NMR Solution Studies of Hamster Galectin-3 and Electron Microscopic Visualization of Surface-Adsorbed Complexes: Evidence for Interactions between the N- and C-Terminal Domains†

Berry Birdsall,‡ James Feeney,*,‡ Ian D. J. Burdett,§ Sulemana Bawumia,|| Eminia A. M. Barboni,|| and R. Colin Hughes*,||

Molecular Structure Division, Membrane Biology Division, and Protein Structure Division, National Institute for Medical Research, Mill Hill, London, NW7 1AA, U.K.

Received December 21, 2000; Revised Manuscript Received February 8, 2001

ABSTRACT: Galectin-3, a β-galactoside binding protein, contains a C-terminal carbohydrate recognition domain (CRD) and an N-terminal domain that includes several repeats of a proline-tyrosine-glycine-rich motif. Earlier work based on a crystal structure of human galectin-3 CRD, and modeling and mutagenesis studies of the closely homologous hamster galectin-3, suggested that N-terminal tail residues immediately preceding the CRD might interfere with the canonical subunit interaction site of dimeric galectin-1 and -2, explaining the monomeric status of galectin-3 in solution. Here we describe high-resolution NMR studies of hamster galectin-3 (residues 1−245) and several of its fragments. The results indicate that the recombinant N-terminal fragment Δ126−245 (residues 1−125) is an unfolded, extended structure. However, in the intact galectin-3 and fragment Δ1−93 (residues 94−245), N-terminal domain residues lying between positions 94 and 113 have significantly reduced mobility values compared with those expected for bulk N-terminal tail residues, consistent with an interaction of this segment with the CRD domain. In contrast to the monomeric status of galectin-3 (and fragment Δ1−93) in solution, electron microscopy of negatively stained and rotary shadowed samples of hamster galectin-3 as well as the CRD fragment Δ1−103 (residues 104−245) show the presence of a significant proportion (up to 30%) of oligomers. Similar imaging of the N-terminal tail fragment Δ126−245 reveals the presence of fibrils formed by intermolecular interactions between extended polypeptide subunits. Oligomerization of substratum-adsorbed galectin-3, through N- and C-terminal domain interactions, could be relevant to the positive cooperativity observed in binding of the lectin to immobilized multiglycosylated proteins such as laminin.

Galectins are a family of carbohydrate-binding proteins implicated increasingly in cell-signaling and cell-adhesion events through binding specifically to various intracellular and extracellular carbohydrate polymers (*1*−*7*). Galectin-3 is a multidomain molecule (Figure 1) containing a C-terminal carbohydrate-recognition domain (CRD)[1] and a unique N-terminal domain of various lengths according to species, which is encoded within a single exon of the genomic sequence (*8*) and consists of repeats of proline-tyrosine-glycine rich motifs (*9*). There are nine such repeats in hamster galectin-3 (Figure 1). The N-terminal domain also contains sites for phosphorylation, which appear to regulate retention of the lectin within the nucleus (*10*, *11*), and other determinants important for the secretion of the lectin by a novel, nonclassical mechanism (*12*−*14*).

The CRD of galectin-3 has close structural homology to the CRDs of other galectins (*15*−*18*). However, the crystal structure of human galectin-3 (*16*) indicated that it does not form stable 2-fold symmetric homodimers characteristic (*19*) of galectin-1 and -2. Gel filtration chromatography studies have consistently failed to detect dimers or oligomers of galectin-3 (*20*, *21*). A comparison of the crystal structures of human galectin-3 and dimeric galectins-1 and -2 (*15*, *16*, *18*), as well as modeling of hamster galectin-3 (*17*, *22*), indicated that the apolar subunit interaction site present in galectin-1 and -2, although conserved to large extent in galectin-3, might in the latter be obstructed by amino acid residues at the end of the N-terminal domain, i.e., the residues immediately proximal to the CRD proper (LTPVY in hamster galectin-1, Figure 1).

Although galectin-3 appears not to form stable dimers or higher oligomers in solution, other evidence strongly suggests that transient oligomerization can occur. Chemical and

† E.A.M.B. thanks University of Rome La Sapienza, Recherche di Ateneo 1998, for support.

* To whom correspondence should be addressed. Phone: 020-8959 3666, ext. 2023. Fax: 020-8906 4477. (R.C.H) E-mail: chughes@nimr.mrc.ac.uk. (J.F.) E-mail: jfeeney@nimr.mrc.ac.uk.

‡ Molecular Structure Division.

§ Membrane Biology Division.

|| Protein Structure Division.

[1] Abbreviations: COSY, correlated spectroscopy; CRD, carbohydrate recognition domain; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; DQF, double quantum filtered; GARP, a broad-band decoupling sequence; HMQC, heteronuclear multiple quantum coherence spectroscopy; HSQC, heteronuclear single quantum coherence spectroscopy; MLEV-17, isotropic mixing sequence; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PBS, phosphate-buffered saline; ROESY, rotating frame Overhauser effect spectroscopy; SCUBA, stimulated cross-peaks under bleached alphas; TOCSY, total correlation spectroscopy; WT, wild-type; 3D, three-dimensional; 2D, two-dimensional.

10.1021/bi002907f CCC: $20.00
Published on Web 03/20/2001

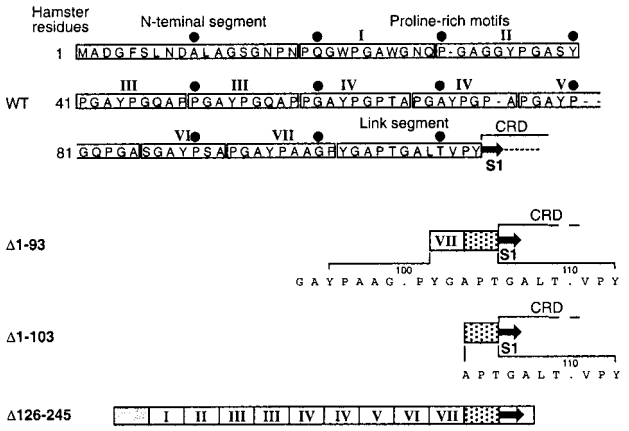

FIGURE 1: Structures of galectin-3 and its C-terminal and N-terminal fragments. Wild-type hamster galectin-3 starting from the N-terminus with an 18 amino acid residue segment of average composition followed by nine proline-tyrosine-glycine-rich repeats classsified as motifs I−VII (*9*), a 12-residue link segment and the first β-strand (residues 114−120, arrow) of the carbohydrate-recognition domain (CRD). The Δ1−93 mutant obtained by deletion of residues 1−93. The amino acid sequence between residues 94 and 113 is shown and comprises most of motif VII (open box) and all of the link segment (stippled box). The Δ1−103 protein obtained from galectin-3 by collagenase digestion. It lacks N-terminal residues 1 to 103 including all of the repeat sequences and a part of the link segment. The Δ126−245 protein comprising all residues of the N-terminal segment, the repeat sequences, the link segment and residues 114−125 including the first β-strand of the CRD.

transglutaminase-catalyzed cross-linking experiments have shown that monomeric galectin-3 is in equilibrium with higher order oligomers in solution at relatively high concentrations (*23*−*26*). Formation of dimers or higher oligomers predominantly involves interactions between N-terminal domains, although interactions between CRD subunits may make some contribution (*22*−*27*). An ability of galectin-3 to form oligomers is also strongly implied by the fact that galectin-3 binds to multi-glycosylated proteins, such as found on a surface coated with the extracellular matrix component laminin, with positive cooperativity (*21*, *23*, *28*). Such behavior suggests that galectin-3 monomers, after binding to a laminin-coated substratum, can recruit from solution additional lectin molecules to form a substrate-attached complex, permitting multivalent interactions. The switch between mono- and multivalency, which is sensitive to concentration, could play an important regulatory role in mediating the biological activity of galectin-3 (*29*). For example, the migration of human breast carcinoma cells through a laminin-rich layer is inhibited in a concentration-dependent manner (*30*), perhaps due to increased adhesion mediated by the cross-linking of cell surface and extracellular matrix receptors by multivalent lectin at high concentrations.

In this paper, we examine the molecular size and domain structure of hamster galectin-3 in solution by equilibrium sedimentation ultracentrifugation and NMR spectroscopy, respectively. The data indicate that galectin-3 exists as a monomer under these conditions and are consistent with there being an interaction of some portion of the N-terminal domain with the CRD, in agreement with proposals that such interactions would obstruct the canonical dimer interface observed in galectin-1 and -2. We also report the remarkable self-association of the N-terminal fragment of galectin-3 into amyloid-like fibrils when sprayed onto a mica surface for electron microscopy, and the self-association of CRD fragments and full-length lectin under the same conditions.

MATERIALS AND METHODS

Materials. Recombinant hamster full-length wild-type (WT) galectin-3 and truncated protein Δ1−93 lacking N-terminal residues 1−93 (Figure 1) were obtained by bacterial expression followed by purification on immobilized galactoside-based affinity columns as described (*24*, *31*). The CRD fragment Δ1−103 (Figure 1) was obtained after digestion of full-length lectin with bacterial collagenase followed by purification as above (*24*). The N-terminal domain protein Δ126−245 (Figure 1) lacking residues 126−245 of the CRD was expressed as a fusion protein with glutathione S-transferase and purified by affinity chromatography on a glutathione-based matrix followed by thrombin cleavage (*25*). These products gave single bands migrating as expected for the calculated molecular weights after SDS−PAGE, as detected by Comassie blue staining or Western blotting with galectin-3 specific antibodies. Samples (1−3 mg/mL) were stored at 2 °C in phosphate-buffered saline (PBS) containing 100 mM lactose. Immediately before use, the samples were dialyzed against PBS, and centrifuged at 18000*g* for 15 min. Supernatants were used for redetermination of protein concentrations. Rat monoclonal Mac-2 antibody, directed against mouse galectin-3 and cross-reactive with hamster galectin-3, and its Fab fragment, was prepared as described (*31*). The A-blood group tetrasaccharide GalNAcα1,3 [Fucα1,2]Galβ1,4GlcNAc was obtained from Sigma, Poole, Dorset.

Sedimentation Equilibrium Experiments. Galectin-3 and the Δ1−93 protein (20 μM) were equilibrated in PBS 5 μM EDTA, pH 7.2, and studied by analytical ultracentrifugation. Equilibrium ultracentrifugation was done at 10 °C in a Beckman Optima XLA analytical ultracentrifuge fitted with absorbance optics. The effective molecular weight was determined from data acquired at three rotor speeds (22 000, 29 000, and 35 000 rpm). Data were acquired in steps of 0.001 mm at 280 nm: 20−50 scans were averaged for each run. Equilibrium was judged to have been reached when the absorbance trace showed no detectable change. The absorbance data were analyzed individually to a simple ideal monomer, an ideal dimer, and globally using all of the data, according to the equation

$$A(r) = A(r_0)\exp[M(1 - \upsilon\rho)\omega^2(r^2 - r_0^2)/2RT]$$

where A is the absorbance at the radial position, $A(r_0)$ is the absorbance at a reference position r_0, ω is the rotor angular speed, M is the molecular weight, and R is the gas constant. The partial specific volume was taken as 0.74 mL g^{-1}. The solution density ρ was calculated from the buffer composition using tabulated values (*32*).

Nuclear Magnetic Resonance. NMR experiments were carried out on Varian UNITY, UNITY INOVA, and UNITY plus spectrometers operating at proton frequencies of 600, 600, and 500 MHz, respectively. The spectra were recorded at 25, 22, and 5 °C on either unlabeled or ^{15}N-labeled protein samples with volumes of 0.6−1.5 mL and at concentrations 0.007−0.2 mM made up either in D_2O or in 90% H_2O/10% D_2O at pH 5.5 to 7.5 (pH values are pH meter readings uncorrected for deuterium isotope effects). The two-

dimensional NMR experiments carried out on galectin-3 and related samples were DQF−COSY (*33*), TOCSY (*34*) with isotropic mixing time of 60 ms [using MLEV-17 sequence (*35*) with a field strength of approximately 8 kHz], NOESY (*36*, *37*) with a mixing time 100 ms and ROESY (*35*, *38*) with a 125 ms mixing time. These spectra were recorded for both H_2O and D_2O solutions of unlabeled galectin-3 and related samples. In the TOCSY and NOESY experiments in D_2O, the SCUBA technique was applied to recover signals from α-CH protons saturated because they resonate near to the presaturated water resonance (*39*).

The heteronuclear ^{1}H-^{15}N experiments were carried out on the uniformly ^{15}N-labeled galectin-3 and the Δ1−93 protein. For the ^{15}N-labeling, plasmids were grown in minimal salts medium supplemented with 1% glucose and 0.2% ^{15}N-labeled ammonium sulfate. The ^{15}N-labeled proteins were dialyzed against phosphate-buffered saline (PBS), pH 7.4, and concentrated to approximately 0.2 mM. The heteronuclear ^{1}H-^{15}N experiments comprised: 2D ^{1}H-^{15}N HSQC (*40*), 3D TOCSY−HMQC (*41*), 3D NOESY−HSQC (*41*, *42*). 2D and 3D ^{1}H-^{15}N heteronuclear experiments on the complete galectin-3 sample did not give successful results because of the low sample concentration (0.007 mM). A 3D NOESY−HMQC was acquired on the Δ1−93 protein complexed with the A-active tetrasaccharide with a mixing time 50 ms. Decoupling during the detection period of the 3D experiments was achieved using the GARP-1 sequence (*43*). Phase sensitive detection in the indirect dimensions was obtained using the method of States and co-workers (*44*). In the 2D and 3D experiments on samples in H_2O, water suppression was achieved either by using selective presaturation with a DANTE sequence (*45*) or by means of a WATERGATE sequence (*46*). The ^{1}H chemical shifts were referenced to DSS and the ^{15}N chemical shifts were referenced to liquid NH_3 calibrated using the γ ratios method (*47*, *48*).

The {^{1}H}-^{15}N NOE HSQC spectra were collected as described previously (*49*) using a variation of the procedure proposed by Kay and co-workers (*50*).

A difference spectroscopy approach was used to facilitate observation of only those signals in the {^{1}H}-^{15}N NOE HSQC corresponding to nuclei in the more mobile parts of the protein. Signals from nuclei in the core of a protein of this size will have a negative heteronuclear NOE of 0.2 at 5 °C resulting in the observed signals having an intensity of 0.8 compared with a control HSQC spectrum obtained without ^{1}H irradiation. Thus, a difference spectrum obtained by subtracting 0.8 times the control ^{1}H-^{15}N HSQC spectrum from the {^{1}H}-^{15}N NOE HSQC spectrum will show only those signals from nuclei in mobile parts of the protein. Values for the NOEs of the latter nuclei were estimated by determining the scaling factors of the control HSQC spectrum required to null each signal in the difference spectrum.

Electron Microscopy. Protein samples were examined initially by negative staining with 2% sodium silicotungstate, pH 7.2. For rotary shadowing, samples were first diluted with PBS to a protein concentration of about 0.1 mg/mL and with glycerol to 30−40% (*51*). The solutions were then sprayed onto the surface of freshly cleaved mica. Samples were rotary shadowed with 2−4 nm of tungsten from an electron gun (Balzers) at an angle of 5° in a Leybold-Heraeus vacuum coating unit and the replica stabilized with a thin carbon film deposited at 90°. The thicknesses of evaporated tungsten and carbon were measured with a quartz film thickness monitor (Nanotech). The mica surface was scored into small square replicas, and these were floated onto the surface of distilled water before transfer to copper/rhodium grids. Micrographs were taken at 80 kV on a JEOL 1200 EX electron microscope. The magnification was calibrated with negatively stained catalase crystals. Dimensions of well spread molecules were measured with an eyepiece graticule or odometer, either from projected negatives or enlarged photographic prints, and expressed as the mean standard deviation (SD) and the number of measurements (*n*).

RESULTS

Monomeric Galectin-3 in Solution. Preliminary analysis of the monomer/oligomer status of hamster galectin-3 and the Δ1−93 protein in solution was carried out by gel-filtration chromatography on calibrated columns of Biogel P60 or Sephadex G50 using protein samples at initial concentrations up to 20 μM. In all cases single peaks eluting as monomeric species were detected by protein determination or Western blotting with galectin-3 specific antibodies of column fractions (results not shown). To eliminate possible shape effects on elution profiles that might obscure detection of oligomeric species, and effects of sample dilution leading to dissociation of weak complexes, analytical ultracentrifugation was performed. At each of three rotor speeds (22000, 29000, and 35000 rpm) both galectin-3 and Δ1−93 behaved as simple monomeric species with no evidence for higher oligomers. The global fit of the data was consistent with molecular weight values of 24 700 ± 1200 for galectin-3 and 16 717 ± 800 for Δ1−93 protein, close to the calculated monomeric values.

NMR Analysis of Galectin-3 and Its Domains. (i) N-Terminal Domain Containing Residues 1−125 (Δ126−245). The chemical shifts observed in the ^{1}H NMR spectra of recombinant Δ126−245 protein recorded at pH 7.5 and 5.5 were similar to those expected for random coil peptides (Figure 2). No ring current shifted signals were observed indicating the absence of a folded globular structure containing aromatic residues. The observed line widths were very narrow, particularly in the spectrum recorded at pH 5.5 where line widths of less than 2 Hz were measured on some signals. This can be seen in Figure 2 which shows the assigned aromatic region of the ^{1}H spectrum of the domain recorded in D_2O solution at pH 5.5. The two tryptophan residues (Trp22 and Trp26) have signals with almost identical chemical shifts and the signals for the 11 tyrosine residues can all be seen to have fairly similar shifts close to the values expected for an unfolded protein. The spectrum had no residual signals for NH protons showing that these had been completely exchanged for deuterium, thus indicating the absence of secondary structure containing strong hydrogen bonds. A 2D NOESY spectrum showed no long-range interresidue cross-peaks consistent with Δ126−245 being a mobile flexible molecule in solution. Thus the NMR data clearly indicate that the recombinant N-terminal domain represented by Δ126−245 protein is unstructured in solution and exists as an interconverting mixture of conformations similar to those seen for random coil peptides.

(ii) Full-Length Galectin-3 Comprising Residues 1−245. A 2D TOCSY spectrum of full-length galectin-3 recorded

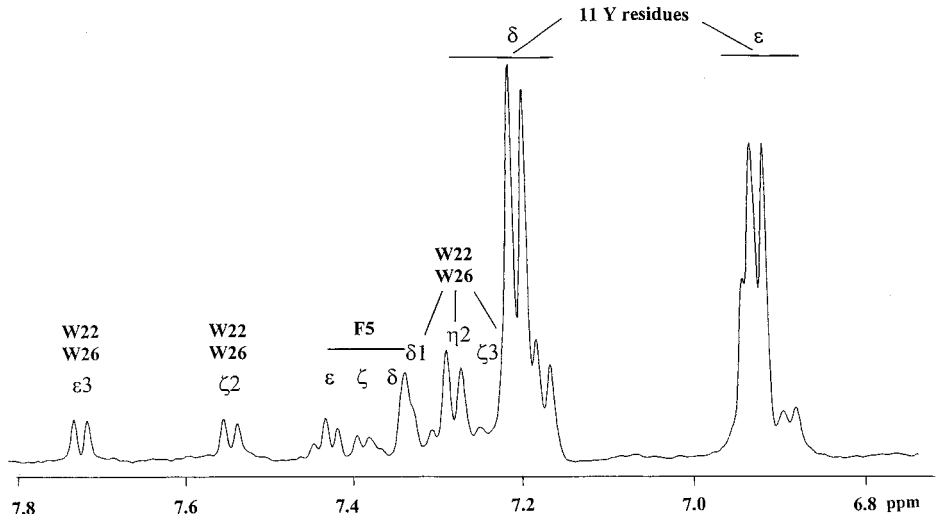

FIGURE 2: Aromatic region of the ^{1}H spectrum of the N-terminal 1−125 residue domain of galectin-3, the Δ126−245 protein in D_2O (∼0.2 mM, 500 MHz, pH 5.5, 25 °C).

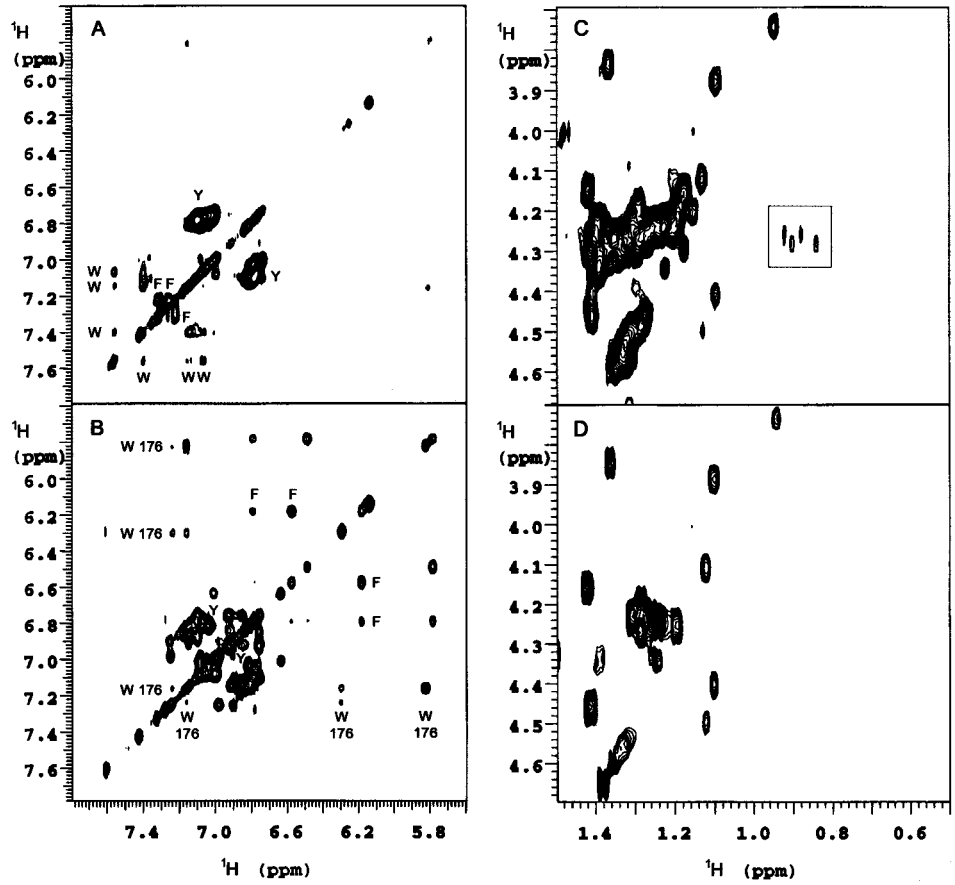

FIGURE 3: (A) The aromatic region of the 2D ^{1}H-^{1}H TOCSY spectrum for full-length galectin-3 (∼0.05 mM, 600 MHz, D_2O, 22 °C) showing signals from Phe5, and Trp22 and Trp26 (overlapped signals), and a cluster of 11 Tyr residues. (B) The corresponding spectral region for the Δ1−103 protein containing residues 104−245 (∼0.2 mM, 500 MHz, D_2O, 22 °C). (C) Part of the aliphatic spectral region for full-length galectin-3 with the signals for the δ/α cross-peaks of the two unstructured Leu residues shown in a box. (D) The corresponding spectral region for the Δ1−103 protein.

in D_2O showed fewer cross-peaks than expected for a protein of this size, approximately 30 kDa, although some of the cross-peaks were very sharp. For example the aromatic region of the spectrum (Figure 3A) showed only the cross-peaks expected for residues in the N-terminal (residues 1−125) domain: thus, only signals for Phe5, Trp22, Trp26, and numerous Tyr residues were observed and these had the ^{1}H chemical shifts of random coil peptides. Because the intact galectin-3 is not very soluble these measurements were made at a concentration of ∼0.1 mM. It is possible that even at this concentration there is some aggregation that could cause the line widths of the core of the protein to be too broad for successful TOCSY signal detection. In contrast, the 2D TOCSY spectrum of the isolated CRD (Δ1−103 protein) showed many well resolved cross-peaks not detected in the spectrum of intact galectin-3. This is clearly seen by

comparing the aromatic region of the spectrum of Δ1−103 protein (see Figure 3B) with that of intact galectin-3 (Figure 3A). Figure 3B shows the absence of the signals previously detected for Phe5, Trp22, Trp26, and most of the Tyr residues and the presence of several other signals such as for the single tryptophan residue Trp176 and Phe and Tyr residues in the CRD core of the protein, many with well-dispersed chemical shifts as expected for a folded protein. A similar comparison can be made for the aliphatic region of the spectrum (Figure 3, panels C and D). In the spectrum of the intact galectin-3, there are two sets of cross-peaks that arise from very sharp Leu δ/α cross-peaks (indicated in Figure 3C) and these have the chemical shifts expected for two Leu residues in an unfolded part of the protein. The δ/γ and α/β cross-peaks of these Leu residues were also clearly seen in other regions of the spectrum of the intact galectin-3 (not shown). None of these signals were observed in the spectrum of the CRD core, Δ1−103 (see Figure 3D). The data indicate that in the intact galectin-3, two Trp, one Phe, two Leu, and many Tyr residues in the N-terminal domain are in an unstructured state similar to that seen in the isolated N-terminal domain, the Δ126−245 protein. If one assumes that the two sets of leucine sharp cross-peaks arise from the N-terminal tail residues Leu7 and Leu11 in intact galectin-3, then this would mean that Leu109 is not an unstructured residue.

(iii) Truncated Protein Containing Residues 94−245 of Galectin-3 (Δ1−93). Heteronuclear {^{1}H}-^{15}N NOE HSQC experiments on the Δ1−93 protein (Figure 4, panels A and B) were carried out to investigate the backbone mobility of this protein, including the neighboring residues to Leu109. For a globular protein of this size, the correlation time would result in negative {^{1}H}-^{15}N NOE values of 0.2 (reducing the observed signal to 0.8 its original intensity). Nuclei showing substantially larger negative NOE values would be expected in parts of the protein having additional motion, for example in flexible tails or loops. Figure 4B shows the {^{1}H}-^{15}N NOE HSQC signals only for the Δ1−93 protein nuclei that have such larger negative NOE values. This is a difference spectrum obtained by subtracting 0.8 times the control HSQC spectrum with no heteronuclear NOE from the HSQC spectrum with the full heteronuclear NOE as described in the Materials and Methods. Signals from ^{15}N nuclei with negative heteronuclear NOE values of 0.20, expected for nuclei in the core of the protein, will be absent from this difference spectrum thus making it easy to visualize those nuclei with larger negative NOEs. There are 11 signals showing larger negative NOEs and these fall in the ^{1}H chemical shift region typical for unfolded residues (~8.0−8.4 ppm). These signals are fully consistent with the N-terminal residues 94−107 having increased mobility and with the observed signals corresponding to 5 alanines (123−125 ppm, ^{15}N chemical shifts), 2 tyrosines (120−121 ppm), 1 threonine (114−115 ppm), and 3 glycine (108−112 ppm) residues in the tail fragment:

$$G^{94}A^{95}Y^{96}P^{97}A^{98}A^{99}G^{100}P^{101}Y^{102}G^{103}A^{104}P^{105}T^{106}G^{107}\text{-}$$
$$\text{ALTVPY}$$

For most of the residues it was possible to confirm the assignments from 3D ^{1}H-^{15}N HSQC−NOESY experiments which showed NOEs from the NH protons to α- and β-protons. The ^{1}H chemical shifts for α- and β-protons

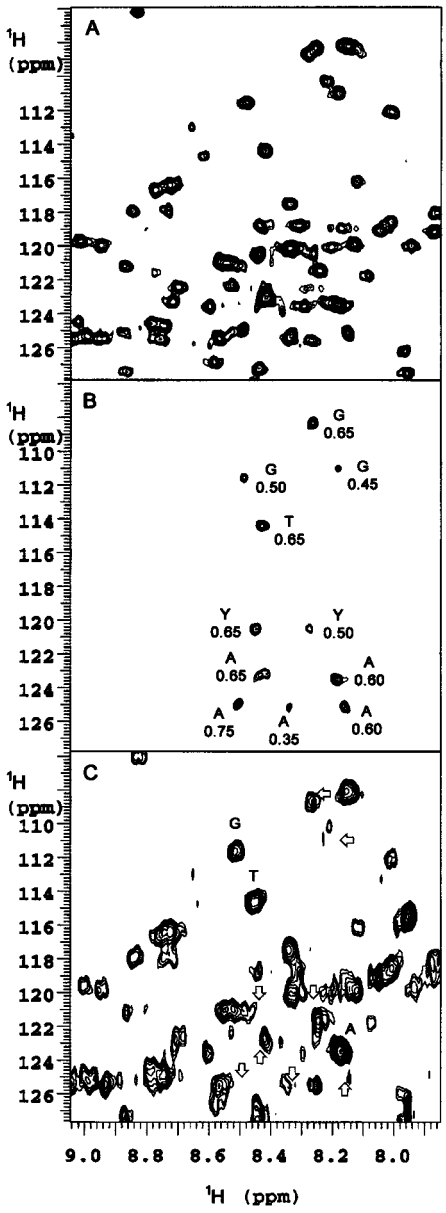

FIGURE 4: (A) 2D ^{1}H-^{15}N HSQC spectrum for the Δ1−93 protein containing residues 94−245 (~0.2 mM, 600 MHz, 90% H_2O/10% D_2O, 5 °C). (B) The corresponding region of the heteronuclear {^{1}H}-^{15}N NOE difference spectrum for Δ1−93 protein. This spectrum was obtained by subtracting 0.8 times the HSQC control spectrum with no heteronuclear NOE (shown in Figure 4A) from the spectrum with the full heteronuclear NOE as described in Materials and Methods. The assignments to residue type and the values of the negative NOEs are indicated on the spectrum. (C) The corresponding region of the ^{1}H-^{15}N HSQC spectrum for the Δ1−103 protein (500 MHz, 90% H_2O/10% D_2O, 5 °C) obtained after collagenase-treatment of the Δ1−93 protein. The positions of missing signals (compared with Figure 4A) are indicated by open-headed arrows on the spectrum.

detected in this way were all typical values for unfolded residues. There are no backbone amide protons for either the Pro residues or the terminal Gly94 residue.

A fragment comprising residues 104−245 (Δ1−103), obtained by collagenase treatment of the Δ1−93 protein was also examined by NMR. Since this sample no longer contains residues 94−103 it could be used to confirm the assignments of the signals in Figure 4B to residues in the tail region of the Δ1−93 protein. The ^{1}H-^{15}N HSQC spectra of the Δ1−93 protein and its collagenase-derived product Δ1−103

(shown in Figure 4, panels A and C, respectively) were compared and the missing signals in Figure 4C (shown by arrows) were found to correspond to the relevant signals of the flexible residues in the spectrum of Δ1−93 protein (Figure 4B). In the spectrum of the collagenase-treated sample one would expect removal of NH signals for two glycines, four alanines (including that of the terminal Ala104) and two tyrosines: signals for Thr106, Gly107, and Ala108 should be retained. The removal of the relevant signals is clearly seen in the spectrum of the collagenase-treated sample (Figure 4C) with signals assigned to Ala104, Thr106, and Gly107 being retained as expected. Thus the ^{1}H-^{15}N HSQC spectra of the Δ1−93 protein and its collagenase-treated product, Δ1−103, provide strong evidence that the signals in Figure 4B showing increased negative heteronuclear NOE values arise from residues in the relatively mobile flexible N-terminal tail of the Δ1−93 protein.

The heteronuclear {^{1}H}-^{15}N negative NOE values of the various signals are indicated in Figure 4B where it can be seen that the negative NOEs range from 0.35 to 0.75. Such measurements have been used previously to characterize motions in flexible tails of proteins. For example, Coles and co-workers (*52*) used this approach to study the dynamics of human neutrophil gelatinase-associated lipocalin: this 175 residue protein was shown to have a flexible 10 residue N-terminal tail which gave negative NOE values of up to 1.75 (at 600 MHz). Interestingly, these negative heteronuclear NOE values are much larger than the values measured here for the 15-residue tail in the Δ1−93 protein (where the largest value is 0.75) indicating somewhat impeded motion. Attenuated motions in protein tails have previously been reported in other proteins such as plasminogen-activator protein staphylokinase (*53*). In this case, a 19-residue N-terminal tail on a 136 residue protein showed negative heteronuclear NOEs of up to 0.8 (at 750 MHz) and ^{1}H-^{1}H NOE effects could be used to show that the tail was interacting with residues in the protein core.

Similar heteronuclear {^{1}H}-^{15}N NOE HSQC measurements were carried out on the Δ1−93 protein in the presence of the A-active tetrasaccharide GalNAcα1,3 [Fucα1,2]Galβ1,-4GlcNAc. Comparisons with the results from the Δ1−93 sample in the absence of sugar showed that there were no significant differences in the chemical shifts or mobilities of the N-terminal tail residues. Evidently, any realignment of the N-terminal tail on sugar binding (*22*) does not change significantly the mobility of the tail residues.

Electron Microscopy of Galectin-3 and Its Domains. Rotary shadowing proved to be the most suitable method for examining the overall structure of galectin-3. The N-terminal domain protein Δ126−245 was visualized as fibrils of varying thickness (Figure 5, panels E−H). In most images, the fibrils were considerably thicker than expected for a single extended polypeptide indicating oligomeric species with side-to-side associations of Δ126−245 subunits. In addition, the lengths of the fibrils varied indicating a head-to-tail association of subunits. Statistical analysis of images indicated a mean length of 158 ± 88 nm for all categories (n = 23). The large standard deviation was reduced somewhat when molecules of length greater than 200 nm were excluded from the analysis: in this case the mean length was 100 ± 34 nm (n = 17). Further refinement came from analysis only of molecules with lengths less than 100 nm:

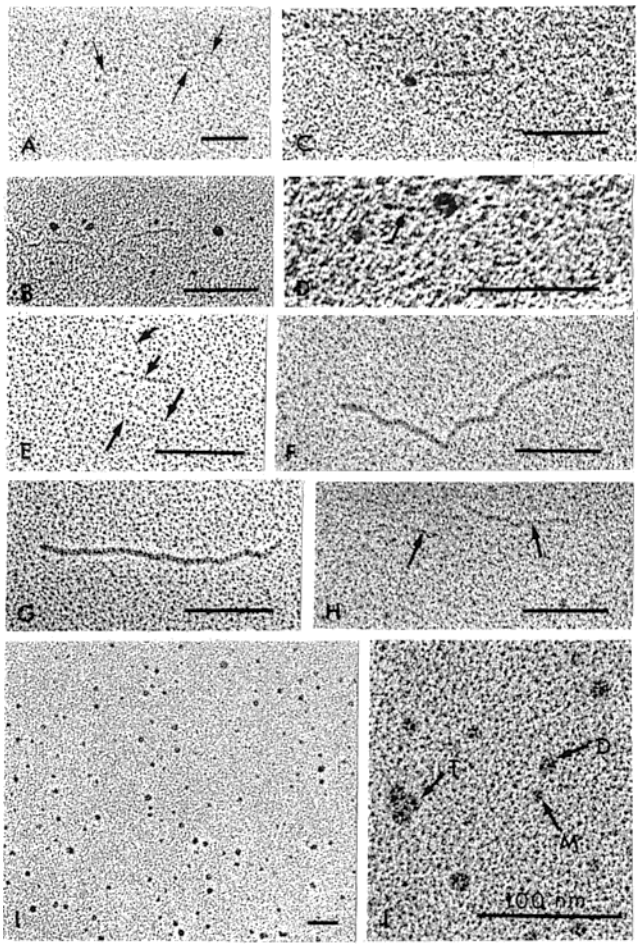

FIGURE 5: Electron micrographs of rotary shadowed preparations of galectin-3(A−D), N-terminal fragment Δ126−245(E−H) and CRD fragment Δ1−103(I, J). Arrows in panel A identify groups of molecules at low magnification, some of which are tail-to-tail dimers (see also in panel B). The formation of fibrils by the N-terminal fragment is indicated by arrows. Monomers (M), dimers (D), and trimers (T) of the CRD fragment are indicated (Panel J).

in this case, the mean length was 52 ± 17 nm (n = 13). These calculations suggest that the fibrils are composed of multiples of a unit length of about 50 nm, close to the theoretical length of a fully extended polypeptide of 125 amino acid residues. In an attempt to characterize single units, we also examined the Δ126−245 protein in its complex with a monoclonal Fab fragment that recognizes an N-terminal epitope of galectin-3 (Figure 6, panels A−C). The Δ126−245 molecules appeared to have Fab tagged at one end of the unit, which were now thinner and more uniform in length. The mean length was calculated as 48 ± 14 nm (n = 27).

The rotary shadowed images of the CRD protein Δ1−103 (Figure 5, panels I and J) showed structures consistent with the presence of monomers as well as dimers and trimers (respectively, 20 and 10% of total images) of the basic unit. Allowing for the thickness of the evaporated metal (about 4 nm) and assuming the molecular unit to be spherical then the average diameter of a monomer would be about 3.2 nm and that of a dimer about 6.4 nm.

The rotary shadowed images of the full-length recombinant galectin-3 showed the presence of complex structures (Figure 5, panels A−D). The simplest image was best interpreted as the sum of the structures obtained for the CRD and the N-terminal domain alone. Thus, the basic shape and pre-

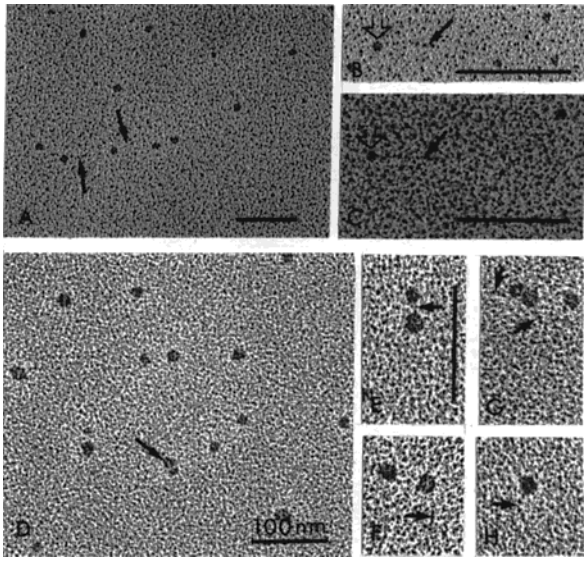

FIGURE 6: Electron micrographs of rotary shadowed preparations of N-terminal fragment Δ126−245 in complex with the Mac2 monoclonal antibody Fab fragment (A−C), and full-length galectin-3 in complex with the A-active tetrasaccharide (D−H). In panels A−C, the open-headed arrow and the black arrows indicate, respectively, the Fab fragment and the extended N-terminal polypeptide. Head-to-head dimers of full-length galectin-3 are indicated by arrows in panels D and E and the extended N-terminal tails of monomeric or dimeric molecules are indicated by arrows in panels G and H.

dominant form of the molecule was that of a globular head of diameter about 3−4 nm attached to a thin tail about 45−50 nm in length. More complex aggregates of this basic unit appeared to form including tail-to-tail dimers, but these were a minor proportion (approximately 10%) of the images. The lectin was preincubated with the A-active tetrasaccharide GalNAcα1,3 [Fucα1,2]Galβ1,4GlcNAc in an attempt to approximate more closely the situation pertaining to the interaction of full-length galectin-3 with a glycoprotein substrate, such as a laminin-coated substratum. The A-active tetrasaccharide is a high-affinity ligand for galectin-3 (*17*, *54*−*57*), as are the glycans of glycoproteins such as laminin (*21*, *56*, *58*). The overall unit structure again corresponded to the expected sum of the images of the individual N-terminal and CRD domains (Figure 6, panels D−H). Also present was a low proportion (approximately 10%) of asymmetric head-to-head dimers, each with a thin tail.

DISCUSSION

The NMR data presented in this paper provide evidence that residues of the N-terminal domain of hamster galectin-3 have great conformational flexibility. The small variations in chemical shifts of the N-terminal domain resonances and their similarity to random coil chemical shifts strongly support other biophysical data using circular dichroism, tryptophan fluorescence spectroscopy, and differential scanning calorimetry (*24*, *59*) that these extensions have little ordered conformation. However, although the observed heteronuclear NOEs indicate increased motion in the N-terminal tail of Δ1−93 protein compared with the CRD core of the protein, the reduced NOE values indicate that their motions are somewhat impeded compared to those observed in the fully flexible tails of other proteins. This indicates that the tail of Δ1−93 protein interacts to some extent with the core of the protein. It should be noted that the ^{1}H chemical shifts for some residues in the N-terminal tail of Δ1−103 (Trp22, Trp26, and Phe5) indicate that their nuclei are slightly more shielded (∼0.1 ppm) for galectin-3 compared with the isolated N-terminal domain, Δ126−245. This also suggests that residues in the N-terminal tail of intact galectin-3 are making transient interactions with the protein core.

Our finding that Leu109 is not detected as a mobile residue, either in full-length hamster galectin-3 or in the Δ1−93 protein is consistent with an interaction of this region of the tail sequence with the CRD. In the crystallographic analysis of human galectin-3 (*16*), the collagenase-derived fragment containing residues 107−250 (equivalent to residues 103−245 of hamster galectin-3, Figure 1) was examined. Leu114 (equivalent to hamster residue Leu109, Figure 1) was the first observable N-terminal residue, and the electron density map strongly indicated that the portion of the N-terminal polypeptide containing this residue and other hydrophobic residues of the sequence LIVPY (LTVPY in hamster, Figure 1) turns into the hydrophobic cleft formed by the first (S1) and last (S12) β-strands of the CRD proper. Such a folding could obstruct the hydrophobic surface of the interface between the S1 and S12 β-strands, utilized in the self-association of galectin-1 or -2 subunits, thus preventing dimerization of galectin-3. In their X-ray study, Seetharaman and co-workers (*16*) observed no electron density for the first six residues GAPAGP (equivalent to GAPTGA in hamster galectin-3), indicating these are disordered at least in the collagenase-derived fragment. Our NMR data suggest that, although these residues are more mobile than the residues LTVPY following on from Leu109, they are measurably more restricted than expected for residues in a fully flexible N-terminal tail, implying some interaction with the relatively less mobile CRD core. This conclusion applies both to residues APTGA of the Δ1−103 protein and additional residues in the longer tail of the Δ1−93 protein (Figure 1) and is consistent with previous modeling and mutagenesis studies (*17*, *22*) suggesting that residues surrounding Tyr102 of hamster galectin-3 or the Δ1−93 protein interact with the CRD core.

Two models for the packing of the N-terminal domain of galectin-3 into the whole structure of galectin-3 have been proposed (*22*). One is the model discussed above and the other is a model in which the N-terminal tail turns away from the S1/S12 β-strand interface to fold across the carbohydrate-binding pocket of the CRD. Recently, we suggested that the stabilization of the alternative configuration of the N-terminal tail by binding to high-affinity carbohydrate ligands might expose hydrophobic surfaces, such as the S1/S12 β-strand interface, for interactions between galectin-3 CRD subunits (*22*). The head-to-head dimers of full-length galectin-3 bound to the A-active tetrasaccharide, as well as the dimers and trimers of the isolated CRD (Δ1−103 protein), shown in the electronmicrographs provide some evidence that these structures do indeed form. We also suggested that CRD−CRD domain associations might contribute to the positive cooperativity observed in the binding of galectin-3 to multivalent ligands (*21*, *22*, *28*). Oligomeric galectin-3 formed in this way would be expected to bind more strongly to a highly glycosylated ligand than would the monomeric lectin, consistent with the

observed two-phase kinetics of binding of galectin-3 to a laminin substratum (*31*) as well as results from oligosaccharide-inhibition assays (*22*). However, it has been shown clearly that binding cooperativity is dependent on the presence of the N-terminal tails of galectin-3 (*21*−*23*, *31*). Our present finding that the recombinant N-terminal fragment can undergo extensive fibrillogenesis is consistent with this proposal, especially inasmuch as these domains appear to associate to some extent in the full-length lectin.

ACKNOWLEDGMENT

We thank Sarah A. Major and Andrew Lane for the ultracentrifugation experiments and analysis, and Vladimir Polshakov and Tom Frenkiel for assistance with the NMR experiments.

REFERENCES

1. Liu, F. T. (1993) *Immunol. Today 14*, 486−490.
2. Barondes, S. H., Cooper, D. N., Gitt, M. A., and Leffler, H. (1994) *J. Biol. Chem. 269*, 20807−20810.
3. Kasai, K.-I., and Hirabayashi, J. (1996) *J. Biochem. Jpn. 119*, 1−8.
4. Hughes, R. C. (1997) *Biochem. Soc. Trans. 25*, 1194−1198.
5. Leffler, H. (1997) *Trends Glycosci. Glycotech. 9*, 9−19.
6. Perillo, N. L., Marcus, M. E., and Baum, L. G. (1998) *J. Mol. Med. 76*, 402−412.
7. Rabinovich, G. A. (1999) *Cell Death Different. 6*, 711−721.
8. Gritzmacher, C. A., Mehl, V. S., and Liu, F. T. (1992) *Biochemistry 31*, 9533−9538.
9. Herrmann, J., Turck, C. W., Atchison, R. E., Huflejt, M. E., Poulter, L., Gitt, M. A., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993) *J. Biol. Chem. 268*, 26704−26711.
10. Wang, J. L., Laing, J. G., and Anderson, R. L. (1991) *Glycobiology 1*, 243−252.
11. Wang, J. L., Werner, E. A., Laing, J. G., and Patterson, R. J. (1992) *Biochem. Trans. 20*, 269−274.
12. Mehul, B., and Hughes, R. C. (1997) *J. Cell Sci. 110*, 1169−1178.
13. Menon, R. P., and Hughes, R. C. (1999) *Eur. J. Biochem. 264*, 569−576.
14. Gong, H. C., Honjo, Y., Nangia-Makker, P., Hogan, V., Mazurak, N., Bresalier, R. S., and Raz, A. (1999) *Cancer Res. 59*, 6239−6245.
15. Lobsanov, Y. D., and Rini, J. M. (1997) *Trends Glycosci. Glycotech. 9*, 145−154.
16. Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998) *J. Biol. Chem. 272*, 13047−13052.
17. Henrick, K., Bawumia, S., Barboni, E. A. M., Mehul, B., and Hughes, R. C. (1998) *Glycobiology 8*, 45−57.
18. Rini, J. M., and Lobsanov, Y. D. (1999) *Curr. Opin. Struct. Biol. 9*, 578−584.
19. Giudicelli, V., Lutomski, D., Levi-Struss, M., Bladier, D., Joubert-Caron, R., and Caron, M. (1997) *Glycobiology 7*, VIII−X.
20. Ochieng, J., Platt, D., Tait, L., Hogan, V., Raz, T., Carmi, P., and Raz, A. (1993) *Biochemistry 32*, 4455−4460.
21. Massa, S. M., Cooper, D. N. W., Leffler, H., and Barondes, S. H. (1993) *Biochemistry 32*, 260−267.
22. Barboni, E. A. M., Bawumia, S., Henrick, K., and Hughes, R. C. (2000) *Glycobiology 10*, 1201−1208.
23. Hsu, D. K., Zuberi, R. I., and Liu, F. T. (1992) *J. Biol. Chem. 267*, 14167−14174.
24. Mehul, B., Bawumia, S., Martin, S. R., and Hughes, R. C. (1994) *J. Biol. Chem. 269*, 18250−18258.
25. Mehul, B., Bawumia, S., and Hughes, R. C. (1995) *FEBS Lett. 360*, 160−164.
26. Yang, R. Y., Hill, P. N., Hsu, D. K., and Liu, F. T. (1998) *Biochemistry 37*, 4086−4092.
27. Kuklinski, S., and Probstmeir, R. (1998) *J. Neurochem. 70*, 814−823.
28. Probstmeier, R., Montag, D., and Schachner, M. (1995) *J. Neurochem. 64*, 2465−2474.
29. Ochieng, J., Leite-Browning, M. L., and Warfield, P. (1998) *Biochem. Biophys. Res. Commun. 246*, 788−791.
30. LeMarer, N., and Hughes, R. C. (1996) *J. Cell. Physiol. 168*, 51−58.
31. Barboni, E. A. M., Bawumia, S. and Hughes, R. C. (1999) *Glycoconjugate J. 16*, 365−372.
32. McRorie, D. K., and Voelker, P. J. (1993) *Self-associating systems in the analytical ultracentrifuge*, Beckman Instruments Inc.
33. Rance, M., Sörensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479−485.
34. Davis, D. G., and Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820−2821.
35. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson. 65*, *355*−*360*.
36. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
37. Macura, S., Huong, Y., Suter, D., and Ernst, R. E. (1981) *J. Magn. Reson. 43*, 259−281.
38. Bothner-By, A. A., Stephens, R. L., Lee, J., Warren, C. D., and Jeanloz, R. W. (1984) *J. Am. Chem. Soc. 106*, 811−813.
39. Brown, S. C., Weber, P. L., and Mueller, L. (1988) *J. Magn. Reson. 77*, 166−169.
40. Bodenhausen, G., and Ruben, D. L. (1980) *Chem. Phys. Lett. 69*, 185−189.
41. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
42. Mori, S., Abeygunawardana, O'Neil Johnson, M., and Van Zijl, P. C. M. (1995) *J. Magn. Reson. 108*, 94−98 (Series B).
43. Shaka, A. J., Barker, P. B., and Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
44. States, D. J., Haberkorn, R. A., and Ruben D. J. (1982) *J. Magn. Reson. 48*, 286−292.
45. Morris, G. A., and Freeman, R. (1978) *J. Magn. Reson. 29*, 433−462.
46. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) *J. Magn. Reson. 102*, 241−245 (Series A).
47. Live, D. H., Davis, D. G., Agosta, W. C., and Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939−1941.
48. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1996) *J. Biomol. NMR 6*, 135−140.
49. Polshakov, V. I., Williams, M. A., Gargaro, A. R., Frenkiel, T. A., Westley, B. R., Chadwick, M. P., May F. E. B., and Feeney, J. (1997) *J. Mol. Biol. 267*, 418−432.
50. Kay, L. E., Torchia, D. A., and Bax, A. (1989) *Biochemistry 28*, 8972−8979.
51. Rutman, A. J., Buxton, R. S., and Burdett, I. D. J. (1994) *FEBS Lett. 353*, 194−196.
52. Coles, M., Diercks, T., Muehlenweg, B., Bartsch, S., Zölzer, V., Tschesche, H., and Kessler, H. (1999) *J. Mol. Biol. 289*, 139−157.
53. Ohlenschläger, O., Ramachandran, R. G. Gührs, K−H., Schlott, B., and Brown, L. (1998) *Biochemistry 37*, 10635−10642.
54. Leffler, H., and Barondes, S. H. (1986) *J. Biol. Chem. 261*, 10119−10126.
55. Sparrow, C. P., Leffler, H., and Barondes, S. H. (1987) *J. Biol. Chem. 262*, 7383−7390.
56. Sato, S., and Hughes, R. C. (1992) *J. Biol. Chem. 267*, 6983−6990.
57. Feizi, T., Solomon, J. C., Yuen, C. T., Jeng, K. C. G., Frigeri, L. G., Hsu, D. K., and Liu, F. T. (1994) *Biochemistry 33*, 5346−6349.
58. Woo, H. J., Shaw, L. M., Messier, J. M., and Mercurio, A. M. (1990) *J. Biol. Chem. 265*, 7097−7099.
59. Agrawal, N., Sun, Q., Wang, S. Y., and Wang, J. L. (1993) *J. Biol. Chem. 268*, 14932−14939.

BI002907F