental interest. In a wider context, we are aiming to understand the basic principles governing protein–ligand interactions, in terms of absolute and relative free energy differences upon ligand binding, entropy–enthalpy correlation, and water dynamics, through the characterization of many novel, systematically varied synthetic inhibitors using a variety of structural and biophysical methods. A detailed experimentally derived description of the geometry of the hydrogen bonds in the ligand binding site of galectin-3 can explain why it is so difficult to tamper with the disaccharide core. Experimentally determined hydrogen atom positions will provide an unambiguous foundation for molecular dynamics simulations, free energy perturbation calculations, and quantum chemical calculations that inform drug design.

The relatively polar, surface-exposed ligand binding site of galectin-3 presents several residues that interact with the ligands through hydrogen bonds (Figure 1). Direct visualization of the H atoms using X-ray crystallography requires very high resolution data, usually beyond 1.1 Å,17 and the visibility of H atoms depends strongly on the mobility of the heavy atoms to which they are attached, as measured by the atomic displacement parameters (ADPs).18 Even in very high-resolution crystal structures at <0.7 Å resolution, only about half of the most ordered H atoms can typically be observed.18

In contrast, neutron crystallography is a powerful experimental technique for providing this information even at more modest resolution. The hydrogen bonding patterns of water molecules are particularly hard to probe with X-ray crystallography, as they are often more mobile than the protein atoms, but an elucidation of water H-bonding is also fundamental to an understanding of the enthalpic and entropic effects of expulsion of ordered water molecules from the binding site. Here again, neutron crystallography is an invaluable tool.

Recently, we presented a method to obtain large crystals of perdeuterated galectin-3C (up to 1.8 mm3) in various states, allowing collection of neutron diffraction data up to 1.6 Å resolution.19 Here, we present the first three neutron crystal structures of galectin-3C, involving unbound (apo) galectin-3C, as well as complexes with the natural ligand lactose and the ligand mimic glycerol, which essentially corresponds to one-half of a galactose molecule. These neutron structures provide novel information pinpointing the geometry and directionality of hydrogen bonds and protonation states of titrating residues. The results are fundamental for our understanding of ligand binding to galectin-3 and will serve to guide future inhibitor design.

### RESULTS

We have determined three neutron crystal structures of galectin-3C: the unbound form (apo, 1.75 Å) and the complexes with lactose (1.7 Å) and glycerol (1.7 Å). All three structures were determined from crystals produced at pH = 7.9 (pH 7.5). Data quality statistics for the lactose and glycerol complexes have been presented previously,19 and statistics for the apo-galectin-3C data set are given in Table 1. All structures were jointly refined against the neutron data and room-temperature X-ray data collected from the same crystal, to much higher resolution. The room-temperature apo X-ray

![Figure 1. Overview of the lactose binding site in galectin-3C. Galectin-3C is shown as a gray cartoon, with the side chains of important residues in the lactose recognition site shown as lines. Lactose is represented by sticks, with the carbon atoms colored yellow and oxygens colored red.](image-url)

| Table 1. X-ray and Neutron Diffraction Data Quality for the Crystal of Apo Galectin-3C Used in This Study |
|---------------------------------|-----------------|-----------------|
| neutrons | X-rays | instrument/beamline | wavelength(s) (Å) | detector | resolution | space group |
| LADI-III | 3.35–4.35 | image plate | 30–1.75 (1.78–1.75) | P212121 |
| ESRF BM30 | 0.98081 | ADSC Q315r | 28–1.03 (1.06–1.03) |
| cell dimensions a, b, c (Å) | 37.2, 58.4, 63.8 | Rmerge (I) (%) | 13.3 (19.0) |
| | | Rfree (I) (%) | 5.3 (11.1) |
| | | CC(1/2) (I) (%) | 0.992 (0.944) |
| | | mean I/σ(I) | 8.3 (4.6) |
| | | completeness (%) | 82.1 (62.1) |
| | | no. unique reflections | 11790 (485) |
| | | multiplicity | 4.9 (3.1) |
| | | Rmerge = 5.2% and multiplicity 13.0. The glycerol data are from the monochromatic instrument BIODIFF at FRM-II, Munich, and are 94.7% complete to 1.65 Å with Rmerge = 8.7% and multiplicity 3.1. The glycerol data set was limited to 1.7 Å for refinement. |

DOI: 10.1021/acs.jmedchem.8b00081
structure was also built and refined to the highest resolution of the X-ray data since these extended to higher resolution (1.03 Å) than what we had achieved before (1.25 Å). The refinement statistics are consistent with the resolution and data quality (Table 2). The backbone traces of the three structures are essentially identical, and none of the protein side chains interacting with ligands differ in conformation between the three structures.

### Hydrogen Bonding in the Lactose Complex

The 2m|F|-|D|F| nuclear density map, calculated on-the-fly in PyMOL, is shown as a gray mesh contoured at a level of 0.6σ. The lactose molecule is shown as thick sticks; protein side chains and water molecules that interact with them and lactose are shown as thin sticks. Deuterium atoms are colored yellow; hydrogen atoms on the nonperdeuterated lactose molecule are colored white. Hydrogen bonds relevant for the discussion are shown as dotted lines with distances between the H atom and the H-bond acceptor. This view is chosen to highlight the interactions made by the outer hydroxyl groups of lactose, as well as Arg162. Another view of the binding pocket, rotated approximately 90° around the horizontal axis with respect to panel A. This view is chosen to highlight the interactions of GalOH4, GalOH6, and GlcOH3' with His158, Gln174, Glu184, and Arg186 on the inner side of the binding pocket. Arg162 is not shown for clarity.

### Table 2. X-ray and Neutron Data and Structure Quality for Four Galectin-3C Structures

<table>
<thead>
<tr>
<th></th>
<th>lactose</th>
<th>glycerol</th>
<th>apo</th>
<th>apo (X-ray only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>resolution range neutron</td>
<td>28.0–1.7 (1.83–1.7)</td>
<td>48.4–1.7 (1.81–1.7)</td>
<td>28.1–1.75 (1.94–1.75)</td>
<td></td>
</tr>
<tr>
<td>resolution range X-ray</td>
<td>32.0–1.1 (1.12–1.1)</td>
<td>43.1–1.1 (1.12–1.1)</td>
<td>29.3–1.03 (1.04–1.03)</td>
<td></td>
</tr>
<tr>
<td>unit cell dimensions a, b, c (Å)</td>
<td>37.0, 58.4, 63.8</td>
<td>37.3, 58.4, 63.9</td>
<td>37.4, 58.6, 64.0 12.8</td>
<td></td>
</tr>
<tr>
<td>Wilson B-factor (Å², from X-ray data)</td>
<td>12.5</td>
<td>11.5</td>
<td>15.8</td>
<td></td>
</tr>
<tr>
<td>no. of reflections used in refinement, neutron</td>
<td>15194 (2778)</td>
<td>15178 (2266)</td>
<td>11090 (1695)</td>
<td></td>
</tr>
<tr>
<td>no. of reflections used in refinement, X-ray</td>
<td>56662 (2773)</td>
<td>54348 (2431)</td>
<td>70166</td>
<td></td>
</tr>
<tr>
<td># of reflections in free set, neutron</td>
<td>769 (135)</td>
<td>782 (112)</td>
<td>697 (106)</td>
<td></td>
</tr>
<tr>
<td># of reflections in free set, X-ray</td>
<td>2815 (146)</td>
<td>2751 (128)</td>
<td>4639 (145)</td>
<td></td>
</tr>
<tr>
<td>Rwork (X-ray)</td>
<td>0.107 (0.177)</td>
<td>0.123 (0.253)</td>
<td>0.128 (0.330)</td>
<td>0.118 (0.339)</td>
</tr>
<tr>
<td>Rfree (X-ray)</td>
<td>0.125 (0.194)</td>
<td>0.135 (0.228)</td>
<td>0.141 (0.330)</td>
<td>0.129 (0.325)</td>
</tr>
<tr>
<td>Rwork (neutron)</td>
<td>0.168 (0.253)</td>
<td>0.152 (0.211)</td>
<td>0.180 (0.238)</td>
<td></td>
</tr>
<tr>
<td>Rfree (neutron)</td>
<td>0.211 (0.328)</td>
<td>0.184 (0.261)</td>
<td>0.210 (0.275)</td>
<td></td>
</tr>
<tr>
<td># of non-D atoms</td>
<td>1364</td>
<td>1314</td>
<td>1276</td>
<td>1395</td>
</tr>
<tr>
<td># of D atoms</td>
<td>1441</td>
<td>1442</td>
<td>1347</td>
<td>1304</td>
</tr>
<tr>
<td>macromolecules</td>
<td>1253</td>
<td>1204</td>
<td>1199</td>
<td></td>
</tr>
<tr>
<td>ligands</td>
<td>lactose</td>
<td>glycerol</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>solvent</td>
<td>111</td>
<td>110</td>
<td>77</td>
<td>142</td>
</tr>
<tr>
<td>protein residues</td>
<td>137</td>
<td>137</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td>rmsd bonds (Å)</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>rmsd angles (deg)</td>
<td>1.7</td>
<td>1.3</td>
<td>1.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>97.1</td>
<td>97.8</td>
<td>97.8</td>
<td>97.8</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>2.9</td>
<td>2.2</td>
<td>2.2</td>
<td>1.5</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>rotamer outliers (%)</td>
<td>0.0</td>
<td>1.5</td>
<td>0.0</td>
<td>1.4</td>
</tr>
<tr>
<td>all-atom clash score</td>
<td>4.0</td>
<td>5.8</td>
<td>1.6</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Figure 2. (a) Complex of perdeuterated galectin-3C with lactose at 1.7 Å resolution. The 2m|F|−|D|F| nuclear density map, calculated on-the-fly in PyMOL, is shown as a gray mesh contoured at a level of 0.6σ. The lactose molecule is shown as thick sticks; protein side chains and water molecules that interact with them and lactose are shown as thin sticks. Deuterium atoms are colored yellow; hydrogen atoms on the nonperdeuterated lactose molecule are colored white. Hydrogen bonds relevant for the discussion are shown as dotted lines with distances between the H atom and the H-bond acceptor. This view is chosen to highlight the interactions made by the outer hydroxyl groups of lactose, as well as Arg162. (b) Another view of the binding pocket, rotated approximately 90° around the horizontal axis with respect to panel A. This view is chosen to highlight the interactions of GalOH4, GalOH6, and GlcOH3’ with His158, Gln174, Glu184, and Arg186 on the inner side of the binding pocket. Arg162 is not shown for clarity.
referring specifically to the oxygen atoms. The three most important polar groups are the 4-OH and 6-OH hydroxyl groups of galactose (GalOH4 and GalOH6) and the O3′ hydroxyl group of glucose (GlcOH3′). The GalO4 oxygen atom accepts an H-bond from NH2 of Arg162 (2.09 Å) and its hydrogen atom is directed toward NE2 of His158 (2.00 Å), which is singly protonated on ND1. GalO6 accepts an H-bond from the amino group of Asn174 (1.81 Å), and the hydrogen atom of GalO6H interacts with the carboxylate group of Glu184 (1.86 Å). GlcO3′ accepts one H-bond from atom NH1 of Arg162 (1.88 Å), and GlcOH3′ donates its H atom to the deprotonated carbonyl group of Glu184 (1.72 Å). All distances in parentheses are from the hydrogen atom to the acceptor atom. Finally, the O5 atom of galactose (GalO5S), enclosed within the sugar ring, accepts an H-bond from atom NH2 of Arg162 (2.21 Å).

Two other polar hydrogen atoms on the outer face of lactose are clearly visible, although they do not directly build H-bonds with the protein, but rather with surrounding water molecules. In particular, we see that the outward-pointing GalO3 accepts an H-bond from a water molecule (1.96 Å), which simultaneously accepts an H-bond from NH2 of Arg144 (1.88 Å), thus bridging an interaction between lactose and the protein. GalOH3 donates its H atom to a water molecule (2.09 Å), which interacts with other crystallographic water molecules further out of the binding site through an ordered H-bond network. The GlcOH2′ hydroxyl group is also clearly oriented, though it does not H-bond to any protein residue or water molecule. In contrast, nuclear density for the OH6′ group of glucose is not well-defined.

Interestingly, binding of lactose to the protein apparently leads to minor conformational strain within lactose. The
geometry imposed by H-bonding to His158 puts the hydroxyl hydrogen atom in GalOH4 into an energetically unfavorable eclipsed conformation (dihedral angle 5°) with respect to the aliphatic hydrogen atom on C4 (Figure 2b). In contrast, the GalOH6 and GlcOH3′ hydrogen atoms are almost completely staggered (dihedral angles 180° and 172°, respectively). On the outer face of lactose, the GalOH3 group is in a gauche conformation (dihedral 43°), and the GalOH2 group is staggered (164°).

**Glycerol Closely Mimics the Inner Side of Lactose.** In the lactose complex, the six heavy atoms of glycerol completely overlap the inner part of the galactose moiety of lactose (Figure 3 and Figure S1b), as previously observed by ourselves and others. However, the present study allows localization of the hydrogen atoms for the first time. Both aliphatic and polar D atoms of the di-glycerol molecule are visible; thus, there are no cancellation effects from the aliphatic H, as in the lactose complex. The interactions of glycerol with galectin-3C are essentially identical to those of galactose: O1 (equivalent to GalO4 in lactose) donates its H atom to NE2 of His158 (1.84 Å) and accepts an H-bond from NH2 of Arg162 (2.14 Å); O3 (equivalent to GalO6) donates its H atom to OE2 of Glu184 (1.82 Å) while accepting an H-bond from ND2 of Asn174 (1.88 Å), with the same geometry as in the lactose complex. Interestingly, the H atom in the remaining hydroxyl group (HO2; equivalent to the ring oxygen GalO5 in lactose) has a very clear directionality, despite not donating an H-bond to any protein functional group or water molecule. This appears to be due to the fact that O2 accepts an H-bond from NH2 of Arg162 (2.17 Å), placing tight restraints on the orbital orientations in O2. The angle between HH22 on Arg162, O2, and HO4 on glycerol is 111°, consistent with very good overlap of the lone pairs on O4 with the σ orbitals of HH22 (Figure 3a).

The position of GlcOH3′ in the lactose structure is occupied in the lactose complex by a well-ordered water molecule, which accepts an H-bond from NH1 of Arg162 (1.86 Å) and donates one to OE2 of Glu184 (1.68 Å). This arrangement strongly mimics the interaction made by GlcOH3′.

**Three Water Molecules in the Apo Form Replace Key H-Bonding Motieties in Lactose.** In the apo (ligand-free) neutron crystal structure at 1.75 Å resolution, only three water molecules are clearly visible in the binding site in the $2\text{m}F_o - D\text{IF}_o$ neutron map (Figure 4 and Figure S1c). The water molecules are bound at the positions of the inner oxygen atoms of bound lactose. Again, this has been observed previously, but hydrogen atom positions are identified for the first time here using neutron diffraction. The GalOH4 position is occupied by W1, GalOH6 by W2, and GlcOH3′ by W3. The isotropic B-factors of the three water molecules are similar (52.0, 41.9, and 47.5 Å$^2$, respectively). At 1.03 Å, the full resolution of the X-ray data, a further three water molecules are observed at lower electron density levels (down to 0.6e) at the positions of GalC6, GalO1, and GalO5 (Figure 4d). These three water molecules were also seen in our previous X-ray structure of apo-galectin-3C at 1.25 Å. The water at GalC6 appears to be a low-occupancy alternative conformation for W2. Further two water molecules were seen in the binding site in the 1.25 Å structure, but these may have been due to traces of lactose in the protein preparations, as noted previously. Here, we will describe only the waters visible in the neutron density maps.

The H-bonding directionality of W2 is clearly determined (Figure 4b), showing a nuclear density peak tilted toward Glu184, similarly to the hydroxyl group GalO6 of lactose. W2 accepts an H-bond from Asn174 (1.95 Å) and donates one to Glu184 (2.47 Å). W3 also shows clear nuclear density, with H-bond orientation similar to that in the glycerol and lactose complexes. It accepts an H-bond from Arg162 (1.90 Å) and donates one to Glu184 (1.81 Å; Figure 4a,b). The nuclear density for the deuterium atoms on W1 is weaker; nevertheless, it is possible to model an orientation in which a single, relatively long H-bond (2.33 Å) is made to His158 (Figure 4a,b). The second proton appears disordered. This water molecule also accepts a hydrogen bond from Arg162 (2.16 Å), and it is clearly visible in the X-ray map (Figure 4c), albeit that it refines with significant anisotropy, disordered in the imidazole plane of His158. Its lack of clarity in the nuclear density map may be due to a combination of this mobility and limitations in the neutron diffraction data.

Interestingly, the hydrogen bond of W2 rotates by about 45° away from OE2 compared to the other two structures, lengthening the H-bond from 1.90 Å in the lactose complex and 1.82 Å in the glycerol complex to 2.47 Å. Likewise, the H-bond of W3 rotates away from Glu184, but the H-bond distance stays approximately the same, at 1.81 Å compared to 1.72 Å in lactose. The H-bond orientation of the corresponding water molecule in the glycerol complex is more similar to that in lactose. This indicates higher conformational freedom in the mobile water molecules than in the ligands.

**Protonation States of Histidine Residues: Correlation between NMR and Neutron Observations.** Histidine side chains have two nitrogen atoms that can be protonated, ND1 and NE2. The pK$\alpha$ values of these atoms are similar (6.6−7.0) and close to neutral pH. Therefore, there are three possible tautomeric states: the H−NE2 tautomer, which is the most stable noncharged state; the rare H−ND1 tautomer, which typically occurs only when stabilized by specific interactions; and the positively charged state, in which both NE2 and ND1 are protonated. It is necessary to determine the protonation state in order to understand function and to provide correct starting models for computational studies. The combination of neutron crystallography and NMR spectroscopy provides a unique opportunity to determine the protonation state of His residues in a protein. The NMR experiments show that the tautomers present in solution at pH 7.5 (the value used for crystallization was pH 7.9, equivalent to pH 7.5) agree well with those observed in the neutron structures (Table 3).
There is complete agreement between all neutron structures and the NMR experiments for His158 in the carbohydrate binding site (Figure S2a) and for His223 (Figure S2b). Both methods unequivocally show that His158 is singly protonated on the ND1 atom, whereas His223 is singly protonated on NE2. The presence of this unusual H-ND1 tautomer is consistent with the fact that NE2 of His158 is a critical H-bond acceptor for HO4 of lactose. Furthermore, the rare tautomer is stabilized by a hydrogen bond to the unprotonated carboxylate side chain of Asp148. His223 is located far from the binding site. Its protonation state is determined by an H-bond from NE2 to the main-chain carbonyl group of Glu205 (Figure S2b). This interaction holds together two loops at the end of the β-sandwich. ND1 of His223 is the H-bond acceptor for the end of a chain of hydrogen-bonded water molecules.

For His208, the NMR experiments show experimentally that its pKₐ is 6.20, which yields a population of the charged state of 5% at pH 7.5. Furthermore, both neutral tautomers are also present, with a major population of NE2 and minor population of ND1. Thus, all three tautomers are present in solution. These observations are augmented by the neutron crystal structures, which show that this residue has two distinct conformations in all structures (Figure S2d), which refine well with approximately equal occupancy. In both conformations, NE2 appears to be protonated, but the overlap of the two conformations prevents us from unambiguously determining the protonation state of the ND1 atom, or indeed which of two possible orientations related by an 180° flip the side chains have. In one of the conformations, NE2 donates a hydrogen bond to the unprotonated carboxylate side chain of Glu205; in the other, it makes no clear interactions. ND1 makes no interactions to surrounding residues in either conformation.

For His217, the NMR results indicate the presence of all three tautomers: one major state with NE2 protonated, a lower-populated state with ND1 protonated, and a very small population (1%) of the charged state. In the lactose neutron crystal structure, His217 appears to be doubly protonated. If this structure is refined with His217 protonated only on the NE2 atom, a positive 2mIFₒ – DIFᵣ difference density peak is seen at the position of the HD1 atom. However, the structure refines equally well with the deuterium atoms at full or half occupancy; thus, the neutron data are not sufficient to quantify the relative populations. In contrast to His208, His217 accommodates the different tautomers without any changes in side chain orientation. The results for the apo and glycerol neutron crystal structures differ from lactose: the residue appears to be protonated only on NE2. In the apo structure, there are some indications of partial protonation of ND1 (Figure S2c), which is also seen in the 0.86 Å X-ray structure at 100 K. In summary, the NMR results broadly agree with the distribution seen in the neutron structures in that the dominant species is protonated on NE2.

DISCUSSION

The three neutron crystal structures of galectin-3C presented here identify for the first time experimentally the positions of key hydrogen atoms in ligand complexes and in the apo protein, thereby highlighting the exquisite fine-tuning of H-bonding interactions in the carbohydrate binding site to exploit all available interaction possibilities in the disaccharide core and providing experimental support for geometry that could only be inferred until now. Even in the highest-resolution X-ray structure obtained to date, the 0.86 Å structure of the lactose complex, the H atom positions are much more ambiguous. The protonation state of His158 is discernible, and some of the aliphatic H atoms on the galactose moiety are visible, but the only well-defined H-bond is the one between GalOH4 and His158. Thus, the combination of neutron and X-ray crystallography, the latter to very high resolution, gives a powerful combination of accuracy in the heavy atom positions while allowing unambiguous localization of hydrogen.

For the most part, the important H-bonds are achieved while maintaining an undistorted, low energy conformation of the ligands, with gauche or staggered conformations around the C–O bond in the ligand hydroxyl groups. The exception is the H-bond made between GalOH4 and His158 in which the hydrogen atom is in an eclipsed geometry with respect to the other end of the C–O bond.

As previously observed, the non-natural ligand glycerol closely mimics the inner side of the more tightly bound of the two sugar moieties in lactose, namely, galactose. The H-bonds made by the OH4 and OH6 groups of glycerol are remarkably similar to those made by GalOH4 and GalOH6 in lactose, in both cases showing high directionality and suggesting that it would be hard to replace these moieties with functional groups that make less directional interactions with the protein, such as halogens, while maintaining tight binding of synthetic inhibitors. The results confirm that inhibitor design has been wise to focus on exploitation of additional nearby binding pockets on the surface of galectin-3 rather than modification of the core carbohydrate binding moieties. The lactose and glycerol complexes also pinpoint for the first time a critical H-bond donation from an arginine side chain (Arg162) to the cyclic oxygen atom O5 of the core recognition motif, namely, the galactose moiety in lactose. Despite not H-bonding to any external group, the H atom on O4 of glycerol is highly directional, emphasizing the necessity to orient the p-orbitals of O4, and thus GalO5 in lactose, toward Arg162. This is a further demonstration of the detailed fine-tuning of protein side chain interactions to exploit every possible H-bonding possibility on the carbohydrate core.

The neutron structures described here have highlighted the central role of Arg162 in maintaining the conserved H-bond network, as it participates in three highly directional H-bonds to lactose. These results are in agreement with NMR experiments, which show that the guanidine group of this residue is not more flexible than the backbone amide groups.

There is excellent agreement between the protonation states of galectin-3C histidine residues observed by neutron crystallography and by NMR. Of these, the functionally most interesting is His158 in the binding site. Both methods agree that the rare ND1-protonated tautomer of His158 is crucial for the recognition of the galactose O4 hydroxyl group. This tautomer is stabilized by an H-bond from ND1 to the unprotonated Asp148.

In the unbound form of galectin-3C, three water molecules occupy the three important sites corresponding to the inner oxygen atoms of bound lactose, which are critical for ligand recognition. The neutron crystal structure shows that two of these water molecules, namely, W2 and W3, build highly directional H-bonds with the protein. Interestingly, these waters, in particular W3, are both replaced in the natural ligands by moieties that are able to build a strong and directional hydrogen bond, while still keeping the interaction with the positive NH donor. It is likely that the release of these waters upon ligand binding is associated with a favorable
change in entropy. W1, although located in the same position as the GalO4 oxygen atom, shows a less well-defined neutron density. This water molecule is not more mobile than the others, but anisotropic atomic displacement parameters indicate that it moves in the same plane as the ring of His158. This suggests that W1 prefers to build an electrostatic interaction with His158, perhaps sacrificing strict H-bond directionality to gain entropy. Similar observations of buried water molecules with high disorder have been reported previously and have been related to increased entropy upon binding, compared to bulk water.23 This result suggests that it might be advantageous not to occupy this water site by ligand oxygens in order not to trigger an entropically unfavorable water release.

**CONCLUSION**

Taken together, the present results provide a detailed picture of the critical H-bonds involved in ligand binding for galectin-3. The high directional specificity of H-bonding interactions in the shallow, exposed binding site of galectin-3 explains the success of disaccharide templates (e.g., thiogalactoside)6,13–16,24,25 as the basis for all potent galectin-3 inhibitors to date and suggests that it moves in the same plane as the ring of His158. This suggests that W1 prefers to build an electrostatic interaction with His158, perhaps sacrificing strict H-bond directionality to gain entropy. Similar observations of buried water molecules with high disorder have been reported previously and have been related to increased entropy upon binding, compared to bulk water.23 This result suggests that it might be advantageous not to occupy this water site by ligand oxygens in order not to trigger an entropically unfavorable water release.

**EXPERIMENTAL SECTION**

Crystallization, Data Collection, and Refinement. Protein expression, perdeuteration, purification, crystallization, generation of the apo form, data collection, and data quality statistics for the lactose and glycerol complexes have been described previously.19 A crystal of the apo protein measuring around 1.0 mm was mounted in a thick-walled quartz capillary (Vitrocom; from CM Scientific, Silsden, UK), with internal diameter of 1.5 mm. A plug of deuterated crystallization mother liquor was left at one end the capillary. Data to 1.7 Å were collected at the LADI-III instrument of the Laue-Langevin, Grenoble, France.20 Twenty-six images with 4° exposures were taken in three different crystal orientations. The wavelength range was 2.74–3.57 Å. The crystal was rotated by 7° between exposures, and the total data collection time was 4.5 days. Data were indexed and integrated using LAUEGEN,27 wavelength-normalized using LSCALE,28 and scaled and merged using SCALA.29 Room-temperature X-ray data were obtained to 1.03 Å resolution from the same crystal at beamline BM30 of the ESRF, Grenoble, France, using a beam size of 0.3 × 0.3 mm. Data were collected in two passes: first, a low-dose pass to 1.3 Å with 250 × 1° rotations and exposure time of 0.5 s, followed by a high-resolution pass to 1.02 Å with 260 × 1° rotations and exposure time of 4 s. The X-ray data were integrated using XDS and merged using XSQALE.30 Merging of MTZ files and other file manipulations were done using programs from the CCP4 package.31 Data quality statistics are presented in Table 1.

All structures were refined jointly against neutron and X-ray data with phenix.refine.32 The starting model for all refinements was a lactose-bound galectin-3C X-ray structure at 0.86 Å resolution,20 omitting water molecules and the ligand. The ligand molecules and restraints were built using the elBoW tool in Phenix.33 The structures, without water molecules but with the ligand (if present), were briefly refined using only X-ray data at the highest resolution of the neutron data until the Rmerge, Rfree, and geometry/atomic displacement parameter weights converged. After this, joint refinement with X-ray and neutron data was done using phenix.refine.34 Oxygen atoms of water molecules were added manually by inspecting only the electron density maps in Coot.35 Waters were added if visible in the electron density map at the neutron data resolution, while those waters that were visible only at higher resolution were not added. When satisfactory Rmerge and Rfree values were obtained, we added the remaining X-ray to the highest possible resolution, which enabled refinement with anisotropic ADPs, even for the D atoms. This resulted in reduction of the neutron Rfree values, demonstrating the validity of this approach. To avoid bias toward the X-ray data due to the much higher number of X-ray reflections, a careful screening of relative weights for the X-ray and neutron terms was carried out. We found that the use of anisotropic ADPs, made possible by using all available X-ray data to the highest possible resolution, produced the best models, with a significant improvement in the Rfree value for the neutron data to 1.7–1.8 Å compared to using only isotropic ADPs. An extensive screening of the relative X-ray and neutron weights for coordinate and ADP refinement showed that there was no risk of overfitting the neutron data, so we chose to weight X-rays and neutron data equally, as is done by default in phenix.refine. Deuterium atoms were added in all positions and refined as individual atoms with bond lengths restrained to ideal values from neutron crystal structures. Finally, since the current apo structure is the highest resolution one to date at room temperature, we also refined the model to the full resolution of the X-ray data, adding some water molecules that were not observed in the neutron maps. In this case, the deuterium atoms were refined in riding positions. Refinement statistics are presented in Table 2. The crystal structures, X-ray, and neutron diffraction data have been deposited in the Protein Data Bank with accession IDs 6EYM (lactose complex), 6EXY (glycerol complex), and 6F2Q (ligand-free).

**NMR Spectroscopy and Determination of Histidine Protonation States.** The NMR samples contained 0.3 mM uniformly 13C/15N-labeled galectin-3C and 150 mM lactose in 5 mM HEPES buffer, originally at pH 7.4, with 5% D2O added for the field-frequency lock. Constant-time1H−13C HSQC experiments were carried out at a static magnetic field strength of 11.7 T and temperature of 301.15 K on an Agilent DirectDrive VNMRS spectrometer. Spectra were acquired with spectral widths of 8012 and 3750 Hz in the 1H and 13C dimensions, sampled over 2240 and 62 points, and with the carriers positioned at 4.75 and 125.9 ppm, respectively. NMR data were processed using NMRPipe36 and analyzed using CCPNM Analysis 2.4.2.37 Spectra were acquired at 15 different pH values, ranging from 5.3 to 8.3. Two samples were used: one sample was titrated down in pH from 7.4 to 5.3 in 12 steps, whereas the other one was titrated up to pH 8.3 in five steps. Acid (0.1 M HCl) or base (0.1 M NaOH) was added in aliquots of 1 μL to adjust the pH at each titration point. The variation in chemical shift of the histidine 13CE1 resonances (δHN) was monitored as a function of pH and fitted, using the Levenberg–Marquardt algorithm38 in MATLAB R2016 (MathWorks), to the following equation:

\[
\delta_{HN} = \delta_{HN}^{\text{ref}} + \frac{\delta_{HN}^{\text{ref}} - \delta_{HN}^{\text{ref}}} {1 + 10^{[(pH - pK) / \alpha]}}
\]

where δHN and δA are the chemical shifts of the charged and noncharged states, respectively. δHN and δA are free parameters of the fit.

**ASSOCIATED CONTENT**

 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.8b00081.

Neutron diffraction analysis of protonation state of His residues in galectin-3C. Schematic diagram of hydrogen bonding patterns in the three neutron crystal structures of galectin-3C (PDF)
1443. recognition domain ADP, atomic displacement parameter; CRD, carbohydrate ■ ABBREVIATIONS USED

We thank Nicolas Coquelle (Institut Laue Langevin) and Jean-Luc Ferrer (Institut de Biologie Structurale) for help with room-temperature X-ray data collection on the crystal of apo-galectin-3C at the BM30 beamline of the ESRF and Ulrich Weininger (now at Universita Galectin-3 at the BM30 beamline of the ESRF and Ulrich room-temperature X-ray data collection on the crystal of apo-


Notes The authors declare no competing financial interest. ○Francesco Manzoni passed away on 12 March 2017.

ACKNOWLEDGMENTS

We thank Nicolas Coquelle (Institut Laue Langevin) and Jean-Luc Ferrer (Institut de Biologie Structurale) for help with room-temperature X-ray data collection on the crystal of apo-galectin-3C at the BM30 beamline of the ESRF and Ulrich Weininger (now at Universitat Halle-Wittenberg) for help with the NMR experiments. We thank the Institut Laue-Langevin for provision of instrument time at LADI-III. Research at the Spallation Neutron Source (SNS) at ORNL was sponsored by the Scientific User Facilities Division, Office of Basic Energy Sciences, U.S. Department of Energy. The work was partly funded by a Ph.D. fellowship from the European Spallation Source ERIIC to F.M., by a grant from the Swedish Research Council (2016-04855) to D.L., and from the Knut and Alice Wallenberg Foundation (KAW 2013.0022) to D.L., M.A., and E.O. This paper is dedicated to the memory of Francesco Manzoni, who sadly passed away on 12 March 2017.

ABBREVIATIONS USED

ADP, atomic displacement parameter; CRD, carbohydrate recognition domain

REFERENCES


