Solution Structure of the CBM10 Cellulose Binding Module from Pseudomonas Xylanase A †, ‡

ARTICLE *in* BIOCHEMISTRY · MARCH 2000

Impact Factor: 3.02 · DOI: 10.1021/bi992163+ · Source: PubMed

6 AUTHORS, INCLUDING:

S. Raghothama
Indian Institute of Science
86 PUBLICATIONS **1,625** CITATIONS
SEE PROFILE

Peter Simpson
Imperial College London
30 PUBLICATIONS **1,266** CITATIONS
SEE PROFILE

Lorand Szabo
Intrexon Corporation
13 PUBLICATIONS **525** CITATIONS
SEE PROFILE

Mike P Williamson
The University of Sheffield
215 PUBLICATIONS **9,886** CITATIONS
SEE PROFILE

Solution Structure of the CBM10 Cellulose Binding Module from *Pseudomonas* Xylanase A[†,‡]

S. Raghothama,[§,||] Peter J. Simpson,[§,||] Lóránd Szabó,[⊥,#] Tibor Nagy,[⊥] Harry J. Gilbert,*[,⊥] and Michael P. Williamson[§]

Department of Molecular Biology and Biotechnology, Krebs Institute, University of Sheffield, Sheffield S10 2TN, U.K., Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, U.K., Department of Biotechnology and Microbial Genetics, Gödöllô University of Agricultural Sciences, Páter Károly u. l., Gödöllô H-2101, Hungary

Received September 17, 1999; Revised Manuscript Received November 18, 1999

ABSTRACT: Plant cell wall hydrolases generally have a modular structure consisting of a catalytic domain linked to one or more noncatalytic carbohydrate-binding modules (CBMs), whose common function is to attach the enzyme to the polymeric substrate. Xylanase A from *Pseudomonas fluorescens* subsp. *cellulosa* (Pf Xyn10A) consists of a family 10 catalytic domain, an N-terminal family IIa cellulose-binding module, and an internal family 10 cellulose-binding module. The structure of the 45-residue family 10 CBM has been determined in solution using NMR. It consists of two antiparallel β-sheets, one with two strands and one with three, with a short α-helix across one face of the three-stranded sheet. There is a high density of aromatic residues on one side of the protein, including three aromatic residues (Tyr8, Trp22, and Trp24), which are exposed and form a flat surface on one face, in a classical polysaccharide-binding arrangement. The fold is closely similar to that of the oligonucleotide/oligosaccharide-binding (OB) fold, but appears to have arisen by convergent evolution, because there is no sequence similarity, and the presumed binding sites are on different faces.

Plant cell wall hydrolases from aerobic microorganisms generally have a modular structure consisting of a catalytic domain linked to one or more noncatalytic domains, whose function is to attach the enzyme to the polymeric substrate and thus to increase the catalytic activity (*1*). These domains are known as carbohydrate-binding modules or CBMs.[1] Deletion of the noncatalytic domains has no effect on the activity against soluble polysaccharides, but reduces activity against complex and insoluble substrates containing crystalline cellulose (*2*). Most CBMs bind to cellulose and have therefore often been described previously as cellulose-binding domains or CBDs (*3*). Although there are domains that bind specifically to xylan (denoted xylan-binding modules or XBMs) (*4*), many xylanases contain domains that bind specifically to cellulose rather than xylan (*5*). It has been suggested that the reason for this is that cellulose is a uniform, crystalline, and ubiquitous ligand, whereas xylan can be very variable depending on the species or differentiation state of the plant. In the plant cell wall, xylan is always closely associated with cellulose, and therefore, domains that bind to cellulose will also locate the enzyme close to xylan. Catalytic domains and CBMs have been separately classified into families, based on sequence similarities (*1*).

Structures have been determined for CBMs from families 1, 2a, 3, 4, and 5 and for an XBM from family 2b (*6*−*12*). Most of these domains bind to crystalline cellulose and have a common structure consisting of a β-sheet core, with two or, more usually, three aromatic residues across one face. These residues are highly exposed and form a colinear and coplanar surface, which is optimally constructed for face-to-face stacking with the coplanar surfaces of glucose residues in crystalline cellulose. The aromatic rings are thought to interact in most cases with every other glucose residue (i.e., residues i, $i + 2$, and $i + 4$) on a single cellulose strand. The family 4 and 2b domains bind to amorphous cellulose and xylan, respectively, and have their aromatic residues in a groove. In the case of the XBM2b, the aromatic rings form a twisted binding site, suitably oriented for interaction with the helical structure of xylan.

The aerobic bacterium *Pseudomonas fluorescens* subsp. *cellulosa* expresses a complex mixture of cell wall hydrolases, one of which is xylanase A (*13*). This enzyme contains a family 10 catalytic domain and is therefore known formally as Pf Xyn10A. The catalytic domain is at the C-terminus and is preceded by an *N*-terminal family 2a CBM followed

[†] Supported by BBSRC Grant 13/P09341 and a CASE award, The British Council and by equipment grants from the BBSRC and the Wellcome Trust. The Krebs Institute is a Centre for Molecular Recognition of the BBSRC/EPSRC.

[‡] Coordinates for the structure described in this paper have been deposited with the Brookhaven Protein Data Bank (accession reference 1QLD).

* To whom correspondence should be addressed. Fax: 0191-2228684. Phone: 0191-2226962. E-mail: H.J.Gilbert@Newcastle.ac.uk.

[§] University of Sheffield.

[||] These two authors made an equal contribution to this work.

[⊥] University of Newcastle upon Tyne.

[#] Gödöllô University of Agricultural Sciences.

[1] Abbreviations: CBM, carbohydrate-binding module; CBM10, internal family 10 CBM from *Pseudomonas fluorescens*; XBM, xylan-binding module; OB, oligonucleotide/oligosaccharide binding; NOE, nuclear Overhauser effect.

10.1021/bi992163+ CCC: $19.00
Published on Web 01/14/2000

```
     1        .         .         .         .
XynA GNQQCNWYGTLYPLCVTTTNGWGWEDQRSCIARSTCAAQPAPFGIVGS
XynE GNCQCNWWGTFYPLCQTQTSGWGWENSRSCISTSTCNSQGTGGGGVVC
CelE SGQQCNWYGTLYPLCSTTTNGWGWENNASCIARATCSGQPAPWGIVGG
CelC GGGQCNWYGTLYPLCVSTTSGWGYENNRSCISPSTCSAQPAPYGIVGG
EglB GAQACNWYGTLTPLCNNTSNGWGYEDGRSCVARTTCSAQPAPYGIVST
EglA GGLRCNWYGTLYPLCVTTQSGWGWENSQSCISASTCSAQPAPYGIVGA
XycA NACQCNWWGTRYPLCTNTASGWGWENNTSCITTSTCNSQGAGGGGVVC
         *** **  ***     *** *   **    **  *    * *
```

FIGURE 1: Sequence alignment of family X CBMs. Completely conserved residues are indicated by an asterisk at the bottom of the alignment, and the postulated binding aromatic residues are listed in bold. The proteins are: XynA, *Ps. cellulosa* XynA (this work), GenBank X15429; XynE, *Ps. cellulosa* XynE, GenBank Z48927; CelE, *Ps. cellulosa* CelE, GenBank X86798; CelC, *Ps. cellulosa* CelC, GenBank X61299; EglB, *Ps. cellulosa* EglB, GenBank X52615; EglA, *Ps. cellulosa* EglA, GenBank X12570; XycA, *Cellvibrio mixtus* XynA, GenBank Z48925. The first residue of CBM10 is residue 1, and there is a dot marking each tenth subsequent residue.

by a family 10 CBM (*5*, *14*). The three domains are separated by serine-rich linkers of approximately 50 and 32 residues, respectively. Both CBMs have been shown to be functional cellulose-binding domains, the affinity of the family 10 domain being approximately six times less than that of the family 2a domain (*15*). Family 10 domains are small (roughly 45 residues), and contain at least four cysteine residues. There is 40% sequence identitiy among the domains identified to date (Figure 1). Currently there is no structure of a family 10 CBM. We have therefore determined the solution structure of the domain from Pf Xyn10A [CBM10], which is shown to contain the typical three exposed aromatic residues. We show that the domain has a very similar structure to the oligonucleotide/oligosaccharide binding (OB) fold. However, there is no sequence similarity with other OB-fold proteins, the topology of one strand is different, the ligand-binding sites are on different faces, and it is concluded that the structural similarity is a consequence of convergent evolution.

MATERIALS AND METHODS

Expression and Purification. CBM10 was expressed in *Escherichia coli* strain BL21(DE3):pLysS (Novagen), using a recombinant pET expression vector containing the appropriate region of the Pf Xyn10A gene. The details of plasmid construction were described previously (*15*). The complete expressed sequence is hhhhhhhhhhssghiagrhmGNQQCNWYGTLYPLCVTTTNGWGWEDQRSCIARSTCAAQPAPFGIVGSGhhhhhh, where the lowercase letters represent the extra residues added in the construct. The first uppercase residue represents Gly120 of the native enzyme and is designated as Gly1 in this work. The *E. coli* strain was grown to mid-exponential phase ($A_{550} = 0.5$) in Luria broth (LB) supplemented with 50 μg/mL ampicillin at 30 °C, induced with 1 mM isopropylthio-β-D-galactopyranoside, and incubated for 4 h at 30 °C. The protein was purified by nickel ion affinity chromatography on a Talon column as described previously (*16*). Approximately two mg of purified protein was obtained per liter of culture.

NMR Studies. Samples for NMR were typically 1 mM CBM10 in 50 mM aqueous sodium phosphate buffer, pH 4.5, in 90% H_2O/10% D_2O. Spectra were obtained at 30, 40, and 50 °C on Bruker DRX500 and DRX 600 spectrometers using a 5 mm probehead equipped with gradient coils. Typical 90° pulse lengths used were 9 and 25 μs for high-power pulses and TOCSY spin locking, respectively. The solvent signal was suppressed by low-power presaturation. ^{1}H chemical shifts were referenced to internal 3-trimethylsilyl-2,2,3,3-d_4-propionate. The homonuclear 2D experiments were generally recorded with 512 t_1 increments of 4096 complex data points over an oversampled spectral width in F_2 of 25 ppm. The time domain data were apodized with Gaussian or phase-shifted sine bell windows. Data were zero filled to 2K in the indirect dimension prior to Fourier transformation. A total of 1024 data points was discarded on either side of the frequency domain in F_2 to yield a final matrix size of 2K × 2K data points. For DQF-COSY, 1024 t_1 increments were acquired over a 13.5 ppm spectral width and the final frequency domain data size was 8K × 2K. For E·COSY, 874 t_1 increments were used and the final data size was 8K × 4K. Quadrature detection in the indirect dimension was achieved using States-TPPI. For amide-exchange experiments, the protein sample in H_2O was lyophilized and redissolved in D_2O just before recording a series of TOCSY spectra at 50 °C. In total, seven TOCSY spectra were collected over a period of 30 h using 400 t_1 increments for each spectrum. Amide signals still observed in the spectra after this time were classified as slowly exchanging. All NMR spectra were processed and visualized using Felix 97 (Molecular Simulations Inc.).

Distance restraints were derived from NOESY spectra recorded using a mixing time of 100 ms. Upper distance bounds were calibrated by counting the number of contour levels in the spectrum, to give a different distance bound for each contour level (up to a maximum of nine contours), calibrated by distances in known secondary structures. Lower bounds were set to 1.9 Å in all cases. The highest upper bound corresponding to one contour was set at 5.3 Å. NOEs to groups of protons were applied using the sum averaging approach, with no further correction made for methyl groups (*17*). NOEs that could not be assigned to a specific proton because of chemical shift degeneracy were handled using ambiguous NOE restraints (*18*).

$^3J_{HN\alpha}$ couplings were measured from a DQF-COSY spectrum acquired with high resolution and were used to impose the following restraints on φ: $^3J < 5.5$ Hz, −90 to −20°; $7.0 \leq {}^3J < 8$ Hz, −60 to −180°; $8 \leq {}^3J < 9.0$, −70 to −170°; and $^3J \geq 9$, −80 to −160°. For some residues for which $^3J_{HN\alpha}$ could not be measured, φ could be limited to negative values in the refinement stages of structure calculations based on NOE intensities. $^3J_{\alpha\beta}$ couplings were measured from an E·COSY spectrum and were used to restrain the side chain to within ±30−50° (depending on the value of J) of the appropriate staggered rotamer. Additionally, where limited or no scalar coupling data were available, yet the ensemble of calculated structures converged consistently on one rotamer, it was constrained to that rotamer in the final calculation. Together with NOESY spectra, these were used to make stereospecific assignments. Hydrogen bond restraints were introduced where the backbone amide proton was slowly exchanging and structure calculations made in the absence of hydrogen bond restraints consistently produced a short N−O distance. Two restraints of $d_{HN-O} = 1.8-2.5$ Å and $d_{N-O} = 2.5-3.3$ Å were used. The disulfide restraints were included as part of the covalent geometry. A full list

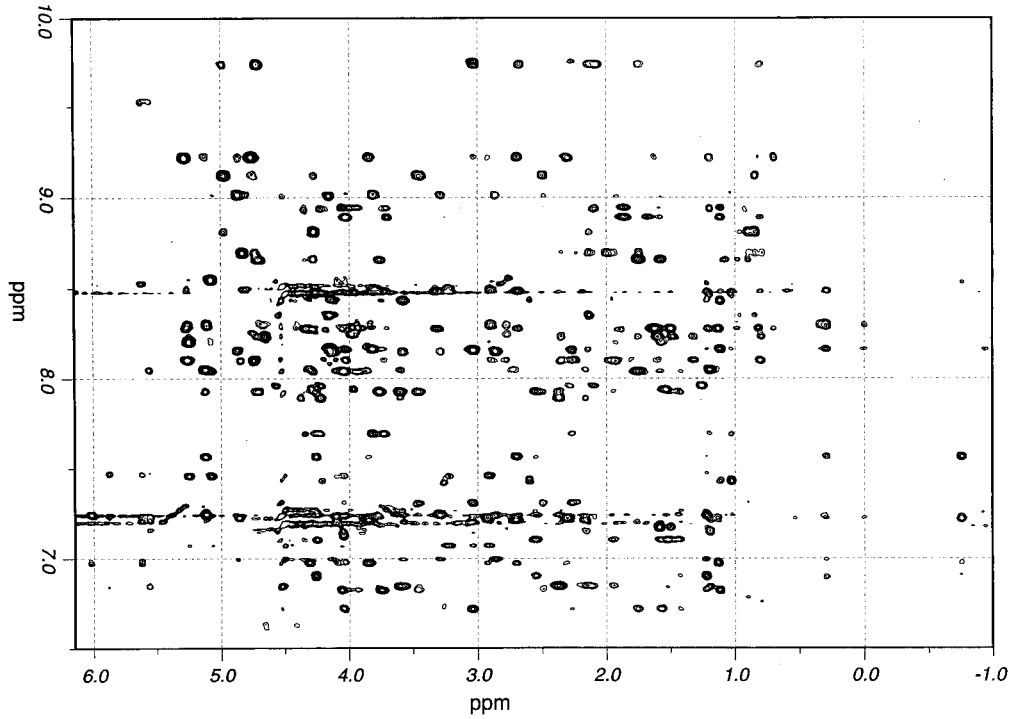

FIGURE 2: NOESY spectrum of the fingerprint region of CBM10, showing the two stripes from the His ring protons.

of restraints can be found in the Protein Data Bank under code 1CT7 or 1QLD.

All structure calculations were made using X-PLOR. Protocols were based on the example scripts in the X-PLOR manual (*19*). The only significant changes from the manual are that the initial temperature was 1500 K rather than 1000 and that 5000 annealing steps were used rather than 2000.

RESULTS

Structure Determination. Initial attempts to express the protein in a soluble form, with a His tag at either the N- or C-terminus, were unsuccessful. However, a soluble form of the protein could be produced by using a His tag at both ends of the CBM. The protein therefore contains 10 histidines at the N-terminus (connected by a 10-residue linker) and six at the C-terminus (tagged directly to the CBM10 sequence). The large number of histidine residues in the protein results in two intense NMR signals from the two ring protons, at approximately 8.5 and 7.2 ppm. These signals produce undesirable t_1 stripes in 2D spectra (Figure 2). To some extent, the problem could be alleviated by running spectra at different temperatures, which causes the amide proton signals to be shifted but not the histidine ring protons. However, the signals from the histidine protons still hide some useful 2D cross-peaks. Therefore, because of the known affinity of polyhistidine for metal ions such as Ni^{2+}, a series of experiments was carried out adding small amounts of paramagnetic metal ions, which might complex preferentially to the histidine rings and, therefore, selectively broaden the histidine ring proton signals. These experiments had some limited success. The addition of Co^{3+}, Cr^{3+}, and Ni^{2+} salts gave little selective broadening, but addition of Cu^{2+} selectively broadened the histidine signals, up to about 1 equiv (Figure 3). However, there was some broadening of other resonances, including the solvent, giving difficulties in achieving a flat baseplane in 2D spectra. Therefore, some samples contained 1 equiv of Cu^{2+}, to resolve signals hidden by the t_1 stripes.

At low temperatures, some signals were broad and caused problems for assignment and structure calculation, possibly due to slow conformational exchange processes. These signals sharpened as the temperature was raised. Most studies were therefore carried out at 50 °C, at which temperature the spectra are sharp and consistent with a single monomeric species. The signal broadening was not affected by dilution of the sample, and therefore is not due to an aggregation phenomenon.

Assignment of the NMR spectrum is essentially complete at 50 °C. The only unassigned signals come from a small number of side-chain amide protons. In addition, signals from the His tags and adjacent linkers were highly overlapped and could not be individually resolved. A table of the assignments has been deposited with BioMagResBank (accession number 4377). There are some signals that are remarkably upfield shifted, including those from Ile31 side-chain protons at 0.3 to −1.0 ppm, Tyr8 Hα at 1.20 ppm, and aromatic protons from Tyr8 and Trp22 at 5.6−5.9 ppm. [The number of the residue refers to its position in CBM10 as given in the alignment of Figure 1: residue 1 corresponds to residue 120 of full-length Pf Xyn10A (*14*).]

The structure of CBM10 was determined using a hybrid distance geometry/simulated annealing protocol in X-PLOR following methods similar to those used by Sorimachi et al. (*20*) and Simpson et al. (*12*). Initially, only about 250 unambiguous nuclear Overhauser effects (NOEs) were used as input, together with a small number of dihedral restraints. Further information was added iteratively during the calculation. Initially, the disulfide-bonding pattern of the four cysteine residues was not known, but preliminary structures showed that the only feasible linkage was from Cys5 to Cys36 and from Cys15 to Cys30. These constraints were therefore added as covalent bonds. Tests using the other

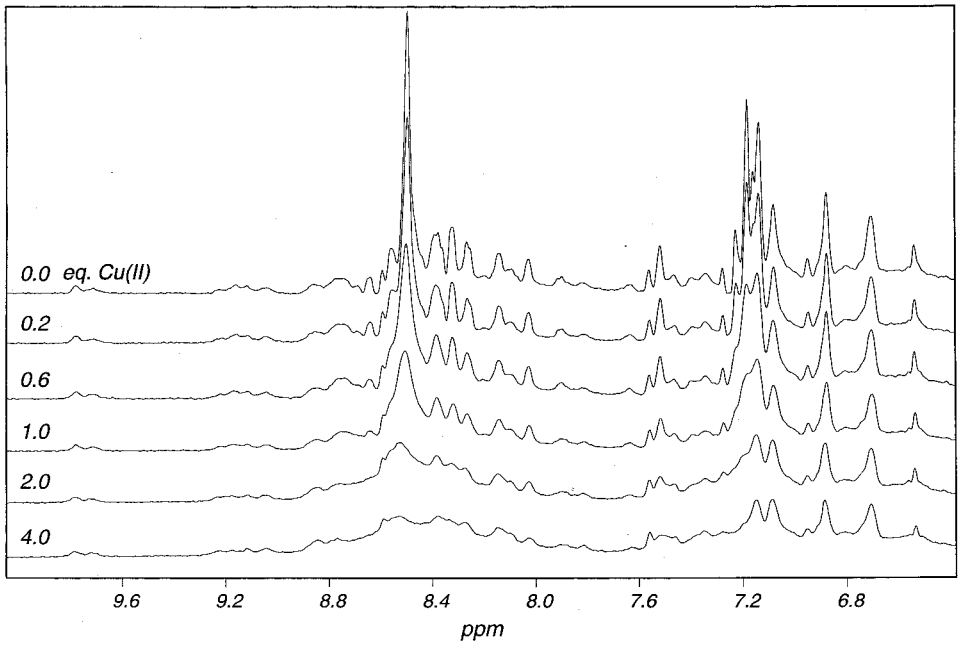

FIGURE 3: Stack plot showing the effect on the NMR spectrum of CBM10 of addition of Cu^{2+} ions.

possible linkages resulted in structures with very high NOE restraint violations.

Some hydrogen bond restraints were used early in the calculation, where the existence of hydrogen bonds across the β-strands was unambiguously evidenced by slow amide proton exchange and cross-strand NOEs. Later in the calculation, these restraints were removed. In subsequent calculations, a large number of close NH···O═C distances was reproducibly generated. Where there was experimental evidence that these short distances corresponded to hydrogen bonds, either by slow amide exchange or by low amide proton temperature coefficients or usually both (*21*), they were added as hydrogen bond restraints in the final round of calculations.The final set of restraints contained 666 unambiguous and 16 ambiguous NOEs, 39 φ and 30 χ_1 restraints, and 18 pairs of hydrogen bond restraints. Stereospecific assignments were made for 11 out of 29 β-methylene pairs and both valine γ-methyl pairs. The φ restraints included three positive angles (for Tyr8, Asp26, and Gln27), indicated by strong intraresidue and sequential $H\alpha$-NH_i NOEs (*22*). Pro42 was constrained as a *cis*-proline, based on a sequential NOE between $Ala41_{\alpha H}$ and $Pro42_{\alpha H}$, and the lack of sequential NOEs to the $Pro42_{\delta H}$ (*23*). All other prolines were trans. The distribution of NOE restraints is shown in Figure 4.

In the final calculation, 50 structures were calculated starting from random coordinates and were subsequently refined. The resulting structures were ranked by total X-PLOR energy and NOE violations. The first 21 models had a very similar energy and had no more than two NOE violations per structure of greater than 0.5 Å and were selected to represent the solution structure of CBM10. In this ensemble, the largest NOE violation was of 0.67 Å, and the largest angle violation was of 6°. The average structure was calculated from these 21 structures (by selecting the lowest energy structure and superimposing the remaining 20 on this one) and was subsequently subjected to restrained energy minimization to yield the minimized average structure. A randomly selected set of five structures and the average structure have been deposited with the Protein Data Bank (PDB codes 1CT7 and 1QLD, respectively).

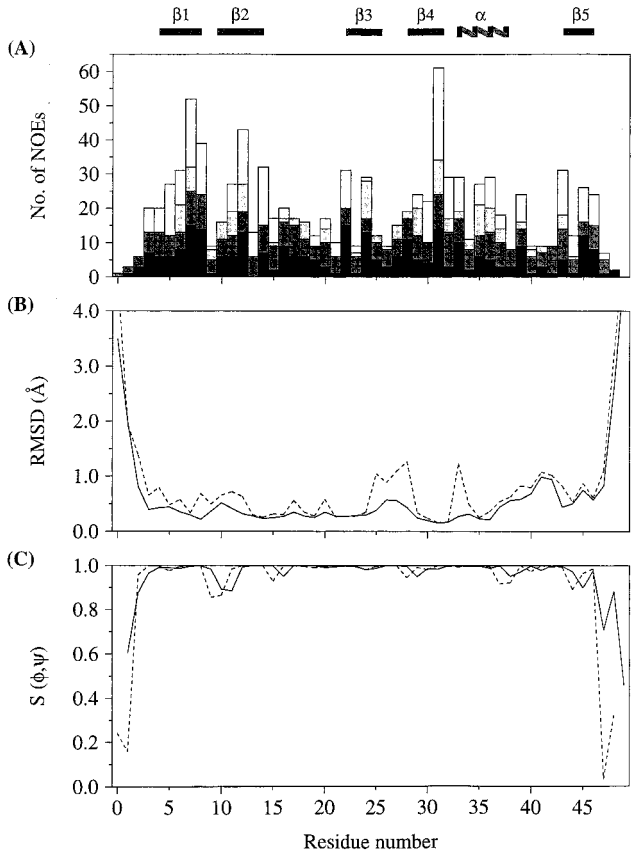

FIGURE 4: Structural parameters for CBM10. (a) Distribution of NOE restraints by residue. From bottom to top, in decreasing blackness, intraresidue, sequential, short-range ($|i - j| \leq 4$), and long-range. (b) Root-mean-square differences from mean structure for backbone atoms (solid) and all heavy atoms (dashed). (c) Angular order parameters of backbone φ (solid) and ψ (dashed) angles. The bars across the top of the Figure represent the locations of β-sheets, and the striped box the α-helix.

FIGURE 5: Backbone structure of CBM10, shown as a stereo diagram of the 21 selected structures. The superposition used the backbone atoms (N, Cα, C′) of residues in regular secondary structure elements (residues 3−14, 22−25, 28−37, 43−46).

Table 1: Structural Statistics for the Family of 21 Calculated Structures of CBM10 and the Minimized Averaged Structure

	⟨CBM10⟩[a]	$CBM10_{av-min}$[a]
rmsd from experimental restraints		
distance restraints, 682 (Å)[b]	0.067 ± 0.0021	0.059
dihedral restraints, 69 (deg)[c]	1.39 ± 0.56	1.01
rmsd from idealized covalent geometry		
bonds (Å)	0.005 ± 0.0	0.005
angles (deg)	0.90 ± 0.012	0.81
impropers (deg)	0.72 ± 0.035	0.62
X-PLOR energies (kcal mol^{-1})		
E_{total}	422.33 ± 15.5	343.8
E_{repel}	43.5 ± 8.1	46.5
E_{NOE}	153.8 ± 9.3	119.9
E_{cdih}	9.3 ± 7.4	4.3
E_{bond}	22.1 ± 2.0	16.7
E_{angle}	161.0 ± 4.3	132.6
$E_{improper}$	32.5 ± 3.2	23.7
Ramachandran analysis		
most favored region (%)	62.9	56.4
additionally allowed regions (%)	34.6	41.0
generously allowed regions (%)	2.6	2.6
disallowed regions (%)	0.0	0.0
precision analysis: rmsd to mean structure		
backbone atoms, secondary structure (residues 3−7, 10−13,22−25, 28−31, 33−37 and 43−46)	0.33 ± 0.08 Å	
all heavy atoms, secondary structure	0.78 ± 0.11 Å	
backbone atoms, residues 2−46	0.41 ± 0.15 Å	
all heavy atoms, residues 2−46	0.80 ± 0.12 Å	

[a] ⟨CBM10 ⟩, ensemble of 21 structures; $CBM10_{av-min}$, minimized average structure. [b] A total of 277 intraresidue, 143 sequential, 65 medium-range, 145 long-range, 16 ambiguous NOEs and 18 pairs of H-bond restraints. [c] 39 ϕ and 30 χ_1 dihedral restraints.

Structural statistics for the ensemble are given in Table 1, and the ensemble is shown in Figure 5.

Description of the Structure. The structure of CBM10 consists of five β-strands, organized as two antiparallel sheets, one of three strands (residues 3−7, 10−14, and 43−46) and one of two (residues 22−25 and 28−31), roughly perpendicular to the first sheet (Figure 6). The structure also contains a short stretch of α-helix, from residues 33−37.

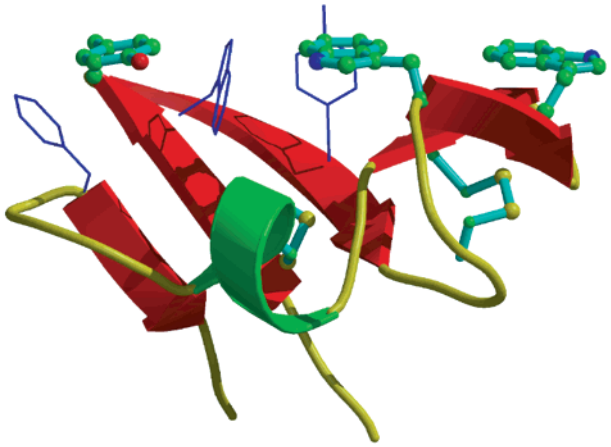

FIGURE 6: Secondary structure of CBM10, shown using MOLSCRIPT (*37*). The three coplanar aromatic rings are shown in ball-and-stick representation, and the other aromatic rings in the protein are shown as bonds. The two disulfide bridges are also shown.

Table 2: Hydrogen Bonds Present in the NMR Ensemble

donor	acceptor	donor	acceptor
Gln4	Val46	Glu25	Arg28
Cys5	Tyr12	Arg28	Glu25
Asn6	Gly44	Cys30	Gly23
Trp7	Thr10	Ile31	Pro13
Thr10	Trp7	Ala32	Gly21
Tyr12	Cys5	Cys36	Ala32
Leu14	Gln3	Ala37	Arg33
Cys15	Ile31	Gly44	Asn6
Gly23	Cys30	Val46	Gln4

Residues 17−20 form a helical (Type I/III) turn. The backbone hydrogen bonds identified in the protein are listed in Table 2. These hydrogen bonds are limited to those expected from the secondary structure, except for Cys15−Ile31 and Ile31−Pro13. Experimental data suggest that Arg33 HN is also hydrogen bonded. In almost all structures, it is within hydrogen-bonding distance of Thr19 Oγ, and it is likely that this hydrogen bond is formed in solution, although it was not included in the restraint list. Most of the structure is well-defined (Figure 4), except for the two termini and the loop between the helix and strand 5, which is facing away

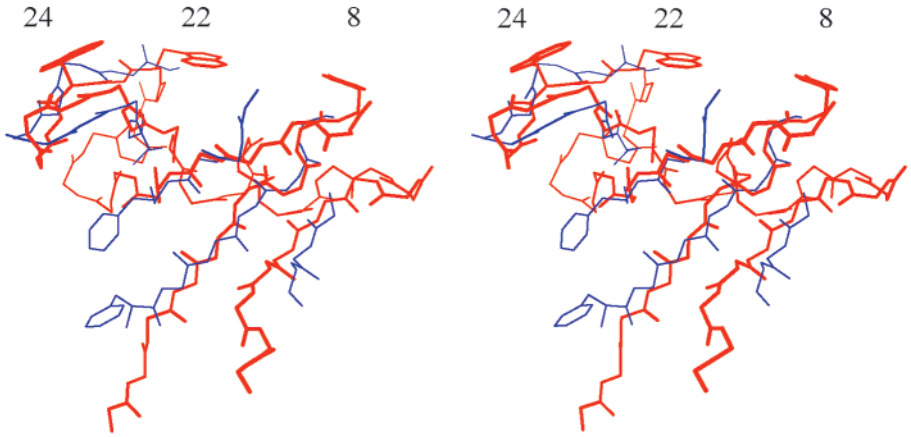

FIGURE 7: Overlay of CBM10 with the *B. subtilis* cold-shock protein, showing backbone atoms as a stereoview. CBM10 is shown in thicker lines, and the liganding aromatic rings are indicated. For clarity, only the structurally similar regions of the cold-shock protein are indicated, together with the two phenylalanine residues that form the binding site, which are on the lower left of the figure.

from the cellulose-binding site (Figure 6). The structured part of the domain therefore consists of, approximately the 45 residues, residues 2−46. The protein is stabilized by two disulfide bridges and is essentially unchanged at temperatures up to at least 50 °C, as judged by NMR spectra.

The protein has a hydrophobic core, consisting of residues Trp7, Leu14, and Ile31. These residues are well-defined in the NMR ensemble. The protein has an unusual distribution of aromatic amino acids, with all the aromatics at one side of the protein (Figure 6). The close contacts made by aromatic residues mean that some NMR signals are strongly shifted from their random coil positions. In particular, the resonances from Ile31 side chain, Tyr8 Hα side chain, and Trp22 side chain are shifted markedly upfield. These shifts are readily explained by the close proximity of aromatic rings in the structure and allow a high degree of confidence in the orientation of the aromatic rings.

The quality of the structure has been further evaluated using PROCHECK-NMR, which suggests that the quality is good (*24*). An analysis of the φ/ψ distribution (Table 1) shows that 2.6% of residues are in the generously allowed region. However, these represent exclusively Tyr8, which has $\varphi \approx 32°$, $\psi \approx 28°$. This residue is one of the three aromatic residues forming the presumed cellulose-binding site and occupies a loop between β-strands 1 and 2 (Figure 6). Its unusual conformation is therefore presumably related to its functional importance.

DISCUSSION

The structure of CBM10 consists largely of β-sheet, in common with those of the other CBMs determined to date. As expected for a domain that binds crystalline cellulose, the domain has three aromatic residues (Tyr8, Trp 22, and Trp24) that are exposed and approximately coplanar and well placed to interact with cellulose. In the following paper in this issue (*25*), it is demonstrated that these residues are indeed the residues that bind to cellulose. A comparison of this domain with that of the other CBMs known shows no similarity in three-dimensional structure other than the general features given above.

A search for similar protein folds was conducted using the program DALI (*26*), which showed that the structure has high similarity with the major cold-shock protein from *Bacillus subtilis* (*27*, *28*). This is an interesting result, because this protein is a member of the oligonuceotide/oligosaccharide binding (OB) fold (*29*), a family that includes staphylococcal nuclease (*30*) and a number of toxins (*31*−*33*) and is found in both eukaryotes and prokaryotes. The OB fold has a five-stranded β-sheet coiled to form a closed β-barrel, capped by an α-helix between the third and fourth strands. All OB fold proteins have the ligand-binding site in the same region, at the loops between strands 1 and 2, strands 4 and 5, and strand 4 and the helix (*29*). A comparison of secondary structure elements showed that the β-sheet residues overlay with an rmsd of 1.77 Å (i.e. residues 18−23, 28−31, 55−64, and 8−10 of the PDB structure of the cold-shock protein 1mjc with residues 2−7, 10−13, 22−31, and 44−46 of CBM10) (Figure 7). This is a low rmsd and would normally indicate homologous proteins. However, there is no sequence similarity between the two proteins (including disulfides and key aromatics), and the binding sites of the proteins are on different faces. Moreover, the C-terminal β-strand of CBM10 corresponds structurally to the N-terminal strand of the cold-shock protein. We therefore suggest that despite the strong structural similarity and common function, these two proteins are not evolutionarily related, and the structural similairity is due to convergent evolution.

In light of the structure of CBM10, it is instructive to compare the sequences of family 10 CBMs (Figure 1). Of the completely conserved residues, the four cysteines are structurally important and Trp22 is the central cellulose-binding residue, as shown in the following paper in this issue (*25*). Gly9, Gly21, and Gly44 all have positive φ angles and would therefore be energetically costly to replace with any other residue. Trp7 and Leu14 form part of the hydrophobic core. Surprisingly, the *cis*-Pro, Pro42, is not completely conserved, although when it is replaced, it is replaced by a glycine. The most interesting conserved residues are Thr10, Tyr12, Ser29, and Gln39. These residues (together with the almost completely conserved residues Gln27 and Arg28) are all on the surface of the protein, close to the three aromatic residues that bind to cellulose. CBMs generally interact with their ligands via hydrophobic stacking of the exposed aromatic rings, together with some specific hydrogen bonds (*9*). Thr10, Tyr12, Ser29, and Gln39 would be good candidates for residues that hydrogen bond specifically to

the substrate. Trp7 Nϵ points directly toward the substrate and could also be involved in hydrogen bonding, even though the rest of the residue forms part of the hydrophobic core.

Two of the family 10 CBMs (XynE and *Cellvibrio mixtus* XynA) have two extra cysteines at the N- and C-termini. From the structure, it is clear that cysteines at these positions could form a disulfide bridge with no conformational strain.

Interestingly, both Trp24 and Tyr8 can be replaced by other aromatics (Tyr and Trp respectively), despite their likely involvement in cellulose binding. A similar substitution occurs in the family 1 CBMs, where Trp in *Trichoderma reesei* endoglucanase EGI is replaced by Tyr in *T. reesei* cellobiohydrolase CBH1, with a loss in binding affinity (*34*). It is relevant that in Pf Xyn10A the family 2a CBM binds roughly six times more strongly to crystalline cellulose than does the family 10 CBM and that deletion of the CBM10 results in no measurable change in overall affinity (*15*). In other cellulases, it has been shown that the catalytic activity is proportional to the binding affinity, providing an obvious rationale for the presence of CBMs (*15*, *35*, *36*). Clearly, the prime function of CBM10 is not merely to attach the enzyme close to cellulose, because this is done more efficiently by the family 2 CBM. It is therefore likely that CBM10 has a function either in binding heterogeneous substrates or in assisting the catalysis indirectly (*15*).

REFERENCES

1. Tomme, P., Warren, R. A. J., and Gilkes, N. R. (1995) *Adv. Microbiol. Physiol. 37*, 1−81.
2. Gilkes, N. R., Warren, R. A. J., Miller, R. C., and Kilburn, D. G. (1988) *J. Biol. Chem. 263*, 10401−10407.
3. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., and Warren, R. A. J. (1991) *Microb. Rev. 55*, 303−315.
4. Black, G. W., Hazlewood, G. P., Millward-Sadler, S. J., Laurie, J. I., and Gilbert, H. J. (1995) *Biochem. J. 307*, 191−195.
5. Ferreira, L. M. A., Durrant, A. J., Hall, J., Hazlewood, G. P., and Gilbert, H. J. (1990) *Biochem. J. 269*, 261−264.
6. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) *Biochemistry 28*, 7241−7257.
7. Mattinen, M.-L., Linder, M., Drakenberg, T., and Annila, A. (1998) *Eur. J. Biochem. 256*, 279−286.
8. Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) *Biochemistry 34*, 6993−7009.
9. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) *EMBO J 15*, 5739−5751.
10. Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G., and McIntosh, L. P. (1996) *Biochemistry 35*, 14381−14394.
11. Brun, E., Moriaud, F., Gans, P. Blackledge, M. J., Barras, F., and Marion, D. (1997) *Biochemistry 36*, 16074−16086.
12. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) *Structure 7*, 853−864.
13. Hazlewood, G. P., and Gilbert, H. J. (1998) *Prog. Nucl. Acid Res. Mol. Biol. 61*, 211−241.
14. Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J., Gilbert, H. J. (1989) *Mol. Microbiol. 3*, 1211−1219.
15. Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., Hazlewood, G. P., and Gilbert, H. J. (1999) *Biochem J. 342*, 473−480.
16. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P., and Gilbert, H. J. (1998) *Biochem. J. 331*, 775−778.
17. Fletcher, C. M., Jones, D. N. M., Diamond, R., and Neuhaus, D. (1996) *J. Biomol. NMR 8*, 292−310.
18. Nilges, M. (1995) *J. Mol. Biol. 245*, 645−660.
19. Brünger, A. T. (1992) *X-PLOR version 3.1. A system for crystallography and NMR*, Yale University, New Haven, CT.
20. Sorimachi, K., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D. B., and Williamson, M. P. (1997) *Structure 5*, 647−661.
21. Baxter, N. J., and Williamson, M. P. (1997) *J. Biomol. NMR 9*, 359−369.
22. Ludvigsen, S., and Poulsen, F. M. (1992) *J. Biomol. NMR 2*, 227−233.
23. Wüthrich, K., Billeter, M., and Braun, W. (1984) *J. Mol. Biol. 180*, 715−740.
24. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
25. Ponyi, T., Szabó, L., Nagy, T., Orosz, L., Simpson, P. J., Williamson, M. P., and Gilbert, H. J. (1999) *Biochemistry 39*, 985−991.
26. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
27. Schindelin, H., Marahiel, M. A., and Heinemann, U. (1993) *Nature 364*, 164−168.
28. Schnuchel, A., Wiltscheck, R., Czisch, M., Herrier, M., Willimsky, G., Graumann, P., Marahiel, M. A., and Holak, T. A. (1993) *Nature 364*, 169−171.
29. Murzin, A. G., and Chothia, C. (1992) *Curr. Opin. Struct. Biol. 2*, 895−903.
30. Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Richardson, D. C., Richardson, J. S., and Yonat, A. (1971) *J. Biol. Chem. 246*, 2302−2316.
31. Sixma, T. K., Pronk, S. E., Kalk, K. H., Wartna, E. S., van Zanten, B. A. M., Witholt, B., and Hol., W. G. J. (1991) *Nature 351*, 371−377.
32. Sixma, T. K., Pronk, S. E., Kalk, K. H., van Zanten, B. A. M., Witholt, B., Berghuis, A. M., and Hol., W. G. J. (1992) *Nature 355*, 561−564.
33. Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L., and Read, R. J. (1992) *Nature 355*, 748−750.
34. Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T. T., and Pettersson, G. (1995) *FEBS Lett. 372*, 96−98.
35. van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R., and Pettersson, G. (1986) *FEBS Lett. 204*, 223−227.
36. Black, G. W., Rixon, J. E., Clarke, J. H., Hazlewood, G. P., Theoderou, M. K., Morris, P., and Gilbert, H. J. (1996) *Biochem. J. 319*, 515−520.
37. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

BI992163+