Site-Specific Characterization of the Association of Xylooligosaccharides with the CBM13 Lectin-like Xylan Binding Domain from *Streptomyces lividans* Xylanase 10A by NMR Spectroscopy†

Manuela Schärpf,[‡,§] Gregory P. Connelly,[‡,§,||] Gregory M. Lee,[‡,§] Alisdair B. Boraston,[§,⊥,#,@] R. Antony J. Warren,[§,⊥] and Lawrence P. McIntosh*[,‡,§,#]

Department of Biochemistry and Molecular Biology, Department of Chemistry, Department of Microbiology and Immunology, Protein Engineering Network of Centres of Excellence, and Biotechnology Laboratory, University of British Columbia, Vancouver, BC, Canada V6T 1Z3

Received October 17, 2001; Revised Manuscript Received January 18, 2002

ABSTRACT: Endo-β-1,4-xylanase 10A (Xyn10A) from *Streptomyces lividans* includes an N-terminal catalytic module and a 130-residue C-terminal family 13 carbohydrate-binding module (CBM13). This latter domain adopts a β-trefoil structure with three potential binding sites (α, β, and γ) for a variety of small sugars, xylooligosaccharides, and xylan polymers. To investigate the role of this multivalency in carbohydrate binding, we have used NMR spectroscopy to characterize the interaction of isolated CBM13 with a series of sugars. We have assigned resonances from the main chain nuclei of CBM13 using heteronuclear NMR experiments. Analysis of ^{15}N NMR relaxation data using the extended model free formalism reveals that CBM13 tumbles as an oblate ellipsoid ($D_{\|}/D_{\perp} = 0.80 \pm 0.02$) and that its backbone is relatively rigid on the sub-nanosecond time scale. In particular, the three binding sites show no distinct patterns of increased internal mobility. Ligand-induced chemical shift changes in the ^{1}H−^{15}N HSQC spectra of CBM13 were monitored as a function of increasing concentrations of L-arabinose, lactose, D-xylose, xylobiose, xylotetraose, and xylohexaose. Patterns of shift perturbations for well-resolved resonances demonstrate that all of these sugars associate independently with the three binding sites of CBM13. On the basis of the site-specific association constants derived from a quantitative analysis of these titration data, we show that L-arabinose, lactose, and D-xylose preferentially bind to the α site of CBM13, xylobiose binds equally well to all three sites, and xylotetraose and xylohexaose prefer binding to the β site. Inspection of the crystallographic structure of CBM13 [Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G., and Rose, D. R. (2002) *Biochemistry 41*, 4246−4254] provides a rationalization for these results.

Protein−carbohydrate interactions are critical for numerous major biological processes ranging from synthesis or hydrolysis of structural and storage polysaccharides, intercellular trafficking and signaling, carbohydrate-based cellular recognition, and host−pathogen adhesion. Consequently, it is important to define the structural, thermodynamic, and dynamic bases for these macromolecular interactions.

Glycoside hydrolases (GHs),[1] for example, are carbohydrate-active enzymes generally characterized by a composite structure featuring a catalytic domain combined with one or more auxiliary carbohydrate-binding modules (CBMs). These CBMs appear to facilitate binding or targeting of the enzyme to its substrate, perhaps being augmented by a role in disrupting this complex material. Currently, 29 families of CBMs are described in the latest update of the CAZy database (*2*). These families have been grouped into three classes based on their structures, functions, and ligand specificities. Type A CBMs bind insoluble surfaces, type B CBMs bind to polysaccharides, and type C CBMs bind to mono- and disaccharides (*3*). Within these types, a wide variation in substrate specificity arises. For example, CBMs belonging to families 1, 2a, 3, 5, and 10 bind mainly to crystalline cellulose, while preferences for xylan are exhibited by members of families 2b, 4, 6, 13, and 22 (*4*−*11*). Three-dimensional structures of members of several CBM families

† This work was funded by the Government of Canada's Network of Centres of Excellence Program supported by the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council (NSERC) through the Protein Engineering Network of Centres of Excellence (PENCE, Inc.). L.P.M. acknowledges the Canadian Institutes of Health Research for a Scientist Award. M.S. thanks the Deutsche Forschungsgemeinschaft (DFG) for a postdoctoral fellowship.

* To whom correspondence should be addressed: Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, BC, Canada V6T 1Z3. Telephone: (604) 822-3341. Fax: (604) 822-5227. E-mail: mcintosh@otter.biochem.ubc.ca.

‡ Department of Biochemistry and Molecular Biology and Department of Chemistry.

§ Protein Engineering Network of Centres of Excellence.

|| Present address: Althexis, Waltham, MA 02451.

⊥ Department of Microbiology and Immunology.

Biotechnology Laboratory.

@ Present address: Department of Chemistry, Structural Biology Laboratory, University of York, Heslington, York YO10 5DD, U.K.

[1] Abbreviations: CBM, carbohydrate-binding module; CAZy, carbohydrate-active enzymes; CSI, chemical shift index; GH, glycoside hydrolase; HSQC, heteronuclear single-quantum coherence; K_a, association constant; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; RTB, ricin toxin B chain; TALOS, torsion angle likelihood obtained from shift and sequence similarity; TOCSY, total correlation spectroscopy.

10.1021/bi015866b CCC: $22.00
Published on Web 03/09/2002

are now available from crystallographic as well as NMR spectroscopic studies (*1*, *10*, *12*−*19*). Most of these modules are formed exclusively from β-strands. Complementary to their substrates, type A CBMs exhibit a flat binding surface, whereas those of types B and C have binding grooves.

The first identified family 13 CBMs were plant lectins such as the ricin toxin B chain (RTB) and *Ricinus communis* agglutinin, both of which preferentially bind galactose (*20*). These modules show a sequential 3-fold internal repeat of ~42 residues, which assemble into a β-trefoil structure (*21*). Each of these three repeats, denoted α, β, and γ, is a structurally distinct unit of four β-strands with a disulfide bridge, arranged to form a potential sugar binding site. In general, these lectin-like CBMs exhibit relatively weak binding to a range of small sugars (*22*).

Streptomyces lividans endo-β-1,4-Xyn10A includes an N-terminal catalytic domain, a short linker, and a C-terminal family 13 CBM denoted herein as CBM13. Crystallographic studies of CBM13 (*1*) and of the closely related module from *S. olivaceoviridis* E-86 β-xylanase (*10*) confirm that these domains adopt a pseudo-3-fold symmetric β-trefoil structure. As a lectin-like module, CBM13 exhibits a relatively broad spectrum of weak binding to a variety of mono-, di-, and oligosaccharides (*22*). However, in contrast to the RTB, CBM13 also exhibits a higher degree of affinity and specificity for polymeric xylans (*22*), thereby enhancing the activity of the holoenzyme toward its substrate (*23*). Paralleling this promiscuous spectrum of binding, crystallographic studies show that lactose and xylopentaose associate with CBM13 in distinctively different structural modes (*1*).

Complementing a crystallographic structural study of CBM13 (*1*), we have used NMR spectroscopy to investigate the solution behavior of this *S. lividans* domain. Spectral assignments for isolated CBM13 reflect the characteristic 3-fold sequence and secondary structure repeat of the β-trefoil fold. Analysis of ^{15}N relaxation data yields an axially symmetric diffusion tensor consistent with the oblate ellipsoidal shape of CBM13, while also indicating that the backbone of the protein is relatively rigid on the subnanosecond time scale. Finally, NMR-monitored titrations of CBM13 with L-arabinose, lactose, D-xylose, and a series of xylooligosaccharides clearly demonstrate independent binding of these sugars to each of the α, β, and γ sites of the protein. Quantitative analysis of these titration data provides site-specific association constants. Preferential interaction of the various sugars with the three binding sites in CBM13 can be rationalized on the basis of the crystallographic analysis of this domain complexed with lactose and xylopentaose (*1*).

MATERIALS AND METHODS

Protein Expression and Purification. Samples of isotopically labeled CBM13 were produced by expression of the plasmid pTugCBM13 in *Escherichia coli* JM101 cells (*22*). Overnight cultures were diluted in minimal M9 medium supplemented with kanamycin (35 μg/mL) and $^{15}NH_4Cl$ (1 g/L), $^{15}NH_4Cl$ (1 g/L) and [$^{13}C_6$]glucose (3 g/L), or $^{15}NH_4Cl$ (1 g/L) and [$^{13}C_6$]glucose (0.3 g/L) and unlabeled glucose (2.7 g/L) for production of ^{15}N-, ^{15}N- and ^{13}C-, or ^{15}N- and 10% ^{13}C-labeled CBM13, respectively. After growth at 37 °C to an OD_{600} of 0.6, isopropyl 1-thio-β-D-galactopyranoside was added to a final concentration of 0.5 mM. The cells were incubated at 30 °C for 30 h, followed by sedimentation (6000*g*) for 15 min at 4 °C. The protein was harvested from the periplasmic fraction by resuspension of the cells in 25 mL of osmotic shock buffer [30 mM Tris-HCl (pH 6.8), 20% sucrose, and 1 mM EDTA], centrifugation (8000*g*) for 10 min, resuspension of the pellet in 25 mL of 5 mM $MgSO_4$, and, finally, centrifugation (10000*g*) for an additional 15 min. CBM13 was purified from the supernatant by Ni^{2+} affinity chromatography using buffers of 50 mM Hepes (pH 7.5), 150 mM NaCl, and 5% (v/v) glycerol with 5, 50, and 500 mM imidazole for binding, washing, and elution, respectively. Protein fractions were analyzed on 20% Phast gels (Pharmacia). Pure CBM13 fractions were pooled, dialyzed against a final buffer of 50 mM potassium phosphate (pH 6.1) and 50 mM KCl, and concentrated by ultrafiltration (Amicon, Beverly, MA). The final product contains the full-length 130-amino acid CBM13 sequence of *S. lividans* [residues 348−477 in Xyn10A (*24*)], with an N-terminal six-His tag and factor Xa cleavage site. Residues are numbered from the start of the binding domain (*HHHHHHIEGRAS*-Glu1-Pro2-Pro3...). The concentration of CBM13 was determined by absorbance spectroscopy (calculated ϵ_{280} = 32 300 M^{-1} cm^{-1}) (*25*).

NMR Spectroscopy and Spectral Assignments. Samples for NMR analysis contained ~1.5 mM uniformly ^{15}N-, ^{15}N- and ^{13}C-, or ^{15}N- and 10% ^{13}C-labeled CBM13. All NMR spectra were collected on a Varian Unity 500 MHz spectrometer equipped with a triple-resonance pulse field gradient probe at 30 °C and processed using Felix95 or Felix2000 software (Molecular Simulations, Inc., San Diego, CA). Experiments with $^1H^N$ detection were recorded using enhanced-sensitivity pulsed field gradient methods (*26*, *27*). Selective water flip-back pulses were incorporated to minimize the perturbation of the bulk water magnetization (*28*, *29*). Chemical shifts are referenced to external 2,2-dimethyl-2-silapentanesulfonic acid (*30*). Assignments of signals from main chain nuclei were obtained as previously described from ^{1}H−^{15}N HSQC, CBCA(CO)NH, HNCACB, and HNCO spectra (*31*). Aliphatic side chain assignments were achieved using ^{15}N-edited TOCSY-HSQC, H(CC)(CO)NH-TOCSY, and (H)CC(CO)-NH-TOCSY experiments (*32*). Stereospecific assignments of the resonances from side chain amide groups of asparagine and glutamine were obtained using a quantitative *J* correlation HMQC pulse scheme (*33*). Stereospecific assignments of the resonances from the diastereotopic methyls of valine and leucine were obtained from ^{1}H−^{13}C HSQC spectra of the 10% nonrandomly fractionally ^{13}C-labeled CBM13 (*31*, *34*).

Protein Dynamics from ^{15}N Relaxation Measurements. ^{15}N NMR relaxation measurements were carried out and the results analyzed as previously described (*35*). ^{15}N T_1 values were measured from the spectra recorded with delays of 11.1, 33.3, 55.5, 111, 222, 333, 499.5, 666, and 943.5 ms. T_2 values were determined from spectra recorded with delays of 16.7, 33.4, 50.1, 66.8, 83.5, 100.2, 116.9, 133.6, and 167 ms. Steady-state heteronuclear {^{1}H}−^{15}N NOE values were measured by recording HSQC-based spectra with and without 3 s of proton saturation and a total recycle delay of 5 s. T_1 and T_2 values were obtained by nonlinear least-squares fitting of the cross-peak heights to the two-parameter equation for an exponential decay. NOE values were taken from the ratio

Table 1: Average Site-Specific Binding Constants Measured for the Association of CBM13 with Soluble Sugar Ligands[a]

	K_a (M^{-1})		
sugar	α site	β site	γ site
L-arabinose	70 ± 7	3 ± 1	10 ± 1
lactose	720 ± 110	20 ± 7	20 ± 3
D-xylose	60 ± 4	10 ± 1	15 ± 1
xylobiose	160 ± 30	130 ± 20	180 ± 30
xylotetraose	250 ± 40	1150 ± 200	660 ± 70
xylohexaose	420 ± 35	1200 ± 100	700 ± 70

[a] Data obtained at 30 °C and pH 6.1 in 50 mM potassium chloride, 50 mM potassium phosphate, and a 10% D_2O/90% H_2O mixture.

of the peak heights recorded with and without proton saturation. Uncertainties in all spectra were estimated according to the method of Farrow et al. (*35*). Analysis of the ^{15}N relaxation data to obtain parameters describing the global tumbling of CBM13 as well as internal dynamics according to the extended model free formalism of Lipari and Szabo (*36*) was carried out with TENSOR2 (*37*). A value of 1.02 Å was used as the NH bond length, and the axially symmetric ^{15}N chemical shift tensor ($\sigma_{||} - \sigma_{\perp} = -170$ ppm) was assumed to be coaxial with the dipolar tensor (*38*).

Carbohydrate Titrations Monitored by NMR Spectroscopy. The binding of soluble sugars (purchased from Megazyme International Ltd. Ireland) to CBM13 at 30 °C and pH 6.1 was measured quantitatively using $^1H-^{15}N$ HSQC NMR spectroscopy. Stock solutions of L-arabinose, lactose, D-xylose, xylobiose, xylotetraose, and xylohexaose ranging from 0.02 to 1.2 M were prepared by weight in a sample buffer of 50 mM potassium phosphate (pH 6.1) and 50 mM KCl. Aliquots of these solutions were added directly to uniformly ^{15}N-labeled CBM13 (starting protein concentration of 100−725 μM), contained within an NMR tube. As summarized in Figure 5, 8 to 12 $^1H-^{15}N$ HSQC spectra were recorded with consecutively increasing sugar concentrations for each titration series.

Equilibrium association constants were determined for each of the α, β, and γ binding sites by nonlinear least-squares fitting of the chemical shift data versus sugar concentration to the Langmuir isotherm describing the binding of one ligand molecule to a single protein site. The program PLOTDATA (TRIUMF, UBC, Vancouver, BC) was used for this analysis. The methodology follows that of Johnson et al. (*39*), with the assumption of three independent binding sites of approximately equal affinity in order to back-calculate the concentration of free sugar during the fitting process. Given the excess of sugar over protein in these titration series due to their generally weak binding, this assumption does not significantly alter the derived site-specific K_a values (i.e., values calculated for one site without correction for changes in sugar concentration due to binding at the remaining two sites lie within a standard deviation of those presented in Table 1).

RESULTS

Spectral Assignments. Nearly complete assignments of the resonances from the main chain $^1H^N$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ nuclei of CBM13 were obtained using a suite of triple-resonance heteronuclear NMR experiments. The only exceptions are Ser88 and the N-terminal six-His tag, the amide

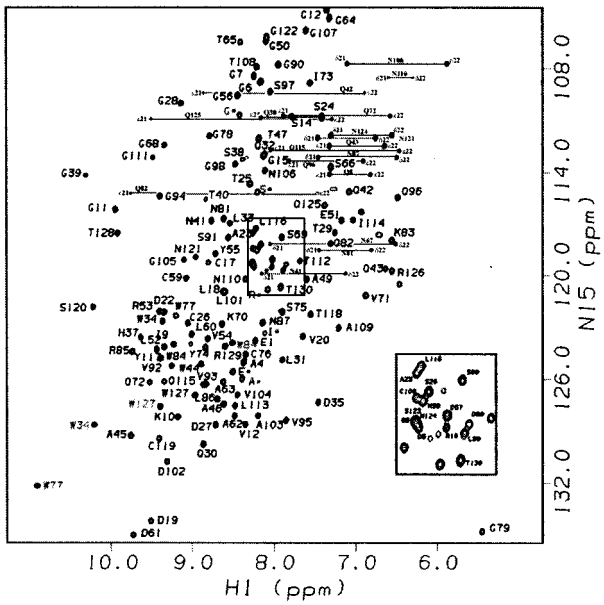

FIGURE 1: $^1H-^{15}N$ HSQC spectrum of CBM13 at 30 °C and pH 6.1, showing the assignments of the resonances from backbone amide (black labels) and tryptophan indole $^{15}N^{\epsilon 1}H$ groups (gray labels), as well as those of the observable signals from the side chain $^{15}N^{\delta}H_2$ and $^{15}N^{\epsilon}H_2$ groups of asparagine and glutamine (lines). Amino acids, N-terminal to Glu1, that were inserted from the cloning vector are denoted with asterisks. The peak from Gly79 is aliased.

resonances of which were not observed, possibly due to conformational or solvent exchange broadening. These assignments, along with those of several additional aliphatic ^1H and ^{13}C nuclei, have been deposited in the BioMagResBank database.

Figure 1 shows the assigned $^1H-^{15}N$ HSQC spectrum of CBM13 at pH 6.1 and 30 °C. The excellent dispersion of both amide proton and nitrogen resonances is indicative of a stable, well-folded protein. Given the availability of a high-resolution crystallographic structure of CBM13, alone and in complex with lactose and xylopentaose (*1*), we did not pursue a detailed conformational analysis by NMR methods. However, a brief analysis of the secondary chemical shift changes ($\Delta\delta = \delta_{observed} - \delta_{random-coil}$) measured for CBM13 using the CSI method ($^1H^{\alpha}$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$) (*30*, *40*) and TALOS ($^1H^{\alpha}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and ^{15}N) (*41*) was consistent with a β-trefoil fold formed by 12 short β-strands (*1*). All 12 strands were identified by TALOS, whereas the first two strands were not clearly predicted by the CSI algorithm. This difference may result from the relatively short β-strands present in CBM13, as TALOS provides an estimate of the main chain (Φ and Ψ) dihedral angles for each residue, whereas the CSI algorithm uses patterns of secondary chemical shifts for neighboring residues to define a consensus secondary structure. More strikingly, inspection of the secondary chemical shifts of CBM13 revealed a repetitive pattern matching the tandem sequence repeat of ~42 residues observed for this protein. An example is shown in Figure S1 of the Supporting Information for the $^1H^{\alpha}$ shifts. This sequence repeat yields the characteristic pseudo-3-fold symmetry of the β-trefoil structure, and creates the three sugar-binding sites (α, β, and γ) in CBM13. Finally, the $^{13}C^{\beta}$ chemical shifts of all six Cys residues were between 42.0 and 45.4 ppm, consistent with their formation of disulfide bridges within each trefoil repeat (*42*).

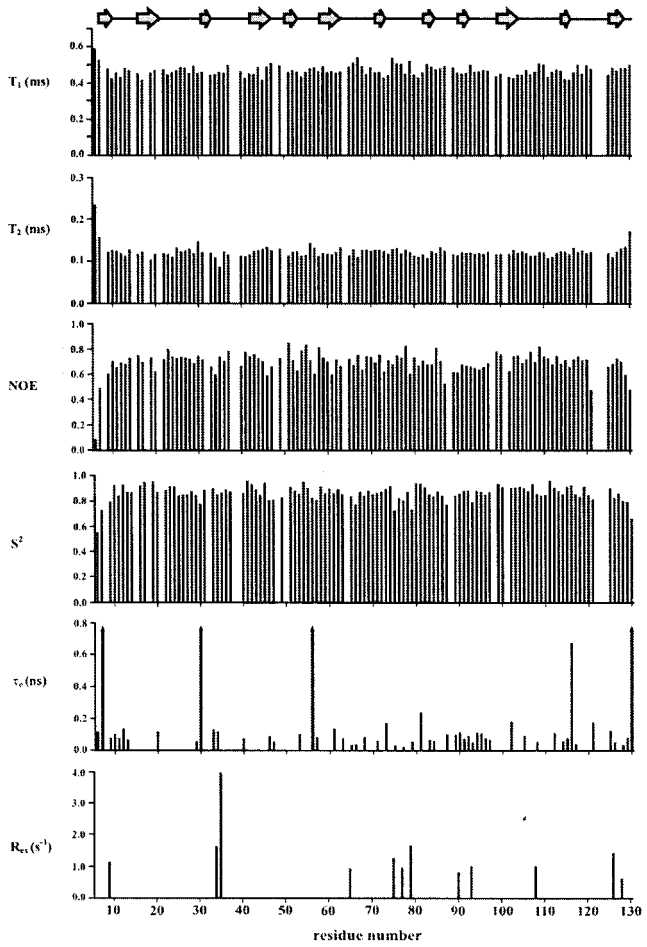

FIGURE 2: Plots of the measured ^{15}N T_1, T_2, and {^{1}H}−^{15}N NOE relaxation parameters and the model free parameters S^2, τ_e, and R_{ex} as a function of residue number for CBM13. The τ_e values, labeled with a vertical arrow in the histogram, are 1.2, 6.2, 2.9, and 1.5 ns for Gly7, Gln30, Gly56, and Thr130, respectively. Empty horizontal arrows indicate the locations of the 12 β-strands in CBM13.

^{15}N Relaxation Measurements. Of the 138 possible correlations involving backbone amide protons, reliable T_1, T_2, and heteronuclear NOE relaxation data were obtained for 110 residues in CBM13 (Figure 2). The remaining residues were not analyzed due to peak overlap (Asp5, Gln8, Gly15, Leu18, Gln32, Ser38, Asp48, Gly50, Gly64, Gly98, Leu101, Gly122, Ser123, and Asn124) or due to extremely weak or missing signals (the N-terminal His-His-His-His-His-His-Ile-Glu-Gly-Arg-Ala-Ser tag, Glu1, Gly39, and Ser88). For these 110 residues, the average T_1 and T_2 values were 467 ± 29 and 122 ± 15 ms, respectively. After the data had been filtered to exclude residues showing a high degree of internal mobility or ^{15}N line broadening (*43*), a fit of these T_1/T_2 ratios with TENSOR2 yielded a correlation time (τ_m) of 6.32 ± 0.03 ns for isotropic global tumbling. This value is indicative of a compact monomeric protein of ~15 kDa (*44*). Using the crystallographic structure of CBM13 (*1*), the T_1 and T_2 relaxation data could be fit in terms of an axially symmetric diffusion tensor with a $D_\parallel/D_\perp$ ratio of 0.80 ± 0.02. The unique axis of the diffusion tensor is approximately collinear with the pseudo-3-fold axis of CBM13. This degree of anisotropy is consistent with that expected from the oblate ellipsoidal shape of this molecule, which has calculated relative moments of inertia of 1.0, 0.81, and 0.78. Use of a fully anisotropic model of global tumbling did not significantly improve the fit of the relaxation data, as judged by an *F* test implemented in TENSOR2.

Having defined the global motional parameters of CBM13, we analyzed the ^{15}N T_1, T_2, and NOE relaxation data using the model free formalism under conditions of anisotropic tumbling. This analysis involved fitting of these data to one of the five standard models of internal mobility, specifically, S^2; S^2, and τ_e; S^2 and R_{ex}; S^2, τ_e, and R_{ex}; and S_s^2, S_f^2, and τ_s (*35*, *45*). On the basis of the model selection criteria implemented in TENSOR2, the average value of the generalized order parameter S^2 for these residues is 0.86 ± 0.06, indicative of little backbone mobility on a sub-nanosecond time scale. With the exception of the N- and C-termini of CBM13, only Gln30, Ser66, Ser75, Gly79, and Asn87 exhibit S^2 values of <0.8. These latter polar residues are located within exposed loop regions of the protein. Four residues (Gly7, Gln30, Gly56, and Thr130) fit to the extended model (S_s^2, S_f^2, and τ_s). A limited number of residues, including Ile9, Trp34, Asp35, Thr65, Ser75, Trp77, Gly79, Gly90, Val93, Thr108, Arg126, and Thr128, also exhibit reduced ^{15}N T_2 values as reflected by the need to include R_{ex} terms in the model free fit. Although these residues generally lie in loop regions, it is interesting that Trp34 and Asp35 form part of the α binding site, whereas Ser75, Trp77, and Gly79 are located in the β site. Since R_{ex} terms may arise from motions on a millisecond to microsecond time scale, it is tempting to ask if this relaxation behavior reflects plasticity in CBM13 that may relate to its promiscuity in binding a range of small sugars. However, more detailed studies to specifically characterize slow time scale motion, such as field-dependent CPMG relaxation measurements (*46*), with and without sugars, are necessary to address this issue. Overall, however, the ^{15}N relaxation study shows that CBM13 is a compact oblate ellipsoid with relatively uniform and limited internal backbone mobility.

Identification of Sugar Binding Sites. The interaction of L-arabinose, lactose, D-xylose, xylobiose, xylotetraose, and xylohexaose with ^{15}N-labeled CBM13 was monitored by ^{1}H−^{15}N HSQC spectroscopy. Chemical shift perturbation mapping is a highly sensitive method for identifying ligand-binding sites within a protein. Perturbations of main chain amide resonances may arise due to the direct interaction with the ligand or indirectly due to conformational changes resulting from the formation of the protein−sugar complex. An overlay of 12 ^{1}H−^{15}N HSQC spectra shows the progressive chemical shift changes of several CBM13 amide resonances upon addition of xylobiose (Figure 3). These changes indicate binding under conditions of fast exchange on the chemical shift time scale. Figure 4 depicts the overall effect of xylobiose binding by mapping the observed main chain and side chain ^{15}N chemical shift changes ($\Delta\delta$) on the crystal structure of CBM13. The largest perturbations are observed for backbone amides, as well as side chain amide and indole groups, of residues within or adjacent to the three binding sites that were identified from crystallographic analysis of CBM13−ligand complexes (*1*). Furthermore, comparable chemical shift changes are observed for all three binding sites, and similar patterns of $\Delta\delta$ values were measured for each titration series (data not shown). This confirms that, in solution, CBM13 binds L-arabinose,

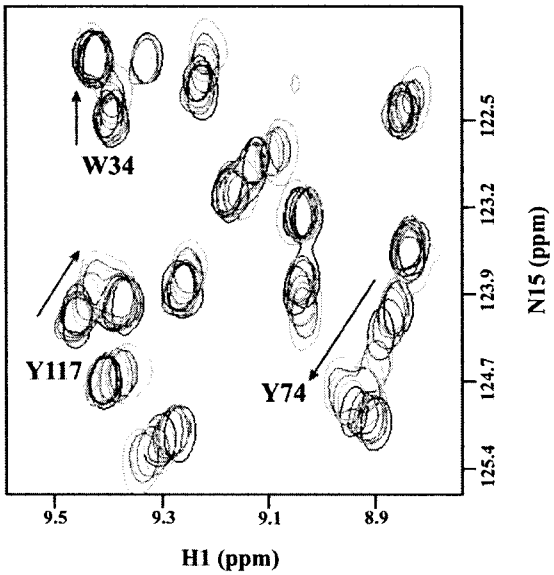

FIGURE 3: CBM13 binds xylooligosaccharides at the α (W34), *β* (Y74), and *γ* (Y117) sites as demonstrated by chemical shift perturbations. Selected regions of 12 $^{1}H-^{15}N$ HSQC spectra of uniformly ^{15}N-labeled CBM13 (100 *μ*M) in the presence of 0, 0.1, 0.2, 0.3, 0.7, 1.5, 3.4, 7.0, 12.3, 22.0, 43.4, and 86.6 mM total xylobiose are shown. The arrows indicate the direction of the chemical shift changes resulting from the addition of sugar to the protein. For clarity, only peaks from the three conserved aromatic residues in the binding sites are labeled.

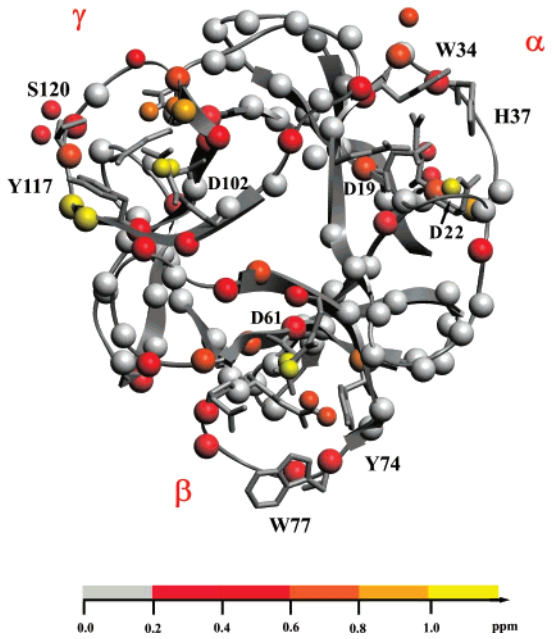

FIGURE 4: Mapping of the ^{15}N chemical shift perturbations, resulting from addition of excess xylobiose, on a ribbon diagram of the CBM13 structure indicates binding of the sugar to the α, *β*, and *γ* sites of this domain. Large and small spheres identify residues for which a change in main chain and side chain ^{15}N resonances, respectively, could be measured due to titration with xylobiose. The color scale indicates the absolute change in the ^{15}N chemical shift due to saturation of CBM13 with this sugar. Three homologous residues for each binding site are labeled, as well as Asp22 in the α site.

lactose, and xylooligosaccharides at three homologous sites within its *β*-trefoil structure.

Site-Specific Binding Constants Determined by NMR Spectroscopy. The association constants (K_a) describing the interactions of L-arabinose, lactose, D-xylose, xylobiose, xylotetraose, and xylohexaose with CBM13 were determined by nonlinear least-squares fitting of the chemical shift titration data to a simple Langmuir binding isotherm. Figure 5 shows examples of the analyzed experimental ^{15}N chemical shift data for the amides of aromatic residues within each CBM13 binding site, namely, Trp34 (α site), Tyr74 and Trp77 (*β* site), and Tyr117 (*γ* site). In a similar manner, association constants were determined from titration curves measured for main chain and side chain ^{15}NH groups of Asp19, Gln32, Trp34, His37, Asn41, and Gln42 within the α site, Asp61, Gln72, Tyr74, Trp77, Asn81, and Gln82 in the *β* site, and Asp102, Gln115, Tyr117, Ser120, Asn124, and Gln125 in the *γ* site. These residues, which exhibited the most significant perturbations in ^{15}N chemical shift upon sugar binding, all lie within or adjacent to the three binding sites of CBM13 identified by crystallographic studies (*1*). A summary of the average site-specific association constant measured for each sugar is given in Table 1. These data were analyzed according to a simple model of each CBM13 binding site interacting with one sugar molecule in an independent or noncooperative manner. This assumption is supported by the following evidence. First, main chain and side chain ^{15}NH groups within a given binding site show monophasic changes in chemical shift upon addition of each sugar (Figure 3), as well as co-incident nomalized titration curves. This discounts the possibility of multiple binding events with differing affinities occurring at any given site. Second, the chemical shift perturbations are localized to the approximate regions of the three crystallographically identified binding sites in CBM13 (Figure 4), indicating that there is no structural coupling between binding events at each distinct site (i.e., as would be expected for homotropic allostery). Third, the titration data follow a hyperbolic dependence of chemical shift change versus ligand concentration, and thus could be adequately fit to the Langmuir binding isotherm describing this simple model (Figure 5). Finally, crystallographic studies (*1*) reveal that the three binding sites in CBM13 are relatively small, discounting the possibility of two or more sugar molecules lying simultaneously within any given site.

DISCUSSION

Spectral Assignments and Secondary Structure. We have assigned the resonances from backbone nuclei of isolated CBM13 using multidimensional heteronuclear NMR spectroscopy. The secondary structure derived from these chemical shifts using CSI and TALOS is consistent with that determined crystallographically for CBM13 (*1*). As shown in Figure 4, this *β*-trefoil structure is composed of three tandem repeats of ~42 residues, each containing four *β*-strands stabilized by a disulfide bridge.

Backbone Dynamics of CBM13. Consistent with the oblate ellipsoidal shape of CBM13, the ^{15}N relaxation data measured for this protein are best fit by a model of axially symmetric global tumbling with a $D_{\parallel}/D_{\perp}$ ratio of 0.8 ± 0.02. Further analysis of these data using the model free formalism in the context of anisotropic tumbling reveals that more than 88% of the residues in CBM13 have order parameter (S^2) values of >0.8. This demonstrates that the protein backbone is relatively rigid on a sub-nanosecond time scale, indicating that CBM13 is a compact and globular protein with no unusually flexible regions, besides its termini. Thus, the

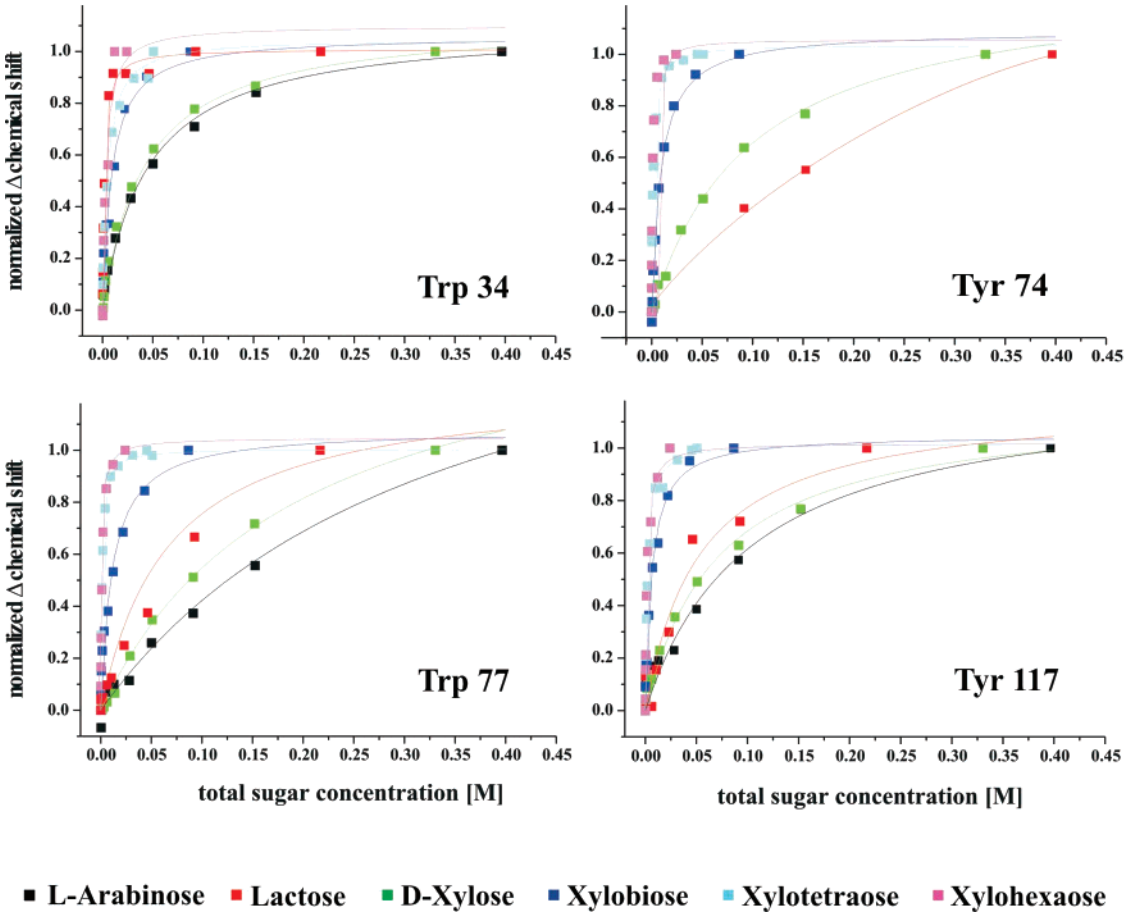

FIGURE 5: Association constants (K_a) for the independent, noncooperative binding of L-arabinose (black), lactose (red), D-xylose (green), xylobiose (blue), xylotetraose (cyan), and xylohexaose (magenta) to sites α (Trp34), β (Tyr74 and Trp77), and γ (Tyr117) were determined by fitting the normalized chemical shift titration data to the Langmuir isotherm.

β-trefoil structure appears to be a rigid scaffold upon which sugar binding sites are grafted.

L-Arabinose, D-Xylose, and Lactose Recognition by CBM13. As shown previously (*22*, *47*), CBM13 exhibits broad specificity for mono- and oligosaccharides with pyranose sugars, as well as a variety of insoluble polymers, including xylan, holocellulose, pachyman, lichenan, arabinogalactan, and laminarin. In this study, we investigated the site-specific binding of L-arabinose, D-xylose, and lactose to CBM13. In each case, we observed a significant preference for the α site, followed by similar affinities for the β and γ sites. Macroscopic association constants measured previously for these sugars by fluorescence spectroscopy are comparable to the values determined by NMR methods for the α site (*22*). Note that this site contains a central tryptophan residue, whereas the β and γ sites have corresponding tyrosine residues. This likely biases the fluorescence data toward binding events occurring at the α site. Small discrepancies in the K_a values may also originate from the slightly different experimental conditions, such as temperature, ionic strength, and pH, that were used for these studies (*22*).

A rationalization for these site-specific binding constants is provided by consideration of the crystallographic structures of CBM13 complexed with lactose and xylopentaose (*1*). Lactose is found to bind with its long axis perpendicular to the surface of CBM13 such that its nonreducing galactose ring contacts residues in the α and γ sites (see Figure 3 of ref *1*). Surprisingly, the glucose unit of a lactose molecule, bound to the γ site of CBM13, also binds to the β site of a neighboring CBM13 molecule within the crystal lattice. NMR studies provide no indication, such as line broadening, of the formation of dimers or oligomers during the titration of lactose to CBM13. We, therefore, suggest that the cross-linking of neighboring molecules by lactose is a result of the appropriate juxtaposition of their β and γ sites within the crystal lattice. We are, however, unable to determine whether lactose binds to any site in CBM13 exclusively by its galactose or glucose moiety. Common to these two modes of interaction is a pair of hydrogen bonds between vicinal equatorial hydroxyl groups of the pyranose ring (O3 and O4 of galactose and O1 and O2 of glucose) and the carboxylate oxygens of a conserved aspartic acid (Asp19, -61, or -102) within the α, β, and γ sites (Figure 4). In contrast to the perpendicular orientation of lactose, xylopentaose is observed to bind with its long axis parallel to the surface of CBM13 in clefts along the α and β sites (see Figure 5 of ref *1*). Although globally distinct, xylopentaose also exhibits the common feature of equatorial hydroxyl groups (O2 and O3) binding to the carboxylate oxygens of the conserved aspartic acids within these sites. We surmise that D-xylose binds to CBM13 in a manner similar to that observed for a single saccharide unit in xylopentaose, i.e., with its C1−C4 axis parallel to the protein surface and its equatorial O2 and O3 hydroxyls hydrogen bonding to aspartate side chains. This is supported by the observation that L-arabinose and D-xylose, which share common hydroxyl configurations at the ring 2

and 3 positions, while differing at the 4 position, exhibit similar K_a values for the corresponding sites in CBM13. Given that this sugar−aspartic acid hydrogen bonding interaction is conserved for all three sites in CBM13, we attribute the higher affinity of the α site for D-xylose and L-arabinose to two possible causes: (i) the presence of Trp34 in the α site, which likely provides more favorable aromatic ring−pyranosyl ring stacking than Tyr74 (*β* site) and Tyr117 (*γ* site), and (ii) the presence of His37 in the α site, which provides sugar hydrogen bonding interactions that are not observed with Trp77 (*β* site) or Ser120 (*γ* site) (*1*).

Conspicuous in Table 1 is the significantly higher K_a of lactose for the α site over that for the *β* or *γ* site, or over those exhibited by D-xylose and L-arabinose for any site. In addition to the two features of the α site described above, crystallographic studies also demonstrate that a nonconserved Asp22 is positioned in this site to accept a hydrogen bond from the glucose moiety of the lactose molecule. In sites *β* and *γ*, glycine residues occupy this corresponding position. A similar aspartic acid is also found in one of the two lactose-binding sites in RTB (*48*). In support of this argument, we note that the amide resonances of Asp22 and Trp34, both in the α site, shift in opposite directions when CBM13 is titrated with lactose compared to the shift with any other sugar studied herein. This could reflect the additional distinct interactions of lactose with Asp22 resulting in tighter binding. Although this hypothesis provides a reasonable explanation for the preferential binding of lactose to the α site, the hydrogen bonding interaction may be relatively weak, as the glucose ring of lactose bound at this site in crystalline CBM13 appears to be disordered with very little observable electron density (*1*).

Binding of Xylooligosaccharides. Although CBM13 exhibits weak binding to a broad range of mono-, di-, and oligosaccharides, as a component of *S. lividans* Xyn10A, its natural target is xylan. To investigate this functionally relevant interaction, we studied the binding of CBM13 to a series of xylooligosaccharides. As shown in Table 1, these sugars bind to all three sites in this module with affinities increasing in the following order: D-xylose < xylobiose < xylotetraose ∼ xylohexaose. On the basis of this series, each binding site appear to span approximately four xylosyl units. This conclusion is consistent with crystallographic results (*1*) in which xylopentaose is observed to lie across the α and *β* sites of CBM13 with direct intermolecular contacts involving three to four xylosyl groups. Although these two xylopentaose molecules bind in opposite chain orientations with respect to the *β*-trefoil repeat, each exhibits the previously discussed central pair of hydrogen bonds to a conserved aspartic acid residue (*1*). Additional interactions are provided primarily by intermolecular hydrogen bonds involving polar side chains along the shallow surfaces of the binding clefts. These include Asp19, Asp22, Gln32, His37, Asn41, and Gln42 in the α site and Asp61, Gln72, Asn81, and Gln82 in the *β* site. While a xylopentaose molecule was not observed in the *γ* site of the crystal complex with CBM13, NMR spectroscopic data clearly indicate binding to this site in solution. Potential hydrogen bonding groups in the *γ* site include Asp102, Asn106, Gln115, Ser120, Asn124, and Gln125.

Although primarily involving polar interactions, binding of sugars to CBM13 also appears to be mediated by van der Waals stacking between pyranosyl rings and aromatic side chains. Such hydrophobic ring stacking interactions are a common feature of protein−carbohydrate complexes and have been shown to be pivotal for CBMs binding to crystalline and amorphous cellulose, xylan, and starch (*6*, *49*−*52*). Central to the three binding sites are the conserved residues Trp34 (α), Tyr74 (*β*), and Tyr117 (*γ*). Importantly, the *β* site also contains a second aromatic ring, Trp77, along the edge of its binding cleft. Crystallographic studies reveal that the side chain rings of Tyr74 and Trp77 are oriented in an approximately perpendicular manner to match the helical twist of the bound xylopentaose (*1*). The presence of this additional nonconserved aromatic residue provides a simple explanation for the tighter binding of xylotetraose and xylohexaose to the *β* site of CBM13. Verifications of these arguments could be obtained through detailed calorimetric studies of the interaction of CBM13 with sugars, as required to determine the enthalpic and entropic contribution to binding (*50*).

Binding of Polymeric Xylan. Previous studies of the binding of CBM13 to soluble xylan revealed a greater than additive decrease in affinity due to the combined effect of mutations in the α and *γ* sites (*22*). This led to a proposal of cooperativity involving the simultaneous binding of a single xylan chain to two sites on one CBM13 molecule. However, in light of our current studies, this original hypothesis does not appear to be correct. Crystallographic studies (*1*) reveal that, although corresponding positions on the α, *β*, and *γ* sites of CBM13 are separated by the length of approximately six xylosyl units, such a mode of binding is not structurally feasible without a considerable bend in the xylan chain. In parallel, NMR titration measurements demonstrate that xylohexaose and xylotetraose exhibit similar K_a values in each repetitive binding site in CBM13. This behavior stands in contrast to an expected increase in affinity if the longer sugar could cooperatively bind two sites on one protein molecule. Furthermore, both xylooligosaccharides also cause similar chemical shift perturbations in CBM13 upon titration, indicative of similar modes of binding (i.e., with no additional protein−sugar contacts, involving residues located between binding sites, observed for xylohexaose). Unfortunately, defined xylooligosaccharides longer than xylohexaose are not readily available for further examination of this hypothesis, and birchwood xylan is of insufficient solubility to carry out NMR-based titration measurements. Note, however, using affinity gel electrophoresis, Boraston et al. (*22*) determined an apparent macroscopic binding constant of 1.13 L/g of soluble birchwood xylan or, based on an average degree of polymerization of 44, a value of 6200 L/mol of polymer chain. In the simplest interpretation, this corresponds to an average association constant on the order of 2000 L/mol of polymer chain for each of the three binding clefts in CBM13. Since each cleft spans approximately four xylosyl units (*1*), a xylan polymer chain could represent 41 potential binding sites for the first bound protein and 11 sites at saturation; these numbers would of course be reduced by excluded volume effects and substitutions or branching of the xylan (*52*). If we conservatively assume a lower limit of only four binding sites per chain, this would translate to an apparent binding constant with an upper limit of $\sim$500 M^{-1} for each CBM13 cleft to its target sites on the xylan chain. This value is

comparable to that measured herein for xylotetraose and xylohexaose (Table 1), indicating that CMB13 does not bind xylan with enhanced microscopic or site-specific affinity relative to these oligosaccharides.

In light of these results, a simple alternative explanation for the mutational studies of apparent cooperativity in CBM13 binding to polymeric xylan reported previously (*22*) is that one protein molecule can associate simultaneously with different strands of sugar. In this respect, the apparent affinity constant measured by electrophoretic methods reflects the contribution of all possible bound states of the protein within the gel matrix, and thus, the effects of multiple mutations appear to be greater than additive due to the elimination of simultaneous binding of one CMB13 molecule to different chains of xylan. This model is clearly supported by our NMR titration data, demonstrating that xylooligosaccharides bind independently to all three sites in CBM13 (Figure 4). In parallel, crystallographic analysis of the CBM13−xylopentaose complex reveals different sugar molecules bound in the α and β sites of a single protein molecule. Interestingly, each xylopentaose is actually bound to the α site of one CBM13 and the β site of another CBM13 within the crystal lattice, indicating that a strand of xylan could be highly saturated by these binding domains. However, in solution, we did not observe any evidence of sugar-mediated oligomerization of CBM13, which could also provide a potential avenue for cooperative association with polymeric xylan. In closing, these studies demonstrate a wide variety of modes by which a multivalent binding module, such as CBM13, can associate with polysaccharide ligands. Understanding the role of these interactions in the degradation of xylan clearly requires a more detailed description of this complex carbohydrate in its natural environment.

NOTE ADDED IN PROOF

Fujimoto et al. (*54*) have recently reported an X-ray crystallographic analysis of xylose, xylobiose, xylotriose, glucose, galactose, and lactose bound to the α- and γ-sites of the homologous family 13 xylan binding domain from *Streptomyces olivaceoviridis* E-86.

ACKNOWLEDGMENT

We thank Douglas Kilburn for support and guidance and Valerie Notenboom and David Rose for helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Figure S1 showing the repetitive pattern of secondary $^1H^\alpha$ chemical shifts of CBM13. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G., and Rose, D. R. (2002) *Biochemistry 41*, 4246−4254.
2. Coutinho, P. M., and Henrissat, B. (1999) in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., Eds.) The Royal Society of Chemistry, Cambridge, U.K.
3. Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G., and Warren, R. A. J. (1999) in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., Eds.) pp 202−211, Royal Society of Chemistry, Cambridge, U.K.
4. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J. 15*, 5739−5751.
5. Brun, E., Moriaud, F., Gans, P., Blackledge, M. J., Barras, F., and Marion, D. (1997) *Biochemistry 36*, 16074−16086.
6. Ponyi, T., Szabo, L., Nagy, T., Orosz, L., Simpson, P. J., Williamson, M. P., and Gilbert, H. J. (2000) *Biochemistry 39*, 985−991.
7. Simpson, P. J., Hefang, X., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000) *J. Biol. Chem. 275*, 41137−41142.
8. Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxå, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000) *Biochem. J. 345*, 53−60.
9. Sun, J. L., Sakka, K., Karita, S., Kimura, T., and Ohmiya, K. (1998) *J. Ferment. Bioeng. 85*, 63−68.
10. Fujimoto, Z., Kuno, A., Kaneko, S., Yoshida, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2000) *J. Mol. Biol. 300*, 575−585.
11. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M. A., Davies, G. J., and Fontes, C. M. G. A. (2000) *Biochemistry 39*, 5013−5021.
12. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) *Structure 7*, 853−864.
13. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
14. Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) *Biochemistry 34*, 6993−7009.
15. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J. 15*, 5739−5751.
16. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) *Nat. Struct. Biol. 4*, 810−818.
17. Brun, E., Johnson, P. E., Creagh, A. L., Tomme, P., Webster, P., Haynes, C. A., and McIntosh, L. P. (2000) *Biochemistry 39*, 2445−2458.
18. Notenboom, V., Boraston, A. B., Kilburn, D. G., and Rose, D. R. (2001) *Biochemistry 40*, 6248−6256.
19. Notenboom, V., Boraston, A. B., Chiu, P., Freelove, A. C. J., Kilburn, D. G., and Rose, D. R. (2001) *J. Mol. Biol. 314*, 797−806.
20. Olsnes, S., Refsnes, K., and Pihl, A. (1974) *Nature 249*, 627−631.
21. Montfort, W., Villafranca, J. E., Monzingo, A. F., Ernst, S., Kazin, B., Rutenber, E., Xuong, N. H., Hamlin, R., and Robertus, J. D. (1987) *J. Biol. Chem. 262*, 5398−5403.
22. Boraston, A. B., Tomme, P., Amandoron, E. A., and Kilburn, D. G. (2000) *Biochem. J. 350*, 933−941.
23. Dupont, C., Roberge, M., Shareck, F., Morosoli, R., and Kluepfel, D. (1998) *Biochem. J. 330*, 41−45.
24. Vincent, P., Shareck, F., Dupont, C., Morosoli, R., and Kluepfel, D. (1997) *Biochem. J. 322*, 845−852.
25. Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) *Anal. Biochem. 200*, 74−80.
26. Kay, L., Keifer, P., and Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
27. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 203−216.
28. Bax, A., and Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131−138.
29. Zhang, O., Kay, L. E., Olivier, J. P., and Foreman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.
30. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
31. Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G., and McIntosh, L. P. (1996) *Biochemistry 35*, 14381−14394.
32. Sattler, M., Schleucher, J., and Griesinger, C. (1999) *Prog. NMR Spectrosc. 34*, 93−158.

33. McIntosh, L. P., Brun, E., and Kay, L. E. (1997) *J. Biomol. NMR 9*, 306−312.
34. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wüthrich, K. (1989) *Protein Sci. 6*, 1197−1209.
35. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
36. Lipari, G., and Szabo, A. (1992) *J. Am. Chem. Soc. 114*, 4546−4559.
37. Dosset, P., Marion, D., and Blackledge, M. (2001) *J. Biomol. NMR 20*, 23−31.
38. Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. (1996) *J. Biomol. NMR 8*, 273−284.
39. Johnson, P. E., Tomme, P., Joshi, M. D., and McIntosh, L. P. (1996) *Biochemistry 35*, 13895−13906.
40. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
41. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR 13*, 289−302.
42. Sharma, D., and Rajarathnam, K. (2000) *J. Biomol. NMR 18*, 165−171.
43. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995) *J. Am. Chem. Soc. 117*, 12562−12566.
44. Wagner, G. (1997) *Nat. Struct. Biol. 4*, 841−845.
45. Mandel, A. M., Akke, M., and Palmer, A. G., III (1995) *J. Mol. Biol. 246*, 144−163.
46. Palmer, A. G., III, Kroenke, C. D., and Loria, J. P. (2001) *Methods Enzymol. 338*, 204−238.
47. Kuno, A., Kaneko, S., Ohtsuki, H., Ito, S., Fujimoto, Z., Mizuno, H., Hasegawa, T., Taira, K., Kusakabe, I., and Hayashi, K. (2000) *FEBS Lett. 482*, 231−236.
48. Rutenber, E., and Robertus, J. D. (1991) *Proteins 10*, 260−269.
49. Nagy, T., Simpson, P. J., Williamson, M. P., Hazlewood, G. P., Gilbert, H. J., and Orosz, L. (1998) *FEBS Lett. 429*, 312−316.
50. Kormos, J., Johnson, P. E., Brun, E., Tomme, P., McIntosh, L. P., Haynes, C. A., and Kilburn, D. G. (2000) *Biochemistry 39*, 8844−8852.
51. Xie, H., Bolam, D. N., Nagy, T., Szabo, L., Cooper, A., Simpson, P. J., Lakey, J. H., Williamson, M. P., and Gilbert, H. J. (2001) *Biochemistry 40*, 5700−5707.
52. Williamson, M. P., Le Gal-Coeffet, M. F., Sorimachi, K., Furniss, C. S., Archer, D. B., and Williamson, G. (1997) *Biochemistry 36*, 7535−7539.
53. Cantor, C. R., and Schimmel, P. R. (1980*) Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*, W. H. Freeman and Co., San Francisco.
54. Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002) *J. Mol. Biol. 316*, 65−78.

BI015866B