Solution Structure of a Cellulose-Binding Domain from *Cellulomonas fimi* by Nuclear Magnetic Resonance Spectroscopy[†,▽]

Guang-Yi Xu,[‡] Edgar Ong,[§] Neil R. Gilkes,[§] Douglas G. Kilburn,[§] D. R. Muhandiram,[‡] Marees Harris-Brandts,[∥] Jeremy P. Carver,[∥] Lewis E. Kay,*[,‡] and Timothy S. Harvey*[,⊥,#]

Protein Engineering Network of Centres of Excellence and Departments of Medical Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 1A8, Protein Engineering Network of Centres of Excellence and Department of Microbiology and Immunology, University of British Columbia, 300-6174 University Boulevard, Vancouver, British Columbia, Canada V6T 1Z3, and Ontario Cancer Institute, Princess Margaret Hospital, 500 Sherbourne Street, Toronto, Ontario, Canada, M4X 1K9

Received February 2, 1995; Revised Manuscript Received March 16, 1995[⊗]

ABSTRACT: Multidimensional, multinuclear nuclear magnetic resonance spectroscopy combined with dynamical simulated annealing has been used to determine the structure of a 110 amino acid cellulose-binding domain (CBD) from C_{ex}, a β-1,4-glycanase from the bacterium *Cellulomonas fimi* (CBD_{cex}). An experimental data set comprising 1795 interproton NOE-derived restraints, 50 ϕ, 34 χ_1, and 106 hydrogen bond restraints was used to calculate 20 final structures. The calculated structures have an average root-mean-square (rms) deviation about the mean structure of 0.41 Å for backbone atoms and 0.67 Å for all heavy atoms when fitted over the secondary structural elements. Chromatography, ultracentrifugation, and ^{15}N NMR relaxation experiments demonstrate that CBD_{cex} is a dimer in solution. While attempts to measure NOEs across the dimer interface were unsuccessful, a computational strategy was employed to generate dimer structures consistent with the derived data set. The results from the dimer calculations indicate that, while the monomer topologies produced in the context of the dimer can be variable, the relative positioning of secondary structural elements and side chains present in the monomer are restored upon dimer formation. CBD_{cex} forms an extensive β-sheet structure with a β-barrel fold. Titration with cellohexaose, [β-D-glucopyranosyl-(1,4)]$_5$-D-glucose, establishes that Trp 54 and 72 participate in cellulose binding. Analysis of the structure shows that these residues are adjacent in space and exposed to solvent. Together with other proximate hydrophilic residues, these residues form a carbohydrate-binding cleft, which appears to be a feature common to all CBDs of the same family.

Many bacterial and fungal cellulases are modular proteins with a distinct N- or C-terminal, cellulose-binding domain (CBD)[1] joined to a catalytic domain (Gilkes et al., 1991a). Attempts to crystallize the intact enzymes for high-resolution analysis have been unsuccessful, possibly because the domains are connected by flexible linkers (Bedarkar et al., 1992); however, small-angle X-ray scattering studies reveal low-resolution structures in which the CBD is an elongated tail-like structure attached to an ellipsoidal catalytic domain (Abuja et al., 1988; Pilz et al., 1990). Other enzymes that hydrolyze insoluble carbohydrate polymers, such as amylases (Svensson et al., 1989) and chitinases (Watanabe et al., 1994), have the same general organization. While the two domains remain functional when separated by proteolysis or genetic manipulation, a glycanase typically has reduced activity on an insoluble substrate when its substrate-binding domain is removed (Gilkes et al., 1988; Tomme et al., 1988). The binding domain may enhance enzyme activity simply by concentrating the enzyme on the substrate surface (Nidetzky et al., 1994), or it could be involved in the disruption of noncovalent interactions between adjacent substrate molecules (Din et al., 1991, 1994a). However, the precise roles of these domains in hydrolysis and their mechanisms of binding to carbohydrates are not well understood.

[†] Supported by the Protein Engineering Network Centres of Excellence (Natural Sciences and Engineering Research Council of Canada and the Medical Research Council of Canada). T.S.H. received a fellowship from NATO.

[▽] The coordinates of the 20 final structures and the mean structure have been deposited with the Protein Data Bank, entries 1EXH and 1EXG.

* Authors to whom correspondence should be addressed.-

[‡] Protein Engineering Network of Centres of Excellence and Departments of Medical Genetics, Biochemistry, and Chemistry, University of Toronto.

[§] University of British Columbia.-

[∥] Protein Engineering Network of Centres of Excellence and Departments of Medical Genetics and Biochemistry, University of Toronto.

[⊥] Ontario Cancer Institute.-

[#] Current Address: Amgen Inc., MailStop 14-2-A-210, 1840 De-Havilland Drive, Thousand Oaks, CA 91320.

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1995.

[1] Abbreviations: CBCA(CO)NH, β-proton to α/β-carbon (via carbonyl carbon) to nitrogen to amide proton correlation; CBD, cellulose-binding domain; (Hβ)Cβ(CγCδ)Hδ, β-proton to β-carbon to γ-carbon to δ-carbon to δ-proton correlation; gd-HCCH-TOCSY, gradient-enhanced proton−carbon−carbon−proton correlation using carbon total correlated spectroscopy; HMQC, heteronuclear multiple quantum coherence; HMQC-J, heteronuclear multiple quantum coherence experiment to measure NH−H$^\alpha$ scalar coupling; HN(CA)HA, amide proton to nitrogen to α-proton correlation (via α-carbon); HNCO, amide proton to nitrogen to carbonyl carbon correlation; HN(COCA)HA, amide proton to nitrogen to α-proton correlation (via carbonyl carbon and α-carbon); NOESY, nuclear Overhauser effect spectroscopy; DIPSI, decoupling in the presence of scalar interactions; HNHB, amide proton to nitrogen to β-proton correlation; HSQC, heteronuclear single quantum coherence; NMR, nuclear magnetic resonance; rms, root-mean-square; Trp(indole-d_5), Trp with all protons bound to aromatic carbons replaced by deuterons; CBD_{cex}, CBD from C_{ex}.

0006-2960/95/0434-6993$09.00/0

CBDs have been identified in more than 100 β-1,4-glucanases, but they also occur in xylanases, perhaps because cellulose and xylans are intimately associated in the cell walls of plants. They can be grouped into eight structural families (I−VIII) on the basis of amino acid sequence similarities (Coutinho et al., 1992; Tomme et al., 1995). Type I CBDs are relatively small domains containing 36 amino acid residues; all 17 known fungal CBDs are of this type. A three-dimensional structure determined by nuclear magnetic resonance spectroscopy for the type I CBD of cellobiohydrolase I from the fungus *Trichoderma reesei* takes the form of an amphiphilic wedge with dimensions 3.0 × 1.8 × 1.0 nm, and the principal structural element is an irregular, triple-stranded antiparallel β-sheet (Kraulis et al., 1989). One face is predominantly hydrophobic; the other is mainly hydrophilic and contains three tyrosine residues that seem to be involved in binding (Reinikainen et al., 1992).

This paper presents an NMR-derived structure for the type II CBD of C_{ex}, a bifunctional β-1,4-xylanase/glucanase from the bacterium *Cellulomonas fimi* (Gilkes et al., 1991b), and it is the first type II structure to be reported. The structure of the catalytic domain of this protein has recently been determined by X-ray crystallography (White et al., 1994). The type II CBD family, presently the largest with over 30 members, has several representatives from *Streptomyces* spp., *Thermomonospora* sp., and other soil actinomycetes; seven of the eight known *C. fimi* β-1,4-glycanases have type II CBDs (Gilkes et al., 1991b; Tomme et al., 1995). All of these CBDs contain about 100 amino acid residues, including cysteine residues toward the N- and C-termini that form a disulfide bond (Gilkes et al., 1991a,b) and several highly conserved aromatic residues that are implicated in substrate binding (Poole et al., 1993; Din et al., 1994b). CBDs from different glycanases of this family have different binding characteristics (P. Tomme, unpublished data), perhaps even different functions. This study provides a structural basis for analyzing the mechanism of substrate binding by type II CBDs and is a step toward understanding the roles of CBDs in cellulose hydrolysis.

MATERIALS AND METHODS

Bacterial Strains and Plasmids. *Escherichia coli* strains JM101 (Yanisch-Perron et al., 1985) and CY15077 *trp*$^-$ (Ross et al., 1992) have been described. Plasmid pTZEO7 (PTIS), an expression vector containing the *lac* promoter and a fragment of *cex* encoding CBD_{cex} (Ong et al., 1993), was used for the production of unlabeled and ^{15}N-labeled CBD_{cex}. It is unsuitable for the production of ^{13}C,^{15}N-labeled CBD_{cex} because the *lac* promoter is repressed in medium containing glucose; therefore, the *cex* fragment encoding CBD_{cex} was transferred to an expression vector containing the *tac* promoter, which is not glucose repressed. Plasmid pTug10-*xyl*, one in a series of expression vectors whose general features have been described (Graham et al., 1994, 1995), contains the *tac* promoter, a kanamycin resistance gene, and a gene (*xyl*) for β-1,4-xylanase. The *xyl* gene was excised from pTug10-*xyl* and replaced with the *cex* fragment to give pTugEO7K3. Both pTZEO7 (PTIS) and pTugEO7K3 encode CBD_{cex} (Ser 3−Gly 110) plus two additional amino acids (Ala 1 and Ser 2) at the N-terminus (Ong et al., 1993), as shown in Figure 7.

Production and Purification of CBD_{cex}. Unlabeled CBD_{cex} was produced by growing *E. coli* JM101/pTZEO7 (PTIS) at 37 °C in Luria−Bertani medium (Sambrook et al., 1989) containing 100 μg of ampicillin mL^{-1}. ^{15}N-labeled CBD_{cex} was produced by growing *E. coli* JM101/pTZEO7 (PTIS) at 37 °C in M9 medium (Sambrook et al., 1989) containing 0.1% $^{15}NH_4Cl$ (Isotec Inc.) and 100 μg of ampicillin mL^{-1}; 0.2% glycerol replaced glucose in the medium. ^{15}N,^{13}C-Labeled CBD_{cex} was produced by growing *E. coli* JM101/pTugEO7K3 at 37 °C in M9 medium containing 50 μg of kanamycin mL^{-1} and supplemented with 0.1% $^{15}NH_4Cl$, 0.1% [$^{13}C_6$]-D-glucose (Isotec Inc.), and 0.1% Celtone-CN (^{13}C, ^{15}N, 98%, Martek Corporation); additional [^{13}C]glucose (0.05%) was added during growth. CBD_{cex} containing deuterium-labeled tryptophan was produced by growing *E. coli* CY15077 *trp*$^-$/pTugEO7K3 at 30 °C in M9 medium containing 50 μg of kanamycin mL^{-1} and supplemented with 0.5% D-glucose, 1.0% casamino acids, and 0.25 mM L-tryptophan (indole-d_5, 98%, Cambridge Isotopes).

Culture supernatants containing CBD_{cex} were obtained by centrifugation and stored at 4 °C. EDTA (1 mM) and phenylmethanesulfonyl fluoride (1 mM) were added to inhibit proteases. CBD_{cex} was purified by affinity chromatography on CF1 cellulose (Ong et al., 1993). Contaminating carbohydrate was removed by subsequent size-exclusion chromatography on Superose-12 (Pharmacia) in 100 mM potassium phosphate buffer (pH 7) and 150 mM NaCl. The eluate was assayed for carbohydrate (Chaplin, 1986); carbohydrate-free fractions were pooled and exchanged into water by ultrafiltration. The concentrations of purified samples were determined from absorbance at 280 nm (Ong et al., 1993). The yields of unlabeled CBD_{cex}, [^{15}N]CBD_{cex}, [^{13}C,^{15}N]CBD_{cex}, and L-Trp(indole-d_5)-CBD_{cex} were 100, 30, 10, and 30 mg L^{-1} culture supernatant, respectively. The molecular weight of CBD_{cex} containing deuterium-labeled tryptophan, determined by matrix-assisted laser desorption ionization mass spectrometry (Kratos Analytical Inc.), was 11 110; the calculated molecular weight was 11 106.

Sedimentation Equilibrium. Sedimentation equilibrium experiments were performed at 20 °C on a Beckman Model E analytical ultracentrifuge equipped with electronic speed control and Rayleigh interference optics. Determinations of molecular weights were made by using the approach described by Chervenka (1969), using a value of 0.73 for the partial specific volume of CBD_{cex}. Equilibrium runs were performed for approximately 48 h. The apparent weight-average molecular weight was calculated from a second-degree polynomial fit of the data cast in the form ln(y) vs r^2, where y is the optical density and r is the distance traveled in the centrifuge cell.

NMR Spectroscopy. All NMR experiments were performed on a Varian Unity 500 MHz spectrometer equipped with three channels, a pseudo fourth channel for carbonyl decoupling, a pulsed field gradient triple-resonance probe with an actively shielded z gradient, and a gradient amplifier unit. Note that the pseudo fourth channel is no longer used for carbonyl decoupling in our laboratory. All ^{15}N or ^{15}N,^{13}C samples of CBD_{cex} were 1.5 mM in concentration and were dissolved in 90% H_2O/10% D_2O (pH 7.0) at 30 °C. The Trp(indole-d_5) sample was dissolved in 99.9% D_2O. ^1H−^{15}N HSQC spectra were recorded using the enhanced sensitivity pulsed field gradient approach described previously (Kay et al., 1992a; Schleucher et al., 1993; Muhandiram et al., 1994). The triple-resonance experiments for correlating backbone chemical shifts that detect NH

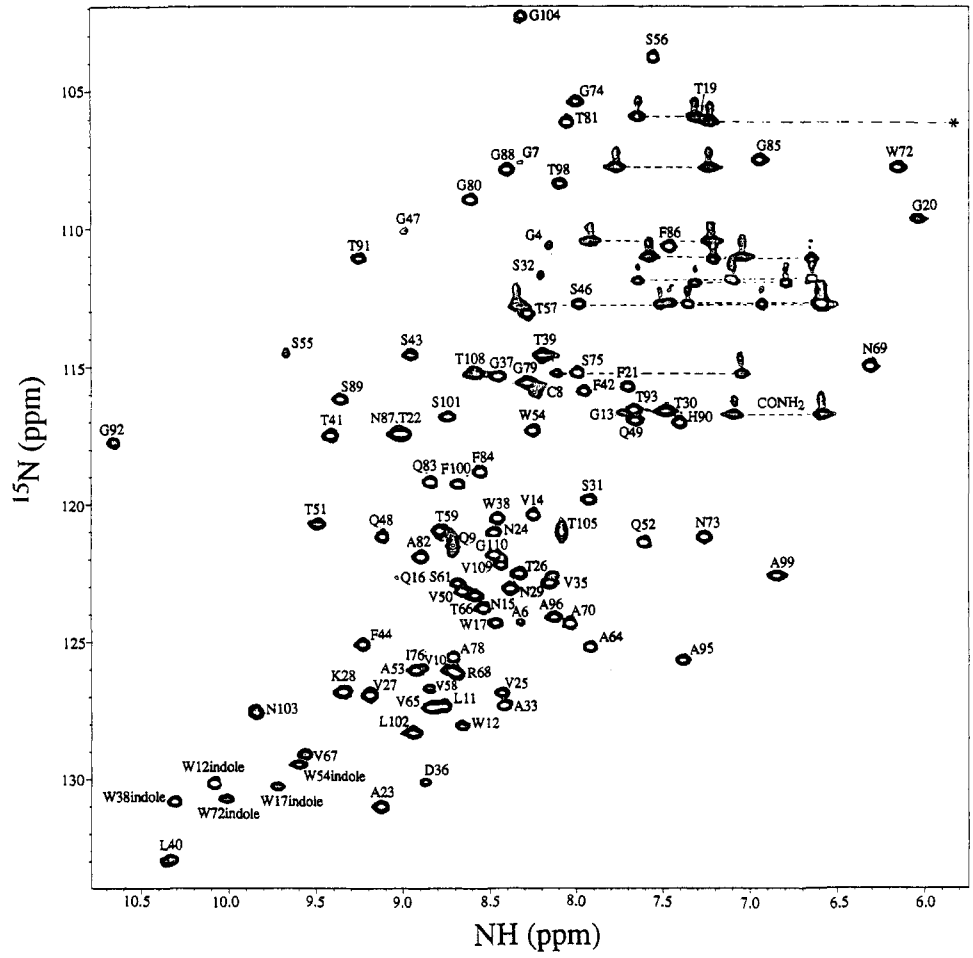

FIGURE 1: 2D ^{1}H−^{15}N HSQC spectrum of CBD$_{cex}$, uniformly labeled with ^{15}N. The spectrum was recorded at 500 MHz, with pulsed field gradient selection of coherence transfer pathways and enhanced sensitivity. Cross peaks connected by dashed lines correspond to Gln and Asn side-chain NH_2 groups.

magnetization (Kay et al., 1990; Grzesiek & Bax, 1992a) were also recorded using enhanced sensitivity pulsed field gradient methods. The following numbers of complex points and acquisition times were employed: HNCO (Kay et al., 1990) ^{13}CO (F_1) 64, 42.7 ms, ^{15}N (F_2) 32, 19.4 ms, NH (F_3) 512, 64 ms (16 transients); HN(COCA)HA (Kay et al., 1992b) HA (F_1) 32, 21.3 ms, ^{15}N (F_2) 32, 19.4 ms, NH (F_3) 512, 64 ms (16 transients); HNCACB (Wittekind & Mueller, 1993) $^{13}C^{\alpha}/^{13}C^{\beta}$ (F_1) 32, 4.3 ms, ^{15}N (F_2) 32, 19.4 ms, NH (F_3) 512, 64 ms (32 transients); CBCA(CO)NH (Grzesiek & Bax, 1992b) $^{13}C^{\alpha}/^{13}C^{\beta}$ (F_1) 52, 6.9 ms, ^{15}N (F_2) 32, 19.4 ms, NH (F_3) 512, 64 ms (32 transients); HN(CA)HA (Clubb et al., 1992; Kay et al., 1992b) HA (F_1) 32, 21.3 ms, ^{15}N (F_2) 32, 19.4 ms, NH (F_3) 512, 64 ms (32 transients). Pulsed field gradient HCCH-TOCSY (Kay et al., 1993) spectra were recorded (16 scans) on an H_2O sample of CBD$_{cex}$ with mixing times of 8 and 16 ms, with the following acquisition parameters: ^{1}H (F_1) 128, 36.6 ms, ^{13}C (F_2) 32, 10.7 ms, ^{1}H (F_3) 512, 64 ms. Simultaneous ^{15}N- and ^{13}C-edited NOESY spectra (Pascal et al., 1994) were recorded (16 scans), employing mixing times of 50 and 150 ms and acquisition parameters of ^{1}H (F_1) 128, 25.6 ms, ^{13}C/^{15}N (F_2) 32, 10.7 ms, ^{1}H (F_3) 512, 64 ms. An (Hβ)Cβ(CγCδ)Hδ experiment (Yamazaki et al., 1993) was recorded with acquisition times of 8 and 64 ms in t_1 and t_2, respectively, and with 1024 scans per complex t_1 point. ^{15}N NOESY-HSQC (Zuiderweg & Fesik, 1989; Marion et al., 1989a) and TOCSY-HSQC (Marion et al., 1989b) spectra were recorded using the enhanced sensitivity pulsed field gradient method (Kay et al., 1992a; Schleucher et al., 1993; Muhandiram et al., 1994a) with acquisition parameters of ^{1}H (F_1) 128, 23.2 ms, ^{15}N (F_2) 32, 19.3 ms, ^{1}H (F_3) 512, 64 ms. Mixing times of 100 and 45 ms were employed in the NOESY and TOCSY spectra, respectively. In the case of the TOCSY spectrum, a DIPSI-2 mixing scheme was employed (Shaka et al., 1988). After mixing, magnetization was placed along the z axis for a delay equal to half of the DIPSI mixing time to eliminate the contributions from cross relaxation in the rotating frame (Griesenger et al., 1989). An HNHB (Archer et al., 1991) spectrum was recorded with acquisition parameters ^{1}H (F_1) 64, 12.8 ms, ^{15}N (F_2) 32, 19.4 ms, ^{1}H (F_3) 512, 64 ms. 2D ^{13}C F_1-edited, F_2-filtered HMQC-NOESY experiments were recorded using the pulse scheme described by Lee et al. (1994a). Mixing times ranging from 150 to 350 ms were employed in these experiments, and delays optimized to purge signals arising from aliphatic carbons (couplings ranging from 125 to 140 Hz) during the filtering portion of the sequence were used. 2D NOESY (Jeener et al., 1979; Macura et al., 1981) spectra were recorded on the CBD$_{cex}$ Trp(indole-d_5) sample (100 ms mixing time) with the following numbers of complex data points and acquisition times: F_1, 400, 66.6 ms; F_2, 512, 85.3 ms. Finally, a 2D HMQC-J (Kay & Bax, 1990) spectrum was recorded to obtain quantitative $^3J_{HN\alpha}$ values. A data set was recorded with 400 complex t_1 points (240 ms acquisition time) and 512 complex t_2 points, with 128 scans per complex t_1 pair.

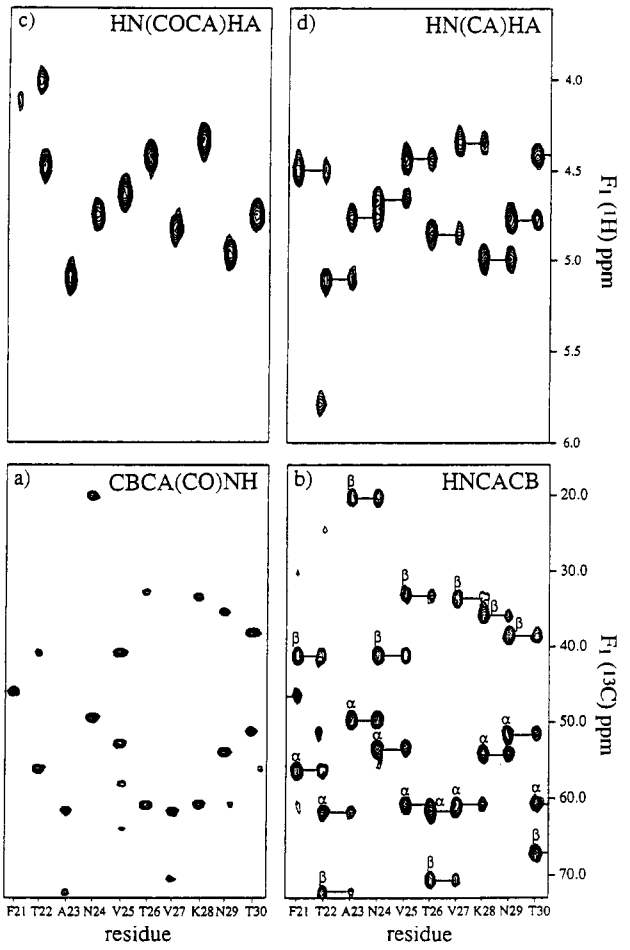

FIGURE 2: F_1 ($^{13}C^{\alpha/\beta}$ or $^1H^{\alpha}$) strips for residues Phe 21−Thr 30 from (a) CBCA(CO)NH, (b) HNCACB, (c) HN(COCA)HA, and (d) HN(CA)HA spectra of CBD_{cex}. Strips are extracted at the F_3/F_2 ($^1H/^{15}N$) frequencies of the backbone amide of each residue.

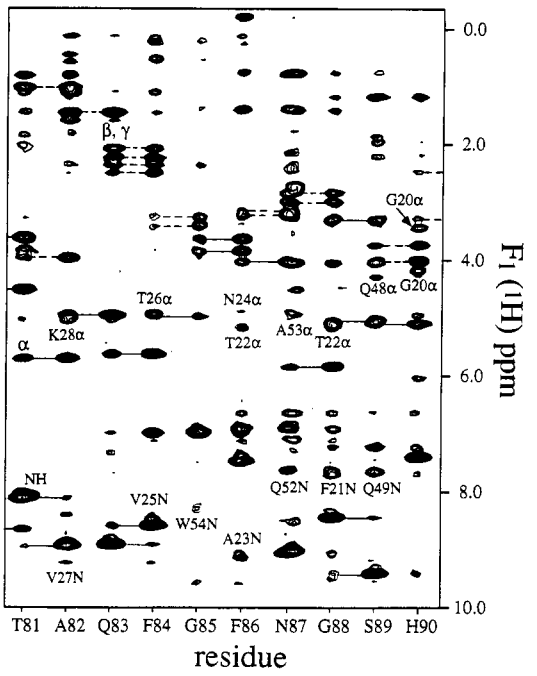

FIGURE 3: NOE strips from the 150 ms mixing time 3D ^{15}N-edited NOESY-HSQC of CBD_{cex} for amides of residues Thr 81−His 90. Selected intraresidue NOEs connecting the NH to the $^1H^{\alpha}$, $^1H^{\beta}$, etc. protons are indicated by α, β, etc., respectively, while interresidue connectivities are labeled with the residue corresponding to the site of origination of magnetization.

^{15}N relaxation data was acquired using the pulse sequences described in Figure 10 of Farrow et al. (1994) and with the acquisition times and delay values indicated in this reference.

Quadrature detection in all of the indirectly detected dimensions was achieved via States−TPPI (Marion et al., 1989c). All spectra were processed on Sun workstations using in-house routines for Fourier transformations (Kay et al., 1989) in conjunction with commercial software from NMRi (New Methods Research, Inc., Syracuse, NY). When software from F. Delaglio (NIH) became available (nmrPipe, nmrDraw), it was used exclusively. Postacquisition solvent suppression was employed in all spectra where NH protons were detected in the acquisition dimension (Marion et al., 1989d). In the case of the simultaneous ^{15}N- and ^{13}C-edited NOESY, a time domain deconvolution procedure described previously (Muhandirum et al. 1993) was employed to minimize the signal from residual water. For constant time ^{15}N evolution periods, mirror image linear prediction was employed to double the time domain signal (Zhu & Bax, 1990). In all spectra, ^1H and ^{13}C chemical shifts were referenced relative to external TSP at 0 ppm, while ^{15}N shifts were referenced relative to liquid ammonia at 0 ppm.

Stereospecific Assignments and Dihedral Angle Restraints. ϕ torsion angle restraints were obtained from an HMQC-J experiment. The measured splittings in the F_1 dimension were corrected for the contribution from dispersive components and finite line widths as described in Kay and Bax (1990). The minimum range allowed for ϕ in structure calculations was ±50°. Partial stereospecific assignments of the β-methylene protons and χ_1 determination were obtained from a short mixing time (50 ms) simultaneous ^{15}N- and ^{13}C-edited NOESY, the 3D ^{15}N-edited TOCSY-HSQC experiment, and an HNHB experiment. The minimum range for χ_1 was ±30°.

Distance Restraints. Interproton distance restraints were obtained from the ^{15}N- and ^{15}N,^{13}C-separated NOESY experiments described earlier, with NOEs classified as strong, medium, or weak and assigned corresponding upper limits of 2.7, 3.3, or 5.0 Å, respectively. In the case of monomer calculations, distances involving methyl groups, non-stereospecifically assigned methylene protons, and Hδ and Hε of aromatic groups were corrected for center averaging (Wüthrich et al., 1983), while an additional 0.5 Å was added to the upper limits for distances involving methyl groups (Wagner et al., 1987). Distance restraint corrections were not used in the dimer calculations since an r^3 potential was employed. Hydrogen bonds, inferred from amide exchange rates and characteristic NOE patterns, were incorporated as upper bound restraints of 3.3 Å between the acceptor and donor heavy atoms and 2.3 Å between the hydrogen and the acceptor heavy atom. All distances were explicitly assigned lower bounds of 0.0 Å.

Structure Calculations. All calculations were carried out using X-PLOR version 3.1 (Brünger, 1993) with the topology and parameter sets topallhdg and parallhdg, respectively. The starting structures for both monomer and dimer calculations were polypeptides with random backbone dihedrals, which were subsequently refined by dynamical simulated annealing (Nilges et al., 1991). Structures were calculated iteratively,

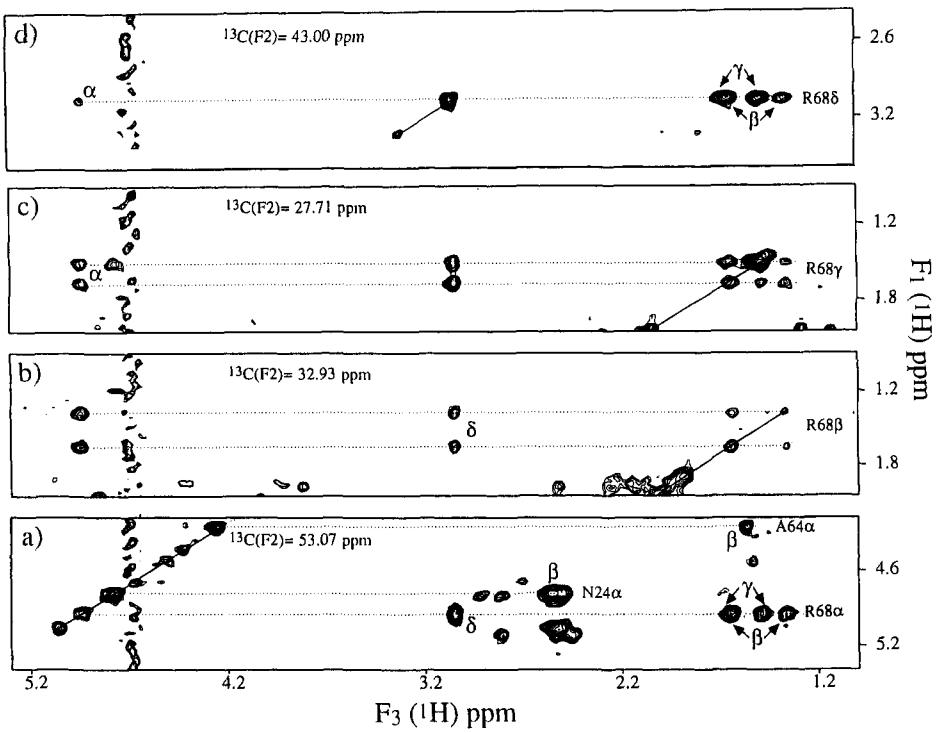

FIGURE 4: F_1-F_3 slices of the gd-HCCH-TOCSY spectrum of CBD_{cex} in 90% H_2O. Slices corresponding to connectivities involving Arg 68 are shown.

with increasing numbers of NOEs incorporated in successive rounds of calculation. Dimer structures were generated, again from random coil peptides, using the protocol described by Nilges (1993), where additional terms to maintain symmetry between the dimers were added. Additionally, a different function was used to restrain the NOE-derived distances. In this case, distances arising from specific restraints are calculated as r^{-3} sum averages of the distances arising from both inter- and intramonomer contributions. This allows the NOE distances to be treated ambiguously, i.e., as arising from intra- or intermonomer interactions or a combination of both. Further rounds of refinement for both monomer and dimer were accomplished using the protocol described previously (Hommel et al., 1992), with modifications for dimer calculations as appropriate. The low convergence properties of the method resulted in only single examples of a particular monomer fold within each dimer. To further explore the conformational space around the different topologies obtained from dimer calculations, five refinements starting from a single fold were performed, using a different velocity distribution in the initial heating stage to produce different refinement trajectories and final coordinates. This enables a more reliable evaluation of the consistent features of each dimer fold.

Secondary structure content, packing analysis, and folding free energy analysis of the dimers were performed with the program vadar (Wishart et al., 1994), utilizing the methods of Richmond (1984) and Richards (1974).

RESULTS AND DISCUSSION

Resonance Assignment. The assignment strategy that has been employed makes use of uniform ^{15}N and ^{13}C labeling of proteins in combination with triple-resonance multidimensional spectroscopy. Briefly, a number of 3D triple-resonance experiments were recorded on CBD_{cex} that correlate a variety of nuclei (^{15}N, ^{13}C, and ^{1}H spins). A comparison of the chemical shifts obtained from the family of triple-resonance experiments provides both the assignment of the intraresidue backbone chemical shifts as well as the linkage between sequential residues. Detailed descriptions of the approach are given in the literature (Clore & Gronenborn, 1991; Bax & Grzesiek, 1993) and will not be repeated herein. Figure 1 shows the ^{15}H−^{1}H HSQC spectrum of CBD, illustrating excellent chemical shift dispersion. Because the pH of the sample was reasonably high by NMR standards (pH 7), and as a result of fast exchange with water (Kay et al., 1992a), several of the backbone amide resonances are extremely weak or missing altogether. Resonances from Ser 2, Ser 3, Gln 60, Gly 62, Ser 63, and Asn 94 are missing, as are the side-chain NH_2 peaks of Asn 18. Note that the HSQC spectrum was recorded using the enhanced sensitivity pulsed field gradient approach, with coherence transfer selection provided by gradients. In this way, water suppression is achieved via gradient dephasing rather than presaturation, and the intensities of (^{15}N,NH) cross peaks are significantly improved. Further improvements can be achieved by the use of water flip back pulses to minimize dephasing of water (Grzesiek & Bax, 1993; Stonehouse et al., 1994; Kay et al., 1994) and by ensuring that water magnetization is along the $+z$ axis prior to detection. This strategy was not employed in the spectrum displayed in Figure 1, however.

Figure 2 illustrates strip plots of the 3D HN(COCA)HA, HN(CA)HA, CBCA(CO)NH, and HNCACB data sets extending from residue Phe 21 to Thr 30. In describing these experiments, we make use of the nomenclature given by Ikura et al. (1990), so that HN(COCA)HA corresponds to an experiment where the NH and ^{15}N chemical shifts of residue i are linked with the H^{α} shift of the previous residue. In a similar fashion, the CBCA(CO)NH experiment correlates the C^{β} and C^{α} shifts of residue i with the NH and ^{15}N shifts

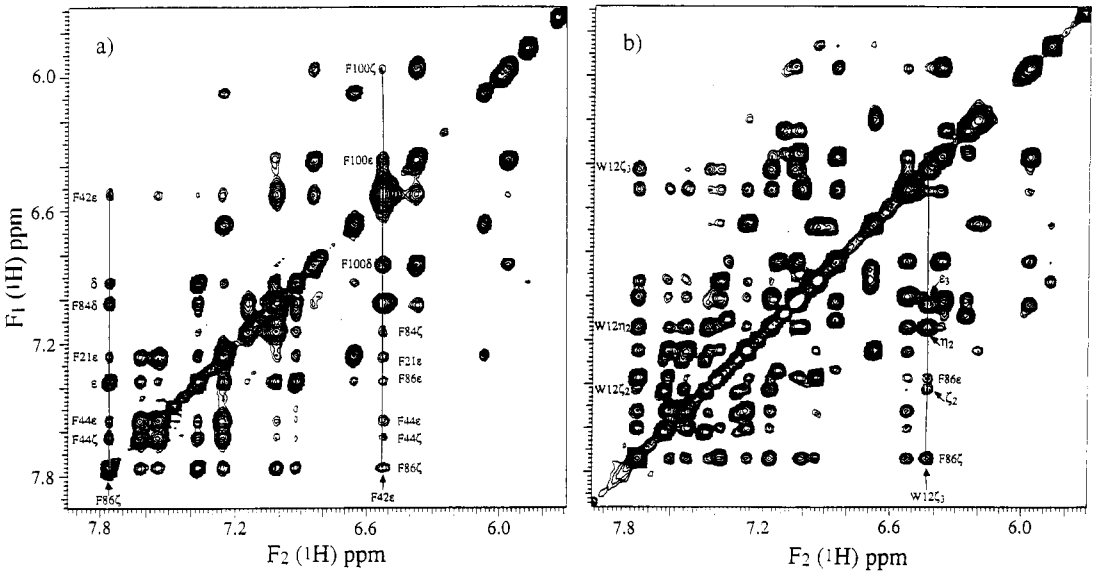

FIGURE 5: Comparison of aromatic−aromatic regions of 2D NOESY spectra of (a) Trp(indole-d_5) CBD_{cex} and (b) fully protonated CBD_{cex} in D_2O. A mixing time of 100 ms was used for both spectra.

of residue $i + 1$. Both the HNCACB and the HN(CA)HA experiments can provide intra- and interresidue correlations by virtue of both one- and two-bond $^{15}N-^{13}C^{\alpha}$ scalar couplings. In this way, the HNCACB experiment correlates intraresidue NH, C^{α}, and C^{β} chemical shifts, as well as providing correlations linking the C^{α} and C^{β} shifts of the previous residue, while the HN(CA)HA experiment provides both intraresidue NH, ^{15}N, and H^{α} correlations as well as cross peaks linking (^{15}N,NH) pairs with the H^{α} of the previous residue. All of the experiments made use of an enhanced sensitivity pulsed field gradient strategy to select for the coherence transfer pathway passing through nitrogen magnetization. In principle, for sufficiently well-resolved spectra, it should be possible to obtain a complete sequential assignment from the HNCACB and CBCA(CO)NH experiments alone. In the case of CBD_{cex}, the degeneracy of some ($^{13}C^{\alpha}$,$^{13}C^{\beta}$) pairs precluded a complete analysis based only on these two experiments. For example, Ser 31 and Ser 32 have identical $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts. In contrast, the H^{α} shifts of these residues are well removed (4.76 and 4.43 ppm). For this reason, the triple-resonance experiments that record H^{α} shifts are useful. The excellent resolution and sensitivity afforded by this family of four 3D triple-resonance experiments was sufficient to obtain the complete backbone assignment of CBD_{cex}. The assignments obtained are in agreement with the sequential NOE analysis using ^{15}N-edited TOCSY and NOESY spectra. Figure 3 illustrates strip plots extending from residue Thr 81 to His 90 extracted from the ^{15}N-edited NOESY-HSQC spectrum of CBD_{cex} recorded using a mixing time of 100 ms. These residues compose strand h of a β-sheet in the structure, and there are strong $d_{\alpha N}(i,i+1)$ connectivities throughout this region, as shown in the figure.

Side-chain assignments were obtained by analysis of both 8 and 16 ms mixing time HCCH-TOCSY experiments. The strategy that has been adopted in this laboratory is to minimize complexities in the assignment by recording as many spectra as possible, making use of only a single H_2O sample. In addition to minimizing the expense associated with the production of uniformly ^{15}N,^{13}C-labeled samples, this approach avoids complexities in assignment introduced

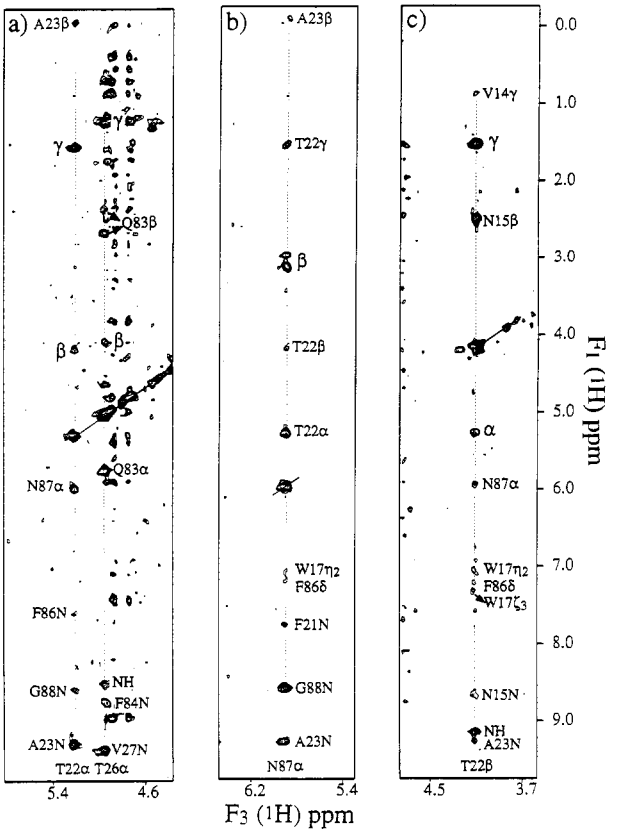

FIGURE 6: Selected F_1-F_3 slices from the 150 ms 3D simultaneous ^{15}N,^{13}C-edited NOESY spectrum of CBD in 90% H_2O: (a) NOE correlations to the H^{α} protons of Thr 22 and Thr 26 at $F_2 = 60.9$ ppm; (b) NOE correlations to the H^{α} proton of Asn 87 at $F_2 = 50.5$ ppm; (c) cross peaks to the H^{β} proton of Thr 22 at $F_2 = 71.6$ ppm. Interresidue correlations are labeled with the residue name from which magnetization originates. Intraresidue NOEs are labeled by α, β, γ, etc.

by isotope shifts and slightly different sample conditions associated with the use of both H_2O and D_2O samples. To this end, HCCH-TOCSY spectra were recorded on the same CBD sample used in all of the triple-resonance experiments, with water suppression achieved with the aid of pulsed field gradients. Figure 4 shows a number of strips from a 16 ms

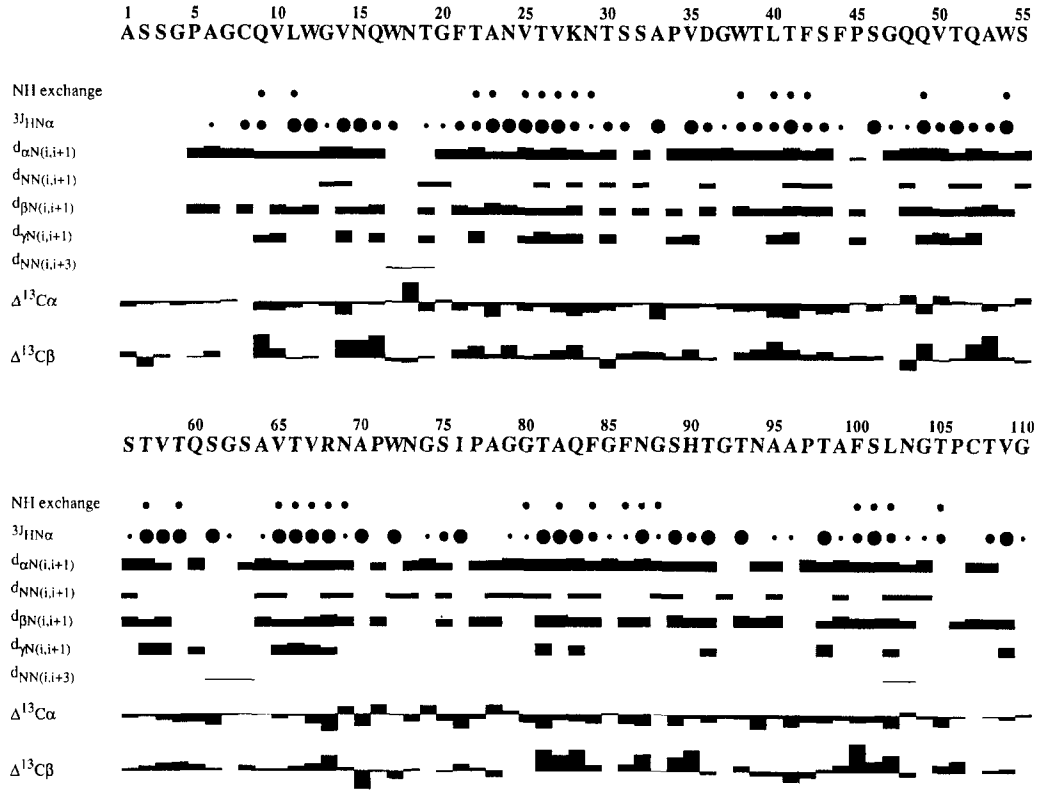

FIGURE 7: Summary of the sequential and short-range ($|i - j| < 5$ residues) NOEs involving NH atoms, $^3J_{HN\alpha}$ values, amide exchange data, and the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts measured for CBD_{cex}. Line thickness for the NOEs reflects the intensity of the NOE connectivities. Amides that produced cross peaks in a $^1H-^{15}N$ correlation experiment 1 h after dissolution of lyophilized protein in D_2O are indicated with circles. The magnitudes of the $^3J_{HN\alpha}$ values measured from HMQC-J spectra are indicated, with the size of the circles related to the size of the measured couplings (large circles, $^3J_{HN\alpha} > 9$ Hz; small circles, $^3J_{HN\alpha} < 5$ Hz).

experiment at ^{13}C chemical shifts corresponding to the side-chain carbons of Arg 68, which illustrates the quality of the data and the excellent level of water suppression achieved.

The large number of aromatic amino acids in CBD_{cex} (one His, seven Phe, and five Trp) complicated the assignment of the side chains for these residues past the C^{β} position. In principle, NOEs linking the ring protons with the H^{β} protons can provide the sequence-specific assignment of aromatic ring ^1H and ^{13}C chemical shifts. Unfortunately, ^{13}C-edited NOESY experiments often do not provide sufficient resolution to link the aromatic and aliphatic parts of the side chain. This is in large part due to the poor dispersion of aromatic carbon chemical shifts, strong coupling effects between adjacent aromatic carbons, and the efficient dipolar relaxation pathway that the ^{13}C spin provides for the attached protons. For this reason, we have made use of a strategy suggested by McIntosh et al. (1990) whereby a sample of CBD_{cex} was prepared with Trp(indole-d_5). Figure 5 compares 2D homonuclear $^1H-^1H$ NOESY spectra recorded on a sample of Trp-(indole-d_5) CBD (a) relative to a fully protonated CBD sample (b). The simplification is considerable and facilitated complete assignment of the aromatic spin systems. In addition, valuable long-range NOE constraints were also obtained from this experiment. For example, at $F_2 = 7.77$ ppm, in addition to cross peaks connecting the ζ,ϵ- and δ-protons of Phe 86, NOEs to Phe 44ζ,ϵ, Phe 21ϵ, Phe 84δ, and Phe 42ϵ were observed (Figure 5a). Recently an alternate strategy for the assignment of NOEs involving aromatic protons was suggested by Vuister et al. (1994). In their approach, termed reverse labeling, all of the amino acids are labeled uniformly with ^{13}C with the exception of an amino acid of interest, say phenylalanine. In this case it is possible to filter out signals from all residues with the exception of Phe using pulse schemes that have been developed over the past several years (Wider et al., 1990; Ikura & Bax, 1992; Gemmecker et al., 1992; Lee et al., 1994a). In addition, a ^{13}C-edited, ^{12}C-filtered 3D NOESY spectrum (Lee et al., 1994a; Vuister et al., 1994) can provide a wealth of long-range NOEs connecting the aromatic Phe protons. In the case of CBD_{cex}, however, incorporation of deuterated Trp proved to be sufficient to sort out the long-range NOE ambiguities that were present when 2D NOE data or 3D ^{13}C NOESY data on the uniformly labeled sample were examined.

Sequence-specific assignments of the ring protons of aromatic residues were verified by recording an ($H\beta$)$C\beta$-($C\gamma C\delta$)$H\delta$ experiment. This experiment makes use of the resolution provided by the $^{13}C^{\beta}$ chemical shifts, which have previously been assigned via triple-resonance experiments, to assign aromatic ring H^{δ} protons in a sequence-specific manner. Because this experiment makes use of magnetization transfer via scalar couplings exclusively, potential ambiguities associated with NOE-based assignment strategies are eliminated. Identical assignments of the H^{δ} protons were obtained via both methods. The backbone and side-chain ^{15}N, ^{13}C, and ^1H chemical shifts of CBD_{cex} obtained using the ensemble of experiments described earlier are available as supplementary material.

Figure 6 illustrates a number of regions from the 3D simultaneous $^{13}C,^{15}N$-edited NOESY-HSQC data set recorded on the $^{13}C,^{15}N$-labeled CBD_{cex} water sample that was used for all of the triple-resonance experiments. Cross

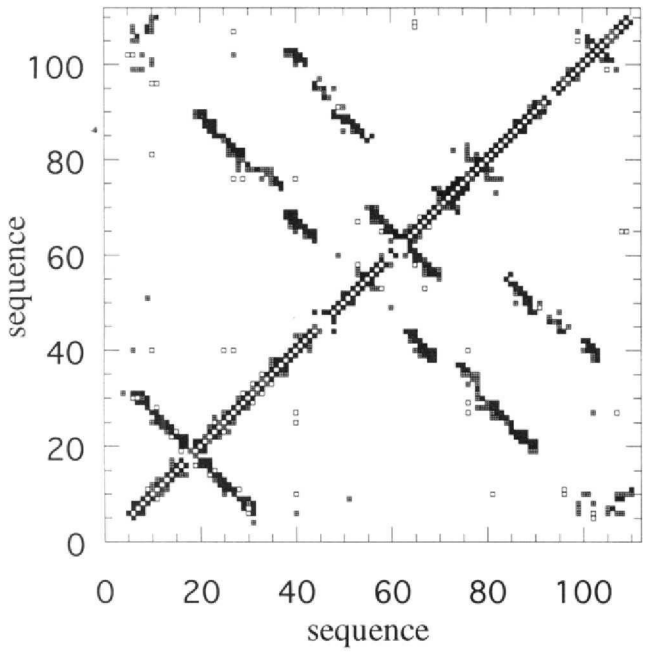

FIGURE 8: Diagonal plot of the observed interresidue NOEs in CBD_{cex}. The solid squares correspond to backbone−backbone NOEs, the shaded squares correspond to backbone−side-chain NOEs and the open squares mark side-chain−side-chain contacts.

sections at the carbon chemical shifts (F_2) 60.9 ppm (a), 50.5 ppm (b), and 71.6 ppm (c) are shown. This data set was particularly useful in the assignment of NOEs to select proton pairs because NOEs to both ^{15}N- and ^{13}C-bound protons are present. In this way, symmetry-related cross peaks are available for NOEs connecting ^{13}C- (or ^{15}N-) bound protons exclusively, as well as NOEs linking carbon- and nitrogen-bound protons. Thus, four chemical shifts, two for the protons and two for the heteroatoms to which the protons are attached, can be derived for each NOE (Pascal et al., 1994).

Secondary Structure and Topology. Figure 7 summarizes the sequential and short-range ($|i - j| < 5$ residues) NOE connectivities that have been obtained from the 2D and 3D data sets recorded on CBD_{cex} and provides qualitative NH exchange rates and $^3J_{HN\alpha}$ coupling constants. In addition, deviations from random coil chemical shift values are also illustrated. The values of $\Delta^{13}C^{\alpha}$ and $\Delta^{13}C^{\beta}$ have been shown to correlate with the secondary structure of the protein (Spera & Bax, 1991; Wishart et al., 1991; Wishart & Sykes, 1994), with negative $\Delta^{13}C^{\alpha}$ and positive $\Delta^{13}C^{\beta}$ values observed for residues in β-sheets. For residues in helical segments of the protein, exactly the opposite trend is observed. Figure 8 depicts the long-range contacts obtained for CBD_{cex}, which, when used in conjunction with the amide exchange data in Figure 7, reveals the almost totally β-sheet character of the molecule. There are 10 antiparallel β-strands, labeled a−j in Figure 9, which form a 9-stranded β-barrel, comprising two faces. The first face consists of strands a, b, h, and d (the abhd face) and the second contains five strands, e, f, c, i, and j (efcij face). The final strand, g, is located at one end of the barrel and is not part of either of the faces.

Strand a (residues 8−17) is connected to strand b (residues 19−29) by a type I turn (Wilmot & Thornton, 1990). These two strands, in conjunction with strands h (residues 80−89) and d (residues 49−54), form one side of a half-open β-barrel, with a classical type β-bulge (Richardson et al., 1978) at residues 51 and 52 in strand d terminating the hydrogen-bonding network across the face of this sheet. The opposing sheet is initiated at strand c (residues 33−44), which is linked to strand b by an unstructured loop. The N-terminal portion of strand c makes contacts with strand g (residues 74−78), but these interactions are halted by a

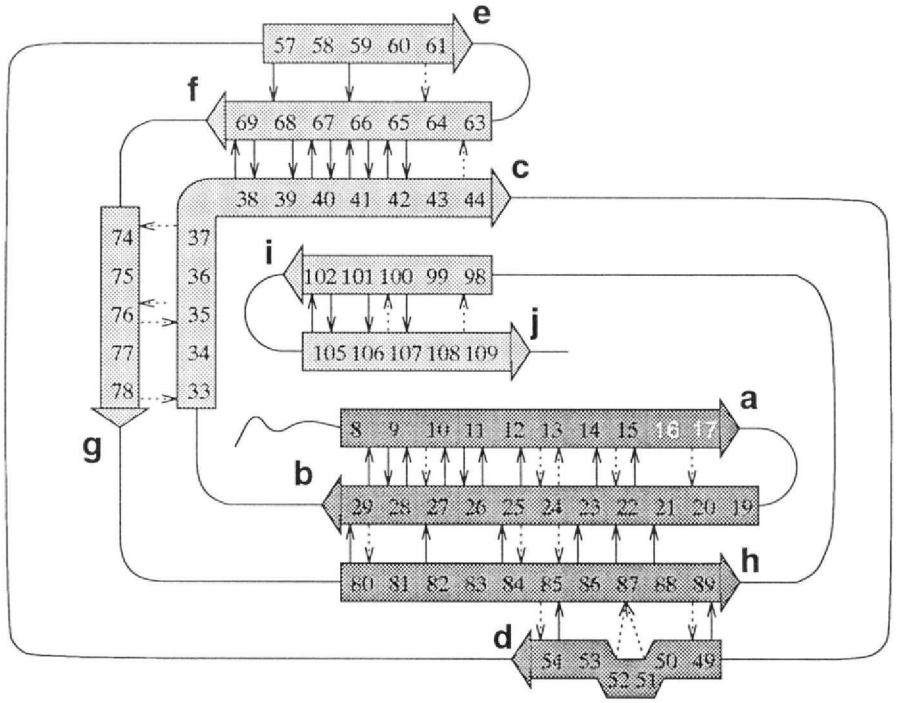

FIGURE 9: Secondary structure of CBD_{cex} with each strand labeled at its C-terminus. The strands are shaded according to the sheet in which they are found. Hydrogen bonds are also indicated: those represented by solid arrows were derived from the results of amide exchange experiments, while those represented by dotted arrows were obtained from analysis of the derived structures and consideration of experimental NOE data.

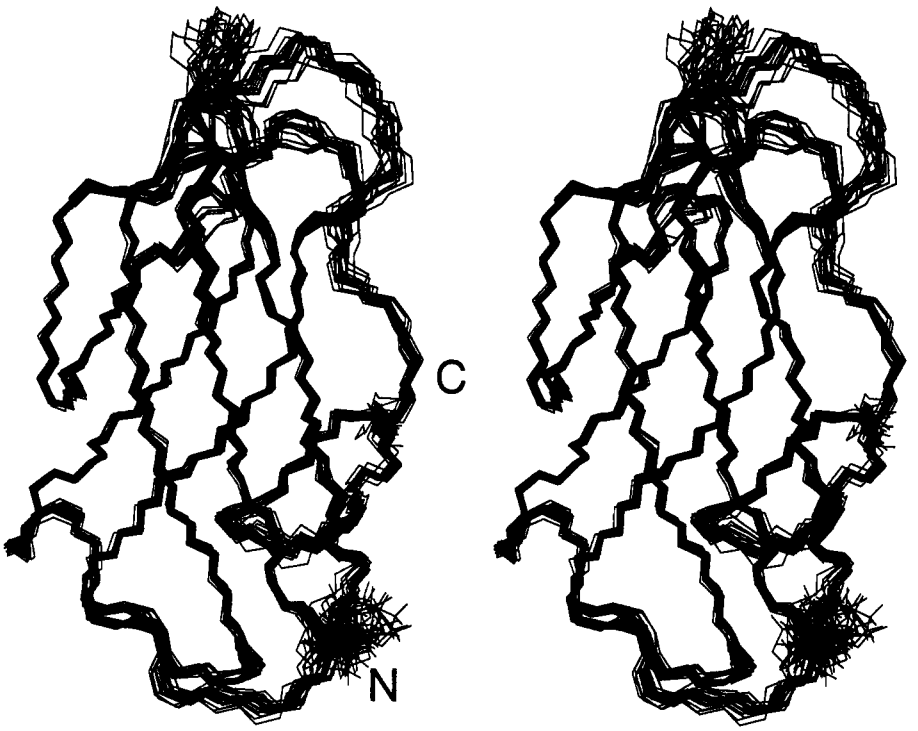

FIGURE 10: Superposition of the backbone atoms of the 20 final simulated annealing structures of CBD_{cex}. The N- and C-termini of the molecule are indicated in the figure, and the first five residues, which are disordered, have been omitted for clarity.

weakly helical turn encompassing residues 70−73 between strands f and g, as indicated by the secondary chemical shifts of several residues in this region. Gly 37, a conserved residue in most of the type II CBD sequences, facilitates the twist in strand c. This results in a shift of four in the strand pairing upon going from the c−g interaction to c−f. Strand c is also hydrogen bonded to strand i (residues 98−102) at its C-terminus. Strands e (residues 57−61) and f (residues 63−69) are connected by a tight turn and nestle between strands c and d, although there are no hydrogen bonds observed between strands d and e, possibly a consequence of the irregular secondary structure present in strand d at this point. The loop between strands c and d leads the chain back to the larger abhd face of the barrel. The β-bulge in strand d is typified by a loss of secondary chemical shift characteristics at residues Thr 51 and Gln 52 and the weakened sequential α-N NOE observed for Thr 51. A very tight turn encompassing residues 54−56 leads to strand e, which returns the chain to the efcij face. The type II′ turn following strand e leads directly to strand f, which interacts in a regular antiparallel fashion with strand e. Following the disruption of strand f, strand g (residues 74−78) leads back to the β-sheet on the opposite face of the molecule and strand h (residues 80−89), sandwiched between strands b and d. The long loop from residues 90−98 between strands h and i can potentially interact with the loop connecting strands c and d, residues 45−48, on the basis of an analysis of the final structures (see the following). The 5-stranded face of CBD is completed by the short stretches of β-sheet comprising strands i and j (residues 105−109), linked by a tight turn including Gly 104, conserved in most type II CBD sequences.

Hydrogen Bonds. One hundred and six distances, representing fifty-three backbone−backbone NH−CO hydrogen bonds, were used in the calculations. Thirty-two of these were obtained after analysis of amide exchange data, while the remainder were deduced from early structures and NOE patterns (Kraulis et al., 1989). Hydrogen bond restraints were only added for those interactions consistently present

Table 1: Structural Statistics and Atomic rms Deviations[a]

	⟨SA⟩	$(SA)_r$
rms deviations from exptl		
distance restraints (Å)		
all (1795)	0.025 ± 0.0032	0.021
intra (501)	0.015 ± 0.0032	0.014
sequential ($\|i-j\|=1$) (360)	0.024 ± 0.0051	0.022
short ($\|i-j\|\leq 5$) (145)	0.031 ± 0.0069	0.022
long ($\|i-j\|>5$) (679)	0.025 ± 0.0056	0.021
H-bonds (108)	0.043 ± 0.0097	0.038
dihedrals (83)	0.37 ± 0.06	0.370
deviations from idealized		
covalent geometry		
bonds (Å)	0.01 ± 0	0.010
angles (deg)	2.4 ± 0.007	2.4
impropers (deg)	0.99 ± 0.01	0.71
energies		
E_{total} (kcal mol^{-1})	2736.6 ± 20.2	2647.8
E_{repel} (kcal mol^{-1})	52.7 ± 4.1	44.4
E_{NOE} (kcal mol^{-1})[b]	56.8 ± 6.3	39.1
E_{cdih} (kcal mol^{-1})[c]	0.86 ± 1.0	0.72
E_{bond} (kcal mol^{-1})[d]	91.5 ± 2.7	83.4
E_{angle} (kcal mol^{-1})[d]	2424.4 ± 9.1	2378.2
$E_{L\text{-}J}$ (kcal mol^{-1})[e]	−300.4 ± 10.7	−317.5
$E_{improper}$ (kcal mol^{-1})[d]	110.2 ± 2.4	101.6

[a] Notation: ⟨SA⟩ represents the 20 final simulated annealing structures; $(SA)_r$ is the restrained minimized mean structure obtained by restrained minimization of the mean structure. The number of each type of restraint used is given in parentheses in column 1. [b] E_{NOE} was calculated by using a force constant of 50 kcal mol^{-1} Å^{-2}, a square well potential, and center averaging. [c] E_{cdih} was calculated by using a force constant of 200 kcal mol^{-1} rad^{-2}. [d] The force constants used to calculate E_{bond}, E_{angle}, and $E_{improper}$ were 1000 kcal mol^{-1} Å^{-2}, 500 kcal mol^{-1} rad^{-2}, and 500 kcal mol^{-1} rad^{-2}, respectively. [e] E_{repel} was calculated by using a final value of 4.0 kcal mol^{-1} Å^{-4} with the van der Waals hard sphere radii set to 0.75 times those in the parameter set parallhsa supplied with X-PLOR (Brünger et al., 1987; Brünger, 1988), with modifications as described by Bagby et al. (1994).

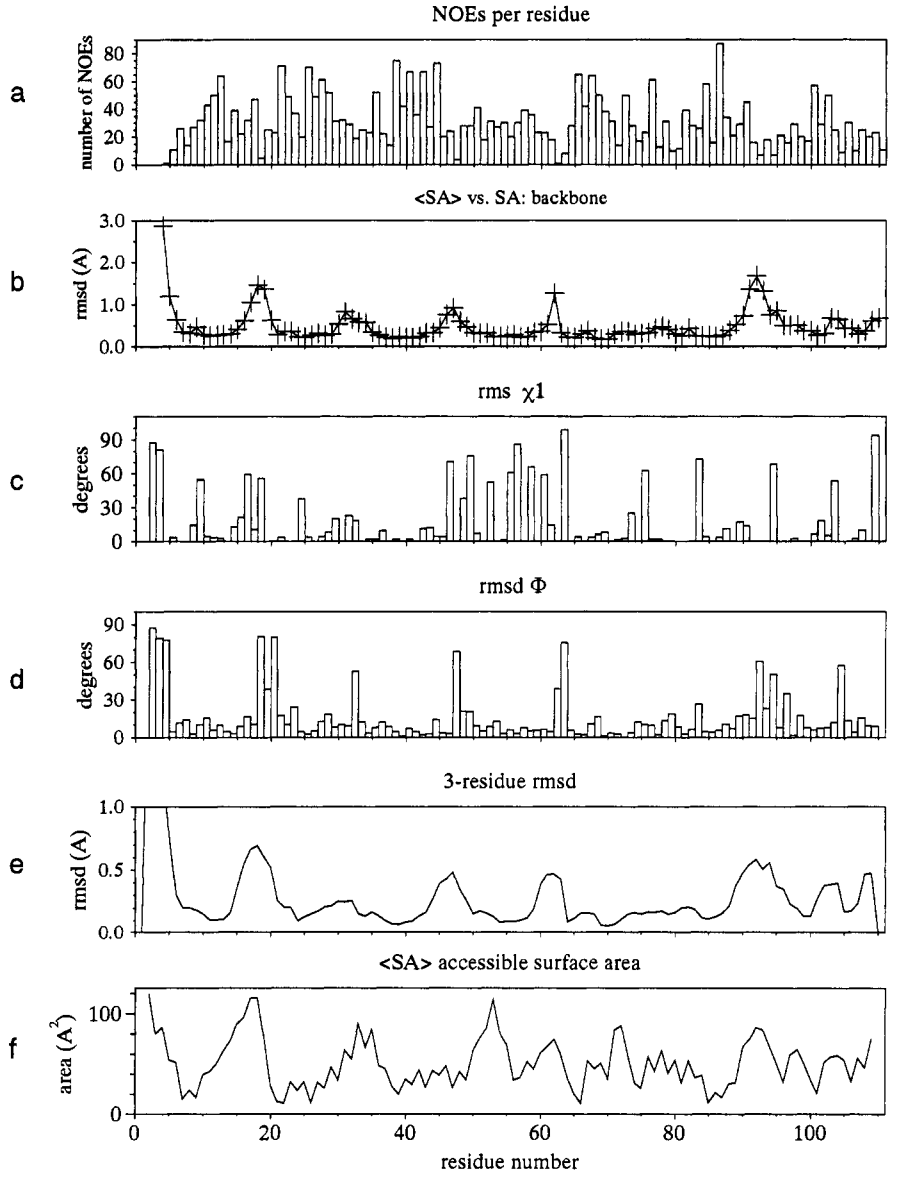

FIGURE 11: Structural data for the 20 final structures shown as a function of residue number: (a) NOE distribution (note that an intraresidue NOE would count twice in the total for that residue); (b) average rms distance of the backbone atoms from the average structure; (c) rms deviation of the backbone dihedral angle χ_1 calculated from the final 20 structures; (d) rms fluctuation of ϕ for the 20 final structures; (e) three-residue rms distance. This is a measure of local definition and is calculated by using a moving three-residue window, with the result displayed above the central residue of the triplet. Regions that show a high rms distance to the average on global superposition will show a small value for the three-residues rms distance if the observed structural variability is a result of rigid body type displacements. (f) The solvent-accessible surface for each residue, calculated using a probe of radius 1.6 Å. The values were then smoothed by calculating a running average over three successive residues.

in both well-converged structures and regular secondary structural elements.

Monomer Structures. Twenty final structures were calculated from a data set comprising 1795 interproton NOE-derived restraints, 50 ϕ, 34 χ_1, and 106 hydrogen bond restraints. Structural statistics are given in Table 1. Figure 10 shows a best-fit superposition of the backbone atoms. The distribution of NOEs along the polypeptide chain, the per residue rms distance relative to the mean structure, the χ_1 rms, and solvent accessibilities are shown in Figure 11. The data presented earlier indicate that good convergence has been achieved and that the structures are of high precision. The structures show small deviations from ideal geometry and have favorable nonbonded contacts, as indicated by the low value for the independently computed van der Waals energy. The 20 final structures have an average atomic rms distance from the mean structure of 0.41 Å for backbone atoms and 0.67 Å for all non-hydrogen atoms, when fitted over the regular secondary structural elements: residues 8−17, 19−29, 33−44, 49−54, 57−61, 63−69, 74−78, 80−89, 98−102, and 105−109. There were no distance restraints violated by more than 0.3 Å and no dihedral angle restraint violations over 3.5°.

Overall, CBD_{cex} is somewhat elongated, with dimensions of approximately 45 × 25 × 25 Å, with the β-sheets oriented along the longest dimension. Unlike the previously reported structure for the CBD of *T. reesei*, CBH I (Kraulis et al., 1989), there is no definite partitioning of the hydrophilic and hydrophobic side chains. Viewed from the exterior, CBD_{cex} has a cone-shaped cavity, approximately 7 Å deep by 5 Å at its widest point. At the apex of the cone lies Val 27, which also composes a part of the hydrophobic core of CBD.

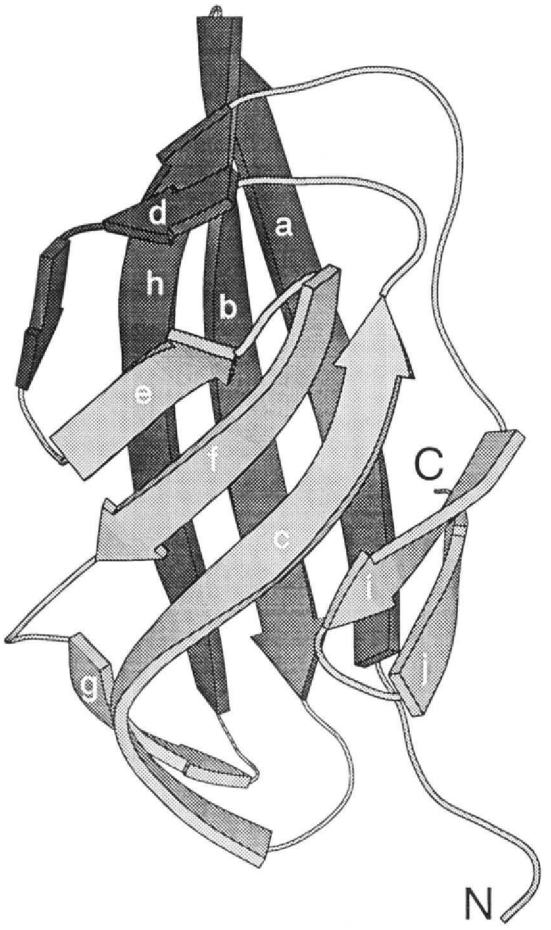

FIGURE 12: Schematic representation of CBD_{cex} showing the locations of the β-strands. The four-stranded sheet comprising strands abhd is shaded darker than the five-stranded face comprising strands efcij (see Figure 9).

The sides of this feature are formed by the side chains of Val 35 and Trp 38 from strand c, Leu 102 from strand i, Ala 6 and Cys 8 from strand a, and Ile 76 and Ala 82 from strand h, as well as the aliphatic portion of the side chain of Asn 29. The function of this cavity is unclear at present.

Figure 12 is a ribbon cartoon of the architecture of CBD_{cex}. The overall topology is similar to that of tendamistat (Pflugrath et al., 1986), except for the insertion of residues 57−69, which form two additional β-strands (e and f, as shown in Figure 13). The nine strands of CBD_{cex} are assembled around an extensive aromatic core, comprising residues Trp 12, Phe 21, Trp 38, Phe 42, Phe 44, Phe 84, and Phe 86, that is functionally conserved in all type II CBDs (Din et al., 1994b). The indole NH of the Trp 12 side chain can possibly interact with the backbone carbonyl groups of either Gly 20 or Thr 22, preventing the burial of the non-hydrogen-bonded NH. Phe 100, also associated with this core, is partially exposed, lying stacked on the side chain of Phe 42. Trp 12 and Phe 21, 44, and 86 lie toward one end of the molecule and are in close contact. The aromatic side chains of Trp 12 and Phe 21 lie edge to edge in the same plane, with Phe 44 approximately midway between the two with its ring perpendicular to this plane. The side chain of Phe 42, on the edge of this cluster, is oriented perpendicular to Trp 12, Phe 21, and Phe 44. Trp 38, Phe 42, and Phe 84 are associated with a larger proportion of aliphatic side chains, namely, Leu 102, Leu 40, and Val 27. Val 10, Val 25, Leu 40, Val 65, and Val 67 are also largely buried in the interior of the molecule. Val 65 lies in a space formed between the side chains of Phe 42 and Phe 44, while Ala 23 occupies a similar position, filling space between Trp 12, Phe 21, and Phe 86. The side chain of Val 25 lies on the opposite side, nestling between the aromatic rings of Phe 84, Phe 86, and Trp 12. All of the branched aliphatic residues mentioned earlier show conservative substitutions throughout type II CBD domains. The side chains of valines 50 and 58 lie on top of this core, between strands e and f. This is illustrated in Figure 14a, which shows all aromatic side chains for the family of 20 final structures superimposed on the backbone of the average structure of CBD_{cex}. Figure 14b shows the aliphatic side chains in a similar manner.

The remaining tryptophan residues are not associated with the hydrophobic core of the molecule, but are located at its surface. This finding is consistent with fluorescence data, indicating that Trp residues in the type II CBD of *C. fimi* endoglucanase A (Trp 14 and Trp 68, respectively) corresponding to Trp 17 and Trp 72 of CBD_{cex} are located at the surface (Din et al., 1994b). The conserved residue Trp 17 is situated at one end of the molecule, lying over the turn connecting strands a and b. There is no obvious reason for the conservation of this residue, but its position at one end of the molecule and its disorder indicate that it may be involved in the interface between the catalytic and binding domains in the intact molecule. One of the notable features of the CBD_{cex} monomer is a ridge, formed by convergence of the loop between strands f and g and the turn between strands d and e, essentially along the edge of the abhd and efcij faces (Figure 12). Each of these loops is surmounted by a tryptophan side chain, Trp 54 in the case of the d−e turn and Trp 72 for the f−g loop (Figure 14a), with a small space in between. Both side chains are oriented with their aromatic functions parallel to the surface of the molecule. The d−e turn and f−g loop are somewhat hydrophilic, with Asn 73 reaching back across the loop toward Ala 70, where contacts are observed between the β-methyl and side-chain NH_2 groups. Asn 69, as a result of NOEs observed from its NH_2 to the H^{α} protons of Gly 74, is also buried beneath Trp 72. Similarly, Ser 55, the side-chain hydroxyl of which may hydrogen bond to the backbone carbonyl of Gln 83, and Ser 56 are buried beneath Trp 54. These loops are not tightly associated with the aromatic core of CBD and could easily accommodate cellohexaose ([β-D-glucopyranosyl-(1,4)]$_5$-D-glucose) in the cleft formed between them or, alternatively, on the surface of the molecule at this location. In this context, it is noteworthy that the residues of this region are highly conserved. Furthermore, the chemical functionalities present in this region, namely, the aromatic and hydrophilic groups, are similar to those found in other carbohydrate-binding proteins, for example, CBH I (Rouvinen et al., 1990), where it is suggested that a Trp side chain specifically interacts with a glucosyl ring while numerous hydrogen bonds are formed between protein and ligand. While it is tempting to construct a more detailed model for CBD_{cex} cellohexaose interactions, the dimerization of CBD in this study (see the following) necessitates a more conservative analysis.

Addition of cellohexaose to the CBD_{cex} samples results in the titration of the side-chain indole NH and nitrogen resonances of Trp 54 and Trp 72, as observed in HSQC experiments. The titration is complete after the addition of

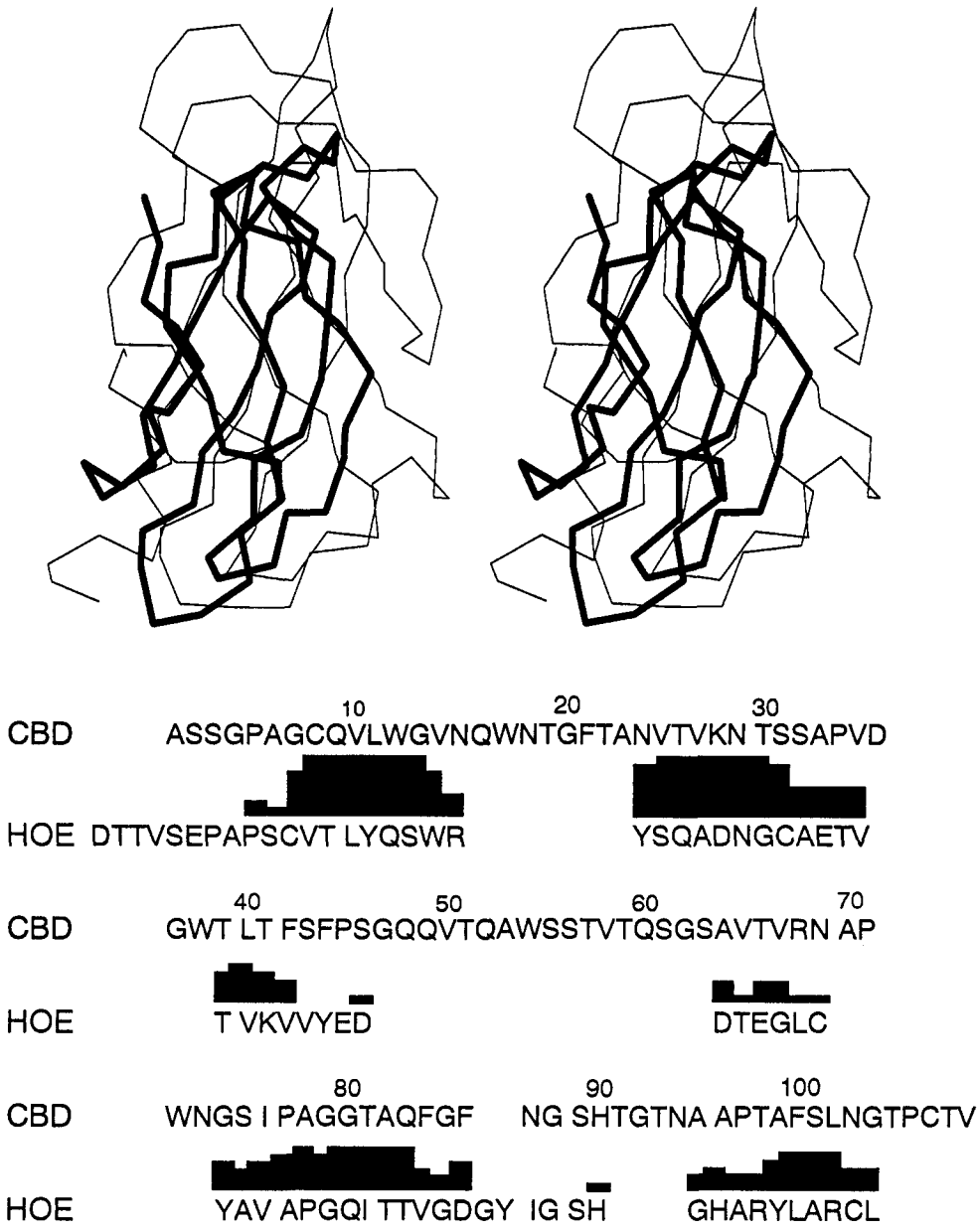

FIGURE 13: (Top) Superposition of the C^{α} coordinates of tendamistat (coordinates taken from Brookhaven entry pdb1hoe.ent are shown in heavy lines) and the average minimized structure of CBD_{cex} (thin lines). (Bottom) An alignment of the sequences of CBD_{cex} and tendamistat is also provided, with the height of the bar between the two sequences indicating the degree of structural similarity at that location. The C^{α} coordinates of these residues were used to superimpose the two coordinate sets.

4 molar equiv of sugar, and during the course of the titration the indole NH of Trp 54 shifts by 0.2 ppm from 9.65 to 9.85 ppm, while the corresponding nitrogen chemical shift changes by 0.4 ppm from 129.5 to 129.9 ppm. The changes for Trp 72 are slightly less, 0.09 and 0.3 ppm for ^{1}H and ^{15}N, respectively, but they occur in concert with those of Trp 54. This indicates either specific interaction with, or change in conformation upon binding of, cellohexaose, probably involving hydrogen bond formation as evidenced by the downfield shift of these resonances. A conformational change is less likely, since Trp 38 in proximity to the base of the d−e and f−g loops (5 Å) shows no changes in chemical shifts. There is a comparatively high concentration of potential hydrogen bond donors and acceptors at this site, a characteristic shared by the conserved residues in CBHI (Kraulis et al., 1989), which may be intuitively expected for a protein that is required to interact with an oligosaccharide molecule such as cellulose. In addition, the surface topology at this site makes it an attractive candidate for binding to cellulose. Alignment of CBD sequences indicates that the residues in this region (Thr 51, Trp 54, Trp 72, Asn 73, and Ser 75) either are conserved or retain hydrogen bond donor/acceptor properties (Din et al., 1994b). These residues are likely retained in all type II CBDs for functional purposes (i.e., cellulose binding). In contrast, all other conserved residues appear to be required for the structure of the molecule.

Data from two sets of site-directed mutagenesis experiments on type II CBDs are relevant to these findings. Replacement of either Trp 14 or Trp 68 in the CBD of *C. fimi* endoglucanase A (corresponding to Trp 17 and Trp 72 of CBD_{cex}, respectively) with Ala resulted in 50- and 30-fold reductions in affinity for cellulose, respectively; conformational changes, relative to the wild-type CBD, were

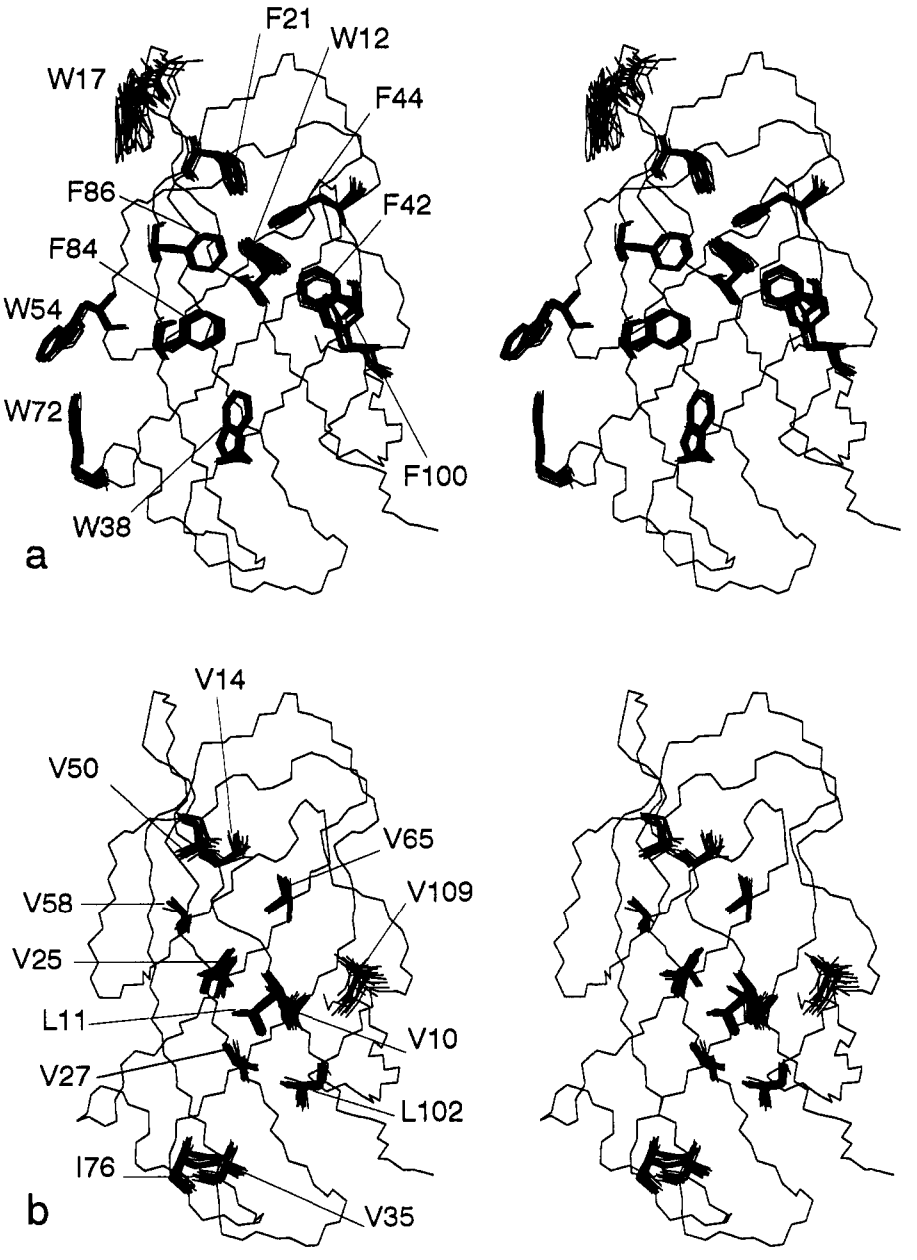

FIGURE 14: (a) Average minimized structure of CBD_{cex}, over which are displayed the side chains of all aromatic residues from the 20 final structures of the molecule. The side chain of Trp 17, at the apex of the molecule, is only poorly defined. Note the positions of Trp 54 and Trp 72, which show the largest chemical shift changes upon the addition of cellohexaose (see text for details). The aliphatic side chains of Val, Leu, and Ile are displayed in (b).

not apparent in either mutant (Din et al., 1994b). In this context, it is of interest to note that, upon titration of CBD_{cex} with cellohexaose, the indole NH and ^{15}N chemical shifts of Trp 17 changed by 0.03 and approximately 0.05 ppm, respectively. These changes are much smaller than those observed for Trp 54 and 72, and in this case, the titration was essentially complete at 0.5 molar equiv of sugar. It should also be noted that the initial rates of change of ^{15}N and NH chemical shifts were much larger for Trp 54 and Trp 72 than for Trp 17. For this reason, we favor a model in which cellohexaose binds to a cleft between the d−e and f−g loops, with Trp 54 (d−e loop) and Trp 72 (f−g loop) participating in the binding. The mutagenesis data are consistent with the involvement of Trp 72 of CBD_{cex} in binding to cellohexaose, and they further suggest, in the context of the cellohexaose titration data, that Trp 17 could participate in the binding of longer β-1,4-glucans in cellulose. In contrast, replacement of Trp 66 in the CBD of *Pseudomonas fluorescens* xylanase A (corresponding to Trp 72 of CBD_{cex}) with Ala or Phe did not significantly affect the affinity for cellulose (Poole et al., 1993). This result suggests that structural differences may exist between CBD_{cex} and the CBD of *P. fluorescens* xylanase A. Replacement of either Trp 34 or Trp 38 of xylanase A (corresponding to Trp 38 and Phe 42 in the hydrophobic core of CBD_{cex}, respectively) with Ala or Phe resulted in a substantial loss of affinity for cellulose. It may be that conformational changes that expose buried Trp residues occur upon binding to insoluble cellulose. However, the possibility that the reduced affinities resulted from improper folding of the mutants cannot be excluded since the structural intergrity of the xylanase A CBD mutants was not examined.

The orientations of the hydrophilic side chains of Asn 29 and Gln 48 are of some interest. The side chain of Asn 29

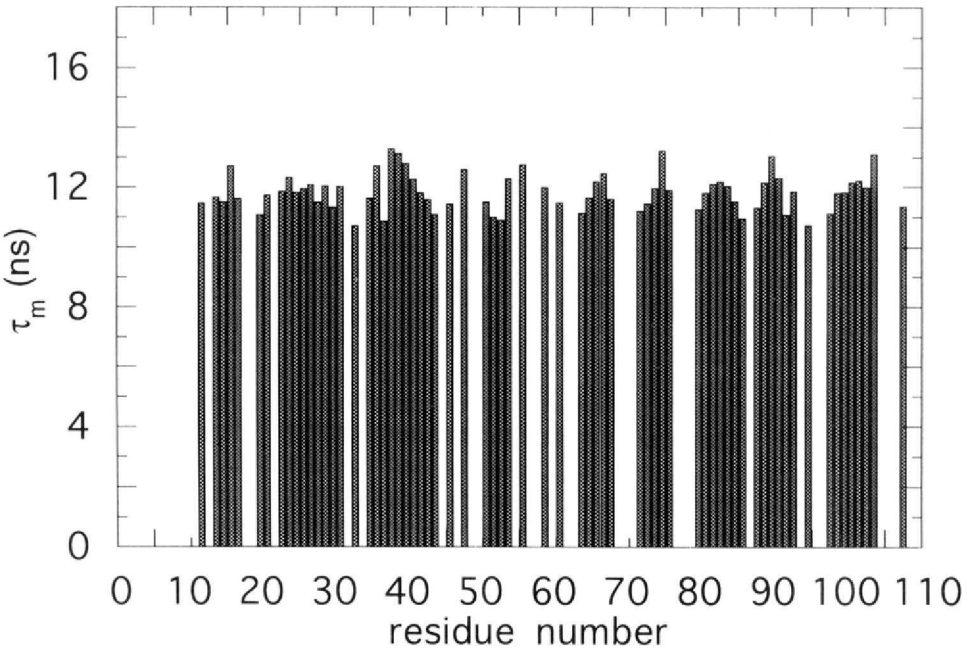

FIGURE 15: Overall correlation time of CBD_{cex} as a function of residue number calculated from ^{15}N T_1 and T_2 and $^1H-^{15}N$ steady state relaxation data using the approach described by Farrow et al. (1994). For a CBD monomer, a correlation time of approximately 6−7 ns would be obtained.

is buried by the loop connecting strands b and c, where the NH_2 moiety has the possibility of forming hydrogen bond interactions with the carbonyl group of Ala 78 and the hydroxyl group of Ser 31. The carbonyl group of the side chain of Asn 29 may act as a hydrogen bond acceptor for the amide group of Gly 7 HN, although this is not a totally consistent feature of the final ensemble of structures. The side chain of Gln 48, similarly, does not project directly into the solvent, but instead buries the hydrophobic portion of its side chain in the interior of the loop formed by residues 90−98 (h−i loop). The NH_2 function of this residue extends into the solvent on the opposite side of the molecule in the region of Phe 21, to which NOEs from the side-chain NH_2 are observed. There are a number of potential interactions between the loops at this end of the molecule, as revealed by analysis of hydrogen bonds in the 20 final structures. Interactions occur between the e−f, c−d, and h−i loops, with the potential formation of weak hydrogen bonds between Val 65 NH and Ser 43 O^γ, Thr 91 NH and Ser 46 O^γ, and Ser 46 NH and Asn 94 carbonyl groups. These may act to stabilize the tertiary structure of CBD_{cex}. Further potential interactions occur between side chains of Thr 41 and Thr 66 across the c and f strands, with the side-chain group of Thr 66 continuing this pattern by interacting with Ser 61. Such an analysis also indicates the presence of an α-helical type hydrogen bond from Gly 74 NH to the carbonyl of Ala 70. Since the force field used to calculate the structures contains no attractive components, electrostatic or otherwise, the hydrogen bonds found by this analysis are not an artifact of the method.

Dimer Structure. Despite the apparent precision of the monomer structure, as described earlier, and the fact that none of the 1795 distance restraints were violated by more than 0.3 Å, the available data indicate that CBD_{cex} is a dimer in solution. First, the quality of the 2D correlation spectra, which rely on magnetization transfer via scalar connectivities (COSY, TOCSY), as well as 3D triple-resonance spectra were noted to be significantly worse for CBD_{cex} than for samples of other proteins with approximately the same numbers of amino acids. In order to investigate the oligomeric state of CBD_{cex}, ^{15}N relaxation experiments were recorded as described by Farrow et al. (1994). Specifically, ^{15}N T_1 and T_2 and $^1H-^{15}N$ steady state NOE relaxation measurements were carried out, and the experimental data were fitted on a per residue basis using the model-free approach of Lipari and Szabo (1982a,b). Figure 15 shows a plot of the overall correlation time, τ_m (assumed to be isotropic), as a function of residue for CBD_{cex}. An average value of $\tau_m = 11.8 \pm 0.6$ ns was obtained. This is approximately a factor of 2 larger than expected for a 110 amino acid protein and strongly suggests that, at the concentrations employed for NMR (∼10−15 mg/mL), CBD_{cex} is a dimer. In order to establish this further, a sedimentation equilibrium study was performed on a sample of CBD at a concentration of 1.1 mg/mL (data not shown). A weight-average molecular weight of 15 200 was obtained, which is approximately 40% larger than the molecular weight calculated on the basis of the primary sequence of the protein. Thus, even at concentrations that are an order of magnitude lower than those used for NMR, CBD is in equilibrium between monomer and dimer. In order to establish whether the dimer was the result of the formation of intermonomer disulfides, SDS−polyacrylamide gel electrophoresis was performed on CBD_{cex} samples that were preincubated at 37, 60, and 100 °C in either the presence or absence of β-mercaptoethanol. All samples ran with an apparent molecular weight of approximately 11 000 (data not shown), and there were no higher molecular weight bands. In addition, there was no evidence of covalent dimers from mass spectrometry.

With these results in mind, we prepared an NMR sample consisting of an equal mixture of ^{13}C-labeled and unlabeled CBD_{cex}, following the approach of Folkers et al. (1993, 1994), in order to identify inter-subunit NOEs. Despite the fact that a number of 2D ^{13}C F_1-edited, F_2-filtered HMQC-NOESY spectra were recorded with a range of mixing times

Table 2: Structural Statistics for Representative Families of Dimer Folds

type[a]	% secondary structure	E_{NOE}[b]	E_{cdih}	E_{total}	E_{L-J}	E_{SA}[c]
A	60 ± 12	29.5 ± 10.0[d]	1.9 ± 1.4	584.7 ± 46.4	−470 ± 46	−176 ± 1
A	57 ± 7	32.3 ± 6.2	1.4 ± 1.0	608.6 ± 22.7	−468 ± 23	−180 ± 5
B	56 ± 8	25.9 ± 3.8	2.8 ± 3.2	614.2 ± 41.6	−473 ± 42	−181 ± 4
B	55 ± 5	30.5 ± 4.2	0.6 ± 0.5	584.8 ± 30.7	−492 ± 31	−186 ± 5
C	54 ± 7	35.9 ± 3.5	1.3 ± 1.3	615.7 ± 17.1	−484 ± 17	−182 ± 8
C	52 ± 5	39.3 ± 4.9	5.3 ± 4.7	705.7 ± 24.8	−452 ± 25	−181 ± 4
C	56 ± 8	49.5 ± 6.7	7.6 ± 5.7	737.3 ± 32.9	−449 ± 33	−186 ± 3
C	57 ± 6	40.2 ± 7.6	2.7 ± 2.0	659.5 ± 45.2	−459 ± 45	−182 ± 5
C	52 ± 2	34.7 ± 6.7	6.8 ± 5.7	637.4 ± 48.3	−464 ± 48	−179 ± 5
C	56 ± 8	57.5 ± 11.2	3.3 ± 2.1	781.1 ± 60.7	−467 ± 61	−192 ± 5
C	58 ± 7	44.0 ± 16.7	2.7 ± 1.2	668.0 ± 60.6	−492 ± 61	−184 ± 7
C	54 ± 1	25.0 ± 4.3	1.9 ± 1.4	555.6 ± 20.5	−489 ± 21	−181 ± 3

[a] From a pool of over 100 structures calculated using the method of Nilges (1993), 12 coordinate sets were selected to provide further evaluation of the folds obtained. This was achieved by using each structure as a template to generate five independent structures with the same fold by refinement with different starting velocities. Each dimer structure is classified according to the following types: A, monomers separated in space; B, dimers with interactions at the side-chain level but no strand interchanges; C, dimers with both strand interchanges and side-chain intermonomer interactions. [b] NOE energies are calculated using SUM averaging within X-PLOR with an r^3 potential. Therefore, the monomer structure and dimer structure NOE energies are not directly comparable. The units of E_{NOE}, E_{cdih}, E_{total}, E_{L-J}, and E_{SA} are kcal mol^{-1}. The force constants for distance restraints and dihedral angle restraints were 50 kcal mol^{-1} Å^{-2} and 200 kcal mol^{-1} rad^{-2}, respectively. [c] E_{SA} = solvation free energy calculated using the method of Eisenberg et al. (1985). A value for E_{SA} of −185 is expected (Chiche et al., 1990). [d] rms deviation values are based on a set of five structures within each class.

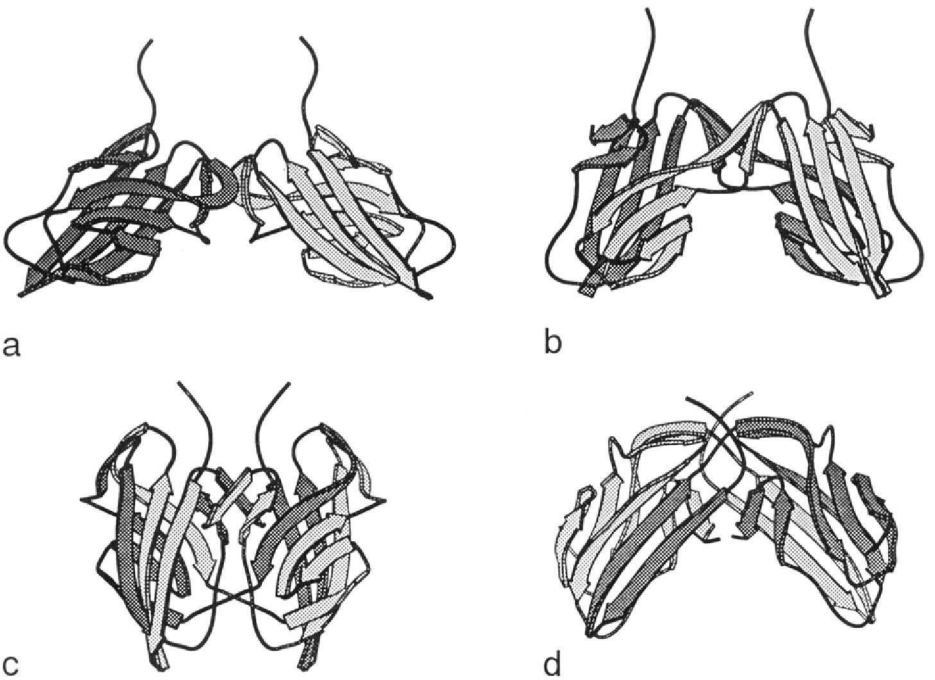

FIGURE 16: Representative CBD_{cex} dimer structures from a number of families, illustrating the range of calculated intermonomer interactions (see text). To facilitate identification of the monomer fold, the two chains are shaded differently. Dimer structures were calculated according to the method of Nilges (1993).

and extensive signal averaging, no inter-subunit NOEs were detected (i.e., no NOEs were detected between protons attached to ^{13}C and protons attached to ^{12}C). This may reflect significant motion on a microsecond to millisecond time scale at the interface between the two monomers, which could broaden any potentially observable interactions. In the absence of experimental data, dimer structures were calculated by allowing every NOE to be either inter- or intramonomer (including hydrogen bonds). In contrast to previous findings (Clore et al., 1990; Lodi et al., 1994), intersubunit NOEs were not trivial to identify on the basis of the secondary structure, since the overall topology was not known *a priori* for this protein. Indeed, it was not possible to identify any NOEs that did not fit with the emerging structure. The difficulty associated with this particular problem for NMR is highlighted by the results from the dimer calculations, in that the topology is essentially undefined, without even the possibility of correctly identifying partners for strands in β-sheets. This problem may not necessarily be the case for proteins that contain a significant proportion of α-helical secondary structure, which would limit the number of possible conformations by the inherently different packing and topology-restricting properties of helices in relation to β-sheets. Nevertheless, the utmost caution must be exercised when interpreting data that come from non-monomeric proteins, since even in cases where the resulting structures appear precise assumptions made at any level may adversely affect the derived structures (Lee et al., 1994b), and the resulting picture of the molecule may be precise but totally inaccurate.

By using a very conservative approach whereby NOEs were not explicitly assigned as arising from either monomer or dimer interactions, over 100 dimer structures were obtained by using the protocol described by Nilges (1993). As reported previously (Lee et al., 1994b), the convergence was very poor due to the lack of specific intermonomer NOE restraints. The best 20 structures, as judged in terms of their symmetry and fit to the experimental data, were selected for further refinement as described in the Materials and Methods section. Table 2 summarizes, for each family, various properties that might be used to assess the validity of the structures, with a view to eliminating various folds. This has proved extremely difficult, given the overlap between the various quantities shown.

Representatives from a number of families are shown in Figure 16 to illustrate the variability observed. The range of calculated intermonomer interactions extends from essentially separate subunits (not shown, but see the comments in Table 2), through subunits interacting at the side-chain level, to those where the subunit topology is virtually unrecognizable from that of the calculated monomer, the fold of which is restored upon assembly of the dimer. Fold a (Figure 16a) has the dimer−dimer interface at the base of strand c, with additional contacts involving strand g. There are several occurrences of this weak interaction within the converged structures that utilize different regions of the molecule. The fold of Figure 16b (fold b) shows an interesting case where the two sheets are separated and the normal sheet−sheet interface becomes that of the dimer. In this case, the dimer integrity would be almost exclusively maintained by hydrophobic contributions, as is the case in the core of the monomer. The folds illustrated in Figure 16c,d illustrate more complicated strand exchanges. The different folds produced here are a direct result of the inherent flexibility of the loops connecting the strands forming CBD. Were it possible to limit the range of conformations accessible to the backbone of these regions, then the observed variability and consequent uncertainty observed almost certainly would not occur for this molecule. Conversely, the correlation between the structurally conserved glycines in the turns and residues in the hydrophobic core is a strong indication that the secondary structural features present in both the monomer and dimer structures are correct.

The results from the dimer calculations indicate that, while the monomer topologies produced can be variable, the relative positioning of secondary structural elements and side chains present in the monomer is restored upon dimer formation. While the structures of the dimers closely resemble those of two adjacent monomers, it is often the case that the individual elements of the secondary structure are contributed by different polypeptide chains. Thus, the relative positioning of side chains within the monomer is preserved in the dimer structure in spite of the changes in backbone tracing between dimer folds. The structures that have been produced therefore provide a starting point for understanding the genetic and biochemical data that have been obtained on this system and, in particular, establish the region in the domain that interacts with cellulose.

ACKNOWLEDGMENT

We thank Julie Forman-Kay, Roger Graham, Lawrence McIntosh, and Tony Warren for helpful discussions, Janet Thonton and Christine Orengo for homology searches, Michael Nilges for providing dimer calculation protocols for X-PLOR, Gary Lesnicki for technical assistance, and Cyril Kay and Les Hicks for sedimentation equilibrium results. pTug10-*xyl* was provided by Roger Graham.

SUPPLEMENTARY MATERIAL AVAILABLE

Table S1 giving ^{1}H, ^{13}C, and ^{15}N NMR assignments for CBD_{cex} at 30 °C and pH 7.0 (3 pages). Ordering information is given on any current masthead page.

REFERENCES

Abuja, P. M., Pilz, I., Claeyssens, M., & Tomme, P. (1988) *Biochem. Biophys. Res. Commun. 156*, 180−185.
Archer, S. J., Ikura, M., Torchia, D. A., & Bax, A. (1991) *J. Magn. Reson. 95*, 636−641.
Bax, A., & Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131−138.
Bedarkar, S., Gilkes, N. R., Kilburn, D. G., Kwan, E., Rose, D. R., Miller, R. C., Jr., Warren, R. A. J., & Withers, S. G. (1992) *J. Mol. Biol.* 693−695.
Brünger, A. T. (1993) *X-PLOR Manual*, Yale University, New Haven, CT.
Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1987) *Protein Eng. 1*, 399−406.
Chaplin, M. F. (1986) in *Carbohydrate analysis: a practical approach* (Chaplin, M. F., & Kennedy, J. F. Eds.) pp 1−36, IRL Press, Oxford, UK.
Chervenka, C. H. (1969) *A Manual for the Analytical Ultracentrifuge*, Spinco Division of Beckman Instruments Inc., Palo Alto, CA.
Chiche, L., Gregoret, L. M., Cohen, F. E., & Kollman, P. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3240−3243.
Clore, G. M., & Gronenborn, A. M. (1991) *Science 252*, 1390−1399.
Clubb, R. T., Tanabal, V., & Wagner, G. (1992) *J. Biomol. NMR 2*, 203−210.
Coutinho, J. B., Gilkes, N. R., Warren, R. A. J., Kilburn, D. G., & Miller, R. C., Jr. (1992) *Mol. Microbiol. 6*, 1243−1252.
Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J., & Kilburn, D. G. (1994a) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11383−11387.
Din, N., Forsythe, I. J., Burtnick, L. D., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J., & Kilburn, D. G. (1994b) *Mol. Microbiol. 11*, 747−755.
Eisenberg, E., & McLachlan, A. D. (1985) *Nature 319*, 199−203.
Farrow, N. A., Muhandiram, D. R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., & Kay, L. E. (1994) *Biochemistry 33*, 5984−6003.
Folkers, P. J. M., Folmer, F. H. A., Konings, R. N. H., & Hilbers, C. W. (1993) *J. Am. Chem. Soc. 115*, 3798−3799.
Folkers, P. J. M., Nilges, M., Folmer, R. H. A., Konings, R. N. H., & Hilbers, C. W. (1994) *J. Mol. Biol. 236*, 229−246.
Gemmecker, G., Olejniczak, E. T., & Fesik, S. (1992) *J. Magn. Reson. 96*, 199−204.
Gilkes, N. R., Warren, R. A. J., Miller, R. C., Jr., & Kilburn, D. G. (1988) *J. Biol. Chem. 263*, 10401−10407.
Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., & Warren, R. A. J. (1991a) *Microbiol. Rev. 55*, 303−315.
Gilkes, N. R., Claeyssens, M., Aebersold, R., Henrissat, B., Meinke, A., Morrison, H. D., Kilburn, D. G., Warren, R. A. J., & Miller, R. C., Jr. (1991b) *Eur. J. Biochemistry 202*, 367−377.
Graham, R. W., Atkinson, T., Kilburn, D. G., Miller, R. C., Jr., & Warren, R. A. J. (1994) *Nucleic Acids Res. 21*, 4923−4928.
Graham, R. W., Greenwood, J. M., Warren, R. A. J., Kilburn, D. G., & Trimbur, D. E. (1995) *Gene* (in press).
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870−7872.
Grzesiek, S., & Bax, A. (1992a) *J. Magn. Reson. 96*, 432−440.
Grzesiek, S., & Bax, A. (1992b) *J. Am. Chem. Soc. 114*, 6291−6293.
Grzesiek, S., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593−12594.

Hommel, U., Harvey, T. S., Driscoll, P. C., & Campbell, I. D. (1992) *J. Mol. Biol. 227*, 271−272.
Ikura, M., & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 2433−2440.
Ikura, M., Kay, L. E., & Bax, A. (1990) *Biochemistry 29*, 4659−4667.
Jeener, J., Meir, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.
Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110−126.
Kay, L. E., Marion, D., & Bax, A. (1989) *J. Magn. Reson. 84*, 72−84.
Kay, L. E., Ikura, M., Tschudin, R., & Bax, A. (1990) *J. Magn. Reson. 89*, 496−514.
Kay, L. E., Keifer, P., & Saarinen, T. (1992a) *J. Am. Chem. Soc. 114*, 1063−1065.
Kay, L. E., Wittekind, M., McCoy, M. A., Friedrichs, M. S., & Mueller, L. (1992b) *J. Magn. Reson. 98*, 443−450.
Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., & Forman-Kay, J. D. (1993) *J. Magn. Reson. B, 101*, 333−337.
Kay, L. E., Xu, G. Y., & Yamazaki, T. (1994) *J. Magn. Reson. A 109*, 129−133.
Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Petersson, G., Knowles, J., & Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
Lee, W., Revington, M. J., Arrowsmith, C. H., & Kay, L. E. (1994a) *FEBS Lett. 350*, 87−90.
Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D., & Arrowsmith, C. H. (1994b) *Struct. Biol. 1*, 877−890.
Lipari, G., & Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4546−4559.
Lipari, G., & Szabo, A. (1982a) *J. Am. Chem. Soc. 104*, 4559−4570.
Lodi, P. J., Garrett, D. S., Kuszewski, J., Tsang, M. L., Weatherbee, J. A., Leonard, W. J., Gronenborn, A. M., & Clore, G. M. (1994) *Science 263*, 1762−1767.
Macura, S., Juang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259−281.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989a) *J. Am. Chem. Soc. 111*, 1515−1517.
Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. (1989b) *Biochemistry 28*, 6150−6156.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989c) *J. Magn. Reson. 85*, 393−399.
Marion, D., Ikura, M., & Bax, A. (1989d) *J. Magn. Reson. 84*, 425−430.
McIntosh, L. P., Wand, A. J., Lowry, D. F., Redfield, A. G., & Dahlquist, F. W. (1990) *Biochemistry 29*, 6341−6362.
Muhandiram, D. R., & Kay, L. E. (1994) *J. Magn. Reson. B 103*, 203−216.
Muhandiram, D. R., Farrow, N. A., Xu, G. Y., Smallcombe, S. H., & Kay, L. E. (1993) *J. Magn. Reson. B 102*, 317−321.
Nidetzky, B., Steiner, W., Hayn, M., & Claeyssens, M. (1994) *Biochem. J. 298*, 705−710.
Nilges, M. (1993) *Proteins 17*, 295−309.
Nilges, M., Kuszewski, J., & Brünger, A. T. (1991) in *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy* (Hoch, J. C., Poulsen, F., & Redfield, C., Eds.) pp 451−455, Plenum Press, New York.
Ong, E., Gilkes, N. R., Miller, R. C., Jr., Warren, R. A. J., & Kilburn, D. G. (1993) *Biotechnol. Bioeng. 42*, 401−409.
Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D., & Kay, L. E. (1994) *J. Magn. Reson. B 103*, 197−201.
Pflugrath, J. W., Wiegand, G., & Huber, R. (1986) *J. Mol. Biol. 189*, 383−386.
Pilz, I., Schwarz, E., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A. J., & Gilkes, N. R. (1990) *Biochem. J. 271*, 277−280.
Poole, D. B., Hazlewood, G. P., Huskisson, N. S., Virden, R., & Gilbert, H. J. (1993) *FEMS Microbiol. Lett. 106*, 77−84.
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.
Richards, F. M. (1974) *J. Mol. Biol. 82*, 1−14.
Richardson, J. S., Getzoff, E. D., & Richardson, D. C. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2574−2578.
Richmond, T. J. (1984) *J. Mol. Biol. 178*, 63−89.
Ross, J. B. A., Senear, D. F., Waxman, E., Koombo, B., Rusinova, E., Huang, Y. T., Lanes, W. R., & Hasselbacher, C. A. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 12023−12027.
Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., & Jones, T. A. (1990) *Science 249*, 380−385.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schleucher, J., Sattler, M., & Griesinger, C. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 1489−1491.
Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274−293.
Spera, S., & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490−5492.
Stonehouse, J., Shaw, G. L., Keeler, J., & Laue, E. D. (1994) *J. Magn. Reson. A 107*, 178−184.
Svensson, B., Jespersen, H., Sierks, M. R., & MacGregor, E. A. (1989) *Biochem. J. 264*, 309−311.
Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., & Claeyssens, M. (1988) *Eur. J. Biochem. 170*, 575−581.
Tomme, P., Warren, R. A. J., & Gilkes, N. R. (1995) *Adv. Microb. Physiol.* (in press).
Vuister, G. W., Kim, S.-J., Orosz, A., Marquardt, J., Wu, C., & Bax, A. (1994) *Struct. Biol. 1*, 605−613.
Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611−639.
Watanabe, T., Ito, Y., Yamada, T., Hashimoto, M., Sekine, S., & Tanaka, H. (1994) *J. Bacteriol. 176*, 4465−4472.
White, A., Withers, S. G., Gilkes, N. R., & Rose, D. R. (1994) *Biochemistry 33*, 12546−12552.
Wider, G., Weber, C., Traber, R., Widmer, H., & Wüthrich, K. (1990) *J. Am. Chem. Soc. 112*, 9015−9016.
Wilmot, C. M., & Thornton, J. M. (1990) *Protein Eng. 3*, 479−493.
Wishart, D. S., & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
Wishart, D. S., Richards, F. M., & Sykes, B. D. (1991) *J. Mol. Biol. 222*, 311−333.
Wishart, D. S., Willard, L., & Sykes, B. D. (1994) VADAR, Dept. of Biochemistry, Univ. of Alberta, Edmonton, AB, Canada.
Wittekind, M., & Mueller, L. (1993) *J. Magn. Reson. B*, 201−205.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−961.
Yamazaki, T., Forman-Kay, J. D., & Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 1054−1055.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103−119.
Zhu, G., & Bax, A. (1990) *J. Magn. Reson. 90*, 405−410.
Zuiderweg, E. R. P., & Fesik, S. W. (1989) *Biochemistry 28*, 2387−2391.

BI950248E