Nuclear Magnetic Resonance Structural and Ligand Binding Studies of BLBC, a Two-Domain Fragment of Barley Lectin†

Jeanne Lim Weaver and J. H. Prestegard*

Department of Chemistry, Yale University, New Haven, Connecticut 06511

Received July 3, 1997; Revised Manuscript Received October 27, 1997⊗

ABSTRACT: Plant lectins are useful targets for biophysical studies of protein−carbohydrate recognition, a process of general interest because of its many roles in human physiology. Here, nuclear magnetic resonance (NMR) based structural and carbohydrate binding data on a two-domain fragment of the normally four-domain barley lectin protein are presented. The structural data, while preliminary, clearly shows that the recombinantly produced simplified model system, called BLBC, retains a nativelike fold. However, unlike the full-length parent protein, which is dimeric, BLBC is shown by pulsed-field gradient NMR diffusion studies to be largely monomeric. Still, the fragment retains nativelike carbohydrate binding properties. These properties are examined in some detail using heteronuclear single quantum coherence (HSQC) NMR spectroscopy on a uniformly ^{15}N-labeled sample. Ligand-induced chemical shift changes in the ^{1}H−^{15}N HSQC spectrum are monitored as ^{15}N-labeled BLBC is titrated with increasing concentrations of the unlabeled carbohydrate, *N*,*N*′,*N*″-triacetylchitotriose. Well-resolved resonances from the individual domains show that BLBC binds ligand at two distinct and independent ligand binding sites, one in each domain. Binding constants of $(1.1 \pm 0.2) \times 10^3$ M^{-1} and $(0.6 \pm 0.2) \times 10^3$ M^{-1} are determined for the B and C domain sites, respectively. These results are discussed in relation to ligand binding studies that have previously been carried out on a highly homologous protein, wheat germ agglutinin.

Lectins are carbohydrate-specific binding proteins found in all categories of living organisms (*1*, *2*). They were first discovered in plants over 100 years ago (*3*), and since then the plant lectins have frequently served as biophysical agents for the study of protein−carbohydrate recognition. In recent years, interest in biophysical application has heightened with the discovery that certain mammalian lectins play key roles in human physiology (*4*, *5*). For instance, both inflammation (*5*) and the innate immune responses (*6*) are initiated by specific carbohydrate−lectin interactions. Although the importance of these interactions is now unequivically acknowledged, carbohydrate recognition remains a difficult target for structural studies. While some structural data exists, they are often complicated by acquisition in states not strictly representative of the environment in which recognition occurs and by the inherent binding site heterogeneity of some of the lectins studied. We present here NMR[1] -based structural and carbohydrate binding data collected in solution on a carbohydrate binding fragment of barley lectin that has been engineered to present a simplified binding model for biophysical studies.

To date, most structural data on lectin−carbohydrate interactions has come from crystallographic sources, but here extrapolation to behavior in solution has sometimes been difficult due to concern about effects of crystal packing. Modern NMR techniques, coupled with advances in molecular biology, now present an alternative for structural characterization which may be pursued directly in solution. Furthermore, NMR techniques are readily applicable to studying a variety of carbohydrates in association with lectin and are extendable to studying interactions at the surface of a model membrane (*7*), where many lectin−carbohydrate interactions are thought to occur. Currently, however, the application of NMR methods is limited to proteins of moderate size. Since most carbohydrate binding proteins are relatively large and frequently multimeric (*8*), quality NMR data on lectin−carbohydrate interactions has been difficult to obtain.

For the purpose of using NMR to study carbohydrate recognition, we have engineered a two-domain fragment of the protein barley lectin (BL). This protein, which has not itself been well-characterized, is a member of the highly conserved class of proteins known as the cereal lectins and shares approximately 95% sequence homology with wheat germ agglutinin (WGA), a lectin that has been extremely well-characterized (see Figure 1) (*9*−*12*). Thus, NMR studies of the BL fragment may not only serve to delineate

† This research was supported by a grant from the National Institutes of Health, GM33225.

* To whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* December 15, 1997.

[1] Abbreviations: BL, barley lectin; BLBC, a B and C domain fragment of barley lectin; WGA, wheat germ agglutinin; WGA-A, isolectin A of wheat germ agglutinin; WGA-B, isolectin B of wheat germ agglutinin; WGA-D, isolectin D of wheat germ agglutinin; HEV, hevein; UDA, *Urtica dioica* agglutinin (aka nettle lectin); Ac-AMP2, lectin from *Amaranthus caudatus*; NMR, nuclear magnetic resonance; 1D, one-dimensional; 2D, two-dimensional; HSQC, heteronuclear single quantum coherence spectroscopy; NOESY-HSQC, nuclear Overhauser enhancement spectroscopy−heteronuclear single quantum coherence spectroscopy; TOCSY-HSQC, total correlation spectroscopy−heteronuclear single quantum coherence spectroscopy.

S0006-2960(97)01619-X CCC: $15.00
Published on Web 01/06/1998

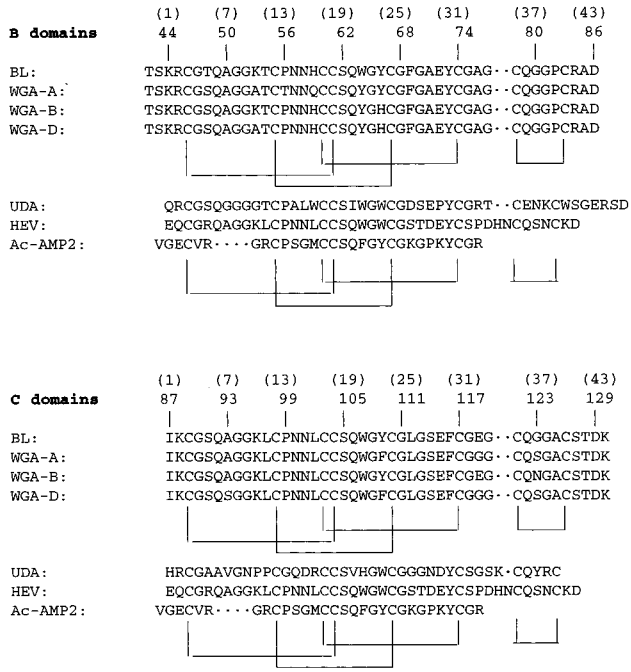

FIGURE 1: Amino acid sequences of homologous lectins. Numbering in parentheses reflects the domain positions of the residues, whereas the numbering below these reflects the sequential order of the barley lectin (BL) and WGA amino acid sequences. For UDA (*Urtica dioica agglutinin*, or nettle lectin), a protein consisting of two hevein-like domains in tandem, the N-terminal and C-terminal domains are presented with the B and C domains of BL, respectively. Hevein and Ac-AMP2 are single-domain proteins; their sequences are repeated under both the B and C domain headings for ease of comparison. The disulfide linkage pattern is indicated below the sequences of each domain. The cloned fragment of barley lectin, BLBC, encompasses residues Thr42-Lys130 of the full-length four-domain protein.

the structural and ligand binding properties of the native protein but also may address questions of carbohydrate recognition that have been raised from work on WGA (*vide infra*).

The molecular structure of BL may be anticipated from its close homology to WGA. Both proteins, which bind selectively to *N*-acetyl-D-neuraminic acid and *N*-acetyl-D-glucosamine moieties, exist as 36 kDa dimers composed of two identical polypeptide chains (*13*, *14*). Each chain consists of four homologous 43-residue domains, termed A, B, C, and D as read from the N to the C terminus, and each domain is stabilized by a network of four disulfide bonds. The chains associate noncovalently in a "head-to-tail" fashion to form a dimeric molecule with inter-subunit contacts between domains A_1 and D_2, B_1 and C_2, C_1 and B_2, and D_1 and A_2 (subscripts indicate the independent polypeptide chains of the dimer) (*15*). According to X-ray studies of WGA, each of these dimer interfaces accommodates two carbohydrate binding sites (*15*−*17*). For each of these sites, one of the domains serves as the "principal" (pr) binding domain which contributes a conserved serine from domain position 19 as well as 2−3 nonconserved aromatic residues from domain positions 21, 23, and 30 (*17*). The opposing domain, if it has an acidic group at domain position 29, may participate in ligand binding by serving as a "helper" (hl) binding domain. Eight independent ligand binding sites, which may be grouped into four unique pairs, can be extracted from crystallographic data—2 A(pr), 2 B(pr)/C(hl), 2 C(pr)/B(hl), and 2 D(pr)/A(hl). Although not all are occupied simultaneously in crystals, ligand has been observed to bind at every location, and all sites are believed to be functional (*15*−*17*).

While multiple crystallographic and solution studies of ligand binding to WGA have been made over the past 2 decades, it has proved extremely challenging to link the results of the different methods to produce a coherent binding picture. From X-ray studies, the most frequently occupied binding sites of WGA have been determined to be the 2 B(pr)/C(hl) sites followed by the 2 D(pr)/A(hl) sites (*15*−*17*). Indeed, until recently, occupation of the A(pr) or the C(pr)/B(hl) sites was not observed (*17*). From solution studies, the number of functional ligand binding sites in WGA has been found to be four, rather than eight, per dimeric lectin, and these appear to be of equal affinities (*14*, *18*−*20*). While it is tempting to say that the infrequently occupied A(pr) and C(pr)/B(hl) sites must be too weak to be detected by solution methods and that the highly occupied B(pr)/C(hl) and D(pr)/A(hl) sites must correspond to the four equivalent ligand binding sites, such a connection between the X-ray and solution data cannot be made deductively. Crystallographically determined site occupation data is complicated by competing crystal-packing interactions (*17*), and the solution data, which so far has not been structural in nature, cannot itself pinpoint the structural locations of the detected binding sites. The most current analysis of WGA binding comes from a theoretical modeling study and suggests that the equal affinity sites detected from solution studies correspond to the 2 B(pr)/C(hl) and 2 C(pr)/B(hl) sites (*21*). However, experimental data in support of this latest ligand binding model is lacking.

The recent focus on the B/C dimer interfaces as the locations of the high-affinity binding sites suggests that an NMR study of isolated BC domains may address questions pertaining to binding in the full-length parent. We have thus engineered a fragment of barley lectin that consists of just the internal B and C domains. This new system, which we call BLBC, is more manageable in terms of size and homogeneity of binding sites for NMR characterization than the native protein. We have used recombinantly produced ^{15}N-labeled product to obtain a preliminary characterization of the BLBC structure, which demonstrates that a nativelike fold has been retained in the fragment. We have further used this product to characterize the number, locations, and binding affinities of the fragment's ligand binding sites. The latter studies rely heavily on ligand-induced chemical shift changes in the protein ^{1}H−^{15}N HSQC NMR spectrum (*22*). While the interaction picture drawn from such data is rather qualitative, it nevertheless reveals valuable information about ligand binding and paves the way for studies of a more diverse set of ligands in the future.

EXPERIMENTAL PROCEDURES

Protein Preparation. The plasmid pET3d-BL encoding the full-length mature barley lectin protein was kindly given to us by Dr. Natasha Raikhel of Michigan State University (*10*, *11*). From this plasmid, the gene segment encoding just the B and C domains of the protein was isolated and subcloned as a maltose binding protein (MAL) fusion in the XmnI/BamHI site of the pMAL-p2 vector (New England

Biolabs; Beverly, MA). Recombinant DNA procedures used to create the new plasmid, pMAL-FXa-BLBC, came from Sambrook et al. (*23*) as well as guidelines given by New England Biolabs.

For the production of recombinant fusion protein (MAL-FXa-BLBC), pMAL-FXa-BLBC was transformed into *Escherichia coli* XL1-Blue (Stratagene; La Jolla, CA). Growth and purification protocols then followed the procedures outlined by New England BioLabs; however, for the production of ^{15}N-labeled protein, a minimal medium (*24*) using 1 g per L of $^{15}NH_4Cl$ (>98% ^{15}N; Cambridge Isotope Laboratories; Andover, MA) as the sole nitrogen source was substituted for the nutrient rich one. Expression of recombinant protein from this system appeared efficient—addition of IPTG to the growing cell cultures induced the production of a dark band at the molecular weight for the fusion product (54 kDa) as determined by SDS-PAGE. However, not all of the product was isolable from the first step of purification, which entailed osmotic shock to release periplasmic proteins followed by centrifugation to pellet the cell debris. After this step, much fusion protein remained associated with the debris. Future modifications to the osmotic shock process may result in higher yields.

To continue purification, supernatant from an osmotically shocked 5 L cell growth was applied to a column of Q-sepharose (1.5 cm wide × 9.5 cm high; Pharmacia; Uppsala, Sweden), pre-equilibrated in buffer containing 20 mM Tris, pH 8.0, 25 mM NaCl, and 1 mM EDTA. Purified fusion protein was eluted from this column using a 200 mL 25−500 mM NaCl gradient in column buffer. The amount of pure MAL-FXa-BLBC collected at this stage, based on the predicted absorbance at 280 nm (*25*), was 81 mg.

Purified MAL-FXa-BLBC was cleaved by addition of the specific protease factor Xa to a ratio of approximately 1:250 protease:fusion protein. The reaction was followed by SDS-PAGE and allowed to progress approximately 24 h, or until most of the starting material had been cleaved. Products from the digestion were separated and purified to greater than 90% purity through the sequential use of Q-sepharose (1.5 cm × 9.5 cm) and SP-tentacle (1.5 cm × 5 cm; EM Scientific; Gibbstown, NJ) columns in buffer containing 20 mM MOPS, pH 6.5, and 25 mM NaCl. Yields of pure BLBC after factor Xa cleavage and purification were typically 13−19% of the theoretical yield from pure fusion protein.

For NMR studies, which were carried out in Shigemi sample tubes (Tokyo, Japan), BLBC was concentrated to a volume of 160 μL and exchanged into NMR buffer (0.1 M potassium phosphate, pH 6.0, 0.15 M NaCl, and 0.1% NaN_3) using a Centricon-3 (Amicon; Beverly, MA). A typical total yield of purified BLBC from a 5 L minimal medium growth was 2 mg, or approximately 1 mM, at the volume used for NMR studies.

Electrospray mass spectrometry of BLBC demonstrated the purity of the sample and showed it to possess the anticipated molecular weight of 9 kDa. An Ellman's test (Pierce; Rockford, IL) assessed it to have few free thiols (none detected in a 0.02 mM solution of BLBC). From 20% SDS-PAGE Phastgels (Pharmacia LKB; Uppsala, Sweden) run in the absence of reducing agent, the protein appeared to exist as 60% monomer with some dimer and trimer species present as well (see Figure 2). The dimer and trimer protein

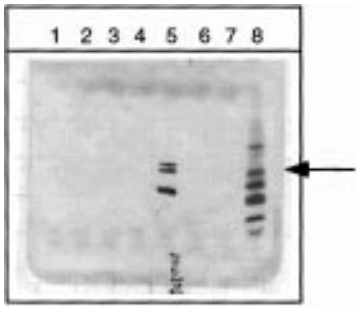

FIGURE 2: SDS Phastgel (20%) (silver stained) showing BLBC purified from the final SP-Tentacle cation exchange column (10 mL resin; EM Science; Gibbstown, NJ) used in its purification. Lane 1, flowthrough collected while the column was being loaded. Lane 2, buffer collected during the column wash that followed the sample loading. Lanes 3−7, column elution fractions 14−22, respectively. (BLBC was eluted from the column using a 100 mL 25−1500 mM NaCl gradient, and collected column volumes were approximately 1 mL). Lane 8, low molecular weight standards from Pharmacia (Uppsala, Sweden). From the band indicated by an arrow and moving downward, the markers indicate the mobilities of 16.9, 14.4, 8.2, 6.2, and 2.5 kDa proteins.

bands collapsed to coincide with the monomer upon dithiothreitol treatment. We believe that the higher order species are not present to a significant degree in the protein sample used for NMR studies but that a significant amount of these species arises from scrambling of disulfide bonds under the denaturing conditions of the SDS gel. This is supported by variable percentages of the minor species observed in different gels from the same preparation, by the fact that no higher order species are observed from electrospray mass spectrometry (which does not destroy disulfide bonds during the ionization process), and by the significantly lower percentage of oligomeric species observed in NMR NOESY data. Anomalous migration of non-reduced proteins by SDS-PAGE has been observed previously for the proteins Ac-AMP1 and Ac-AMP2 (*26*); as discussed later in this paper, these proteins are highly homologous to BLBC and also possess an unusually high cysteine content.

NMR Samples. Homonuclear experiments were collected on a 1 mM sample in a solvent of 100% D_2O, and ^{15}N-edited NMR experiments were collected on a 1.24 mM ^{15}N-labeled (>98%) BLBC sample in a solvent of 90% H_2O/10% D_2O. This latter sample was also subsequently used for the isotope-edited ligand binding studies. *N*,*N*′,*N*″-Triacetylchitotriose (2-acetamido-2-deoxy-β-D-glucopyranosyl (1 → 4)-2-acetamido-2-deoxy-β-D-glucopyranosyl(1 → 4)-2-acetamido-2-deoxy-β-D-glucopyranose; Sigma, St. Louis, MO) was aliquoted into the protein sample from a 32 mM stock solution prepared in the NMR buffer. This resulted in some dilution of the protein during the titration; however, dimerization of protein was known to be small at initial concentrations (*vide infra*) and minimal direct effects on chemical shift were expected from the dilution.

NMR Experiments. All NMR experiments were collected at 30 °C using an Omega 500, a Varian Unity 500, or a Varian Unity Plus 600 spectrometer equipped with a triple resonance probe and pulsed field gradients. ^{1}H data were typically collected using 11 ppm spectral widths centered at the frequency of H_2O. ^{15}N spectral widths ranged from 1368 to 5000 Hz depending on field strength and desired resolution and were centered at 115 ppm. All multidimensional experiments were collected in States·TPPI mode. For all

^{15}N-edited experiments, sensitivity enhanced pulsed-field gradient sequences were employed (*27*, *28*), with selective "flip-back" pulses to avoid saturation of amide signals (*29*) and WATERGATE modules for the suppression of solvent magnetization (*30*).

The following isotope-edited NMR experiments were acquired. Gradient selected HSQCs (*27*), collected over 512 complex data points in the ^{1}H dimension and 256 complex points in the ^{15}N dimension, were acquired for the initial structure determination of BLBC and after each aliquot of carbohydrate during the ligand titration study. A TOCSY-HSQC (*31*, *32*) was acquired using an 80 ms DIPSI-3 mixing scheme and 128, 45, and 512 complex data points in *t1* (^{1}H), *t2* (^{15}N), and *t3* (^{15}N-H), respectively. Two NOESY-HSQCs (*31*, *32*) were recorded with 100 and 200 ms mixing times. A total of 128 *t*1, 38 *t*2, and 512 *t*3 complex points were collected for the 100 ms data set, and 128 *t*1, 22 *t*2, and 512 *t*3 complex points were collected for the 200 ms data set. Acquisition times, using repetition rates of approximately 1.6 s, ranged from 4 h for the HSQCs to 48 h for the 3D experiments.

For the homonuclear experiments in D_2O, no solvent suppression was used. A clean-TOCSY (*33*−*35*) was acquired with an 80 ms MLEV16 + 60° spin lock mixing scheme over 2048 complex points in the directly detected ^{1}H dimension and 512 complex points in the indirectly detected ^{1}H dimension. Also collected was a 100 ms mixing time NOESY (*36*). This was acquired over 2048 and 310 complex points for the directly and indirectly detected ^{1}H dimensions, respectively. Acquisition times for the 2D experiments using repetition rates of 1.9 s were 24 h.

Data were processed using Felix 95 software (Biosym/Molecular Simulations; San Diego, CA). Typical apodizations included 90°-shifted sinebells spanning the collected size in all dimensions, and data sets were typically zero filled to two−four times the collected size after apodization. Most spectra were referenced indirectly to TMS (^{1}H dimension) and liquid ammonia (^{15}N dimension), both at 0 ppm (*37*). The sugar-containing HSQC spectra in the titration study were referenced to the Ala125 peak in the sugar-free spectrum. This peak resonated at 7.06 and 127.88 ppm in the ^{1}H and ^{15}N dimensions, respectively.

Pulsed-Field Gradient NMR Diffusion Measurements. The pulsed-field gradient NMR experiment described by Altieri et al. (*38*), with modifications to include an HSQC filter for the selective detection of ^{15}N-labeled proteins in H_2O, was used to measure the self-diffusion constant, D_s, for BLBC and, thus, derive information about its aggregation state. These measurements were taken several times over the span of several months for both labeled and unlabeled protein as well as for protein in the presence of increasing concentrations of carbohydrate ligand. The delay between defocusing and refocusing gradients (Δ) was set to 70 ms, and the duration of these gradients (δ) was set to 3 ms. A total of 32 experiments were collected for each D_s assessment with all parameters held constant from experiment to experiment except the gradient strength. This parameter took values from 0 to its maximum value (32 G/cm) and was incremented in each successive experiment. Further experimental details and explanations of how assessments of aggregation state are made have been given by Altieri et al. The analysis of pulsed-field gradient diffusion data has been described by Stejskal and Tanner (*39*). When possible, the accuracy of each measured value was assessed by measuring the diffusion constant for a sample of lysozyme, a protein of 14.3 kDa that is known to be monomeric and whose literature D_s value is 1.35×10^{-6} cm^2/s at 30 °C (*40*). Additionally, theoretical D_s values of BLBC monomer and dimer were calculated using the program HYDRO (*41*). For these calculations, the BLBC monomer was modeled as two beads of 10.5 Å separated by a 10 Å linker. These dimensions were taken from crude measurements of the X-ray structure of WGA-B (*42*) (PDB accession code 1WGT). HYDRO-calculated values for the BLBC monomer and dimer, respectively, at 30 °C were 1.77 and 1.40×10^{-6} cm^2/s. These values may be slightly overestimated as solution viscosities were assumed to be those of water.

Preliminary X-PLOR Structure Generation. NOEs were converted to distance constraints according to the following equations (*43*),

$$d = d_{\mathrm{std}}\left(\frac{V_{\mathrm{std}}}{V}\right)^{1/6} \quad (1)$$

$$d+ = 0.06d^2 \quad (2)$$

where d denotes the distance calculated from a particular NOE volume, V. The upper limit to deviations from the distance d is given by $d+$, and the lower limit was set to the van der Waals distance between the two nuclei. d_{std} and V_{std} denote a distance standard and an associated NOE volume standard, respectively. The distance standards were taken from Wüthrich (*44*) and are typical of cross-strand NH-NH (3.3 Å), sequential NH-Hα (2.2 Å), and cross-strand Hα-Hα (2.3 Å) distances found in anti-parallel β-sheets. To determine V_{std} for each type of distance in each separate data set, the volumes of NOEs between likely anti-parallel β-sheet residues were averaged. For intraresidue NH-Hα distances only, the upper bounds for the distance constraints were increased by 1.5 Å because ~35% of these constraints were seen to be short in relation to X-ray values when calculated with interresidue distance standards. Intraresidue transfers may be somewhat anomalous due to coherent transfer contributions and lesser effects of spin diffusion. Increasing upper bounds minimizes structural impact of the possible anomalies.

NMR structures of BLBC were calculated using X-PLOR (version 3.840) (*45*), beginning with a torsion angle dynamics (TAD) protocol (*46*, *47*) and following with a round of simulated annealing refinement. A modified version of the CHARMM22 forcefield (MacKerrell et al., in preparation) using the repel function rather than Leonard-Jones parameters and a soft square potential in the restraining function for the effective NOE energy term were applied. NOE restraints were given a scaling factor of 100 kcal/mol except during the final Powell minimization stage of the TAD protocol; at this stage these restraints were scaled at 75 kcal/mol. All other parameters carried the standard values given in the X-PLOR tutorial files. Stereospecific assignments of resonances were not made for these preliminary structure calculations; thus center averaging and correction to pseudo-atoms was applied to all methylene ^{1}H pairs, aromatic ^{1}H pairs, and methyl ^{1}H triplets (*48*). Structures were deemed acceptable after the refinement procedure if they had no NOE

Table 1: Average Fractions of Protein Bound at Each Domain

$[P_t]^a$ (M)	$[S_t]^b$ (M)	$\bar{f}_b{}^c$	$\bar{f}_c{}^d$
0.00120	0.00112	0.129	0.082
0.00112	0.00317	0.769	0.441
0.00088	0.00853	0.992	0.940
0.00075	0.01190	1.000	1.000

[a] The total protein concentration. [b] The total sugar concentration. [c] The average fraction of protein bound at the B domain according to both NH and ^{15}N chemical shift changes. [d] The average fraction of protein bound at the C domain according to both NH and ^{15}N chemical shift changes.

distance violations greater than 0.4 Å, no root mean square differences for bond and angle deviations from ideality greater than 0.01 Å and 1°, respectively, and were under 225 kcal/mol in total energy.

Calculation of Association Constants. Chemical shift changes for several residues in the protein were monitored as a function of ligand and protein concentrations. Using eqs 3a and 3b, these chemical shift changes were used to calculate the fractions of occupied protein binding sites (f_B and f_C) for the separate domains at every sugar concentration.

$$f_B = \frac{[P_B]}{[P_t]} = \frac{\nu_{B(obsd)} - \nu_{B(free)}}{\nu_{B(bound)} - \nu_{B(free)}} \quad (3a)$$

$$f_C = \frac{[P_C]}{[P_t]} = \frac{\nu_{C(obsd)} - \nu_{C(free)}}{\nu_{C(bound)} - \nu_{C(free)}} \quad (3b)$$

In these equations, the subscripts B and C refer to residues in the BLBC B and C domains, respectively. ν_{free} denotes the chemical shift of a residue in the absence of ligand, and ν_{bound} denotes its chemical shift when the protein is fully bound. The latter values were measured from the HSQC spectrum acquired at the ultimate sugar concentration since almost no changes in chemical shifts were observed between this and the penultimate titration point, indicating a saturation of binding sites. ν_{obsd} denotes the actual chemical shift observed for a particular residue at any given concentration of sugar, $[P_B]$ and $[P_C]$ are the concentrations of protein with ligand bound at the B and C domains, respectively, and, finally, $[P_t]$ is the total protein concentration. Once calculations were completed for each monitored residue, all f_B and all f_C values at each sugar concentration were averaged to $\bar{f}_B$ and $\bar{f}_C$ values, respectively. Both individual and average fractions were used in calculations of the separate association constants for each domain. The average values, as well as total protein and sugar concentrations used at each stage of the titration, are listed in Table 1.

The concentration of free ligand in the NMR sample, $[S_f]$, could be calculated at each point in the titration according to eq 4, where $[S_t]$ denotes the total sugar concentration.

$$[S_f] = [S_t] - [P_B] - [P_C] \quad (4)$$

Similarly, the concentration of protein that remained unliganded at each domain, $[P_{fB}]$ and $[P_{fc}]$, could be calculated according to eqs 5a and 5b.

$$[P_{fB}] = [P_t] - [P_B] \quad (5a)$$

$$[P_{fC}] = [P_t] - [P_C] \quad (5b)$$

$[P_B]$, $[P_C]$, $[P_{fB}]$, $[P_{fC}]$, and $[S_f]$ are further related by binding constants for each domain, K_B and K_C. The final form of the equation used to solve for the equilibrium constant characterizing ligand binding to the B domain (K_B) is shown in eq 6. A similar equation exists to solve for K_C (see Results for further detail).

$$f_B = \frac{1}{2}\left\{-\left(\bar{f}_C - 1 - \frac{S_t}{P_t} - \frac{1}{K_B[P_t]}\right) - \sqrt{\left(\bar{f}_C - 1 - \frac{[S_t]}{[P_t]} - \frac{1}{K_B[P_t]}\right)^2 - 4\left(\frac{[S_t]}{[P_t]} - \bar{f}_C\right)}\right\} \quad (6)$$

Comparison of NMR Binding Data to X-ray Binding Data. In addition to determining binding constants from chemical shifts, several attempts were made to correlate shifts with structural changes. One structural feature of interest was hydrogen bonding. PDB coordinate files for free and bound WGA-A (PDB accession codes 7WGA and 1WGC, respectively) were used to estimate hydrogen bond lengths in both states. Hydrogen bonds were assumed to exist between all N−O pairs (except an amide bond pair) separated by less than or equal to 3.5 Å. For the B domain (the C domain does not bind ligand in these structures of WGA), the average change in hydrogen bond length upon complexation was found to be 0.16 ± 0.18 Å. Only changes of this approximate magnitude or greater were considered further.

Hydrogen bonding of both the amide proton and the carbonyl oxygen of a given amide bond can affect ^{15}N chemical shifts as described in the Discussion. As an example of a hydrogen bond based prediction of a chemical shift change, consider the following prediction for the C domain residue W107 in BLBC. This residue was first related to Y64, the residue in the homologous position in the B domain of WGA-A. Hydrogen bond length changes greater than or equal to 0.15 Å (*vide supra*) involving the amide nitrogen of Y64 or the carbonyl oxygen of Q63 were compared between the free and bound WGA-A structures. Three relevant bonds were found, one between Y64 NH and S62 OG, one between Q63 O and Q79 HE2, and one between Q63 O and Q49 HE2. Upon ligand binding, the first bond increased in length by 0.29 Å whereas the second and third bonds decreased by 0.18 and 0.15 Å, respectively. Thus, the net change in H-bond lengths was a decrease of 0.04 Å, indicating that the ^{15}N chemical shift of the W107 residue in BLBC should move downfield during the ligand titration study. In this case, the prediction matched the experimental data.

Other factors explored for their effects on chemical shift perturbations were changes in dihedral angles, ring current effects from the tryptophan residues of each domain, and direct magnetic shielding of protein residues by ligand groups. The first of these factors was examined using Procheck (*49*) to compare the free and bound WGA-A structures and to compute circular variance values for each residue in the B and C domains. Circular variance is a measure of how closely the free and bound dihedral angles for each residue overlap, with smaller values indicating better overlap. The values were averaged, and only those residues with circular variance values greater than 1 standard deviation from the mean were considered in the data analysis. To

examine ring current and magnetic shielding effects, the "zone" command in MidasPlus (*50*) was used to determine which residues of BLBC were close in space to tryptophan side chains or bound ligand atoms, respectively.

RESULTS

BLBC, the combined B and C domains of barley lectin, has been successfully expressed, refolded, and isolated through the application of a gene fusion and periplasmic secretion strategy. Previously, attempts had been made to isolate this fragment based on a post-expression refolding protocol (*11*), but in our hands, the yield of protein with properly formed intradomain disulfide linkages proved to be too low for utilization. Expression and purification from a vector that directs secretion of recombinant fusion protein to the periplasmic space, where refolding and intradomain disulfide formation can occur *in vivo*, appears more efficient. Overall yields are still low due to losses during the outer membrane lysis and factor Xa cleavage steps; however, adequate yields are obtained to pursue an NMR study. On the basis of electrospray mass spectrometry and an Ellman's test, the product is one with few improper interdomain disulfide linkages and no free thiols.

Pulsed-Field Gradient Diffusion Measurements. The question of whether BLBC functions as a monomer or noncovalent dimer is important for both elucidation of the ligand binding mode and for practical aspects of NMR studies. We attempted to assess the level of aggregation by determining effective diffusion constants for several of our samples. Diffusion constants decrease approximately as the inverse cube root of an aggregate's effective molecular weight. Thus, diffusion coefficient measurement can provide a convenient, in sample monitor of aggregation. For measurement we used the pulsed-field gradient NMR method described by Altieri et al (*38*). In these experiments, a 90° ^{1}H pulse tips bulk ^{1}H magnetization into the transverse plane where it is immediately dephased by the application of a pulsed-field gradient. After a time delta (Δ), a second pulsed-field gradient is applied to rephase the magnetization and the rephased signal is collected. However, since the field during the gradient application is position dependent, this second gradient will only rephase that magnetization from molecules which have not diffused away from their starting points. By repeating the experiment with all parameters held constant except the gradient strength, which should be incremented from 0 to its maximal value, a series of spectra with decreasing signal amplitudes may be collected. From these spectra the diffusion constant of the protein may be calculated as described by Stejskal and Tanner (*39*).

The diffusion constants of BLBC at 30 °C for samples in the 1 mM range averaged to $(1.5 \pm 0.2) \times 10^{-6}$ cm^2/s. They moved to slightly higher values for more dilute samples. These constants were compared to the equivalent measured and literature D_s values for lysozyme of 1.35×10^{-6} cm^2/s and to the HYDRO-calculated (*41*) D_s values of 1.77 and 1.40×10^{-6} cm^2/s for the BLBC monomer and dimer, respectively. Scaling the lysozyme value to that expected for a protein equivalent in size to monomeric BLBC, one calculates 1.54×10^{-6} cm^2/s, a value in perfect agreement with our measured BLBC value. The measured value for BLBC lies between that of the modeled monomer and dimer, but these values are likely to be overestimated because of our use of pure water viscosities in the HYDRO calculations. Thus, we concluded that in the absence of ligand, BLBC was mostly monomeric.

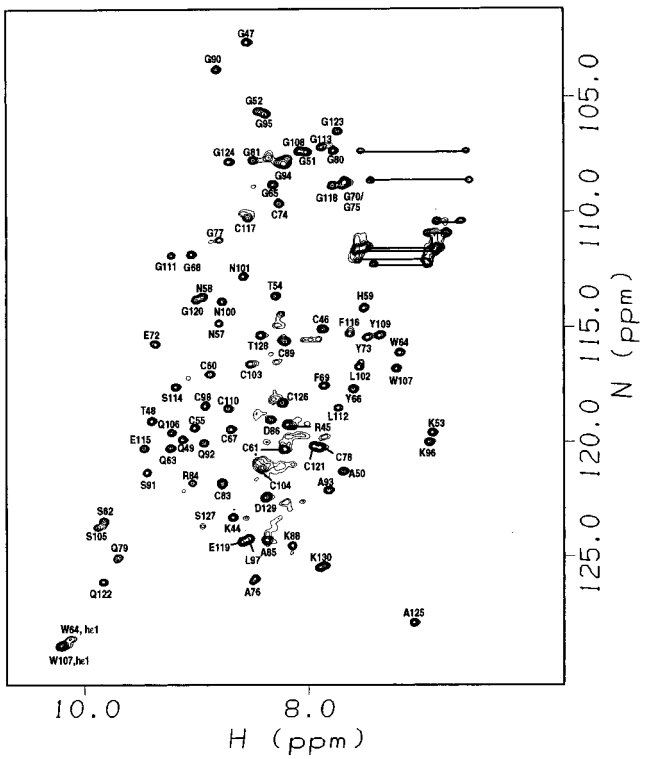

FIGURE 3: ^{1}H$-^{15}$N HSQC of BLBC. The protein sample is 1.24 mM and is in 0.1 M potassium phosphate pH 6.0, 0.15 M NaCl, 0.1% NaN_3, 90% H_2O/10% D_2O (30 °C).

Since X-ray studies have always shown the binding sites of WGA to lie at the dimer interfaces of opposing domains and to usually involve residues from both protomers (*15, 17*), we also investigated the possibility that the presence of ligand could induce dimerization. To do this we repeated the pulsed-field gradient diffusion measurements after each aliquot of *N*,*N*′,*N*′′-triacetylchitotriose in the ligand binding studies. The magnitudes of the measured diffusion constants remained constant and near that predicted for the monomer, indicating that BLBC binds carbohydrate ligand in monomeric fashion.

^{1}H and ^{15}N Assignments of BLBC. A prerequisite to structural analysis of BLBC using solution state NMR is assignment of the NMR spectra. The ^{1}H$-^{15}$N HSQC spectrum of BLBC is shown in Figure 3. BLBC contains 89 amino acids, of which three are prolines and two are arginines. For the former, no amide resonances will appear in ^{15}N-edited experiments, and for the latter, additional side chain amides may contribute. Thus, if BLBC is a homogeneous well-structured protein, one expects to see a maximum of 88 amide crosspeaks in the central region of the ^{1}H$-^{15}$N HSQC spectrum. In fact, 81 strong and several more weak amide crosspeaks are observed. Additional side chain crosspeaks arising from NHs of 2 Trp residues and NH2s from 6 Gln and 4 Asn residues appear in the low-field and high-field regions of the spectrum, respectively. Also observed in the HSQC spectrum are small, less well-defined crosspeaks that lie primarily in the center of the amide region, usually near cysteine residues. While an

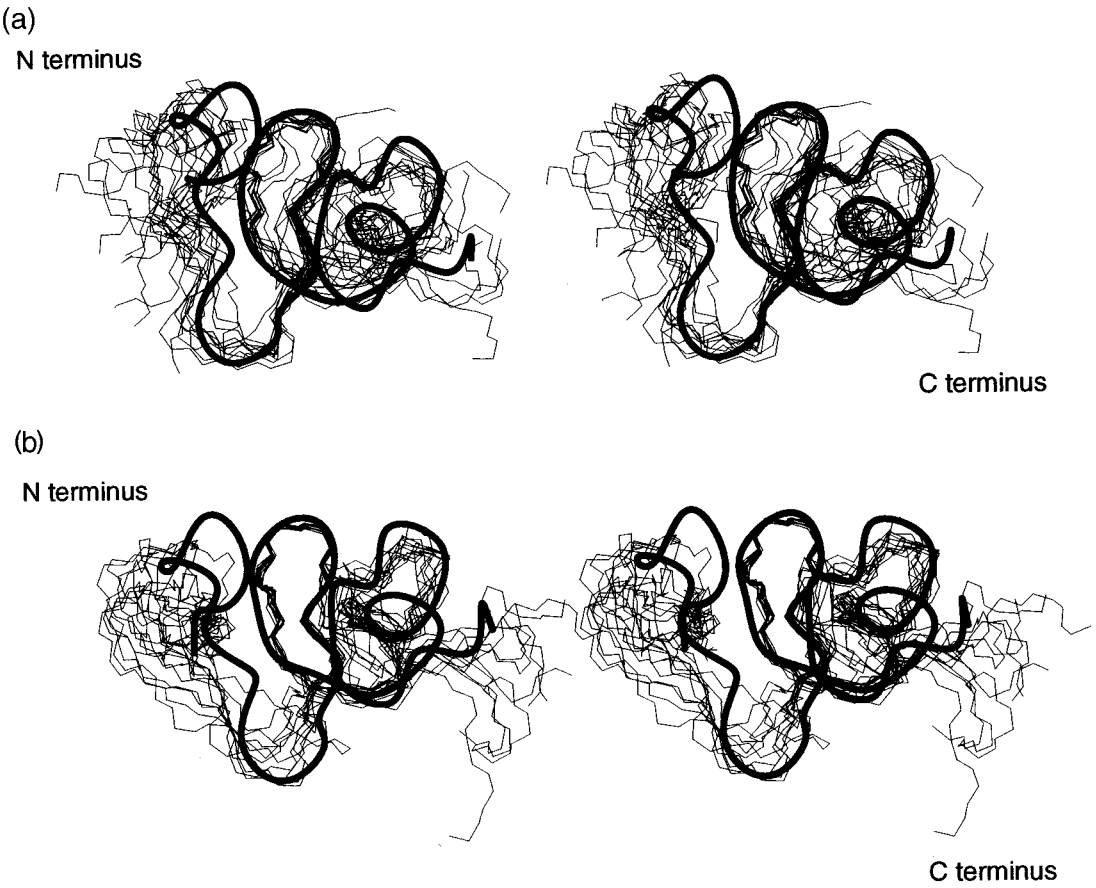

FIGURE 4: Stereoview of the overlayed backbone atoms of the 11 final accepted structures from X-PLOR calculations. The backbone structure of WGA-B is also overlayed and is depicted in its ribbon representation. (a) Overlayed B domain structures. (b) Overlayed C domain structures. The structures were generated using the molecular graphics programs MidasPlus from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081) and MOLMOL (MOL analysis and MOLecular display), a program written as part of a collaboration project between BRUKER/Spectrospin and the group of Prof. Wüthrich at the ETH Zürich. Information on obtaining the PDB files used to generate these figures can be found on the Prestegard Lab homepage at http://www.ccrc.uga.edu/~jpresteg.

Ellman's test showed no significant level of free thiols, it is possible that these less well-defined crosspeaks reflect a small percentage of mispaired disulfides. On the basis of qualitative intensity estimates, the percentages of these species total to less than 10%.

^{1}H and ^{15}N chemical shift assignments were made using a number of heteronuclear NMR experiments. The well-dispersed HSQC spectrum served as a useful starting point and aid in the assignment process. A 3D TOCSY-HSQC (80 ms) was used to assign individual spin systems to amino acid classes from through-bond heteronuclear correlations. 3D NOESY-HSQCs (100 and 200 ms), with intraresidue peaks eliminated on the basis of comparison to the TOCSY-HSQC, were then used to connect the amino acid types in sequential order from through-space proton−proton interactions. Assignments are given in the Supporting Information.

Structure Calculation. Distance restraints used for the preliminary structure calculations were found in the 3D NOESY-HSQC and 2D homonuclear NOESY data sets. In all, 279 intraresidue and 265 interresidue distance restraints, the latter including 84 restraints between proton pairs greater than or equal to 4 residues apart, were included in the X-PLOR structure calculations. An iterative structure determination process suggested a pairing of cysteines to form disulfides in a pattern consistent with the WGA X-ray structure. Disulfide bonds were thus included as restraints in the final stages of the structure calculations. A total of 50 structures were generated, and of these, 11 met our predefined limits of acceptability (see Procedures section). These 11 structures are shown overlayed with each other and with a ribbon representation of the crystal structure of WGA-B B and C domains (PDB accession code 1WGT) in Figure 4.

The domains are shown separately because at this early stage in the structure determination lack of restraints in the domain linker region prevents the elucidation of an absolute orientation of the domains with respect to one another. The only interdomain NOEs observed were one between Cys83 and Ile87 and two between Tyr109 and Asp86. Notably, all three of these contacts involve residues that lie at the domain interfaces and thus may not be indicative of true interdomain interactions. The lack of true interdomain contacts supports the absence of significant dimerization and suggests that the domains in the monomer are tethered loosely. Nevertheless, the convergence within domains is quite good. In both domains, a β-sheet plus short α-helix motif are clearly visible. This motif, which is similarly observed in WGA as well as lectins of noncereal origin that have homologous sequences (*51*−*53*), contains residues which are believed to play prominant roles in ligand binding. For our preliminary NMR structure of BLBC, the residues in this region (i.e., residues His59-Cys74 in the B domain and residues Leu102-Cys117 in the C domain) show RMSDs for the backbone atoms of 1.05 ± 0.40 and 0.44 ± 0.08 Å

for the B and C domains, respectively. RMSDs for the entire B domain (residues 44−86) and C domain (residues 87−129) structures are 1.95 ± 0.37 and 1.47 ± 0.16 Å, respectively.

Convergence between BLBC and WGA domains is also very high. For the β-sheet plus short α-helix region the RMS errors between the WGA-B structure and the average of the 11 BLBC structures are 1.47 and 1.17 Å for the B and C domain backbone atoms, respectively. Thus the NMR data provide a good argument for homology with WGA and a good justification for using the higher resolution X-ray data in rationalizing ligand-induced changes in NMR parameters as discussed below.

NMR Titration Studies. Having determined that BLBC retained a nativelike fold, we turned next to answering questions concerning its ligand binding characteristics. ^{15}N-Labeling and 2D resolution of protein resonances offered the possibility of monitoring ligand binding in great detail and independently at both domains. Addition of carbohydrate ligand to an isotopically labeled BLBC sample caused some line broadening of resonances and numerous ligand concentration dependent chemical shift changes. This suggested a fast to intermediate exchange process between free and bound protein. The primary perturbations observed are given in Tables 2 and 3, and several of these are depicted in the ^{1}H-^{15}N HSQC spectrum in Figure 5.

While most ligand concentration dependent resonance shifts supported a simple rapid exchange between unliganded and liganded protein, a few resonances showed evidence for a more complex dynamic process. For instance, the Trp64 resonance broadened beyond detection in the presence of 3.17 mM chitotriose. In Figure 5, no resonance for Trp64 can be identified for the liganded protein. Yet, unlike a number of other resonances, the Trp64 resonance showed almost no change in its chemical shift at lower ligand concentrations (data not shown). Intermediate rather than fast exchange rates between free and bound states can produce such broadening but should show a dependence on the magnitude of ligand-induced shift squared. This is clearly inconsistent with the small chemical shift changes at lower ligand concentrations. In addition to the Trp64 anomaly, a few other residues showed two sets of resonances after addition of ligand, with each member of the set being different from the resonance for unliganded protein. An example can be seen in the shape of the peaks for Tyr73 in the complexed species (dotted lines) in Figure 5. This observation suggested a slow exchange process, but this slow process must be distinct from that of ligand binding and must occur preferentially in the bound state. In support of this suggestion, the "slow exchange" and "fast exchange" resonances showed different saturation profiles. The fast exchange process was saturable, as it should be for changes which result directly from the binding process, and in fact, shifts reflecting a fast exchange process had reached saturation by approximately 8.5 mM chitotriose. The slow exchange pairs of resonances showed approximately constant relative intensities throughout the ligand concentration range observed. The most logical interpretation is that two slightly different forms of the fully liganded protein exist.

Analysis of the primary ligand-induced perturbations in BLBC shows that these may be localized to two regions, one in each domain. Figure 6 shows plots of the magnitude of induced shift changes for NH and ^{15}N chemical shifts as a function of sequence. Two clusters, one near Q63-E72 and one near Q106-E115 are observed. These regions correspond to homologous domain positions. The magnitudes of the shifts for homologous residues in each domain are generally different, but the directions of the shifts are often the same (see also Tables 2 and 3). For example, while the NH resonance of Q63 in domain B shifts by 0.115 ppm, that of Q106, which resides in the C domain at the corresponding position to Q63, shifts by only 0.088 ppm. Nevertheless, both of these resonances are among those with the largest shift changes in each domain, and both resonances shift downfield upon addition of ligand. The behavior observed here for BLBC is more consistent with a model in which ligand binds to both domains in similar independent fashion than a model in which one domain interacts with or operates cooperatively with the other domain. This independence of sites is also supported by previously acquired binding data on native WGA/BL (*14*, *18*, *20*). However, in contrast to the case in the native dimeric proteins, the independence of sites observed for BLBC, which diffusion

Table 2: Chemical Shift Perturbations in the BLBC B Domain and Their Relation to Ligand-Induced Changes Observed in the X-ray Structure of the WGA-A B Domain[a]

affected residue	ligand-induced chem shift change (ppm)[b]	proximity to bound ligand (Å)[c]	net H-bond length changes (Å)[d]	Φ, Ψ dihedral angle changes[e]	proximity to W64[f]
		^{15}N Chemical Shift Changes			
W64 (Y64)	broadened	<4	0.11	affected	—
Q63	−0.686	<6	0.21	—	affected
E72	−0.564	<4	0.19	—	—
A76	+0.375	<9	0.2	—	—
F69	−0.349	<8	0.17	—	—
		NH Chemical Shift Changes			
E72	+0.182	<4	—	—	—
Q63	+0.115	<6	—	—	—
G77	+0.097	>15	—	—	—
Y66	−0.096	<4	—	—	affected
K44	+0.091	<7	—	—	—
Q79	+0.085	<8	—	affected	
C74	+0.074	<8	—	—	—
N58	+0.073	>15	—	—	—
Q49	+0.070	<8	—	affected	—

[a] The mean changes in chemical shifts (for all residues in BLBC) upon going from free to completely bound protein are 0.14 ± 0.16 ppm and 0.036 ± 0.038 ppm for ^{15}N chemical shifts and NH chemical shifts, respectively. This table only includes residues that show changes greater than 1 standard deviation from the mean. The concentration of BLBC in the free sample is 1.24 mM, and that in the fully bound sample is 0.75 mM (+11.9 mM carbohydrate). If the B domains of BLBC and WGA-A differ in amino acid sequence at a specified domain position, the corresponding WGA-A residue is indicated in parentheses. [b] Upfield chemical shift changes are negative, and downfield chemical shift changes are positive. [c] Any atom of the sialic acid residue. [d] Longer net hydrogen bond lengths upon ligand complexation to WGA-A's B domain are positive, and shorter net bond lengths are negative. [e] If a significant dihedral angle change is observed for this residue upon binding in the X-ray structure of WGA-A, then this column is marked "affected" to indicate that a dihedral angle change may contribute to the observed chemical shift change. [f] If the perturbed residue is within 4 Å of W64 in the average NMR structure, then it is marked "affected" to indicate that its chemical shift may be affected by changes in the orientation of this aromatic sidechain. For all columns, "—" indicates no basis for comparison or no effect from ligand binding.

Table 3: Chemical Shift Perturbations in the BLBC C Domain and Their Relation to Ligand-Induced Changes Observed in the X-ray Structure of the WGA-A B Domain[a]

affected residue	ligand-induced chem shift change (ppm)[b]	proximity to bound ligand (Å)[c]	net H-bond length changes (Å)[d]	Φ, Ψ dihedral angle changes[e]	proximity to W107[f]
		^{15}N Chemical Shift Changes			
W107 (Y64)	+0.752	<4	0.11	affected	—
F116 (Y73)	−0.576	<4	0.19	—	—
Q106 (Q63)	−0.467	<6	0.21	—	affected
G111 (G68)	−0.348	<8	0.25	—	—
L102 (Q59)	−0.319	>15	—	—	—
C117 (C74)	−0.297	<8	−0.2	—	—
C110 (C67)	−0.292	<6	—	—	—
		NH Chemical Shift Changes			
W107 (Y64)	+0.196	<4	—	affected	—
E115 (E72)	+0.105	<4	—	—	—
G108 (G65)	+0.090	<6	—	—	affected
Q106 (Q63)	+0.088	<6	—	—	affected

[a] The mean changes in chemical shifts (for all residues in BLBC) upon going from free to completely bound protein are 0.14 ± 0.16 and 0.036 ± 0.038 ppm for ^{15}N chemical shifts and NH chemical shifts, respectively. This table only includes residues that show changes greater than 1 standard deviation from the mean. The concentration of BLBC in the free sample is 1.24 mM, and that in the fully bound sample is 0.75 mM (+11.9 mM carbohydrate). If the B domains of BLBC and WGA-A differ in amino acid sequence at a specified domain position, the corresponding WGA-A residue is indicated in parentheses. [b] Upfield chemical shift changes are negative, and downfield chemical shift changes are positive. [c] Any atom of the sialic acid residue. [d] Longer net hydrogen bond lengths upon ligand complexation to WGA-A's B domain are positive, and shorter net bond lengths are negative. [e] If a significant dihedral angle change is observed for this residue upon binding in the X-ray structure of WGA-A, then this column is marked "affected" to indicate that a dihedral angle change may contribute to the observed chemical shift change. [f] If the perturbed residue is within 4 Å of W64 in the average NMR structure, then it is marked "affected" to indicate that its chemical shift may be affected by changes in the orientation of this aromatic side chain. For all columns, "—" indicates no basis for comparison or no effect from ligand binding.

studies demonstrate remains largely monomeric even in the presence of ligand (*vide supra)*, indicates that no helper domain interactions participate in ligand binding to the modified protein.

The ability to observe discrete shift perturbations for each domain allowed an independent determination of binding constants. For instance, K_B, the association constant for the interaction of ligand with the BLBC B domain, was evaluated using eq 6 from the Experimental Procedures section and a method that minimized the squared residual between the calculated and experimental values of f_B with respect to K_B. A curve of the squared residuals was generated using Mathematica (*54*). The minimum in the curve was taken to be the value of the association constant. A completely analogous calculation was made to determine K_C, the association constant for the interaction of ligand with the BLBC C domain. Repeating the calculation for each significantly perturbed resonance in the separate domains, we have determined values for K_B and K_C of 1.1 ± 0.2 × 10^3 and 0.6 ± 0.2 × 10^3 M^{-1}, respectively.

DISCUSSION

The preliminary structure calculated for BLBC shows that the recombinantly produced fragment has retained a native-like fold. Diffusion studies of the protein, both alone and in the presence of carbohydrate ligand, show that the two-domain fragment is largely monomeric in contrast to its native four-domain parent molecule. Observed chemical shift perturbations and the binding constants extracted from them demonstrate that the two-domain BLBC protein retains nativelike binding ability and carries two distinct ligand binding sites, one in each domain. Isotope-edited 2D NMR methods have been especially useful in these ligand binding studies, for they have provided the spectral resolution necessary to observe the presence or absence of ligand-induced chemical shift changes for almost every resonance in the protein spectrum. Here we try to correlate the largest of the observed changes with sites of protein−ligand interaction or ligand-induced changes in protein structure. The existing NMR-based ligand binding studies of the homologous lectins hevein (*51*), Ac-AMP2 (*52*), and nettle lectin (*53*), as well as the X-ray structures of complexed and uncomplexed WGA-A, can be coupled with some basic understanding of the origin of chemical shifts to form the basis for such analyses.

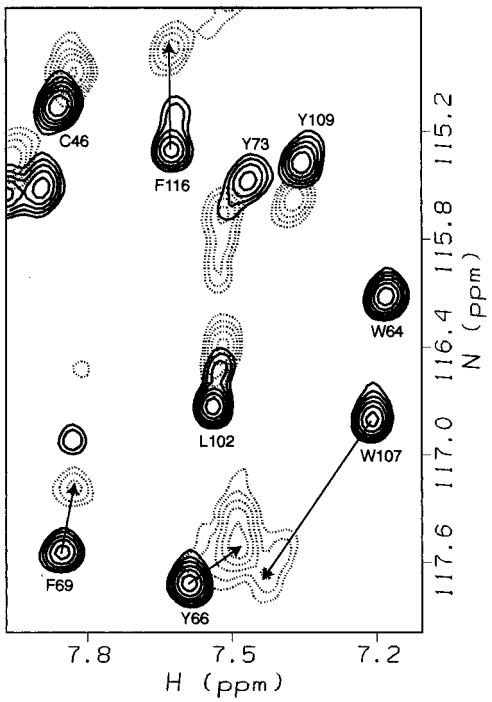

FIGURE 5: Overlay of HSQCs acquired of free (solid lines) and completely bound (dotted lines) BLBC. The ligand-free spectrum was acquired at 1.24 mM BLBC, and the completely bound spectrum was acquired at 0.75 mM BLBC/11.9 mM sugar. Both spectra display the region between 8.0 and 7.1 ppm in the HN (D1) dimension and 114.6−118.0 ppm in the ^{15}N (D2) dimension.

First, the residues affected by ligand binding in BLBC were compared to those affected in similar studies of the lectins hevein, Ac-AMP2, and nettle lectin. These proteins, which consist of one (hevein and Ac-AMP2) or two (nettle lectin and BLBC) homologous domains, have been shown to bind ligand monomerically (*51*−*53*). Figure 7 shows a graphical representation of the comparisons. All studies show perturbations in two regions, around domain positions 22 and 30. This suggests that the proteins or protein domains bind ligand in a similar fashion. Furthermore, the domain positions found to be most affected by ligand binding,

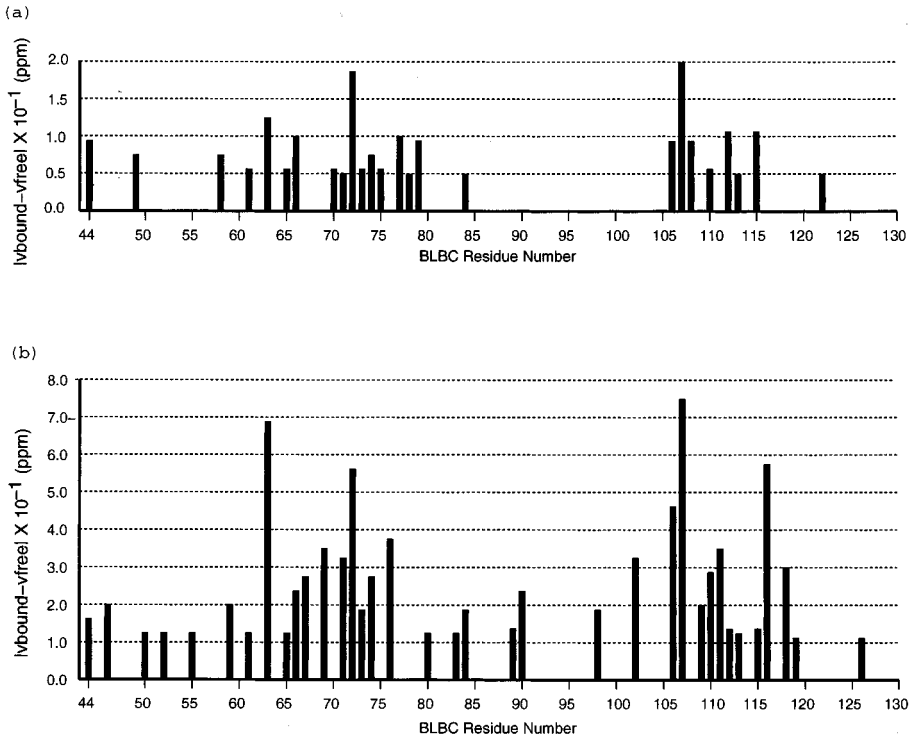

FIGURE 6: Magnitudes of the ligand-induced perturbations at the B and C domains of BLBC. Two sets of residues, one set in the B domain (residues 44−86) and a homologous set in the C domain (residues 87−129), are perturbed upon ligand binding. (a) NH chemical shift perturbations. (b) ^{15}N chemical shift perturbations.

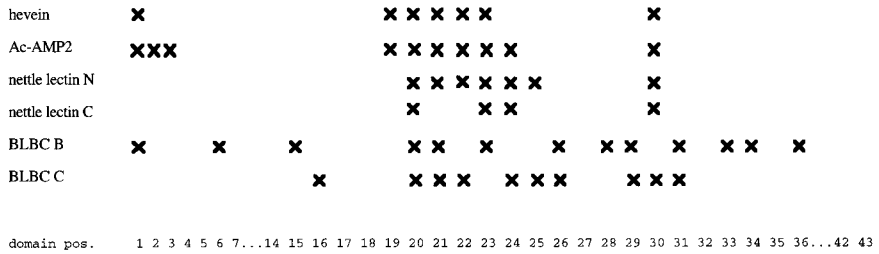

FIGURE 7: Comparison of the ligand-affected residues from NMR titration studies of hevein (*51*), Ac-AMP2 (*52*), nettle lectin (*53*), and BLBC. All titration studies used *N*,*N*′,*N*″-triacetylchitotriose as the ligand, and X's mark the residues that are reported to shift in the presence of ligand. For nettle lectin, N and C refer to the N-terminal and C-terminal domains of this protein, respectively. For BLBC, B and C signify changes observed in the B and C domains, respectively; furthermore, only ligand-induced chemical shift changes greater than 1 standard deviation from the mean ligand-induced chemical shift change are denoted. The relatively larger number of positions that appear affected by ligand binding in the ^{15}N-labeled BLBC domain may reflect the greater ability of isotope-edited NMR techniques to detect ligand-induced changes in individual protein resonances. However, some of the uncorrelated perturbations may also reflect a secondary ligand-induced process in the B domain, such as a slow conformation change.

positions 20−24 (e.g., Q63-C67 and Q106-C110 in the BLBC B and C domains, respectively) and position 30 (e.g., Y73 and F116 in the BLBC B and C domains, respectively) coincide with or are adjacent to the domain positions identified as important for ligand binding in X-ray structures of complexed WGA-A (*16*, *17*). The crystallographically determined ligand binding residues in the B domain principal site of WGA-A are S62 (domain position 19) and the aromatic residues Y64, Y66, and Y73 (domain positions 21, 23, and 30 respectively). A fifth important residue comes from the helper binding domain; it is E115 (domain position 29) in the C domain helper site. Since this residue is directly adjacent to residues involved in principal site binding it is not possible to use its chemical shift perturbation as evidence for helper domain involvement. Furthermore, the homology in shift perturbations for residues in domain positions 20−25 in both the B and C domains of BLBC suggests that both domains function as principal sites, not one as an exclusive principle domain and one as an exclusive helper domain.

Shifts are more widespread in our studies of the BC domains of barley lectin than has been observed in hevein, Ac-AMP2, or nettle lectin. Whether this indicates less well-defined sites or more extensive structural perturbations cannot be said. We remind the reader that previous studies were based on proton only spectra, and it may have been difficult to resolve resonance perturbations for all parts of the protein.

Chemical shift changes could in principle be used to give a more detailed picture of ligand interaction if their origin could be tied to either direct perturbations by magnetic and electrostatic properties of the ligand or to ligand-induced conformational changes in the protein. Magnetic shielding and electrostatic effects, which are both expected to show sharp dependence on proximity of perturbed nuclei to bound ligand, were assessed first. One-third of the ^{15}N resonances shifted by more than 0.29 ppm and ^{1}H resonances shifted

by more than 0.07 ppm belonged to residues within 4 Å of the sialic acid moiety in the ligand-bound WGA-A structure (see Tables 2 and 3). Such localized effects may be due, at least in part, to magnetic shielding of protein residues (*55*). For instance, if *N*,*N*′,*N*″-triacetylchitotriose is presumed to bind BLBC in a site and orientation similar to that observed for sialic acid in X-ray structures of complexed WGA-A, then the shifting of the Tyr73 hϵ1/hϵ2 resonances observed in 1D NMR titration studies (data not shown) might be due to shielding by the magnetically anisotropic acetamido group of the terminal nonreducing carbohydrate ring. However, magnetic shielding explanations for observed chemical shift changes are limited by the facts that only the acetamido group of chitotriose is significantly anisotropic and only a few of the resonances observed to shift considerably are hypothesized to lie near this group. Analogously, electric field dependent mechanisms arising from charged ligand groups, which have also been shown to affect chemical shifts (*55*), are limited in their explanation of the observed chemical shift changes by the fact that the *N*-acetylglucosamine rings of chitotriose do not carry charged moieties. However, the relatively large dipole moment of the acetamido's carbonyl group could have a significant effect. Again, such effects could explain only a very limited number of shift changes.

Since a complete explanation for observed chemical shift changes could not be made on the basis of direct ligand interaction with BLBC, the possibility of locally induced structural changes was examined next. Oldfield and coworkers have established a correlation between Φ_i, and Ψ_{i-1} angles in proteins and ^{15}N chemical shifts (*56*). A detailed comparison of Φ and Ψ angles for complexed and uncomplexed protein was therefore worthy of examination. PROCHECK (*49*) was used to compare the Φ and Ψ angles of free and complexed WGA-A and to thus determine the most significantly ligand-affected protein residues. These were Q49, Y64, C78, and Q79 for the B domain, which in the X-ray structure served as the principal binding site for the ligand. Only Y64 correlated with residues in the BLBC B and C domains that showed significant ligand-dependent ^{15}N chemical shift perturbations (see Tables 2 and 3). The corresponding BLBC residues were W64 (B domain) and W107 (C domain). Comparson of the Φ_i and Ψ_{i-1} angles for Y64 in free and complexed WGA-A showed that the downfield chemical shift change of BLBC's W107 was consistent with predictions from torsional changes. W64, however, did not undergo much change in chemical shift but instead broadened out. Thus, it appears that while dihedral angle changes might correlate with ^{15}N chemical shifts, they do not easily explain all ^{15}N chemical shift changes observed for ligand binding to BLBC.

Another factor considered for its effects on ligand-dependent chemical shift changes was the shielding or deshielding effects of the tryptophan residue in each domain (W64 and W107). The aromatic ring of the tryptophan side chain is highly anisotropic and is capable of producing geometry dependent ring current shifts for nearby ^{1}H and ^{15}N resonances. In the B domain of WGA-A, Y64 is the residue corresponding to BLBC's W64 and W107, and it has been identified as an important residue for ligand binding. Y64 also shows significant changes in its side chain orientation upon complexation of ligand (*15, 17*). To predict which residues of BLBC might be affected by the tryptophan side chains, residues within 4 Å of W64 (B domain) and W107 (C domain) were determined from the average NMR structures of each domain using the program MidasPlus (*50*). The results indicate that S62, Q63, G65, and Y66 in the BLBC B domain and K88, C104, S105, Q106, G108, and Q122 in the BLBC C domain are candidates for ring current induced chemical shift changes upon reorientation of W64 or W107 side chains (see Tables 2 and 3). Q63, Y66, Q106, and G108 all show large shifts; the others do not. An exact correlation would require an actual knowledge of how the tryptophan side chains reorient as well as a calculation of ring current shift change. Since the corresponding residue in WGA is a tyrosine and not a tryptophan, we did not feel justified in carrying out a detailed calculation, but we do point to the possibility that ring current effects contribute to the ligand-induced chemical shift changes in BLBC.

A final factor considered was whether ligand-induced changes in the hydrogen-bonding pattern of WGA-A's B domain could explain the chemical shift perturbations we observed in BLBC. Live and co-workers (*37*) found that a decrease in hydrogen bond lengths involving an amide nitrogen or the adjacent carbonyl oxygen would result in a downfield chemical shift change of the nitrogen resonance. Conversely, an increase in these hydrogen bond lengths would result in an upfield chemical shift change of the nitrogen resonance. Furthermore, the effects of longer and shorter hydrogen bonds should be additive. Predictions for the direction of chemical shift changes in BLBC upon ligand binding were made for each residue, identifying hydrogen bond length changes as discussed here and in the Procedures section. Considering both the B and C domains, the 12 ^{15}N resonances most perturbed by ligand binding, i.e., those of residues listed in Tables 2 and 3, do show significant hydrogen-bonding changes in 10 cases. However, only five of these lead to predictions of chemical shift displacements in the proper direction. A similar marginal level of success results when attempting to correlate changes in proton chemical shifts with hydrogen-bonding changes.

Thus, it does not seem possible to quantitatively ascribe large ligand-induced chemical shift changes to any single mechanism. However, in 20 of the 25 cases of large change in chemical shift listed in Tables 2 and 3, changes can be qualitatively correlated with either proximity to the bound ligand or proximity to protein sites that are structurally perturbed upon ligand binding. It is also noteworthy that in 21 of the 25 cases, the shifted nuclei are within 8 Å of the bound ligand. The latter fact supports recent use of chemical shift changes to qualitatively localize sites of ligand binding in less well-characterized systems (*22*). In our case, clustering of perturbed residues provides strong evidence for ligand occupancies of sites in both B and C domains of BLBC that are analogous to principal sites in the WGA-A crystal structure.

The association constants measured for BLBC by 2D NMR methods may be translated to dissociation constants (K_D) of approximately 0.94 mM (B domain) and 1.6 mM (C domain). Even the strongest of these binding constants is more than an order of magnitude weaker than the constants of 0.09 mM reported by Bains et al. from microcalorimetry studies of a mixture of WGA isolectins (*19*) and 0.02 mM reported by Nagahora et al. from equilibrium dialysis studies

of WGA-D (*20*). However, the K_D's for BLBC with chitotriose are similar in magnitude to the K_D's measured for the interactions of the smaller, monomerically acting proteins hevein, nettle lectin, and Ac-AMP2 (K_D's = 0.12, 0.1−2.3, and 1 mM for hevein, nettle lectin, and Ac-AMP2, respectively) (*51*−*53*). This similarity once again suggests that BLBC binds ligand in a monomeric fashion, in agreement with the results of pulsed-field gradient NMR studies which showed the protein to remain largely monomeric in the presence of ligand.

The results of our studies allow us to comment on the hypothesized equivalence of the B(pr)/C(hl) and C(pr)/B(hl) binding sites of native BL and WGA (*21*). According to our findings, complete equivalence is unlikely since the independent B and C domains show different affinities for ligand. However, there are several factors to consider in making comparisons between binding results of different experimental methods. For instance, it is highly unlikely that the previously applied solution methods for assessing binding constants can detect affinities that differ by less than a factor of 2, especially since deconvoluting binding data from different physical locations is not possible with these techniques (*14*, *18*−*20*). Furthermore, dimeric binding may introduce contacts that balance the inherently unequal ligand affinities in the 2 domains. Given these considerations, the B(pr)/C(hl) and C(pr)/B(hl) sites may indeed look equivalent and may in fact be the four equal affinity sites that have been repeatedly detected. However, until a study that specifically probes perturbations of A and D domains can be conducted it is difficult to rule out the possibility that the 2 B(pr)/C(hl) and 2 D(pr)/A(hl) sites constitute the four equivalent ligand binding sites. These two unique sites are the only ones occupied in WGA crystals complexed with *N*,*N*′-diacetylchitobioside, although occupancies are unequal. The B site is the most highly occupied site in complexes prepared with native crystals (*17*).

Concluding Remarks. The objectives of our study of the barley lectin protein were 2-fold. The first objective was to create a simplified model system for studying carbohydrate recognition by NMR. The second was to use this model system to answer questions relating to ligand binding by the cereal lectins and homologous proteins. Molecular biology techniques allowed the isolation and isotopically labeled production of a fragment of the BL protein that is smaller than intact BL and is more homogeneous in terms of binding sites. However, removal of the A and D domains resulted in loss of the native dimeric structure. NMR studies allowed structural and ligand binding characterization of the individual sites in domains B and C of the fragment. Indeed, the system has been used to show that the B and C domains of BL-like proteins possess slightly unequal ligand affinities. The latter characterization is one that is possible from NMR methods because of the ability to separately observe interactions for each domain.

ACKNOWLEDGMENT

Clones encoding the full length barley lectin gene, as well as clones encoding several other homologous plant lectin genes, were generously provided by Prof. Natasha V. Raikhel at Michigan State University. Members of Prof. Donald M. Engelman's laboratory at the Center for Structural Biology at Yale University, Drs. Shy Arkin, John Flanagan, Mark Lemmon, and Kevin MacKenzie, assisted with recombinant procedures, as did Dr. David Paolella of Prof. Alana Schepartz laboratory in the Department of Chemistry at Yale University. Much assistance in protein NMR spectroscopy and assistance with data fitting and the use of Mathematica came from Kai Huang and Michael Andrec, respectively, both of Prof. Prestegard's laboratory at Yale University. We also acknowledge stimulating discussions with Prof. Christine Schubert Wright of Virginia Commonwealth University in the early phases of this project.

SUPPORTING INFORMATION AVAILABLE

The chemical shift assignments for BLBC are available (2 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Goldstein, I. J., & Poretz, R. D. (1986) in *The Lectins* (Sharon, I. E., & Goldstein, I. J., Ed.) pp 33−247, Academic Press, Orlando.
2. Sharon, N., & Lis, H. (1989) *Science 246*, 227−234.
3. Stillmark, H. (1888) *Über rizin, ein giftiges ferment aus dem samen von Ricinus communis L. und einigen anderen euphorbiaceen*, Inaug. Diss, Dorpat.
4. Rademacher, T. W., Parekh, R. B., & Dwek, R. A. (1988) *Annu. Rev. Biochem. 57*, 785−838.
5. Lasky, L. (1992) *Science 258*, 964−969.
6. Hoppe, H.-J., & Reid, K. B. M. (1994) *Protein Science 3*, 1143−1148.
7. Hare, B. J., Rise, F., Aubin, Y., & Prestegard, J. H. (1994) *Biochemistry 33*, 10137−10148.
8. Sharon, N. (1993) *Trends Biochem. Sci. 18*, 221−6.
9. Peumans, W. J., Stinissen, H. M., & Carlier, A. R. (1982) *Biochem. J. 203*, 239−243.
10. Lerner, D. R., & Raikhel, N. V. (1989) *Plant Physiol. 91*, 124−129.
11. Schroeder, M. R., & Raikhel, N. V. (1992) *Protein Expression Purif. 3*, 508−511.
12. Wright, C. S., Schroeder, M. R., & Raikhel, N. V. (1993) *J. Mol. Biol. 233*, 322−324.
13. Wright, C. S. (1977) *J. Mol. Biol. 111*, 439−457.
14. Nagata, Y., & Burger, M. M. (1974) *J. Biol. Chem. 249*, 3116−3122.
15. Wright, C. S. (1984) *J. Mol. Biol. 178*, 91−104.
16. Wright, C. S. (1990) *J. Mol. Biol. 215*, 635−651.
17. Wright, C. S. (1992) *J. Biol. Chem. 267*, 14345−14352.
18. Kronis, K. A., & Carver, J. P. (1985) *Biochemistry 24*, 826−833.
19. Bains, G., Lee, R. T., Lee, Y. C., & Freire, E. (1992) *Biochemistry 31*, 12624−12628.
20. Nagahora, H., Harata, K., Muraki, M., & Jigami, Y. (1995) *Eur. J. Biochem. 233*, 27−34.
21. Wright, C. S., & Kellogg, G. E. (1996) *Protein Sci. 5*, 1466−1476.
22. Shuker, S. B., Hajduk, P. J., Meadows, R. P., & Fesik, S. W. (1996) *Science 274*, 1531−4.
23. Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24. Neidhardt, F. C., Bloch, P. L., & Smith, D. F. (1974) *J. Bacteriol. 119*, 736−747.
25. Perkins, S. J. (1986) *Eur. J. Biochem. 157*, 169−180.
26. Broekaert, W. F., Marien, W., Terras, F. R. G., De Bolle, M. F. C., Proost, P., Van Damme, J., Dillen, L., Claeys, M., Rees, S. B., Vanderleyden, J., & Cammue, B. P. A. (1992) *Biochemistry 31*, 4308−4314.
27. Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.

28. Muhandiram, D., & Kay, L. E. (1994) *J. Magn. Reson. Ser. B 103,* 203−216.
29. Grzesiek, S., & Bax, A. (1993) *J. Am. Chem. Soc. 115,* 12593−12594.
30. Piotto, M., Saudek, V., & Skelnar, V. (1992) *J. Biomol. NMR 1,* 661−665.
31. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989) *Biochemistry 28,* 6150−6156.
32. Zhang, O., Kay, L. E., Olivier, J. P., & Foreman-Kay, J. D. (1994) *J. Biomol. NMR 4,* 845−858.
33. Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110,* 7870−7872.
34. Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65,* 355−360.
35. Levitt, M., Freeman, R., & Frenkel, T. (1982) *J. Magn. Reson. 47,* 328−330.
36. Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95,* 1−6.
37. Live, D. H., Davis, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106,* 1939−1941.
38. Altieri, A. S., Hinton, D. P., & Byrd, R. A. (1995) *J. Am. Chem. Soc. 117,* 7566−7567.
39. Stejskal, E. O., & Tanner, J. E. (1965) *J. Chem. Phys. 42,* 288−292.
40. Cantor, C. R., & Schimmel, P. R. (1980) *Biophys. Chem.* W. H. Freeman & Co., New York.
41. Garcia de la Torre, J., Navarro, S., Lopez Martinez, M., Diaz, F., & Lopez Cascales, J. (1994) *Biophys. J. 67,* 530−531.
42. Harata, K., Nagahora, H., & Jigami, Y. (1995) *Acta Crystallogr., Sect. D: Biol. Crystallogr. 51,* 1013−1019.
43. Oswood, M. C., Kim, Y., Ohlrogge, J. B., & Prestegard, J. H. (1997) *Proteins 27,* 131−43.
44. Wüthrich, K. (1986) *NMR of proteins and nucleic acids,* J. Wiley & Sons, New York.
45. Brünger, A. T. (1992) *X-PLOR: a system for X-ray Crystallography and NMR,* Yale University Press, New Haven.
46. Rice, L. M., & Brunger, A. T. (1994) *Proteins: Structure, Function, & Genetics 19,* 277−290.
47. Stein, E. G., Rice, L. M., & Brunger, A. T. (1996) *J. Magn. Reson. submitted,*
48. Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169,* 949−961.
49. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26,* 283−291.
50. Ferrin, T. E., Huang, C. C., Jarvis, L. E., & Langridge, R. (1988) *J. Mol. Graph. 6,* 13−27.
51. Asensio, J. L., Canada, F. J., Bruix, M., Rodriguez-Romero, A., & Jimenez-Barbero, J. (1995) *Eur. J. Biochem. 230,* 621−33.
52. Verheyden, P., Pletinckx, J., Maes, D., Pepermans, H. A. M., Wyns, L., Willem, R., & Martins, J. C. (1995) *FEBS Lett. 370,* 245−249.
53. Hom, K., Gochin, M., Peumans, W. J., & Shine, N. (1995) *FEBS Lett. 361,* 157−161.
54. Wolfram, S. (1991) *Mathematica: a system for doing mathematics by computer,* Addison-Wesley Publishing Co., New York.
55. Zürcher, R. F. (1967) in *Progress in NMR Spectroscopy* (Emsley, J. W., Feeney, J., & Sutcliffe, L. H., Ed.) pp 205−257, Pergamen Press, New York.
56. Le, H., & Oldfield, E. (1994) *J. Biomol. NMR 4,* 341−348.

BI971619P