Conformational Differences in Liganded and Unliganded States of Galectin-3[†]

Kimiko Umemoto,[‡] Hakon Leffler,[§] Andre Venot,[||] Homay Valafar,[||] and J. H. Prestegard*,[||]

Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, Department of Chemistry, International Christian University, Tokyo, Japan 1818585, and Section MIG (Microbiology, Immunology, Glycobiology), Institute of Laboratory Medicine, Lund University, Lund, Sweden S-22362

Received August 19, 2002; Revised Manuscript Received December 9, 2002

ABSTRACT: The conformation of the carbohydrate recognition domain of Galectin-3, a lectin known to bind galactose containing oligosaccharides in mammalian systems, has been investigated in the absence of ligand and in the presence of *N*-acetylactosamine. A new methodology based on the measurement of residual dipolar couplings from NMR spectra has been used to characterize differences in protein structure along the backbone in the presence and absence of ligand, as well as the binding geometry of the ligand itself. The data on the ligand are consistent with the ligand binding geometry found in a crystal structure of the complexed state. However, a significant rearrangement of backbone loops near the binding site appears to occur in the absence of ligand. The implications for ligand specificity and protein functionality are discussed.

The binding of ligands by proteins is fundamental to processes that range from enzyme catalyzed reactions to signaling in response to effectors of cellular function. In each case a molecular level description of both the liganded and unliganded state of the protein is critical to an understanding of function. Today atomic-level structures of many proteins, in some form, are known through the efforts of X-ray crystallography or NMR spectroscopy, but it is often the case that structures in both the presence and absence of particular ligands of interest are not available. Methods that can use available structural information to efficiently extend definition to the missing states are of particular utility. Here we use residual dipolar couplings between $^{15}N-^{1}H$ amide pairs along the backbone of the carbohydrate recognition domain (CRD) of the human lectin, galectin-3, to gain a rapid assessment of the relationship of a crystal structure to a structure in solution and the relationship of the crystal structure, which contains the disaccharide ligand, *N*-acetyllactosamine (LacNAc), to the structure in solution in the absence of ligand. We also extend this evaluation to the bound geometry of the LacNAc itself, as it exists in solution, using the same technology.

Galectin-3 is a member of a protein family defined by carbohydrate recognition domains (CRDs) of about 130 aa with affinity for β-galactoside-residues and conserved sequence elements. Galectin-3 has one CRD at the C-terminus, whereas the rest of the molecule has an unrelated structure. Other galectins can contain one or two CRDs in different arrangements; 14 galectins have been reported in mammals and genome sequences suggest a few more (*1*−*4*).

The biological function of galectins is not known, but important roles for galectin-3 in inflammation and cancer are supported by many reports (*2*, *5*, *6*). The affinity for β-galactosides, typically found at the cell surface, drew the initial attention to possible galectin functional roles in cell adhesion and in induction of signals from the cell surface (*5*). However, galectins are unusual among lectins in that they have properties more typical of cytosolic than extracellular proteins. They have no secretion signal peptide and reside in the cytosol for long times; they may be targeted for secretion by nonclassical (non-ER-Golgi) pathways but also appear in the nucleus. Recent evidence supports intracellular roles of galectins, e.g., an anti-apoptotic effect of galectin-3 (*7*, *8*). Since β-galactosides are not present in most intracellular compartments, the galectins would have to interact with other ligands in these instances. These facts make the study of galectins in liganded and unliganded states particularly important.

The crystal structures of the galectins-1, -2, -7 and a fragment of galectin-3 (galectin-3C) containing the CRD have been determined in complex with disaccharide ligands (*9*). In all cases they consist of a sandwich with a six-strand sheet (named S1−S6) and a five-strand sheet (F1−F5) and demonstrate a conserved binding site for the disaccharide ligand (lactose or *N*-acetyllactosamine) in a shallow groove over β-strands S4−S6. The β-galactose moieties of these disaccharides are most tightly bound and interact via hydrogen bonds with side chains provided by strands S4−S5 (His-158, Asn-160, Arg-162, and Asn-174 in galectin-3) and in addition via a hydrogen bond with Glu-184 and a stacking interaction between Gal H3−H5 and the side chain of Trp 181 (*10*). The latter two residues are in the first part of β-strand S6 (S6a) that can be regarded equally well as part of the loop 177−184 (between S5 and S6b) because it does not interact extensively with the rest of the protein and deviates from the typical β strand structure. Because of interactions with residues on a potentially deformable loop

[†] This work was supported by NIH grants GM33225 and RR05351 to J.H.P. and grant 12165 from the Swedish Research Council to H.L.

* Corresponding author.
[‡] International Christian University.
[§] Lund University.
[||] University of Georgia.

10.1021/bi026671m CCC: $25.00
Published on Web 03/13/2003

and because hydrophobic residues such as Trp 181will not prefer exposure to solvent in an unliganded state, it would not be unreasonable to expect this loop, or other residues near this loop, to rearrange in the unliganded structure. Such rearrangements may be important to the kinetics of interaction, the thermodynamics of interaction, and the diverse ligands that seem to be accommodated by this class of protein. Moreover, the 177−184 loop contains the NWGR sequence motif. This sequence is found in Bcl-2 and other apoptosis regulating proteins and has been implicated in the intracellular anti-apoptotic effect of galectin-3 (*7*, *8*). Modulation of exposure of this loop on structural rearrangement may be a part of a regulatory mechanism.

The previous assignment of the NMR resonances for the ^{13}C,^{15}N-labeled galectin-3C provide an essential underpinning for structural investigation of this protein in solution (*11*). Here we exploit residual dipolar coupling data as an efficient means for the investigation of structural departures from existing X-ray structures when the ligand is removed. Residual dipolar coupling data have become an important ingredient of protein structural investigations over the last five years, and numerous applications to both proteins and carbohydrates have been reported (*12*, *13*). In a few cases, applications have even been extended to a study of carbohydrate geometry in protein bound states (*14*−*16*). We will extend applications to *N*-acetyllactosamine bound to galectin-3 in the following study.

Residual dipolar couplings are both easily acquired and exquisitely sensitive to structural variations from known protein structures. They arise when the dipole−dipole interaction between a pair of magnetically active nuclear spins does not average to zero during molecular tumbling. The interaction in terms of a contribution to splitting of resonance lines in Hz can be expressed as in eq 1, where θ is the angle between the interaction vector and the direction of the magnetic field of the spectrometer, r is the internuclear distance, and the γ's are magnetogyric ratios for the particular pair of spins. The angular brackets denote averaging over molecular tumbling. In normal solutions where sampling of orientations is nearly isotropic, the averaging reduces the expression to zero, leaving no contribution that can be measured as a perturbation of resonance frequencies. To restore a measurable contribution, special media are used to promote preferential orientation. In this work we have used an aqueous dispersion of discoidal lipid bilayer fragments (bicelles), doped with a negatively charged amphiphile, to promote order (*17*):

$$D_{ij}^{\mathrm{res}} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij}^{\ 3}}\frac{\langle 3\cos^2\theta_{ij} - 1\rangle}{2} \qquad (1)$$

When the pair of nuclei is directly bonded, the distance r is well defined, and the expression provides primarily angular restraints. One can easily imagine that a restraint on the direction of every amide $^{15}N-^{1}H$ bond vector along a protein backbone would provide a powerful assessment of the validity of any proposed structure, and even a means of quantitatively describing any departures from that structure. $^{15}N-^{1}H$ dipolar couplings are easy to measure because they simply add to splittings of cross-peaks in heteronuclear single quantum coherence (HSQC) spectra when one omits proton decoupling during evolution in the nitrogen domain. Similar arguments hold for $^{13}C-^{1}H$ bond vectors in carbohydrates and their measurement in coupled HSQC spectra. There are, of course, more sophisticated and precise ways of measuring residual dipolar couplings. We have taken the opportunity to compare two of these methods in the course of this work (*18*, *19*).

There are complications in deriving constraints from the simple measurements described above; for example, the anisotropic averaging is quite complex, being dictated by a combination of steric and electrostatic interactions of the protein (or carbohydrate) with the medium. A description of the averaging process normally requires specification of five parameters. These can be taken to be the five independent elements of a symmetric, traceless, three-by-three order matrix, S_{kl}, or they can be taken to be a principal order parameter (S_{zz}), an asymmetry parameter (η), and three Euler angles that define the orientation of a molecule fixed frame with respect to an ordering frame (α, β, γ). We will derive a best-fit set of parameters using a set of $^{15}N-^{1}H$ data and an order matrix description (*20*). Equation 1 can be rewritten to give an expression for each residual dipolar coupling in terms of the molecular frame direction cosines for its interaction vector, $\cos(\alpha_k)$, and the order matrix elements to give eq 2. Given a molecular structure that defines the direction cosines and a set of residual dipolar couplings dominated by conserved parts of the structure, the set of equations can be used to solve for the S_{kl}:

$$D_{ij}^{\mathrm{res}} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij}^{\ 3}}\sum_{kl} S_{kl}\cos(\alpha_k)\cos(\alpha_l) \qquad (2)$$

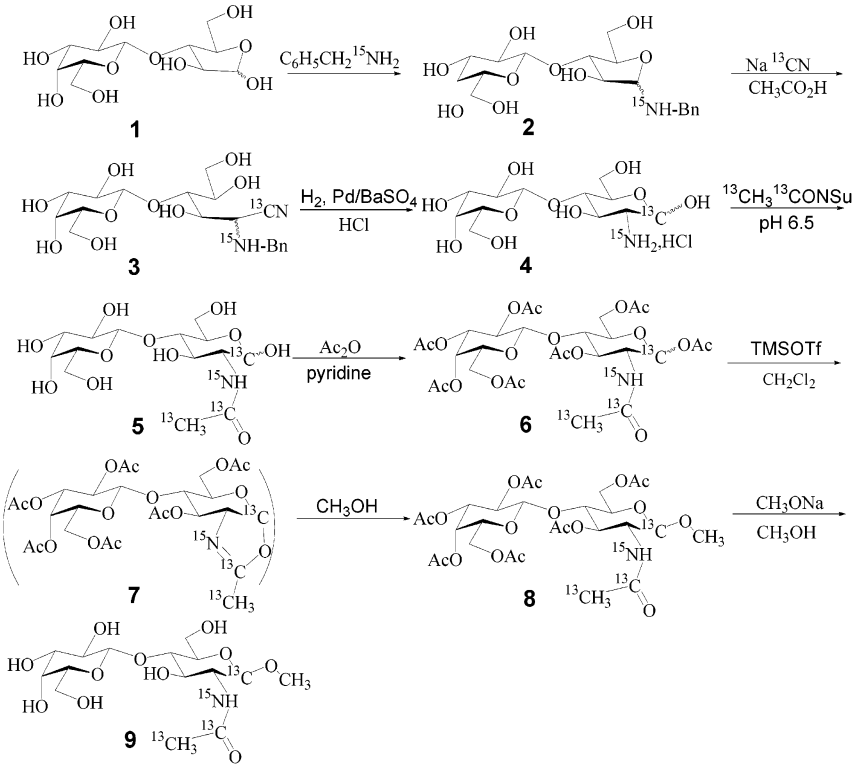
Once the order parameters are known, values for any $^{15}N-^{1}H$ pair can be predicted by using eq 2. Assuming that the majority of data agree with a predicted structure, a simple plot of predicted versus measured values will produce a visual assessment of the accuracy of the model. Points falling on a line of slope one would indicate exact agreement; outliers become targets for further investigation in terms of experimental error or structural deviations.

Dipolar couplings from nuclear spin pairs of ligands bound to protein must reflect the same order and averaging as dipolar couplings from nuclear spin pairs of the protein. In our case the ligand has only one $^{15}N-^{1}H$ pair but additional dipolar information can come from $^{13}C-^{1}H$ pairs. These other interactions also follow equation one but have different internuclear distances and different magnetogyric ratios, γ_{ij}. There is an additional complication here in that ligands of lectins are not normally tightly bound and observed couplings are actually an average of couplings in bound and free states. This makes it essential to measure coupling for both the exchanging ligand-protein complex and the free ligand. With the aid of a molecular dissociation constant, couplings for the bound ligand can be extracted and compared to models in much the same way that couplings for the protein can be compared to couplings predicted for a known X-ray or NMR structure.

MATERIALS AND METHODS

^{15}N-Labeled Galectin-3C. Uniformly ^{15}N-labeled galectin-3 was produced using the expression system described by

Scheme 1: Methyl *O*-(*β*-D-Galactopyranosyl)-(1→4)-2-[1,2-^{13}C,2-^{15}N]-acetamido-2-deoxy-*β*-[1-^{13}C]-D-glucopyranoside **9**



Massa et al. (*21*) and labeling conditions described by Umemoto and Leffler (*11*). The C-terminal fragment (aa 108−250) was generated by collagenase digestion followed by purification on a lactosyl-Sepharose column. Lactose left in the sample after elution with a lactose-containing buffer was removed by repeated ultrafiltration in a Centricon concentrator 10 (Amicon) using a 20mM phosphate buffer at pH 6.8.

Isotopically Labeled LacNAc. Partially double labeled *N*-acetyllactosamine (methyl-*O*-(*β*-D-galactopyranosyl)(1−4)-2-[1,2-^{13}C]-acetyl-[^{15}N]-amino-2-deoxy-*β*-D-[1-^{13}C]-glucopyranoside, was used as the ligand in studies of the complexed protein (Scheme 1). The synthetic procedure used to produce this material was derived from that of Kuhn and Kirschenlohr (*22*, *23*). Several ^{13}C- and ^{15}N-labeled monosaccharide D-glucosamine derivatives have also been synthesized in a similar fashion (*24*−*27*). ^{15}N-benzylamine prepared according to Horneman (*24*) from ammonium-^{15}N nitrate (98% ^{15}N, Sigma) was reacted with 3-O-(*β*-D-galactopyranosyl)-D-arabinose, providing 1-[^{15}N]-benzyl-3-O-(*β*-D-galactopyranosyl)-D-arabinosylamine 2. Intermediate **2** was then reacted with Na^{13}CN (99% ^{13}C, Cambridge Isotope Laboratories) in the presence of acetic acid (*28*) and the reaction product **3** reduced in acidic conditions, providing the [1-^{13}C,2-^{15}N]-lactosamine hydrochloride **4**. After neutralization, this product was reacted with N-[1,2-$^{13}C_2$]-acetoxy-succinimide in water, giving the labeled *N*-acetyl-lactosamine **5**. The succinamide was prepared following procedures described by Heidlas et al. (*29*) using 90% ^{13}C labeled acetic acid (1,2-^{13}C, 99% Cambridge Isotope Laboratories). Purification and acetylation of **5** gave an anomeric mixture **6** of peracetylated derivatives. The *β*-methyl glycoside was obtained from **6** using a trimethylsilyl trifluoromethane sulfonate promoted *β*-glycosidation (*30*) that proceeds through the oxazoline intermediate **7**. The disaccharide **8** was de-O-acetylated and finally purified through Biogel P2, providing **9**.

Bicelle Preparation. Bicelle media were prepared as described previously (*17*). In short, a 10% (w/v) lipid solution was prepared by mixing in a molar ratio of 2.8:1 dimyristoyl-L-α-phosphatidylcholine (DMPC, Sigma, St. Louis, MO) and 1,2-dicaproyl-sn-glycero-3-phosphocholine (DHPC, Avanti Polar Lipids, Birmingham, Al) in 20mM phosphate buffer (pH 6.8 with 10% ^2H_2O). This was mixed with an equal volume of sample solution to produce a 5% (w/v) bicelle preparation. Typical concentrations of protein were 0.1 or 0.2 mM and 0.25 or 0.5 mM ligand. To improve stability in the presence of protein, we found it necessary to add 2 mM SDS to the samples. This medium was used for collection of all data under field alignment conditions. The alignments of the bicelle solutions were checked by monitoring the quadrupole splitting of the water resonance using ^{2}H NMR (10.8 ± 0.2 Hz at 38 °C). Data collection under isotropic conditions was done in bicelle media at 25 °C for the protein containing samples and in nonbicelle containing buffer for the free ligand.

NMR Experiments. Experiments were performed primarily on a Varian INOVA 600 MHz spectrometer using a triple resonance probe equipped for *z*-axis pulsed field gradients. One bond ^{15}N−^{1}H coupling data were collected using two types of experiments, a phase encoded (or intensity based) heteronuclear single quantum coherence experiment (PE-HSQC) (*18*), and a frequency resolved coupling enhanced heteronuclear single quantum coherence experiment (CE-HSQC) (*19*). Experiments were conducted on both aligned and unaligned samples, and the difference in couplings measured under the two conditions was taken to be the residual dipolar coupling contribution. Typically, for the PE-HSQC, a total of 1024 complex point were acquired over

10 000 Hz in the directly detected proton dimension and 128 complex points were acquired over 2200 Hz in the indirect nitrogen dimension. Three sets of data were acquired with constant time delays of 64.5, 60.0, and 58.2 ms.. Coupling measurements in the three sets agree within 1 Hz, and only the average coupling is reported. For the CE-HSQC collection parameters for the proton dimension were the same, but 256 points were collected in the indirect dimension to improve spectral resolution. The coupling enhancement factor used (1.0) resulted in doubling of the apparent coupling. Values agreed with the PE-HSQC data within 2 Hz. In both cases 32 transients were collected for each indirect point. Similar protocols were applied to N−H couplings for both the protein and ligand.

For the studies of the LacNAc ligand, one bond $^{13}C-^{1}H$ couplings from the anomeric site and acetyl-methyl site of the *N*-acetylglucosamine moiety were measured with a CE-HSQC experiment similar to that described above. The data were collected on a 500 MHz INOVA system. The number of points acquired was 1152 over 5000 Hz in the directly observed proton dimension and 256 points over 12.5 kHz in the indirect dimension. Typically, 32 transients per indirect point were collected. Studies were conducted in both isotropic and aligned states for both free ligand and protein-bound ligand samples.

Data Analysis. The FELIX data processing package (Molecular Simulations, version 98.0) was used for initial transformation and display of data. A Bayesian statistical analysis package (*31*) was used to extract couplings from the CE-HSQC $^{13}C-^{1}H$ data on the ligand. This allowed both precise measurement and an estimate of errors (±1 Hz).

For ligand couplings measured in the presence of protein, values must be apportioned between bound and free states. This was done using a literature value of the ligand binding constant, $K_b = 5238.0\ M^{-1}$ for 25 °C, and an enthalpy measured at 25 °C ($\Delta H° = -35.89$ kJ/mol) (*32*) to adjust the binding constant to the temperature of data collection (38 °C). This corrected binding constant (2859 M^{-1}) predicts the relative contributions of free and bound populations to coupling to be approximately 85:15 at the 0.1 mM protein and 0.25 mM ligand concentrations used for measurements of couplings in the ligand.

Structural Analysis. Order tensors for the free and complexed protein were calculated using a singular value decomposition approach (*20*), $^{15}N-^{1}H$ dipolar data for sites in areas of well-defined secondary structure, and coordinates from the X-ray structure of the complex as reported by Seetharaman et al. (*10*). The principal values of the order tensor determined were −0.001 331, 0.000 944, and 0.000 387 for the free protein and −0.001 251, 0.000 805, and 0.000 445 for the protein in the presence of ligand. Back calculation of residual dipolar couplings for the all N−H vectors were done with the aide of a program BACKCALC_RDC written by Homay Valafar (http://tesla.ccrc.uga.edu), and comparative analysis was performed on EXCELL spread sheets.

Structural Refinement. Structural refinement was accomplished with the aide of the CNS program (*33*). Dipolar couplings for NH vectors were entered as constraints represented as flat-well potentials having well widths of 1.0 Hz. In addition, NOE constraints from a previous study (Umemoto and Leffler, unpublished data) were entered using flat well potentials with bounds depending on classification of NOEs into strong, medium, and weak connectivities as described in the literature (*34*). Starting with coordinates from the crystal structure of LacNAc-bound galectin-3C, simulated annealing was carried out in four steps with the following heating protocols: A constant high-temperature dynamics stage at 5000 K over 500 cycles (7.5 ps), followed by a torsion dynamics cooling stage from 5000 to 0 K in 5000 steps (75 ps), a second Cartesian dynamics cooling from 2000 to 0 K in 5000 steps (25 ps), and the final minimization over 50 cycles in 1000 steps each. As the constraints, 1145 NOEs were used with the scaling factors decreasing from 150 to 50, and 111 residual dipolar couplings with the scaling factors increasing from 1 to 100 during the course of the simulated annealing. The NOE data were previously acquired for galectin-3C bound to lactose (Umemoto and Leffler, unpublished), and the two sets of dipolar coupling data were for galectin-3C in the presence and absence of LacNAc. The above scaling protocol was planned so that the final structures mainly reflect ones consistent with the dipolar coupling data. Twenty structures were generated each for free and complexed galectin-3C. The resulting structures have no violations over 0.5 Hz for the dipolar constraints. They have rmsd deviation of backbone atoms from a mean position of atoms in the various structures of 0.3 Å. The mean positions of atoms in strands not in contact with ligand deviate by an average of 1.4 Å from the initial coordinates in the starting crystal structure.

RESULTS

$^{15}N-^{1}H$ dipolar couplings were measured by both frequency resolved and intensity based (or phase encoded) variations of HSQC spectra as described above. In all, 113 and 114 amide backbone couplings were measured with frequency based methods, and 117 and 115 amide backbone couplings were measured with intensity based methods for galectin-3C in the absence and presence of ligand, respectively. The couplings are assigned to sequence specific positions based on data in ref *11*. In all, 113 corresponded to backbone amide groups seen in the crystal structure and, hence, could be used for structural comparison. The data for each site in both the presence and absence of ligand have been deposited in the BMRB data bank (deposition # 6823), using in most cases the average of three intensity based measurements.

The coupling (RDC) measurements range from 17.97 Hz to −30.25 Hz, in accord with the positive and negative offsets expected from the θ dependence of eq 1. The methods used to determine these couplings are based on quite different principles, and one might expect them to deviate from one another due to various systematic errors, such as those due to spin relaxation, in addition to random errors associated with noise in the measurements. The root-mean-square deviation between measurements using the two methods is 1.1 Hz for isotropic and 2.1 Hz for aligned samples, numbers only slightly larger than expectations for precision based on noise in individual spectra. This suggests that the measurements are both precise and accurate.

Protein Geometry. Assuming that solution structures would be identical to the crystal structure over most of the protein, one can also assess the accuracy of the measurements by plotting measured couplings versus calculated couplings.

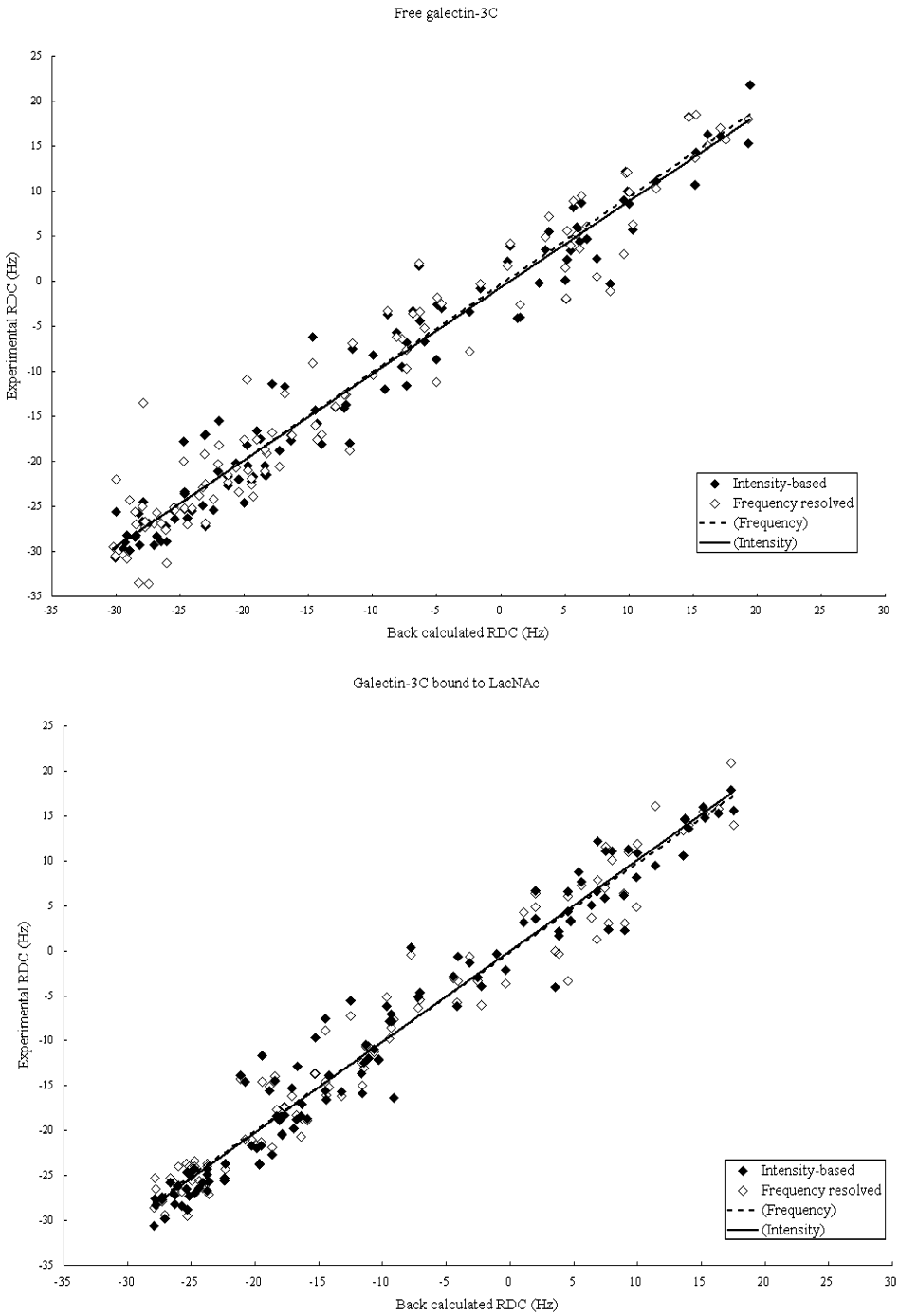


FIGURE 1: Residual dipolar coupling data obtained by frequency-resolved (open marks) and intensity-based (solid marks) experiments for galectin-3C in the absence (A) and presence (B) of ligand, plotted against the values that are back-calculated on the basis of the liganded crystal structure (Seetharaman et al., 1998).

Calculating couplings requires the determination of elements of an order tensor. This symmetric and traceless 3×3 tensor has five independent elements. These elements were determined using a singular value decomposition method to reach a best least-squares fit to the data. Using the best order tensor solution and the coordinates from N−H pairs in the crystal structure of the liganded protein as modified by addition of protons using standard peptide bond geometries, residual dipolar couplings were back-calculated for all sites using data collected in both the presence and absence of ligand. The plot of back-calculated values versus measured values is shown in Figure 1. Data from the frequency resolved data are shown as open points, and the data from intensity-based data are shown as solid points. Considering data from both types of experiment, the standard deviation between calculated and measured points is 3.3 Hz for the protein in the absence of ligand and 2.7 Hz in the presence of ligand, values similar to the sum of the deviations found above. The deviations considering data from the two types of experiment separately are nearly identical. The similarity suggests that use of one experiment as opposed to the other cannot be justified based on the level of agreement with the crystal structure. It is significant, however, that the intensity-based data were collected in a shorter time (5.5 vs 8.5 h), and a slightly larger number of measurements were possible because of the reduced overlap when peaks are not split by

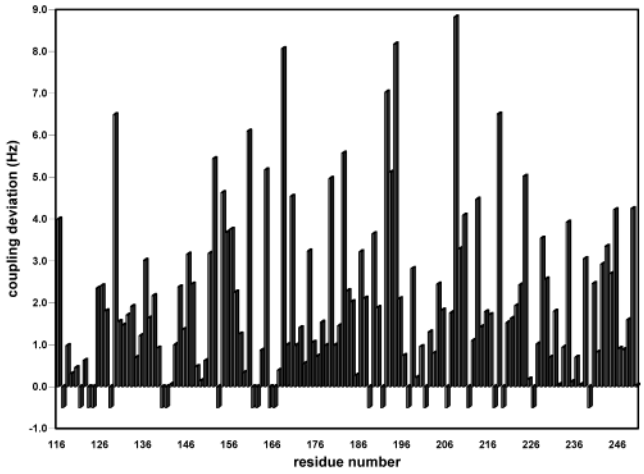


FIGURE 2: Residue specific deviations of dipolar couplings measured in the absence of ligand from those back-calculated from the liganded crystal structure. All deviations are presented as absolute values and residues missing dipolar data are indicated by an assigned negative value.

frequency domain couplings. The overlap problem for frequency resolved data can, however, be avoided by using methods that present the two components of a doublet in different spectra (*35*, *36*).

The fact that agreement between measurements and back-calculated couplings is good, particularly in the presence of ligand, can be seen by converting deviations in Hz to structural deviations; the rmsds given in Hz correspond to an average angular deviation for N−H vectors of 2−10°, depending on their orientation with respect to the principal alignment axis. The most likely explanation for the slightly larger deviations in the unliganded state is a small variation in structure on introduction of the ligand, and/or the existence of enhanced local motions in the unliganded state that reduce sizes of dipolar couplings to various extents.

The clear outliers in the unliganded state have been checked for possible assignment errors; deviations are still found to exist. It, therefore, seems logical to seek a structural explanation. This can most easily be done starting with a residue specific depiction of deviations between couplings for the unliganded molecule in solution and couplings back-calculated from the liganded crystal structure. Such a depiction is shown in Figure 2. Setting a threshold of 5.0 Hz or greater 13 residues show significant deviations. Most can be rationalized on the basis of changes in structural elements that directly or indirectly contribute to ligand binding as seen in the crystal structure (1A3K). Two residues (164 and 168) are in the loop connecting S4 and S5, a loop which is adjacent to a residue shown to interact with the ligand (Arg 162); two residues (179 and 182) are in the loop connecting S5 and S6a, which, together with the end of S6a, provides the side chains of Trp181 and Glu184 that also interact directly with the ligand; one residue (160) is at the base of the β strand groove and interacts directly with the ligand; two residues (152 and 224) are in loops that underlie the S5−S6a loop and make side chain contacts with the S5−S6a loop (<3.5 Å for 183;O-224;NH2 and <4.0 Å for 178;N-154;OD1); three residues (192, 193, and,194) are in a loop that underlies the S4−S5 loop and makes side chain contacts with the S4−S5 loop (<3.5 Å for 169, NH1−192, O; and <4.0 Å for 163, CE2−194, OG). The remaining three residues are also in loops but are remote from the binding site. Somewhat smaller, but still very significant, deviations occur for NMR data in the presence of ligand. These deviations may well reflect the fact that, even with 0.2 mM protein and 0.5 mM ligand, the affinity at 38° predicts that about half the protein is in the unliganded state. The deviations seen both in the presence and absence of ligand, thus, provide strong evidence that the unliganded protein undergoes a concerted structural change to form a ligand-binding site similar to that seen in the crystal structure.

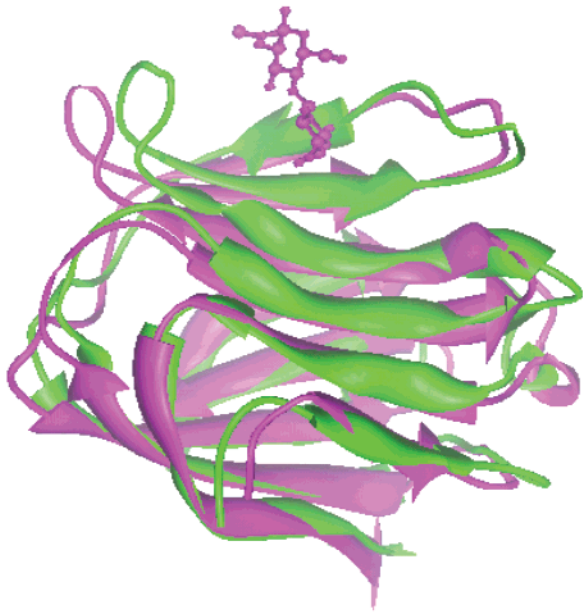


FIGURE 3: Solution structure of galectin-3C in the absence of ligand (green ribbon) compared to the X-ray crystal structure with LacNAc bound (magenta ribbon). The loops to the left of the ligand contain residues 163−169, and the loops to the right contain residues 175−184. Both move toward the base of the binding site in the absence of ligand. The ligand (magenta ball-and-stick model) is shown with the GlcNAc residue above the Gal residue. The N-terminus (aa 114) is at the bottom middle of the diagram indicating a bend before the first β-strand. The molecular graphics image was produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081).

A structural model for Galectin-3C, based on dipolar couplings measured in the absence of ligand, can be generated using couplings as constraints in a simulated annealing protocol. The details of the procedure are described in the methods section. In Figure 3 we present the results (green ribbon) in comparison to the liganded crystal structure (magenta ribbon). It is clear that the binding site has closed up in the absence of ligand, with both the S4−S5 loop (top left) and the S5−S6a loop (top right) moving inward toward the base of the binding cleft. The supporting loops, most clearly seen on the left portion of the figure, clearly follow this movement.

Ligand Geometry. It is possible to collect residual dipolar data on bound ligands by taking advantage of $^{13}C-^{1}H$ couplings in addition to the $^{1}H-^{15}N$ couplings used in analyzing protein data. The crystal structure we used in the comparison above has an *N*-acetyllactosamine (galactosyl-(β1−4)-*N*-acetylglucosamine) ligand attached. Our synthesized analogue has a single $^{15}N-^{1}H$ vector in the *N*-acetyl group, but collecting sufficient data to test for consistency with ligand orientation requires more interaction vectors. It

Table 1: Dipolar Coupling Data for Bound *N*-Acetyllactosamine[a]

dipolar couplings (Hz)	N−H	C1−H1	methyl C−H
free ligand measured	−4.6	8.4	−0.4
complex measured	−0.7	−1.0	−2.6
bound ligand extracted	21.4	−36.4	−10.9
bound ligand calculated	16.8	−30.8	−7.0

[a] Data were collected on samples with bound to free ligand populations varying from 15:85 to 25:75.

is becoming possible to collect data on $^{13}C-^{1}H$ vectors of carbohydrates without the need for isotope enrichment (*37*), at least when they are free in solution and can be examined at relatively high concentrations. In our case, both broadening of resonances on binding to the protein and the limited solubility of the protein suggested that isotope enrichment at ^{13}C sites would be advisable. Given that the labeled *N*-acetyllactosamine would require a multiple step synthesis, we chose to enrich a limited number of sites using relatively inexpensive starting materials. As detailed in the methods section the sites enriched included the nitrogen and C1 sites on *N*-acetylglucosamine ring and the carbonyl and methyl sites in the *N*-acetyl group itself. Among the possible couplings that could be measured, the most useful proved to be C1−H1 (from the anomeric group of the glucosamine), H−N (from the *N*-acetyl group), and C8−H8 (from the acetyl methyl group). These couplings are reported in Table 1.

Interpretation of ligand couplings requires consideration of several motional averaging effects. First, free rotation of the methyl group results in projection of C−H couplings on to the C7−C8 axis of the acetyl group; this requires correcting for rotational averaging and redefinition of the interaction vector. These corrections are taken into account when reporting back-calculated couplings. Second, the values reported in Table 1 are actually averages of couplings seen for free and bound *N*-acetyl-lactosamine. To extract couplings for the bound state, measurements also need to be made for free carbohydrate in the same orienting medium, and a binding constant needs to be used to apportion contributions. The free state data are reported in Table 1, along with a set of bound state values extracted using a binding constant adjusted to the experimental temperature of 38 °C using thermodynamic data reported by Bachhawat-Sikder et al. (*32*).

The order tensor found from the N−H dipolar data on the protein in the presence of ligand can, to a good approximation, be used to back-calculate dipolar couplings for the ligand using the geometry of N−H and C−H vectors in LacNAc taken from the crystal structure. These back-calculated values are also reported in Table 1. The differences between the measured and calculated values are significant, but considering the range of binding constants reported in the literature, and the only approximate concentration of the protein in the sample, the agreement is excellent. This suggests that the orientation of the ligand seen in the crystal structure is an adequate representation of the bound geometry in solution.

DISCUSSION

The data presented above illustrate a relatively efficient procedure for the comparison of solution and crystal structures of proteins and protein−ligand complexes. Residual dipolar data on $^{15}N-^{1}H$ pairs in proteins and $^{13}C-^{1}H$ pairs in ligands are easily generated in no more than a few hours for each sample when isotope enrichment and concentrations approaching 1 mM in protein or 1 mM in ligand are used. These times will be reduced dramatically with the availability of cryogenic NMR probes. We have tested two experimental protocols for acquiring data. These appear to vary slightly in efficiency of measurement. However, both suffer from loss of data when spectral overlap is severe. Fortunately, the cases leading to loss are not always the same, and using both experiments results in a larger data set, as well as a check on internal consistency of measurements.

For the data presented on galectin-3C, there is generally good agreement between predictions based on the crystal structure and solution data. Using the rmsd of dipolar couplings as a guide, average angular deviations of N−H vectors would be no more than 10° when the *N*-acetyllactosamine complexed crystal structure is compared to the free state of the protein in solution. The deviations are somewhat less when the complexed state in the crystal is compared to that found in solution in the presence of ligand. Also, the residual dipolar coupling for the bound ligand agrees well with data back-calculated from the ligand geometry as found in the crystal structure of the complex. There are, however, local structural variations when structures produced from solution data collected in the presence or absence of ligand are compared to the crystal structure (Figure 3 for the absence of ligand). For example, the position of the alpha carbon of Asn 166 in the 163−169 segment deviates by 2.8 and 5.6 Å in the presence and absence of ligand, respectively; the alpha carbon of Trp 181 in the 177−184 segment deviates by 5.7 and 6.1 Å in the presence and absence of ligand, respectively. These figures compare to rmsd values of 1.4 Å when backbone atoms in the strands of all beta sheets except the three underlying the binding site are compared.

The above deviations are not hard to rationalize for the structure produced from solution data in the absence of ligand. Changes bring loops at both ends of the ligand binding site closer together, filling space that was formerly occupied by ligand. The deviations between the liganded crystal structure and the structure produced from solution data in the presence of ligand seem more difficult to understand. However, as we discussed earlier, at the concentrations of ligand and protein used, and given the estimated binding constants, only about half the protein would have been complexed in the presence of ligand, and the data used to produce the structure would have represented averages between liganded and unliganded states. The effects of partial complexation may be seen in the structure of the 164−169 segment, which appears to sit midway between positions found in the fully liganded state of the crystal structure and the solution structure determined in the absence of ligand. In comparison to the structure in the absence of ligand, Asn166 actually moves 3.1 Å back toward its position in the crystal structure.

Average structures must, however, be interpreted with caution. Average dipolar couplings do not have to be interpretable in terms of a single rigid structural model. Moreover, it is often difficult to acquire data on segments undergoing dynamic averaging because of exchange broadening of resonances. The primary pieces of residual dipolar data driving the conformational change in the model of the

163−169 segment, for example, are the dipolar couplings for residues 164 and 168. In the absence of ligand these couplings deviate by 5.2 and 8.1 Hz, respectively, from values back-calculated from the crystal structure.. Resonances from amide pairs adjacent to and between these residues are broadened, and dipolar data could not actually be collected on residues 161, 162, 165, and 166. It is also significant that the peak from the ring N−H of Trp 181 could only be seen in the presence of ligand, possibly because ring motion or proton exchange at this site is slowed in the presence of ligand.

Both the changes in structure and the mobility of segments near the ligand binding site may be functionally important. The 177−184 segment contains a number of residues involved in ligand binding, including the more hydrophobic residue, Trp 181. The 163−169 segment contains no residues directly involved in ligand binding, but an adjacent residue, 162, is, and it is possible that the removal of its interaction with the ligand allows increased motion of beta strand S4 and the following loop (163−169). Changes seen in comparing the crystal structure to the unliganded state in solution bring both segments to positions that partially cover the cleft in which the ligand binds. This may prevent unfavorable exposure of Trp 181 and other hydrophobic surfaces implicated in the binding process. Loop 177−184 also contains the NWGR sequence motif implicated in the intracellular anti-apoptotic effect of galectin-3 (*7*, *8*). Movement of this segment or shielding of it by the 163−169 segment may be essential for regulation of interactions with other proteins in the apoptotic pathway. Mobility of segments near the binding site may also be an important part of galectin's ability to bind a variety of ligands terminated with galactose (*4*) as well as other ligands (*8*). The efficiency of the experiments described should allow investigations with a number of different ligands in solution. Such studies may clarify the ways in which binding sites can adapt to accommodate a variety of ligands and alternate functions of proteins such as galectin-3.

ACKNOWLEDGMENT

We thank Inga Svensson and Barbro Kahl-Knutsson for excellent help in preparing ^{15}N-enriched galectin-3C.

REFERENCES

1. Cooper, D. N. W., and Barondes, S. H. (1999) *Glycobiology 9*, 979−984.
2. Rabinovich, G. A., Rubinstein, N., and Fainboim, L. (2002) *J. Leukocyte Biol. 71*, 741−752.
3. Dunphy, J. L., Barcham, G. J., Bischof, R. J., Young, A. R., Nash, A., and Meeusen, E. N. T. (2002) *J. Biol. Chem. 277*, 14916−14924.
4. Leffler, H. (1997) *Trends Glycosci. Glycotechnol. 9*, 9−19.
5. Hughes, R. C. (2001) *Biochimie 83*, 667−676.
6. Matarrese, P., Fusco, O., Tinari, N., Natoli, C., Liu, F. T., Semeraro, M. L., Malorni, W., and Iacobelli, S. (2000) *Int. J. Cancer 85*, 545−554.
7. Song, Y. K., Billiar, T. R., and Lee, Y. J. (2002) *Am. J. Pathol. 160*, 1069−1075.
8. Akahani, S., NangiaMakker, P., Inohara, H., Kim, H. R. C., and Raz, A. (1997) *Cancer Res. 57*, 5272−5276.
9. Rini, J. M., and Lobsanov, Y. D. (1999) *Curr. Opin. Struct. Biol. 9*, 578−584.
10. Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998) *J. Biol. Chem. 273*, 13047−13052.
11. Umemoto, K., and Leffler, H. (2001) *J. Biomol. NMR 20*, 91−92.
12. Prestegard, J. H., Al-Hashimi, H. M., and Tolman, J. R. (2000) *Q. Rev. Biophys. 33*, 371−424.
13. Bax, A., Kontaxis, G., and Tjandra, N. (2001) *Nucl. Magn. Reson. Biol. Macromol., Pt. B* pp 127−174.
14. Thompson, G. S., Shimizu, H., Homans, S. W., and Donohue-Rolfe, A. (2000) *Biochemistry 39*, 13153−13156.
15. Bewley, C. A. (2001) *Structure 9*, 931−940.
16. Bolon, P. J., Al-Hashimi, H. M., and Prestegard, J. H. (1999) *J. Mol. Biol. 293*, 107−115.
17. Losonczi, J. A., and Prestegard, J. H. (1998) *J. Biomol. NMR 12*, 447−451.
18. Tolman, J. R., and Prestegard, J. H. (1996) *J. Magn. Reson., Ser. B 112*, 245−252.
19. Tolman, J. R., and Prestegard, J. H. (1996) *J. Magn. Reson., Ser. B 112*, 269−274.
20. Losonczi, J. A., Andrec, M., Fischer, M. W. F., and Prestegard, J. H. (1999) *J. Magn. Reson. 138*, 334−342.
21. Massa, S. M., Cooper, D. N. W., Leffler, H., and Barondes, S. H. (1993) *Biochemistry 32*, 260−267.
22. Kuhn, R., and Kirschenlohr, W. (1965) *Liebigs Ann. Chem. 600*, 115−120.
23. Kuhn, R., and Kirschenlohr, W. (1965) *Liebigs Ann. Chem. 600*, 135−143.
24. Horneman, U. (1973) *Carbohydr. Res. 28*, 171−174.
25. Taniguchi, M., Nystrom, R. F., and Rinehart, K. L. (1982) *Carbohydr. Res. 109*, 161−166.
26. Palaniswamy, V. A., and Gould, S. J. (1988) *J. Chem. Soc., Perkin Trans. 1*, 2283−2286.
27. Zhou, Z. Y., Sakuda, S., and Yamada, Y. (1992) *J. Chem. Soc., Perkin Trans. 1*, 1649−1652.
28. Alais, J., and Veyrieres, A. (1981) *Carbohydr. Res. 93*, 164−165.
29. Heidlas, J. E., Lees, W. J., Pale, P., and Whitesides, G. M. (1992) *J. Org. Chem. 57*, 146−151.
30. Yu, B., Ouyang, Q. Q., Li, C., and Hui, Y. Z. (1996) *J. Carbohydr. Chem. 15*, 297−302.
31. Andrec, M., and Prestegard, J. H. (1998) *J. Magn. Reson. 130*, 217−232.
32. Bachhawat-Sikder, K., Thomas, C. J., and Surolia, A. (2001) *Febs Lett. 500*, 75−79.
33. Brunger, A., Adams, P., Clore, G., Delano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T., and Warren, G. (2000) *CNS*, Yale University, New Haven, CT.
34. Brunger, A. T., and Nilges, M. (1993) *Q. Rev. Biophys. 26*, 49−125.
35. Lerche, M. H., Meissner, A., Poulsen, F. M., and Sorensen, O. W. (1999) *J. Magn. Reson. 140*, 259−263.
36. Ottiger, M., Delaglio, F., and Bax, A. (1998) *J. Magn. Reson. 131*, 373−378.
37. Tian, F., Al-Hashimi, H. M., Craighead, J. L., and Prestegard, J. H. (2001) *J. Am. Chem. Soc. 123*, 485−492.

BI026671M