Interaction of Soluble Cellooligosaccharides with the N-Terminal Cellulose-Binding Domain of *Cellulomonas fimi* CenC. 2. NMR and Ultraviolet Absorption Spectroscopy†

Philip E. Johnson,‡ Peter Tomme,§ Manish D. Joshi,‡ and Lawrence P. McIntosh*,‡

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

Received May 17, 1996⊗

ABSTRACT: The N-terminal cellulose-binding domain (CBD_{N1}) from *Cellulomonas fimi* β-1,4-glucanase CenC binds amorphous but not crystalline cellulose. To investigate the structural and thermodynamic bases of cellulose binding, NMR and difference ultraviolet absorbance spectroscopy were used in parallel with calorimetry (Tomme, P., Creagh, A. L., Kilburn, D. G., & Haynes, C. A., (1996) *Biochemistry 35*, 13885−13894] to characterize the interaction of soluble cellooligosaccharides with CBD_{N1}. Association constants, determined from the dependence of the amide ^1H and ^{15}N chemical shifts of CBD_{N1} upon added sugar, increase from 180 ± 60 M^{-1} for cellotriose to 4 200 ± 720 M^{-1} for cellotetraose, 34 000 ± 7 600 M^{-1} for cellopentaose, and an estimate of 50 000 M^{-1} for cellohexaose. This implies that the CBD_{N1} cellulose-binding site spans approximately five glucosyl units. On the basis of the observed patterns of amide chemical shift changes, the cellooligosaccharides bind along a five-stranded β-sheet that forms a concave face of the jelly-roll β-sandwich structure of CBD_{N1}. This β-sheet contains a strip of hydrophobic side chains flanked on both sides by polar residues. NMR and difference ultraviolet absorbance measurements also demonstrate that tyrosine, but not tryptophan, side chains may be involved in oligosaccharide binding. These results lead to a model in which CBD_{N1} interacts with soluble cellooligosaccharides and, by inference, with single polysaccharide chains in regions of amorphous cellulose, primarily through hydrogen bonding to the equatorial hydroxyl groups of the pyranose rings. Van der Waals stacking of the sugar rings against the apolar side chains may augment binding. CBD_{N1} stands in marked contrast to previously characterized CBDs that absorb to crystalline cellulose via a flat binding surface dominated by exposed aromatic rings.

Cellulolytic bacteria and fungi produce a plethora of enzymes to efficiently degrade cellulose (Beguin & Aubert, 1994; Henrissat, 1994; Tomme et al., 1995a). A common feature of these cellulases is their modular design. The binding of the enzymes to the substrate is generally mediated by a cellulose-binding domain (CBD),[1] while hydrolysis is carried out by a distinct catalytic domain. Over 100 different CBDs have been identified and grouped into ten families based primarily on their amino acid sequence similarities (Tomme et al., 1995b). It is thought that CBDs potentially play two roles in cellulose degradation. The most obvious of these is to concentrate the enzyme near the surface of cellulose. However, the CBDs may have a more active function in disrupting the cellulose surface, thereby making the polysaccharide chains more accessible for hydrolysis (Knowles et al., 1987; Din et al., 1991; Teeri et al., 1992).

Cellulose is a polymer of β-1,4-D-glucose and, in its naturally occurring state, is often very heterogeneous. Cellulose occurs in several crystalline phases, as well as in allomorphs resulting from the disruption of the non-covalent forces that normally bind the individual carbohydrate strands together within the crystal lattice. A region of cellulose that is highly disrupted, and thus low in crystallinity, is referred to as amorphous cellulose (Atalla, 1993; Tomme et al., 1995a). Amorphous cellulose is generated experimentally as phosphoric acid-swollen cellulose. While chemically identical with crystalline cellulose, amorphous cellulose presents a highly irregular surface for a protein to bind.

Currently there are two published structures of CBDs: those of *Trichoderma reesi* cellobiohydrolase 1 (CBD_{CBH1};

† This work was supported by a grant from the Protein Engineering Network of Centres of Excellence (L.P.M.). P.E.J. is the recipient of a MacMillan-Bloedel Limited Roger and Wiewel Fellowship in Wood Chemistry.

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

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

§ Biotechnology Laboratory and Department of Microbiology and Immunology.

⊗ Abstract published in *Advance ACS Abstracts,* October 1, 1996.

[1] Abbreviations: CBD, cellulose-binding domain; CBD_{CBH1}, the cellulose-binding domain from *T. reesi* cellobiohydrolase 1; CBD_{Cex}, the cellulose-binding domain from *C. fimi* xylanase−glucanase Cex; CBD_{N1}, the N-terminal cellulose-binding domain from *C. fimi* β-1,4-glucanase CenC; DQF-COSY, double quantum filtered correlation spectroscopy; DSS, 2,2-dimethyl-2-silapentane-5-sulfonic acid, sodium salt; HSQC, heteronuclear single quantum correlation; IPTG, isopropyl β-D-thiogalactopyranoside; kb, kilobase; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; PCR, polymerase chain reaction; pH*, the observed pH meter reading without correction for isotope effects; ppm, parts per million; TOCSY, total correlation spectroscopy.

S0006-2960(96)01186-5 CCC: $12.00

Kraulis et al., 1989) and *Cellulomonas fimi* xylanase−glucanase Cex (CBD_{Cex}; Xu et al., 1995). In both of these CBDs, the postulated cellulose-binding face is a flat surface highlighted by three conspicuously exposed aromatic side chains. The three tyrosine residues on the surface of CBD_{CBH1} are important for the ability of the protein to bind crystalline cellulose (Reinikainen et al., 1992, 1995; Linder et al., 1995). Similarly, a recent chemical modification study has demonstrated that the three surface tryptophans of CBD_{Cex} are necessary for its binding activity (Bray et al., in press). Previous mutational analyses of CBDs in the same family as CBD_{Cex} also support the conclusion that the exposed tryptophan side chains are involved in cellulose recognition (Poole et al., 1993; Din et al., 1994). The aromatic tyrosine or tryptophan rings are thought to provide a hydrophobic driving force for binding by stacking directly against the pyranose rings of crystalline cellulose (Creagh et al., 1996). Such aromatic ring−sugar interactions are a common theme for many carbohydrate-binding proteins (Quiocho, 1986; Vyas, 1991).

The β-1,4-glucanase CenC from the bacterium *C. fimi* is unusual in that it has two CBDs (CBD_{N1} and CBD_{N2}) arranged in tandem at its amino terminus (Coutinho et al., 1991, 1992). Both are members of CBD family IV (Tomme et al., 1995b). CBD_{N1} and CBD_{N1N2} have the distinctive ability to bind amorphous (e.g., phosphoric acid-swollen cellulose or Avicel) but not crystalline cellulose (Coutinho et al., 1992; Tomme et al., 1996). This affinity for amorphous cellulose led us to investigate the interactions of CBD_{N1} with cellooligosaccharides that may mimic the natural polymeric substrate. The advantages in studying these cellooligosaccharides are that they are soluble and provide well-defined ligands for structural and thermodynamic characterization of cellulose binding by CBD_{N1}.

In this report, and in the accompanying paper (Tomme et al., 1996), we demonstrate that CBD_{N1} does indeed bind the soluble cellooligosaccharides cellotriose, cellotetraose, cellopentaose, and cellohexaose, forming 1:1 complexes. The association constants for these sugars, determined using heteronuclear NMR spectroscopy, follow the order cellotriose $<$ cellotetraose $<$ cellopentaose $\approx$ cellohexaose. Therefore the CBD_{N1} binding site appears to span approximately five glucopyranosyl units along the cellulose chain. On the basis of perturbations of the chemical shifts of ^{1}H and ^{15}N nuclei in the main chain amides and aromatic rings of CBD_{N1}, the binding site for the cellooligosaccharides is localized to a five-stranded antiparallel β-sheet that forms one face of the jelly-roll β-sandwich structure of this cellulose-binding domain. NMR and ultraviolet difference spectroscopy revealed that this binding site contains tyrosine, but not tryptophan residues. Together, these studies demonstrate that the cellulose-binding site of CBD_{N1} differs from that of other CBDs in that it is not dominated by the presence of aromatic amino acids but rather contains many polar and nonpolar aliphatic side chains. This structural difference correlates with the distinct specificity of CBD_{N1} for amorphous cellulose and highlights further the diverse mechanisms by which proteins recognize carbohydrate ligands and substrates.

EXPERIMENTAL PROCEDURES

Plasmid Expression System

The gene fragment encoding the 152 residue CBD_{N1} from *C. fimi* β-1,4-glucanase CenC (Coutinho et al., 1992) was recloned in the high-expression vector pTug (Figure 1; Graham et al., 1995). This vector was necessary to achieve efficient isotopic labeling of the protein using *E. coli* grown in minimal media. Appropriate restriction sites were introduced at the 5′ and 3′ ends of the CBD_{N1} gene fragment by PCR. Each PCR mixture (50 μL total) contained 25 ng of template DNA (pTZ-JC2; Coutinho et al., 1991), 25−50 pmol of primers, 10% dimethyl sulfoxide, 0.4 mM 2′-deoxynucleotide 5′-triphosphates, and 1 unit *Vent* DNA polymerase in buffer (New England BioLabs). A protocol of twenty successive cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 30 s, and primer extension at 72 °C for 1.5 min was followed. An *Nhe*I site (underlined), coinciding with the amino terminus (Ala-Ser) of the mature protein, was introduced as a silent mutation at the 5′ end of the CBD_{N1} gene, using the oligonucleotide 5′-TTACCT-**CATATG**GCTAGCCCGATCGGGGAGGGAACGT-3′ as primer. An ATG start codon, replacing the *C. fimi* translational start codon GTG (Coutinho et al., 1991), was also introduced as part of the *Nde*I site (bold). *Eco*R1 (bold) and *Hin*dIII (underlined) restriction sites were introduced at the 3′ end of the CBD_{N1} sequence using the oligonucleotide 5′-AGAAT**GAATTC**AAGCTTAGAGCTCGAC-CTCGGAGTC-3′ as primer. The resulting 0.48 kb PCR fragment was digested with *Nde*I and *Eco*RI and subcloned between the *Nde*I-*Eco*RI restriction sites in the polylinker of pEAD4 (Doering et al., 1992) to give pEADN1. The 0.37 kb *Apa*I-*Mlu*I PCR fragment encoding most of CBD_{N1}, was removed by digestion of pEADN1 with *Apa*I and *Mlu*I, and replaced with the *Apa*I-*Mlu*I fragment from pTZ-JC2 to give pEADN1n. pEADN1n and pTugEO7K3 were cut with *Nhe*I and *Hin*dIII, and the entire 0.46 kb fragment obtained from pEADN1n was recloned in frame with the sequence encoding the Cex leader peptide in pTugEO7K3 to give the final construct pTugN1n. This construct was verified by sequencing of the double-stranded DNA using the dideoxy chain termination method with modified T7 DNA polymerase and [^{35}S]α-dATP (Sanger, 1977; Tabor & Richardson, 1987). An error was detected in the amino acid sequence predicted originally for CBD_{N1}. Specifically, residue 139 is leucine, not phenylalanine as reported by Coutinho et al. (1991).

Expression and Purification of CBD_{N1}

The gene fragment encoding CBD_{N1} was expressed in *E. coli* JM101 cells (Yanish-Perron et al., 1983). Unlabeled protein was produced in liquid tryptone−yeast extract−phosphate medium (TYP) (Sambrook, 1989). Biosynthetically ^{15}N-labeled protein was prepared using M9 media (Miller, 1972) containing 1g of 99% $^{15}NH_4Cl$/L (Cambridge Isotopes) and 1 g of 99% ^{15}N-labeled Isogro/L (Isotec Inc.) as the sole sources of nitrogen. The Isogro (algal extract) supplement is necessary for efficient growth of *E. coli* JM101 and protein production in M9 media. The CBD_{N1} samples with selectively deuterated aromatic rings were obtained from a synthetic medium containing 100 mg/L of L-δ_1,ϵ_2,$\zeta_{1,2}$,η_2-[$^2H_5$]tryptophan and 100 mg/L of either L-$\delta_{1,2}$,$\epsilon_{1,2}$,ζ-[$^2H_5$]-phenylalanine or $\delta_{1,2}$,$\epsilon_{1,2}$-[$^2H_4$]tyrosine (Cambridge Isotope

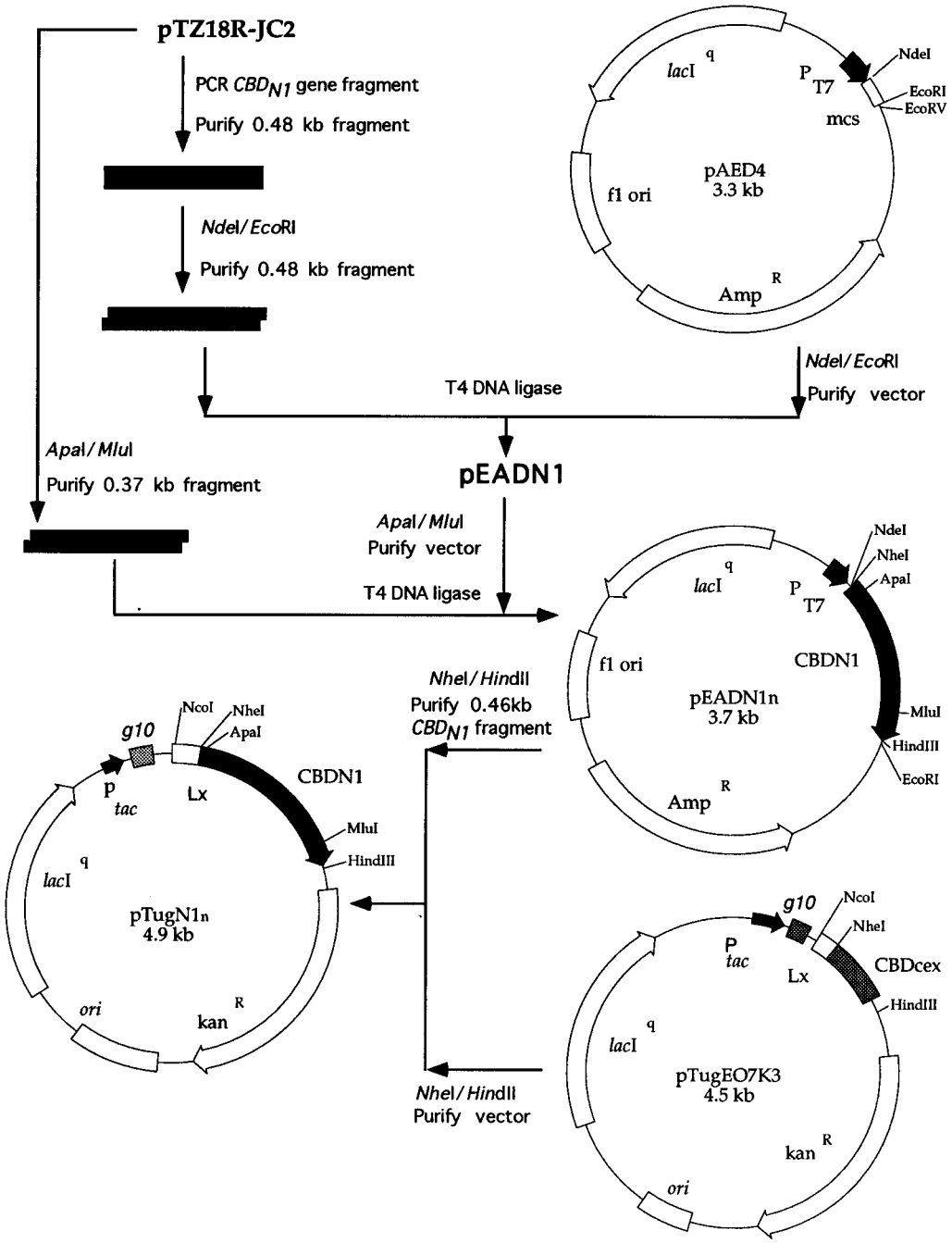

FIGURE 1: Construction of the pTugN1n expression vector containing the gene encoding CBD_{N1} fused to the Cex leader peptide. High-level expression of secreted CBD_{N1} is obtained with this expression system using *E. coli* JM101 cells. *Nde*I-*Nhe*I and *Hin*dIII-*Eco*RI restriction sites were introduced at the 5′ and 3′ ends of the CBD_{N1} gene, respectively, by PCR. The resulting fragment was cloned at the *Nde*I and *Eco*RI sites in the polylinker of pEAD4 to give pEADN1. The 0.37 kb *Apa*I-*Mlu*I PCR fragment in pEADN1 was replaced by the corresponding native fragment obtained from pTZ-JC2 to give pEADN1n. pEADN1n was then treated with *Nhe*I and *Hin*dIII, and the resulting CBD_{N1} gene fragment was subcloned at the *Nhe*I-*Hin*dIII sites of pTugEO7K3 behind the Cex leader peptide to give the final construct pTugN1n.

Laboratories and Isotec Inc.) (McIntosh et al., 1990; McIntosh & Dahlquist 1990). The 1 L bacterial cultures were grown in 1.8 L Fernbach flasks at 30 °C to an A_{600} of 0.6, induced with 0.5 mM IPTG, and incubated with shaking for approximately 24 h. The leader peptide, which is cleaved upon translocation, targets the expressed CBD_{N1} to the periplasm of *E. coli*. However, over this extended incubation period, CBD_{N1} leaks out into the culture supernatant (Ong et al., 1993).

CBD_{N1} was purified by adding 35 g of dry Avicel PH-101 (Fluka Chemika) to each liter of culture supernatant, adjusted to 1 M sodium chloride and 50 mM potassium phosphate at pH 7. When labeled protein was prepared, the cell pellet was also subjected to osmotic shock by resuspending the cells in 30 mM Tris·HCl (pH 8.0) buffer containing 20% sucrose. After 15 min, the cells were centrifuged, resuspended in 5 mM $MgSO_4$ at 4 °C, stored on ice for an additional 15 min, and finally repelleted. The supernatants from this procedure were combined with the culture supernatant before the addition of Avicel and salt. After standing at 4 °C for 4 h, the Avicel was collected by vacuum filtration on a Whatman GF/A glass filter and washed with 250 mL of 1 M sodium chloride, 50 mM potassium phosphate buffer at pH 7. The CBD_{N1} was eluted from the Avicel with 250 mL of distilled water. To increase the recovery of labeled protein, the initial filtrate and the salt wash solution were

recombined with the Avicel and left overnight at 4 °C. The above procedure was repeated, and the two protein fractions were combined. The CBD_{N1} was purified further by anion exchange FPLC at pH 5 and then pH 7 using Q Sepharose (Pharmacia). In both cases, CBD_{N1} was eluted with a gradient of 0−1 M sodium chloride in 50 mM potassium phosphate buffer. Finally, the CBD_{N1} was de-salted, exchanged into sample buffer, and concentrated by ultrafiltration using non-cellulose membranes (Filtron Technology Corp., Northborough, MA). The yields of the unlabeled, ^{15}N-labeled, and selectively deuterated CBD_{N1} samples were approximately 80, 25, and 30 mg/L of culture supernatant, respectively.

Protein concentrations were measured using absorption spectroscopy. The molar absorptivity ϵ_{280} of CBD_{N1} was determined to be 21 370 M^{-1} cm^{-1} (or 1.39 mL mg^{-1} cm^{-1}) using the method of Edelhoch (1967), as reviewed by Gill and von Hippel (1989) and Pace et al. (1995). The small contribution of a disulfide was included in the calculations. This molar absorptivity value, which differs from that originally published by Coutinho et al. (1992), was confirmed by quantitative amino acid analysis. The average molar absorptivity determined from three amino acid analyses, each run in duplicate, agreed within 4% of that obtained using the Edelhoch method.

The molecular mass of the unlabeled CBD_{N1} is 15 425.3 ± 0.9 Da as measured by electrospray mass spectroscopy. This is in excellent agreement with the expected value of 15 425.8 Da based on the sequence of CBD_{N1}, after post-translational cleavage of the secretory leader peptide and corrected for the presence of a disulfide bond. The disulfide bond was identified by the lack of reactivity of CBD_{N1} with 5,5′-dithiobis(2-nitrobenzoic acid) and subsequently confirmed by NMR $^{13}C^{\beta}$ chemical shifts (Wishart & Sykes, 1995). Using Edman degradation, the ten N-terminal residues of CBD_{N1} expressed in *E. coli* were confirmed to be ASPIGEGTFD. This matches that of native CenC from *C. fimi*.

On the basis of sedimentation equilibrium measurements, CBD_{N1} (0.19 mM) in 50 mM sodium chloride, 50 mM potassium phosphate, 0.02% sodium azide, pH 5.9 at 20 °C is a monomeric protein with an apparent mass of 14 130 Da (assuming a partial specific volume of 0.73 cm^3/gm; Chervenka, 1969). No evidence of higher-order association was detected under the conditions employed for the ultracentrifugation analysis. This conclusion is supported qualitatively by the relatively narrow line widths observed in the NMR spectra of CBD_{N1}.

NMR Spectroscopy

The CBD_{N1} samples were exchanged into 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), 0.02% sodium azide, 10% D_2O/90% H_2O. NMR spectra were recorded on a Varian Unity 500 MHz spectrometer equipped with a pulsed field gradient triple-resonance probe. All spectra were collected at 35 °C and processed using FELIX v2.30 (Biosym Technologies; San Diego, CA). ^{1}H chemical shifts were referenced to an internal standard of DSS at 0.00 ppm, and ^{15}N chemical shifts were referenced to external 2.9 M $^{15}NH_4Cl$ in 1 M HCl at 24.93 ppm (Levy & Lichter, 1979).

^{1}H−^{15}N HSQC spectra were recorded using the enhanced sensitivity pulsed field gradient experiment of Kay et al. (1992). A selective water flip-back pulse was incorporated to ensure minimum perturbation of the water magnetization (Zhang et al., 1994). The assignments of the amide ^{1}H and ^{15}N resonances of CBD_{N1} were obtained using triple-resonance correlation experiments (Bax & Grzesiek, 1993). The ^{1}H resonances from the aromatic rings were identified from DQF-COSY, TOCSY, and NOESY spectra recorded with samples of 1.9 mM unlabeled, 2.0 mM ($\delta_{1,2}$, $\epsilon_{1,2}$-[$^2H_4$]-Tyr and δ_1, ϵ_2, $\zeta_{1,2}$, η_2-[$^2H_5$]Trp)-labeled, and 1.9 mM ($\delta_{1,2}$,$\epsilon_{1,2}$,ζ-[$^2H_5$]Phe and δ_1,ϵ_2,$\zeta_{1,2}$,η_2-[$^2H_5$]Trp)-labeled CBD_{N1}, alone and in the presence of 80 mM cellotetraose. These assignments were confirmed using the (Hβ)Cβ(CγCδ)Hδ and (Hβ)Cβ(CγCδCϵ)Hϵ experiments to connect the ring ^{1}H resonances to the corresponding $^{13}C^{\beta}$ resonance of each aromatic residue in uniformly ^{13}C/^{15}N enriched CBD_{N1} (Yamakazi et al., 1993). The protein samples were lyophilized from the above sample buffer and resolubilized in 99% D_2O for several of these measurements. The NMR assignments and structural analysis of CBD_{N1} will be reported elsewhere.

Titration of CBD_{N1} with Cellooligosaccharides Monitored by NMR

The binding of soluble cellooligosaccharides to CBD_{N1} at 35 °C and pH* 5.9 was measured quantitatively using ^{1}H−^{15}N NMR spectroscopy. Stock weight per volume solutions of cellotriose, cellotetraose, cellopentaose, and cellohexaose (Seikagaku Corp.) were prepared in the identical buffer used for the CBD_{N1}. The initial concentration of protein was 0.5 mM, except in the case of the titration with cellotriose for which a 0.18 mM sample was used. Aliquots of the sugar solution were added directly to the protein in a NMR tube and mixed using 1 mm inner diameter PE-160 polyethylene tubing attached to a Gilson Pipetman. To avoid excessive dilution of the CBD_{N1}, additions of greater than approximately 1 mg of sugar were made by removing the protein from the NMR tube and dissolving the solid oligosaccharide directly in the sample solution. For each titration, eight to ten ^{1}H−^{15}N HSQC spectra were recorded consecutively with increasing concentrations of sugar. The spectra were measured in 1.5 h with 1024 and 128 complex points obtained in the ^{1}H and ^{15}N dimensions (spectral widths 6000 and 1450 Hz), and processed with zero filling to a final digital resolution of 2.93 and 1.42 Hz/point, respectively.

Equilibrium association constants were determined 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 the analyses. The binding equation was derived as follows. The association of soluble oligosaccharides to CBD_{N1} is in the fast exchange limit on the NMR chemical shift time scale. Therefore, the fraction of protein in the bound form at each point (i) in the titration (f_{bi}) is calculated using the observed ^{1}H or ^{15}N chemical shift (δ_i) compared to that of the free (δ_f) and fully bound forms (δ_b) by eq 1:

$$f_{bi} = (\delta_i - \delta_f)/(\delta_b - \delta_f) \quad (1)$$

δ_f was determined from the HSQC spectrum of the protein without added sugar, and δ_b obtained from the fitting routine. For a single binding site, the association constant K_a for the

binding of free protein (P_f) and free ligand (L_f) to yield the bound complex (PL) is given by eq 2:

$$K_a = [PL]/([P_f][L_f]) \quad (2)$$

By definition, for each titration point, the fraction of protein in the bound form is given by eq 3, the total protein concentration $[P_t]$ by eq 4, and the total ligand concentration $[L_t]$ by eq 5.

$$f_b = [PL]/([P_f] + [PL]) \quad (3)$$

$$[P_t] = [P_f] + [PL] \quad (4)$$

$$[L_t] = [L_f] + [PL] \quad (5)$$

Rearranging eqs 4 and 5 using eq 2, followed by substitution into eq 3 results in eq 6.

$$f_{bi} = K_a([L_t]_i - f_{bi}[P_t]_i)/(1 + K_a[L_t]_i - K_a f_{bi}[P_t]_i) \quad (6)$$

Equation 6 was solved using the real root of the quadratic equation, allowing eq 1 to be recast in terms of the experimental parameters $[P_t]_i$, $[L_t]_i$, δ_f, and δ_i and the variables K_a and δ_b. The latter two terms were determined by the fitting routine. An initial estimate for δ_b was supplied from the ^1H or ^{15}N chemical shift at the final titration point, and the concentration of protein at each point (i) was corrected for dilution resulting from the addition of ligand. In the case of cellohexaose, an estimate of K_a was made by simulation of the binding isotherm.

UV Difference Spectroscopy

UV difference spectra were recorded at 25 °C with a Hitachi U2000 double-beam spectrophotometer and two thermostated standard 1.0 × 1.0 × 4.5 cm cuvettes. Both cuvettes were filled with equal volumes of CBD_{N1} (21.6 μM) in 50 mM potassium phosphate, pH 7.0. After the base line scan was obtained and stored into the instrument memory, 5−10 μL aliquots of a buffered 2 mM ligand solution were added to the sample cuvette and equal volumes of buffer were added to the reference cuvette. For each ligand concentration, five spectra were recorded from 250−350 nm at 200 nm/min and a spectral bandwidth of 2 nm. The spectra were averaged and corrected for buffer and sugar contributions. For solvent pertubation measurements of the *N*-acetyl-L-tyrosine ethyl ester, the sample cuvette contained 50 mM potassium phosphate, pH 7.0. After the base line was obtained and stored into the instrument memory, dimethylsulfoxide (DMSO, 20% v/v final) was added to both cuvettes. Spectra were recorded and processed as described above.

RESULTS

CBD_{N1} Binds Soluble Cellooligosaccharides

The binding of CBD_{N1} to cellotriose, cellotetraose, cellopentaose, and cellohexaose was detected initially by the observation of numerous changes in the ^1H NMR spectrum of the protein resulting from the addition of these soluble sugars (data not shown). In contrast, the spectrum of CBD_{N1} remained unperturbed in the presence of cellobiose, indicating that the protein does not bind this disaccharide appreciably. Cellooligosaccharides longer than cellohexaose were not investigated as these compounds are not commercially available and have limited solubility in aqueous buffers.

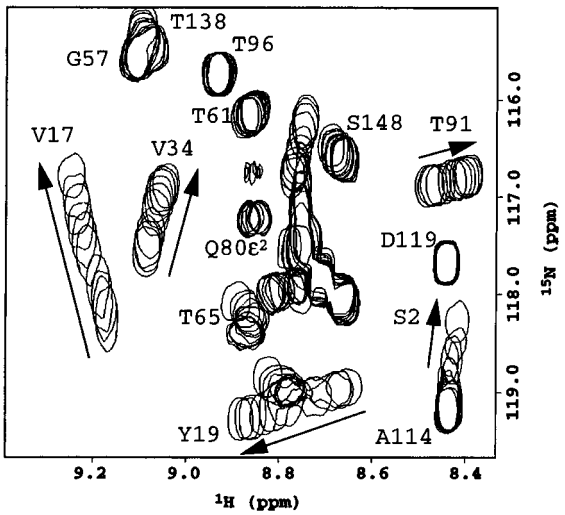

FIGURE 2: Cellotetraose binds to CBD_{N1}. A portion of ten $^1H-^{15}N$ HSQC spectra of uniformly ^{15}N-labeled protein in the presence 0, 0.04, 0.08, 0.16, 0.36, 0.51, 0.76, 1.26, 2.24, and 10.5 mM total cellotetraose is overlaid. The arrows indicate the directions in which the amide $^1H-^{15}N$ peaks shift with added sugar. Although all $^1H-^{15}N$ resonances have been assigned, the peaks in the crowded region of the spectrum are not labeled for clarity. Q80$^{\epsilon 2}$ designates the $^{15}NH^{\epsilon 2}$ of Gln80. The sample buffer was 50 mM sodium chloride, 50 mM potassium phosphate (pH* 5.9), and 0.02% sodium azide in 10% D_2O/90% H_2O at 35 °C.

The interaction of CBD_{N1} with cellotriose, cellotetraose, cellopentaose, and cellohexaose was quantified using two-dimensional $^1H-^{15}N$ correlation spectroscopy. Figure 2 shows an overlay of ten $^1H-^{15}N$ HSQC spectra recorded as uniformly ^{15}N-enriched CBD_{N1} was titrated with cellotetraose. It is readily seen that many peaks, arising from the backbone amide groups of the protein, show significant changes in both ^1H and ^{15}N chemical shifts with the progressive addition of the sugar. In the case of each cellooligosaccharide investigated, the free and bound forms of the protein are in fast exchange on the NMR time scale, resulting in the observation of population weighted average chemical shifts throughout the titration series.

Binding Stoichiometry

Cellotriose, cellotetraose, cellopentaose, and cellohexaose bind to CBD_{N1} with a stoichiometry of one sugar molecule per protein molecule (an exception is observed with high concentrations of cellotetraose, as discussed below). This conclusion is supported by the following evidence. First, as exemplified in Figure 3 for cellotetraose, all of the amides with ^1H and ^{15}N chemical shifts that are perturbed by sugar binding show coincident titration curves. This discounts the possibility of multiple binding sites on the CBD with differing affinities for the cellooligosaccharides. Given that CBD_{N1} is monomeric in solution, it is unlikely to have two or more distinct binding sites with equal affinities for the sugar ligands. Therefore the simplest interpretation of the coincident titration curves is that cellooligosaccharides bind to CBD_{N1} at a single site and that each perturbed $^1H-^{15}N$ group reports the same association event. Second, plots of CBD_{N1} amide chemical shifts versus added cellopentaose or cellohexaose show a plateau at approximately equal con-

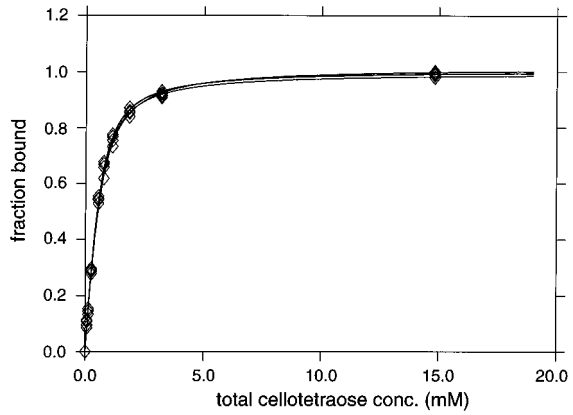

FIGURE 3: Cellotetraose binds to CBD_{N1} at a single site. The coincident plots of the normalized H^N chemical shift changes for residues Tyr19, Gly44, Thr87, and Gly130 versus total added cellotetraose demonstrate that each amide group in CBD_{N1} monitors the same binding event. The solid lines represent the titration isotherms obtained by fitting the observed data points (◊) to the equation describing the association of cellotetraose and CBD_{N1} to form a 1:1 protein−sugar complex. For clarity, the overlapping data and fits for the four residues are not labeled individually.

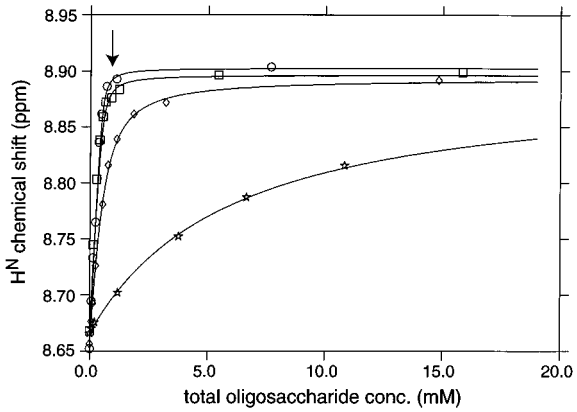

FIGURE 4: The association constants of CBD_{N1} for soluble cellooligosaccharides were determined from titration curves monitored by $^1H-^{15}N$ NMR spectroscopy. The amide H^N chemical shift of Tyr19 is plotted as a function of the total concentration of added cellohexaose (○), cellopentaose (□), cellotetraose (◊), and cellotriose (☆). The solid lines represent the best fit of the experimental data to the equilibrium equation describing binding to a single protein site or, in the case of cellohexaose, a simulated titration curve based on an estimation of the association constant. The arrow marking the plateau in the titration curves for cellopentaose and cellohexaose falls at the point where the total sugar concentration equals the total protein concentration (~0.5 mM CBD_{N1}), indicating a 1:1 binding stoichiometry.

centrations of total sugar and protein, indicating a binding stoichiometry of 1:1 (Figure 4). Cellotetraose and cellotriose do not exhibit such pronounced titration end points due to their lower affinities for CBD_{N1}. However, all four sugars cause similar changes in the $^1H-^{15}N$ spectrum of the labeled protein, strongly suggesting that each binds to CBD_{N1} with the same stoichiometry and at the same site. Third, the titration curves measured for the four cellooligosaccharides are adequately fit to the binding isotherm describing the simple equilibrium expressed by eq 2 (Figures 3 and 4). Finally, this conclusion is confirmed by isothermal titration microcalorimetry (Tomme et al., 1996).

Association Constants Determined by NMR Spectroscopy

The association constants (K_a) describing the interactions of cellotriose, cellotetraose, and cellopentaose to CBD_{N1} were determined by nonlinear least-squares fitting of the chemical shift titration data to the binding isotherm for a protein with a single ligand recognition site. Figure 4 shows an example of the analysis of data measured for Tyr19 as well as a comparison of the binding curves for each cellooligosaccharide. Two K_a's were determined independently for each amide showing a significant chemical shift change upon sugar binding, one using the data for the amide H^N and the other for the ^{15}N nucleus. In this manner, association constants were determined from titration curves measured for 10−16 amides in the protein. These included Val17, Tyr19, Val34, Tyr43, Gly44, Val45, Gly46, Leu49, Asn81, Gly82, Thr87, Ala126, Gly130, Leu139, Leu141, and Ala145. The average association constants for the three cellooligosaccharides were calculated using the values obtained individually from the analyses of the H^N and ^{15}N data recorded for each amide (Table 1). In the case of cellohexaose, binding to the CBD_{N1} was sufficiently tight that, given the concentration of protein required for NMR analysis, it was not possible to determine accurately the K_a. Therefore an estimate of the upper limit of the association constant was obtained by a visual comparison of simulated titration curves with the observed NMR data (Figure 4 and Table 1).

Table 1: Association Constants K_a for the Binding of Soluble Cellooligosaccharides to CBD_{N1}[a]

ligand	K_a (M^{-1})
cellotriose	180 ± 60
cellotetraose	4 200 ± 720
cellopentaose	34 000 ± 7 600
cellohexaose	(50 000)

[a] Data obtained at 35 °C and pH* 5.9 in 50 mM sodium chloride, 50 mM potassium phosphate, 0.02% sodium azide, and 10% D_2O/90% H_2O. The reported K_a values are the average of the those determined from the ^1H and ^{15}N chemical shift perturbations of Val17, Tyr19, Val34, Tyr43, Gly44, Val45, Gly46, Leu49, Asn81, Gly82, Thr87, Ala126, Gly130, Leu139, Leu141, and Ala145. The error range is one standard deviation. The K_a reported for cellohexaose is an estimate of the upper limit of the association constant based upon simulations of the titration curves.

The association constants of CBD_{N1} for the cellooligosaccharides measured by NMR agree with those determined independently by isothermal titration calorimetry under similar experimental conditions (Tomme et al., 1996). Due to the different concentration windows necessary for each method, NMR provided a measure of the relatively weak binding of cellotriose, whereas calorimetry yielded a more accurate value of the association constant of cellohexaose with the CBD.

Identification of the Cellooligosaccharide Binding Site

In addition to providing a means for quantitating the association of the cellooligosaccharides to CBD_{N1}, NMR also yields information regarding the location and structure of the binding site. Perturbations of the resonances of main chain H^N and ^{15}N nuclei upon sugar binding may arise due to the direct interaction of an amide with the ligand, or indirectly due to conformational changes resulting from the formation of the protein−sugar complex. These conformational changes are likely subtle as circular dichroism spectra reveal that the global secondary structure of CBD_{N1} is unaffected in the presence of saturating quanities of cel-

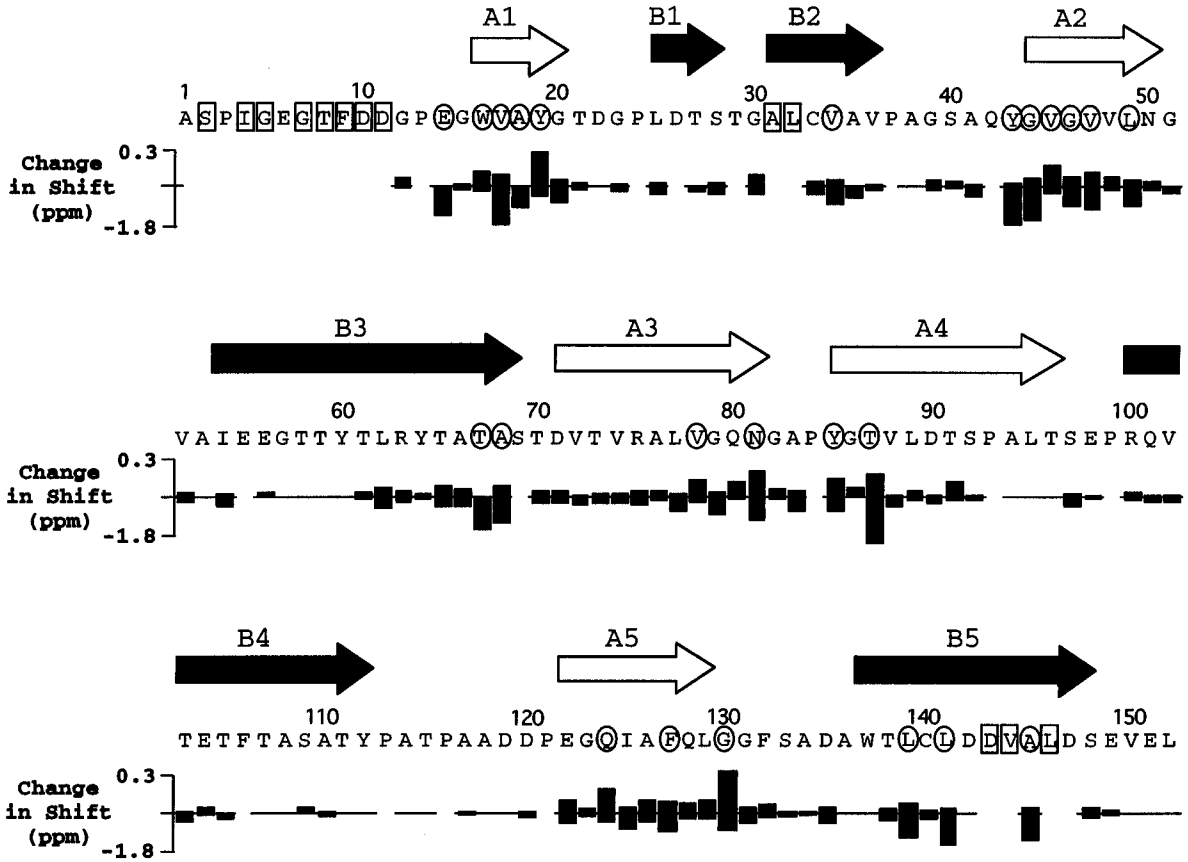

FIGURE 5: Patterns of NMR chemical shift perturbations due to sugar binding demonstrate that the cellooligosaccharide-recognition site of CBD_{N1} is formed by a five-stranded antiparallel β-sheet. The absolute values of the differences between the H^N and ^{15}N chemical shifts of the main chain amides in the free and cellotetraose-bound forms of CBD_{N1} are indicated as positive and negative numbers, respectively. Blank spaces identify residues for which no difference could be determined. The locations of the ten β-strands in the protein are shown by the arrows above the amino acid sequence. Two β-sheets identified in CBD_{N1} are composed of strands A1−A5 (open) and B1−B5 (solid). Circled residues have a change in H^N or ^{15}N chemical shift upon binding greater than one standard deviation from the mean absolute value change observed for all measured residues. On the basis of the pattern of chemical shift perturbations, the oligosaccharide-binding site lies across the face of the protein formed by strands A1−A5. Residues not detected in the $^1H-^{15}N$ HSQC spectrum of CBD_{N1} in the absence of added cellotetraose are boxed.

lotetraose (data not shown). Bearing this in mind, structural insights into the binding site are provided by the pattern of NMR chemical shift changes observed with sugar binding. Figure 5 summarizes the changes of the H^N and ^{15}N chemical shifts of each amide in CBD_{N1} due to the binding of cellotetraose. Similar shift perturbations result from the binding of cellotriose, cellopentaose, and cellohexaose (not shown).

Currently we have determined the secondary structure and a low-resolution tertiary fold of the CBD_{N1} using NMR chemical shift, $^3J_{HN-H\alpha}$ coupling constant, amide hydrogen exchange, and NOE information (manuscript in preparation). The protein is composed of ten β-strands, folded into two antiparallel β-sheets identified as A and B (Figure 5). As illustrated schematically in Figure 6, the global topology of CBD_{N1} is that of a jelly-roll β-sandwich. Residues with exposed side chains that form β-sheet A are identified on this diagram.

The average change in the absolute value of the H^N chemical shifts of all resolved amides in CBD_{N1} due to cellotetraose binding was 0.04 ppm with a standard deviation of 0.05 ppm. In the ^{15}N dimension, the average absolute change in shift was 0.3 ppm, with a standard deviation of 0.4 ppm. Of the 24 residues with a perturbation in amide H^N or ^{15}N chemical shift greater than one standard deviation above the average, 18 lie within or immediately adjacent to

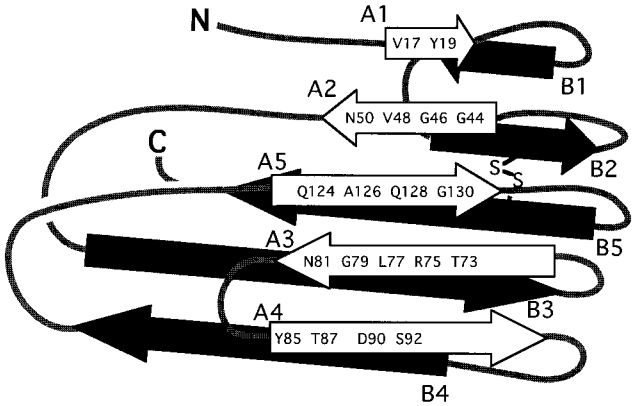

FIGURE 6: Schematic representation of the jelly-roll β-sandwich fold of CBD_{N1}. On the basis of the observed patterns of chemical shift perturbations resulting from sugar binding, the recognition site for the cellooligosaccharides lies across the face of the protein formed by β-strands A1−A5 (open arrows). Residues in these β-strands that have exposed side chains are labeled. Cys33 and Cys140 in strands B2 and B5, respectively, form a disulfide bridge. The loops connecting the β-strands are not drawn to scale.

the β-sheet of CBD_{N1} composed of strands A1−A5 (Figures 5 and 6). This strongly indicates that the binding site for the cellooligosaccharides lies on the face of the CBD formed by these strands. Furthermore, residues showing pronounced chemical shift changes upon the addition of sugar are located

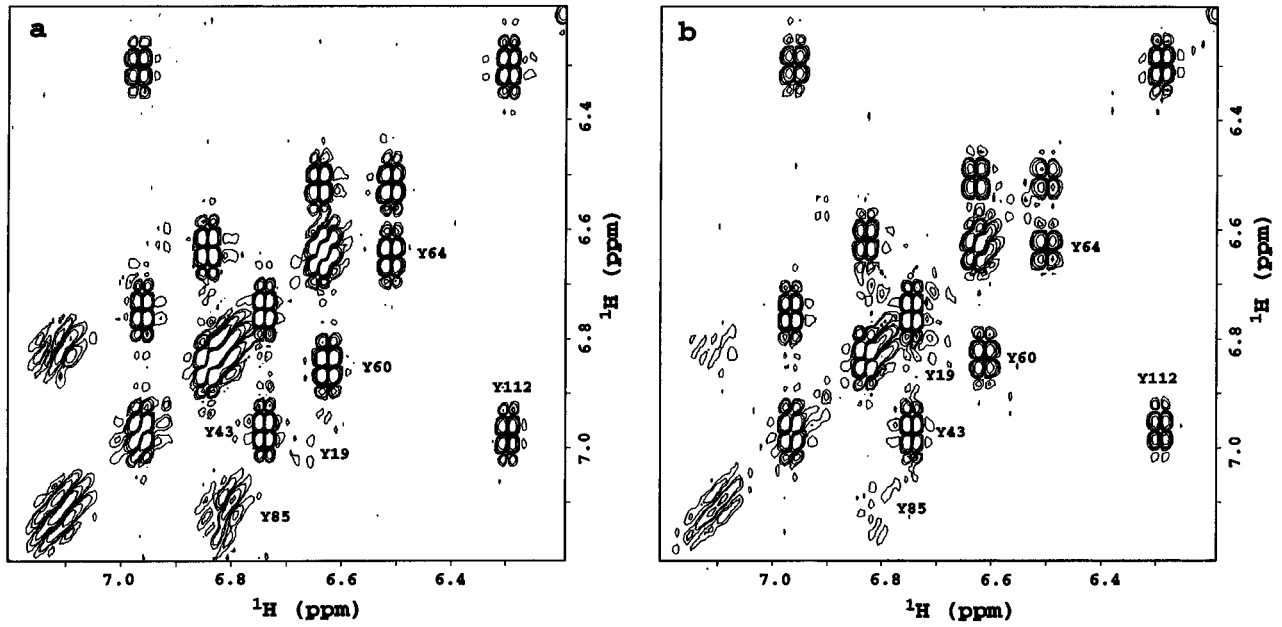

FIGURE 7: Aromatic region of the DQF-COSY spectrum of [$^2H_5$]Phe-, [$^2H_5$]Trp-labeled CBD_{N1} in the (a) absence and (b) presence of 80 mM cellotetraose. Only the resonances from the six tyrosine residues are detected because the aromatic rings of the tryptophan and phenylalanine residues in the protein were biosynthetically deuterated and the amide H^N signals were eliminated by reversibly unfolding the protein in 99% D_2O buffer.

in all five of the β-strands that form this sheet. Therefore, the cellooligosaccharides are likely to bind across, and not parallel to, strands A1−A5. This conclusion is supported by the observation of intermolecular proton NOE interactions between bound cellotetraose and protein side chains located within β-sheet A in CBD_{N1} (data not shown).

A limited number of amides located outside of β-strands A1-A5 also show changes in H^N or ^{15}N chemical shift upon sugar binding that are greater than one standard deviation from the average. All six of these amides either flank the Cys33-Cys140 disulfide bond (Val34, Leu139, Leu141, and Ala145) or lie in β-strand B3, which is adjacent to the strands containing this cystine group (Thr67 and Ala68). It is possible that sugar-binding indirectly influences a structural feature of the disulfide bond, such as disulfide isomerism, resulting in the observed chemical shift perturbations.

The changes in chemical shifts of remaining amides located in the β-sheets and loops of the protein are less than one standard deviation from the average. It is likely that these chemical shift differences reflect subtle structural perturbations associated with ligand binding that are propagated through the core of the protein.

Effect of Cellotetraose Binding on the Aromatic Residues in CBD_{N1}

To investigate the possible roles that the aromatic side chains play in sugar binding, the ^{1}H resonances of the phenylalanine, tyrosine, and tryptophan residues in CBD_{N1} were assigned in the absence and presence of saturating amounts of cellotetraose. Samples of the protein in which either tryptophan and tyrosine or tryptophan and phenylalanine were uniformly deuterated at all ring positions were prepared to simplify the NMR spectra of CBD_{N1} and to identify unambiguously the ^{1}H resonances from the phenylalanine or tyrosine residues, respectively. Near-complete aromatic ^{1}H assignments were obtained for both the unbound and bound forms of CBD_{N1} (Figure 7). An exception is

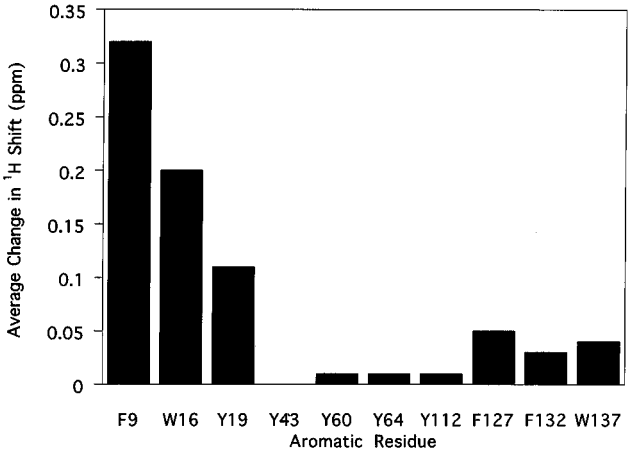

FIGURE 8: Histogram summarizing the perturbations of the ^{1}H chemical shifts of the aromatic rings of the tryptophan, tyrosine, and phenylalanine residues in CBD_{N1} due to the binding of cellotetraose. The data represent the average of the absolute value of the chemical shift changes observed for all protons within the aromatic side chain of each residue. The average change for all assigned aromatic rings is 0.07 ppm with a standard deviation of 0.09 ppm. The chemical shifts of Tyr85 could not be confidently determined due to linebroadening and are not included in the figure. The resonances of Phe106 were not observed in the ^{1}H-NMR spectra of CBD_{N1}.

Phe106, for which no resonances are observed in the homonuclear spectra of the protein. This residue, as well as Tyr19 and Tyr85, which show extensive line broadening, may undergo conformational averaging on a time scale of intermediate exchange due to partially hindered ring flipping (Wüthrich, 1986).

The differences between the aromatic ^{1}H chemical shifts of the free and cellotetraose-bound forms of the protein are summarized in Figure 8. The average change in the absolute values of the ^{1}H chemical shifts for all aromatic side chains was 0.07 ppm with a standard deviation of 0.09 ppm. Of the assigned aromatic residues, only Phe9, Trp16, and Tyr19 show above-average ring chemical shift changes upon

binding the oligosaccharide. In addition, the line shapes of the aromatic proton resonances of Tyr85 are clearly altered in the presence of cellotetraose (Figure 7). As illustrated in Figure 5, Trp16, Tyr19, and Tyr85 are located within β-sheet A and also exhibit significant changes in amide ^1H and ^{15}N chemical shifts upon sugar binding. This provides further support for the identification of the sugar binding site in CBD_{N1}. Although Phe9 and Trp16 show pronounced NMR chemical shift changes upon sugar binding, preliminary structural studies reveal that the side chains of these residues are located within the interior of CBD_{N1}. This leaves Tyr19 and Tyr85 as the only two residues with exposed aromatic rings that are likely to be involved in the binding of cellooligosaccharides.

Difference Ultraviolet Absorbance Spectra of CBD_{N1} Induced by Cellooligosaccharides

Difference ultraviolet absorbance spectroscopy was used to investigate further the possible involvement of aromatic residues in cellooligosaccharide binding by CBD_{N1}. Addition of cellopentaose and cellohexaose to CBD_{N1} produced specific spectral changes characterized by a positive peak at 286−287 nm, a small minimum at 280 nm, and a large negative peak from 250 to 278 nm (Figure 9). These difference spectra were almost superimposible with the solvent perturbation spectrum obtained upon addition of 20% v/v dimethyl sulfoxide to the model compound *N*-acetyl-L-tyrosine ethyl ester (Figure 9). This shows that the binding of CBD_{N1} to cellopentaose and cellohexaose perturbs the environment around tyrosine residues in the protein (Herkovits & Sorensen 1968; Bailey et al., 1968). On the basis of the NMR data presented in Figure 8, we tentatively assign these as Tyr19 and Tyr85. In addition, the absence of a maximum in the 293−295 nm region of the difference absorption spectra excludes the involvment of tryptophan indole rings in the binding process. This supports the previous conclusion that neither Trp16 nor Trp137 is involved directly in ligand binding. Therefore, although the chemical shifts of Trp16 are sensitive to the addition of cellooligosaccharides, the electronic environment of its aromatic side chain is not significantly perturbed by the presence of bound sugar. Finally, due to its spectral properties, these difference ultraviolet absorbance measurements provide no information about the possible involvement of phenylalanine residues in the binding of cellooligosaccharides (Herkovits & Sorensen 1968; Bailey et al., 1968).

In contrast to cellopentaose and cellohexaose, interaction of CBD_{N1} with cellotriose (not shown) and cellotetraose (Figure 9) did not produce the characteristic maxima at 286−287 nm indicative of the perturbation of tyrosine residues. This could reflect a difference in the binding interactions between CBD_{N1} and the cellooligosaccharides with a degree of polymerization of 3 or 4 and those with a degree of polymerization of 5 or greater. Cellopentaose and cellohexaose most likely make contacts with the entire binding face of CBD_{N1}, whereas, cellotriose or cellotetraose may interact with a subsite(s) of the protein. This is consistent with the observation that cellopentaose and cellohexaose bind to CBD_{N1} with greater affinity than do cellotriose and cellotetraose. As illustrated in Figure 6, Tyr19 and Tyr85 are located in β-strands at the edges of β-sheet A and may only show absorbance changes upon binding of the longer oligosaccharides. Note, however, that all four sugars induce

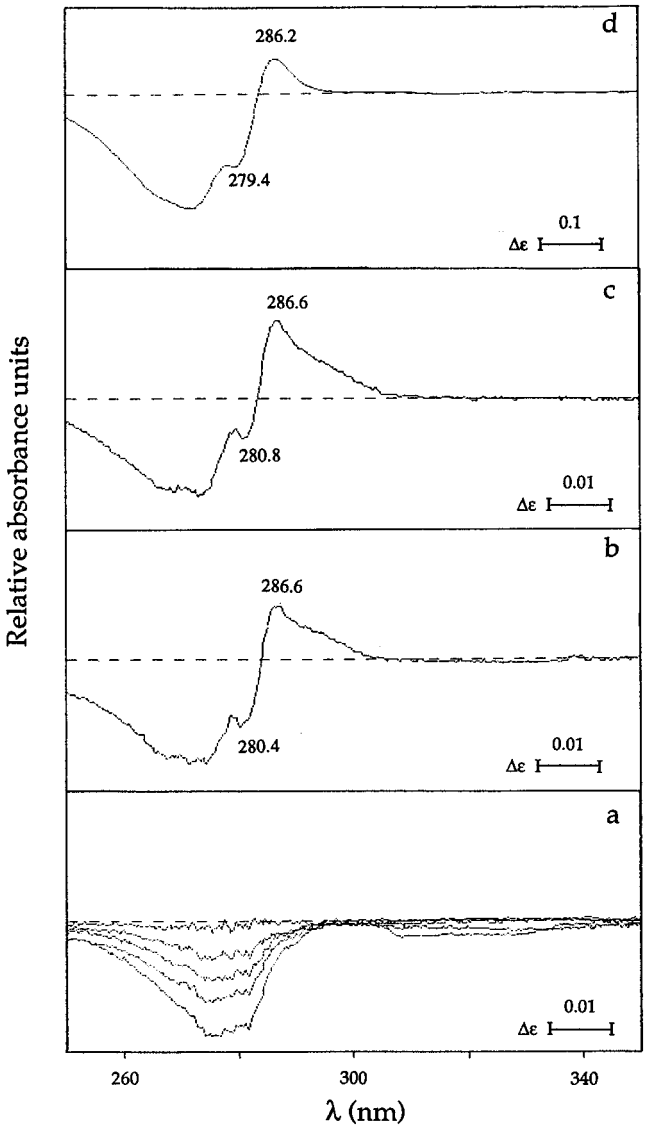

FIGURE 9: Difference ultraviolet absorbance spectra of CBD_{N1} illustrate that binding of soluble cellooligosaccharides perturbs tyrosine residues in the protein. The difference spectra were measured at 25 °C in 50 mM potassium phosphate buffer, pH 7.0, with CBD_{N1} (22 μM) in the presence of (a) cellotetraose (20−500 μM), (b) cellopentaose (65 μM), and (c) cellohexaose (65 μM). (d) The solvent perturbation spectrum (25 °C) for N-acetyl-L-tyrosine ethyl ester (50 μM, pH 7.0) induced with 20% v/v dimethyl sulfoxide shows that the spectral changes observed with CBD_{N1} arise from tyrosine, and not tryptophan, side chains.

similar perturbations in the ^1H and ^{15}N chemical shifts of the main chain amides of CBD_{N1} and that cellotetraose binding alters the NMR chemical shifts or line widths of these two tyrosine side chains (Figures 7 and 8). The chemical shift perturbations may reflect a common structural change in the protein due to the binding of any cellooligosaccharide, or perhaps the rapid interconversion of cellotriose or cellotetraose between multiple subsites within the full CBD_{N1} binding cleft.

Effect of High Concentrations of Cellotetraose

In the absence of added oligosaccharide, resonances from the amides of 13 residues are not observed in the $^1H-^{15}N$ HSQC spectrum of CBD_{N1}. These residues are located at the N-terminus of the protein and in β-strands B2 and B5, next to the disulfide bond (Figure 5). The absence of signals

from these amides is not due to rapid hydrogen exchange with the solvent as the gradient-enhanced HSQC spectra were recorded at pH* 5.9 using selective flip-back pulses to minimize any perturbation of the bulk water magnetization. Therefore, we tentatively attribute this linebroadening to conformational averaging of these amides on a millisecond time scale. Consistent with this explanation, many of these peaks are observed in HSQC spectra of the protein recorded at elevated temperatures (data not shown).

The H^N and ^{15}N resonances from the amides in question are observed in the presence of high concentrations of cellotetraose. The peaks from several of these amides appear initially in the $^1H-^{15}N$ HSQC spectra of CBD_{N1} recorded when cellotetraose is in approximately a 1−2-fold molar ratio to total protein. All 13 peaks are observed when the concentration of cellotetraose is in 50-fold excess. This behavior is exemplified by the appearance of the ^1H and ^{15}N resonances from Ser2 at the fourth titration point in the HSQC spectra shown in Figure 2. Similar ligand to protein ratios of cellopentaose and cellohexaose, and even higher concentration ratios of cellotriose, do not lead to the appearance of these peaks. It is also interesting that the amide ^1H and ^{15}N resonances of Thr65 do not move in a linear fashion during the titration of CBD_{N1} but rather follow a curved trajectory (Figure 2). This behavior, which is also indicative of two or more binding events, occurs only upon addition of cellotetraose. In the spectra recorded to monitor the titrations with the other three cellooligosaccharides, the ^1H and ^{15}N resonances of Thr65 shift linearly with added sugar. Note that Thr65 is in β-strand B3, adjacent to the disulfide bond and on the opposite face of CBD_{N1} as the ligand binding site (Figure 8).

A possible explanation for these anomalous observations is that one or more additional low-affinity sites for cellotetraose may exist in CBD_{N1} such that the binding of this sugar restricts or enhances the conformational mobility of residues near the N-terminus and/or the disulfide bond. However, without knowledge of the chemical shifts of many of these residues in the absence of added sugar or the lifetime of the oligosaccharide−protein complex, we cannot exclude alternative explanations based upon the dependence of exchange broadening on the magnitudes of the chemical shift changes and the kinetics of ligand association and dissociation. Regardless, the association constant reported for cellotetraose in Table 1 was determined from the titrations of amides that do not show any anomalous behavior and corresponds to the primary high-affinity binding site common to each of the four cellooligosaccharides investigated in this study.

DISCUSSION

CBD_{N1} Binds Soluble Cellooligosaccharides

We have shown using heteronuclear NMR and difference ultraviolet absorbance spectroscopy that CBD_{N1} binds soluble cellooligosaccharides to form 1:1 protein−carbohydrate complexes. From quantitative titration measurements, it was found that the affinity of CBD_{N1} for cellooligosaccharides increases in the order cellotriose < cellotetraose < cellopentaose ≈ cellohexaose. The basis for the differences in the affinities for these sugars likely reflects the size of the CBD_{N1} binding site. If this site is the same length as cellopentaose, then cellopentaose and cellohexaose would be expected to bind approximately equally well, while shorter ligands would exhibit diminished affinity. Indeed, on the basis of a preliminary tertiary structure of CBD_{N1}, the length of the binding face is approximately the same as that of an extended chain of cellopentaose.

Binding Site Is Formed by a 5-Stranded β-Sheet

The chemical shift of a nucleus is very sensitive to changes in its local environment, thus providing an avenue for identifying the binding site of a ligand on a protein (Otting, 1993). When mapped onto the schematic model of CBD_{N1}, it is clear that most amides with chemical shifts perturbed significantly due to oligosaccharide binding are located within one of the two β-sheets of this jelly-roll β-sandwich protein (Figure 5). Furthermore, changes in chemical shifts greater than one standard deviation above average occur for amides in each of the five strands of this sheet. We therefore conclude that the binding site for the cellooligosaccharides, and by inference, amorphous cellulose, lies across the face of CBD_{N1} that is composed of strands A1−A5 (Figure 6). A preliminary tertiary structure of CBD_{N1} also reveals that this β-sheet face is concave, thereby forming a cleft into which the cellooligosaccharides, but not crystalline cellulose, could lie.

With the exception of the anomalous behavior observed for a few amides in the presence of high concentrations of cellotetraose, the effects of all four cellooligosaccharides on the $^1H-^{15}N$ HSQC spectra of CBD_{N1} are very similar. That is, saturating amounts of cellohexaose and cellotriose shift approximately equally the amide H^N and ^{15}N resonances of the protein. This implies that the four sugars bind CBD_{N1} at the same location and that binding results in a common perturbation of the backbone structure of the CBD. In contrast, difference ultraviolet absorbance measurements indicate that cellopentaose and cellohexaose perturb the environments of the tyrosine side chains in CBD_{N1} differently than do cellotetraose or cellotriose (Figure 9). This suggests that the two shorter oligosaccharides do not interact with the full complement of side chains that contribute to the binding site of the CBD. However, without further structural and kinetic information, we cannot determine if the four cellooligosaccharides bind CBD_{N1} in a single orientation at a single multidentate site or if they bind in fast exchange at more than one possible subsite or in more than one possible orientation across β-strands A1−A5.

Binding Face Is Composed of Hydrophobic and Hydrophilic Residues

Ultraviolet difference absorbance measurements demonstrate that tyrosine, but not tryptophan, residues are affected by the interaction of the sugars with CBD_{N1}. To identify specifically the aromatic residues involved in sugar binding, we examined the NMR spectra of CBD_{N1} recorded in the absence and presence of cellotetraose. Of the twelve aromatic groups in the CBD, only Phe9, Trp16, and Tyr19 show above-average changes in chemical shift upon addition of this oligosaccharide, while Tyr85 shows a distinct change in line shape. Therefore, these residues are potentially involved in ligand binding. The participation of the side chain of Phe106 cannot be discounted as it remain unassigned. However, Phe9 and Phe106 are unlikely to play a direct role in cellotetraose binding as they do not lie on the

face of the protein formed by β-strands A1−A5 (Figure 6). The changes in the chemical shifts of the aromatic ring protons of Phe9 are attributed tentatively to the same process that results in the appearance of the $^1H-^{15}N$ resonances from amides in the N-terminal region of CBD_{N1} upon addition of excess cellotetraose. On the basis of the preliminary structure of CBD_{N1}, the indole side chain of Trp16 lies in the interior of the protein, and thus also cannot be involved directly in binding. The changes in the amide and side chain chemical shifts of Trp16 probably stem from the fact that this residue lies within β-strand A1 and thus is influenced indirectly by the presence of the bound sugar. However, the perturbation of Trp16 is likely subtle as difference ultraviolet absorbance measurements demonstrate that the environment of the indole does not change significantly upon addition of sugar.

This leaves Tyr19 and possibly Tyr85 as the only residues with aromatic side chains that are likely to be involved directly in cellooligosaccharide binding by CBD_{N1}. These two tyrosines lie in β-strands A1 and A5, respectively, and, on the basis of the preliminary tertiary structure of the protein, both have their side chains exposed to interact with the sugar molecules (Figure 6). It is interesting that the two tyrosines are located at the extreme ends of the CBD_{N1} binding site. This may provide an explanation for different perturbations of the difference ultraviolet absorbance spectrum of the CBD induced by cellohexaose and cellopentaose versus cellotetraose or cellotriose (Figure 9). That is, the shorter oligosaccharides may not contact one or both of the tyrosine rings to the same extent as do the longer oligosaccharides.

In addition to the two candidate tyrosines, numerous nonaromatic residues must also be involved in the binding of the cellooligosaccharides to CBD_{N1}. As shown in Figure 6, an apolar or hydrophobic strip consisting of Val17, Tyr19, Val48, Ala126, Leu77, and Tyr85 extends across the β-sheet binding of the CBD. This strip is flanked by the hydrophilic side chains of Asn50, Gln124, Qln128, Asn81, Arg75, Thr87, Asp90, and Ser92. The H^N and ^{15}N resonances of many of these residues show significant changes upon addition of sugar to CBD_{N1} (Figure 5). The structure of the binding cleft of CBD_{N1} leads to a model in which cellooligosaccharides lie along the cleft formed by β-sheet A. The pyranose rings are stacked against the hydrophobic strip, and the flanking hydrophilic residues provide hydrogen bonds to the equatorial hydroxyl groups of the oligosaccharide. This model is consistent with the significant enthalpic component in the measured free energy of binding of soluble cellooligosaccharides to the protein (Tomme et al., 1996). The lack of absorption of CBD_{N1} to crystalline cellulose may result from a requirement to form hydrogen bonds to the equatorial hydroxyls of an isolated polysaccharide chain lying within the concave binding groove of CBD_{N1}.

Comparison to Other CBDs

CBD_{N1} has the distinctive feature in binding solely to amorphous cellulose, whereas the well-characterized CBDs from *T. reesi* cellobiohydrolase 1 and *C. fimi* xylanase−glucanase Cex absorb preferentially to the crystalline polysaccharide (Coutinho et al., 1992; Tomme et al., 1995a,b). The compositions of the binding sites of CBD_{CBH1} and CBD_{Cex} are in marked contrast to that identified in this study for CBD_{N1}. Specifically, the binding face of the CBD_{CBH1} is a flat surface containing three conspicuously exposed tyrosine side chains (Kraulis et al., 1989), while that of CBD_{Cex} consists of a row of three aligned tryptophans (Xu et al., 1995). A similar pattern of aromatic side chains is postulated to form the binding face of the CBD from *Erwinia chrysanthemi* (E. Brun, personal communication). As exemplified by a recent calorimetric study of CBD_{Cex}, the binding of these latter CBDs to cellulose is entropically driven and thus appears to be mediated primarily by hydrophobic interactions between the exposed aromatic side chains with the pyranosyl rings of the polysaccharide (Creagh et al., 1996). However, the precise structural details of the protein−sugar interactions, such as the orientation of the CBDs along or across the cellulose chains on the face or edge of the crystalline lattice, remains to be elucidated.

The binding of CBD_{N1} to short soluble cellooligosaccharides is distinct among CBDs characterized to date. Previously, evidence for the role of the tryptophan residues of CBD_{Cex} in sugar binding was obtained using NMR spectroscopy to monitor the effects of cellohexaose on the spectrum of the protein (Xu et al., 1995). When 4 equiv of this ligand was added to a uniformly ^{15}N-labeled sample of CBD_{Cex}, the indole $H^{\epsilon 1}$ and $^{15}N^{\epsilon 1}$ resonances of two exposed tryptophans were perturbed slightly. Although this supported the role of the surface tryptophans in cellulose binding, it also demonstrated that CBD_{Cex} interacts only weakly with soluble cellooligosaccharides. In contrast, the titration of the CBD_{N1} with cellohexaose is complete upon the addition of 1 equiv of oligosaccharide, and the effect of ligand binding on the protein is extensive. As shown in Figure 5, approximately 40% of the main chain amides in CBD_{N1} are perturbed upon addition of the soluble sugars. Clearly the CBDs from Cex and CenC differ structurally and thermodynamically in their interactions with cellulosic ligands. We are currently examining by NMR complexes of CBD_{N1} with bound cellooligosaccharides in order to provide a more detailed mechanism for the activity of this cellulose-binding domain.

ACKNOWLEDGMENT

We are grateful to Lewis Kay (University of Toronto) for generously providing the NMR pulse sequences used for this study, to Cyril Kay and Les Hicks (University of Alberta) for performing the ultracentrifgation measurements, and to Linda Cuddeford for help with thiol titrations. Quantitative amino acid analyses were carried out at the Alberta Peptide Institute (Edmonton, Alberta). We thank Emmanuel Brun, Rick Dahlquist, Logan Donaldson, Charles Haynes, Douglas Kilburn, Buon Latte, and Tony Warren for advice and support.

REFERENCES

Atalla, R. H. (1993) in *Trichoderma reesei Cellulases and Other Hydrolases* (Suominen P., & Reinikainen, T., Eds.) pp 25−39, Foundation For Biotechnical and Industrial Fermentation, Helsinki.

Bailey, Y. E., Beaven, G. H., Chignell, D. A., & Gratzler, W. B. (1968) *Eur. J. Biochem. 7,* 5−14

Bax, A., & Grzesiek, S. (1993) *Acc. Chem. Res. 26,* 131−138.

Béguin, P., & Aubert, J.-P. (1994) *FEMS Microbiol. Rev. 13,* 25−58.

Bray, M. R., Johnson, P. E., Gilkes, N. R., McIntosh, L. P., Kilburn, D. G., & Warren, R. A. J. (1996) *Protein Sci*. (in press).

Chervenka, C. H. (1969) in *A Manual of Methods for the Analytical Ultracentrifuge*, Spinco Division of Beckman Instruments Inc., Palo Alto, CA.
Coutinho, J. B., Moser, B., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1991) *Mol. Microbiol. 5*, 1221−1233.
Coutinho, J. B., Gilkes, N. R., Warren, R. A. J., Kilburn, D. G., & Miller, R. C., Jr. (1992) *Molec. Microbiol. 6*, 1243−1252.
Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., & Haynes, C. A. (1996) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Din, N., Gilkes, N. R., Tekant, B., Miller, R. C. J., Warren, R. A. J., & Kilburn, D. G. (1991) *Bio/Technology 9*, 1096−1099.
Din, N., Forsythe, I. J., Burtnick, L. D., Gilkes, N. R., Miller, R. C. J., Warren, R.A. J., & Kilburn, D. G. (1994) *Mol. Microbiol. 11*, 747−755.
Doering D. S. (1992) *Functional and Structural Studies of a Small f-Actin Binding Domain*, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
Edelhoch, H. (1967) *Biochemistry 6*, 1948−1954.
Gill, S., & von Hippel, P. (1989) *Anal. Biochem. 182*, 319−326.
Graham, R. W., Greenwood, J. M., Warren, R. A. J., Kilburn, D. G., & Trimbur, D. E. (1995) *Gene 158*, 51−54.
Henrissat, B. (1994) *Cellulose 1*, 169−196.
Herkovits, T. T., & Sorensen, S. M. (1968) *Biochemistry 7*, 2523−2532.
Kay, L., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
Knowles, J., Lehtovaara, P., & Teeri, T. (1987) *Trends Biotechnol. 5*, 255−261.
Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Petterson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
Levy, G. C., & Lichter, R. L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy*, Wiley, New York.
Linder, M., Mattinen, M.-L., Kontteli, M., Lindberg, G., Ståhlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., & Annila, A. (1995) *Protein Sci. 4*, 1056−1064.
McIntosh, L., & Dahlquist, F. (1990) *Q. Rev. Biophys. 23*, 1−38.
McIntosh, L. P., Wand, A. J., Lowry, D. F., Redfield, A. G., & Dahlquist, F. W. (1990) *Biochemistry 29*, 6341−62.
Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory Press, NY.
Ong, E., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J., & Kilburn, D. G. (1993) *Biotechnol. Bioeng. 42*, 401−409.
Otting, G. (1993) *Curr. Opin. Struct. Biol. 3*, 760−768.
Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995) *Protein Sci. 4*, 2411−2423.
Poole, D. B., Hazlewood, G. P., Huskisson, N. S., Virden, R., & Gilbert, H. J. (1993) *FEMS Microbiol. Lett. 106*, 77−84.
Quiocho, F. A. (1986) *Annu. Rev. Biochem. 55*, 287−315.
Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., Knowles, J. K. C., & Teeri, T. T. (1992) *Proteins 14*, 475−482.
Reinikainen, T., Teleman, O., & Teeri, T. T. (1995) *Proteins 22*, 392−403.
Sambrook, J., Fritch, E. F., Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767−4771.
Teeri, T., Reinikainen, T., Ruohonen, L., Jones, T. A., & Knowles, J. K. C. (1992) *J. Biotechnol. 24*, 169−176.
Tomme, P., Warren, R. A. J., & Gilkes, N. R. (1995a) *Adv. Microb. Physiol. 37*, 1−81.
Tomme, P., Warren, R. A. J., Miller, R. C., Jr., Kilburn, D. G., & Gilkes, N. R. (1995b) in *Enzymatic Degradation of Insoluble Polsaccharides* (Saddler, J. M., & Penner, M., Eds.) pp 142−161, American Chemical Society, Washington, DC.
Tomme, P., Creagh, A. L., Kilburn, D. G., & Haynes, C. A. (1996) *Biochemistry 35*, 13885−13894.
Vyas, N. K. (1991) *Curr. Opin. Struct. Biol. 1*, 732−740.
Wishart, D. S., & Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Xu, G.-Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., & Harvey, T. S. (1995) *Biochemistry 34*, 6993−7009.
Yamazaki, T., Foreman-Kay, J. D., & Kay, L. (1993) *J. Am. Chem. Soc. 115*, 11054−11055.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103−119.
Zhang, O., Kay, L. E., Olivier, J. P., & Foreman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845−858.

BI961186A