Enzyme systems that attack the plant cell wall contain noncatalytic carbohydrate-binding modules (CBMs) that mediate attachment to this composite structure and play a pivotal role in maximizing the hydrolytic process. Although xyloglucan, which includes a backbone of \( \beta-1,4 \)-glucan decorated primarily with xylose residues, is a key component of the plant cell wall, CBMs that bind to this polymer have not been identified. Here we showed that the C-terminal domain of the modular \textit{Clostridium thermocellum} enzyme CtCe1D-Ce1A4A (formerly known as CelI) comprises a novel CBM (designated CBM44) that binds with equal affinity to cellulose and xyloglucan. We also showed that accommodation of xyloglucan side chains is a general feature of CBMs that bind to single cellulose chains.

The crystal structures of CBM44 and the other CBM (CBM30) in \( \text{CtCe1D-Ce1A4A} \) display a \( \beta \)-sandwich fold. The concave face of both CBMs contains a hydrophobic platform comprising three tryptophan residues that can accommodate up to five glucose residues. The orientation of these aromatic residues is such that the bound ligand would adopt the twisted conformation displayed by cello-oligosaccharides in solution. Mutagenesis studies confirmed that the hydrophobic platform located on the concave face of both CBMs mediates ligand recognition. In contrast to other CBMs that bind to single polysaccharide chains, the polar residues in the binding cleft of CBM44 play only a minor role in ligand recognition. The mechanism by which these proteins are able to recognize and decorated \( \beta-1,4 \)-glucans is discussed based on the structures of CBM44 and the other CBMs that bind single cellulose chains.

The recycling of photosynthetically fixed carbon by the action of microbial plant cell wall hydrolases is a fundamental biological process that is integral to one of the major geochemical cycles and, in addition, has considerable industrial potential (1). The complex chemical and physical structure of the plant cell wall restricts enzyme access to the polysaccharides, primarily cellulose and hemicellulose, that comprise this composite macromolecule. Microbial cellulosomes and hemicellulases generally contain noncatalytic carbohydrate-binding modules (CBMs) that target these enzymes to specific sites within the plant cell wall (2). CBMs potentiate the hydrolytic activity of these biocatalysts by bringing the appended catalytic modules into intimate contact with their target substrates, thereby reducing the “accessibility problem” (2–4). Based on primary structure similarity, CBMs have been grouped into 43 sequence-based families (5).

Structural studies on representatives of the majority of CBM families demonstrate that these protein modules display a \( \beta \)-sandwich fold, which has led to the identification of a CBM superfamily (2). This fold consists of two \( \beta \)-sheets each consisting of 3–6 antiparallel \( \beta \)-strands. The topology of CBM ligand-binding sites, which may vary in proteins that are members of the same family, complements the conformation of the target polysaccharide. Thus, in type A modules, which interact with the flat surfaces of crystalline polysaccharides, the binding site consists of a planar hydrophobic platform that contains three exposed aromatic amino acids (6–8). These CBMs show no significant affinity for soluble polysaccharides, and the ligand specificity of CBM families that contain type A modules is usually invariant. In contrast, type B CBMs, which bind to single polysaccharide chains, accommodate the ligands within extended clefts of varying depths (9–12). In these CBMs, polysaccharide recognition is variable within different members of these families and normally reflects the catalytic specificity of their cognate catalytic modules. Indeed, type B CBM families display plasticity in their capacity to accommodate heterogeneity both in the branches that may decorate the sugar backbone and the composition and linkage of the sugar backbone. This variation in ligand recognition is exemplified in CBM family 6 (CBM6), which contains proteins that recognize xylan, cellulose (\( \beta-1,4 \)-linked glucose homopolymer), laminarin (\( \beta-1,3 \)-linked glucose homopolymer), and \( \beta-1,4 \)- and \( \beta-1,3 \)-mixed linked \( \beta \)-glucans such as lichenan (13–15). In addition, plasticity in ligand recognition may result from subtle molecular determinants in a single CBM-binding site. Thus, polar residues in CBM29 are able to hydrogen bond to the axial O-2 in mannose and the equatorial O-2 in glucose enabling the protein to bind cellulose, glucomannan, and mannan (16). Finally, studies on CBMs that recognize the branched hemicelluloses xylan and galactomannan indicate that the side chains of decorated polysaccharides are usually solvent-exposed and do not restrict ligand binding (13, 16–18) or repre-

---

*§ This work was supported by Grant POCT/BIA-PRO/59118/2004 from Fundação para a Ciência e Tecnologia (Portugal) and by the individual grants SFRH/BPD/20357/2004 (to S. N.), SFRH/BPD/16731/2004 (to C. G.), and SFRH/BPD/23784/2005 (to M. C.) from Fundação para a Ciência e Tecnologia (Portugal). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**1** The atomic coordinates and structure factors (codes 2c24, 2c26, and 2c4x) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

**2** The on-line version of this article (available at http://www.jbc.org) contains Table S1.

**3** To whom correspondence may be addressed. E-mail: D.N.Bolam@newcastle.ac.uk.

**4** To whom correspondence may be addressed. E-mail: cafontes@fmv.utl.pt.

**5** The abbreviations used are: CBM, carbohydrate-binding module; PKD, polycystic kidney disease; HEC, hydroxyethylcellulose; CMC, carboxymethylcellulose; AGE, affinity gel electrophoresis; ITC, isothermal titration calorimetry; BSA, bovine serum albumin; CAPS, 3-(cyclohexylammonium)propanesulfonic acid; GH, glycoside hydrolase; CBM4-1, C. fim CBM4-1; PeCBM29-2, P. equi CBM29-2; CbCBM11, C. thermocellum CBM11; CcCBM17, C. cellulovorans CBM17; BspCBM28, Bacillus spp. CBM28; SeMET, selenomethionine; MAD, multi-wavelength anomalous dispersion.
sent specificity determinants. It is believed that this feature of type B CBMs enables these proteins to interact with similar affinities to decorated and undecorated polysaccharides. The generality of this hypothesis, however, remains to be established particularly with respect to the key polysaccharides that contain a backbone of β-1,4-linked glucose polymers that can be undecorated, exemplified by cellulose, or contain extensive side chains such as xylglucan. Both these polysaccharides are abundant in plant cell walls, particularly in the primary cell wall of dicotyledons, where they provide tensile strength and thus undergo modification in rapidly expanding cells.

To investigate whether CBMs can display high affinity for heavily branched β-1,4-glucans such as xylglucan, and to interrogate the molecular determinants that govern this binding specificity, we have studied the structure and function of the N- and C-terminal domains of CtCel9D-Cel44A from Clostridium thermocellum, which is a component of the bacterium’s cellulosome (19). This enzyme contains internal glycoside hydrolase (GH) family 9 and 44 catalytic domains in addition to an N-terminal CBM30, a type I dockerin, and a C-terminal module of unknown function. Here we show that the C-terminal domain of CtCel9D-Cel44A comprises a novel CBM (designated CBM44) that displays high affinity for β-1,4-glucose polymers, including xylglucan. The three-dimensional structures of CBM30 and CBM44, combined with mutagenesis studies, reveal that both modules contain a single ligand cleft with a strip of three aromatic residues that are displayed to the edge of the respective ligand-binding site. The mechanism by which CBMs that recognize cellulose are also able to accommodate the side chains of xylglucan is discussed.

**EXPERIMENTAL PROCEDURES**

*Protein Expression and Purification—*CtCel9D-Cel44A is a modular enzyme containing an N-terminal CBM30, followed by two catalytic domains belonging to GH9 and GH44, a dockerin, and a C-terminal domain of unknown function (Fig. 1). The internal region of the C-terminal domain of CtCel9D-Cel44A displays homology with polysaccharide-binding domains (PKD) domains found in a variety of eukaryotic and prokaryotic proteins. DNA encoding CBM30, PKD-CBM44, CBM44, Cel44A, and Cel44A-CBM44 was amplified by PCR from *C. thermocellum* YS genomic DNA, using the primers listed in supplemental Table S1, and cloned into NheI- and XhoI-restricted pET21a to generate recombinant plasmids were sequenced to ensure that no mutations had occurred during the PCR that used the thermostable DNA polymerase _Pfu_ Turbo (Stratagene). The recombinant proteins, which contain a C-terminal His6 tag, were expressed in _Escherichia coli_ BL21 cells, harboring the appropriate recombinant plasmid, and cultured in LB containing 100 μg/ml ampicillin at 37 °C to mid-exponential phase (A_{500} 0.6). Isopropyl β-D-thiogalactoside was then added to a final concentration of 1 mM, and the cultures were incubated for a further 3–5 h. The His6-tagged recombinant proteins were purified by cell-free extracts by immobilized metal ion affinity chromatography as described previously (20). Preparation of _E. coli_ to generate seleno-methionine (SeMet) PKD-CBM44 was performed as described by Carvalho et al. (21), and the protein was purified by using the same procedures employed for the native CBMs, except that the buffer also contained 2 mM β-mercaptoethanol. For crystallization, CBM30 and PKD-CBM44 were further purified by size exclusion chromatography. Following immobilized metal ion affinity chromatography, the proteins were buffer-exchanged into 50 mM Hepes-HCl buffer, pH 7.5, containing 200 mM NaCl (Buffer A or Buffer A + 5 mM DTT for SeMet protein) and then subjected to gel filtration by using a HiLoad 16/60 Superdex 75 column (Amersham Biosciences) with protein eluted at 1 ml/min in Buffer A. Purified CBM30 and PKD-CBM44 were concentrated using an Amicon 10-kDa molecular mass centrifugal concentrator and washed three times with 5 mM dithiothreitol (for the SeMet proteins) or water (for the native proteins). _Cellulomonas_ fimi CBM4-1 (CCBM4-1), _C. thermo- cellulum_ CBM11 (CCBM11), _Clostridium cellulovorans_ CBM17 (CCBM17), _Bacillus_ spp. CBM28 (BspCBM28), and Pirnomyces equii CBM29-2 (PeCBM29-2) were all prepared as described previously (see Refs. 16 and 21–24, respectively).

*Source of Sugars Used*—All soluble polysaccharides were purchased from Megazyme International (Bray, County Wicklow, Ireland), except oat spelt xylan, laminarin, and HEC, which were obtained from Sigma, and pustullan, which was obtained from Calbiochem. Cello-oligosaccharides were from Seikagaku Corp. (Japan). Avicel (PH101) was obtained from Serva, and acid-swollen cellulose was prepared as described previously (25).

*Binding Assays*—The affinity of CBM30, PKD-CBM44, and CBM44 for a range of soluble polysaccharides was determined by affinity gel electrophoresis (AGE). The method was essentially as described by Tomme et al. (26) using the polysaccharide ligands at a concentration of 0.1% (w/v) unless stated otherwise. Electrophoresis was carried out for 4 h at room temperature in native polyacrylamide gels containing 10% (w/v) acrylamide. The nonbinding negative control protein was bovine serum albumin (BSA). Quantitative assessment of binding was carried out as described previously (27), using polysaccharide concentrations ranging from 0.001 to 0.5% (w/v). Qualitative assessment of PKD-CBM44 and CBM44 binding to Avicel and acid-swollen cellulose was carried out as follows: 30 μg of protein in 50 mM Tris–HCl buffer, pH 7.5, containing 0.05% (w/v) Tween 20 and 5 mM CaCl2 (Buffer A) were mixed with 1 mg of ligand in a final reaction volume of 200 μl. The reaction mixture was incubated for 2 h at room temperature with gentle shaking, after which time the insoluble ligand was precipitated by centrifugation at 13,000 × g for 5 min. The supernatant, comprising the unbound fraction, was removed, and the pellet was washed three times with 200 μl of Buffer A. The bound protein was eluted by boiling the polysaccharides in 200 μl of 10% (w/v) SDS containing 10% (v/v) β-mercaptoethanol for 10 min. Bound and unbound fractions were analyzed by SDS-PAGE using a 14% (w/v) acrylamide gel. Controls containing protein but no polysaccharide were performed in parallel to ensure that no precipitation of the CBM occurred during the experiment.

*Isothermal Titration Calorimetry (ITC)*—ITC measurements were made at 25 °C following standard procedures (28) using a Microcal Omega titration calorimeter. Proteins were dialyzed, extensively, against either 50 mM Hepes-HCl buffer, pH 8.0, or 50 mM sodium phosphate buffer, pH 7.0, and the ligand was dissolved in the same buffer to minimize the heat of dilution. During a titration experiment, the protein sample (40–250 μM), stirred at 300 rpm in a 1.4331-ml reaction cell was injected with a single 1-μl aliquot, followed by 29 successive 10-μl aliquots of ligand comprising 2–5 mg/ml polysaccharide or 0.5–5 mM oligosaccharide at 200-s intervals. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single site binding model (Microcal Origin, version 7.0). The molar concentration of CBM-binding sites present in the polysaccharide ligands was determined as described previously (29). The fitted data yield the association constant (K_a) and the enthalpy of binding (ΔH).

Other thermodynamic parameters were calculated using the standard thermodynamic Equation 1.

\[-RT\ln K_a = ΔG = ΔH − TΔS\]  
*(Eq. 1)*
**TABLE 1**

Data collection and structure statistics

| Data set                  | Native CBM30 | SeMet (peak)* | SeMet (edge)* | SeMet (higher edge energy)* | SeMet (high energy remote)* | Native PKD-CBM44 |
|---------------------------|--------------|---------------|---------------|----------------------------|----------------------------|------------------|
| Wavelength (Å)            | 1.072        | 0.97934       | 0.97936       | 0.97936                    | 0.94645                   | 1.2915           |
| Space group               | P2₁2₁2₁      | P4₁2₂       | P4₁2₂       | P4₁2₂                    | P₄2₂                    | P₄2₂             |
| Unit cell parameters      |              |              |              |                           |                           |                  |
| a (Å)                     | 66.6         | 87.10        | 86.87        | 86.92                    | 86.90                    | 87.47            |
| b (Å)                     | 85.5         | 87.10        | 86.87        | 86.92                    | 86.90                    | 87.47            |
| c (Å)                     | 88.9         | 107.67       | 108.31       | 108.39                   | 108.24                   | 103.19           |
| Resolution limits (Å)     | 50–2.27      | 67.7–2.8     | 67.7–2.5     | 67.7–2.10                 | 67.7–2.1                 | 66.7–2.17        |
| No. observations          | 152,855 (22,036) | 127,725 (17,834) | 265,692 (16,590) | 270,229 (14,500) | 188,468 (26,191) | 139,247 (12,180) |
| No. unique observations   | 24,116 (3,452)| 19,840 (2,884) | 22,678 (2,816) | 23,856 (2,729)       | 24,857 (3,557)       | 21,826 (3,121)   |
| Multiplicity              | 6.3 (6.4)    | 6.4 (6.2)    | 11.7 (5.9)   | 11.3 (5.3)               | 7.6 (7.4)                | 6.4 (3.9)        |
| Completeness (%)          | 99.9 (100)   | 99.8 (99.7)  | 97.6 (85.7)  | 95.7 (77.2)             | 99.9 (99.9)            | 99.9 (99.8)      |
| (I/σ(I))                  | 18.6 (5.9)   | 20.3 (2.6)   | 31.6 (2.2)   | 33.2 (2.4)              | 24.4 (3.0)              | 16.4 (2.1)       |
| Rmerge*                   | 7.6 (31.8)   | 9.1 (56.0)   | 4.9 (60.8)   | 4.5 (52.0)              | 5.3 (62.2)              | 6.6 (54.9)       |

| Structure refinement      | Native CBM30 | SeMet* (high energy remote) | Native PKD-CBM44 |
|---------------------------|--------------|----------------------------|------------------|
| Resolution range          | 20.0–2.27 (2.328–2.27) | 67.7–2.1 (2.155–2.1) | 66.7–2.17 (2.226–2.17) |
| Reflections used          | 21,633 (1,545) | 23,573 (1,729)   | 20,683 (1,483) |
| Rwork (%)                 | 21.6 (29.5)  | 17.4 (24.5)     | 16.9 (26.5)    |
| Rmerge (%)                | 26.5 (35.4)  | 21.0 (28.4)     | 21.6 (28.5)    |
| No. protein atoms         | 2778         | 1935           | 1935           |
| No. water molecules       | 218          | 381            | 380            |
| No. calcium ions          | 0            | 2              | 2              |
| No. ethylene glycol molecules | 0          | 3              | 12             |
| Root mean square deviation from ideal geometry | 0.007 | 0.009 | 0.010 |

* The SeMet data are for the PKD-CBM44 incorporated with seleno-methionine. The peak data were collected from a different crystal from the drop. It suffered radiation damage during edge data collection, so another crystal from the same drop was used. For the higher energy edge data collection, the beam attenuation was increased from 20 to 40%.

Site-directed Mutagenesis—Mutants of CBM30 and CBM44 were generated using the PCR-based QuickChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions, using pCG1 and pCG3 as the template DNA, respectively. The sequences of the primers used to generate these mutants are displayed in supplemental Table S1. The mutated DNA sequences were sequenced to ensure that only the appropriate mutations had been incorporated into the nucleic acid.

Enzyme Assays—The activity of truncated derivatives of Cel9D-Cel44A against various polysaccharides was determined as described previously (20) by measuring the rate of release of reducing sugars using the Somogyi-Nelson reagent (see Ref. 30). Assays were carried out at 50 °C in 50 mM sodium phosphate buffer, pH 5.0, containing 1 mg/ml BSA. To determine the pH profile of Cel44A, 50 mM sodium acetate, pH 4–6, sodium phosphate/citrate, pH 6–7.5, Tris-HCl, pH 7.5–8.5, and CAPS, pH 8.5–10, buffers were used in the enzyme assays. The linearity of the reactions was confirmed by measuring the release of reducing sugars at three time points. All reported results are the mean of three separate experiments.

Crystallization and Data Collection—Crystallization conditions for both PKD-CBM44 and CBM30 were screened using the Hampton crystal screen, crystal screen 2, and PEG/ison (Hampton Research, Alison Viejo, CA). Native or SeMet crystals of PKD-CBM44 were grown by vapor-phase diffusion using the hanging drop method with an equal volume (1 μl) of protein (50 mg/ml in water or 5 mM dithiothreitol, respectively) and reservoir solution (0.2 M CaCl₂, 0.1 M sodium acetate, pH 4.5, and 22.5% ethanol (v/v)). Crystals, which grew over a period of 1 week, were stabilized by adding 1 μl of cryoprotectant solutions containing 30% (v/v) ethylene glycol or glycerol in the crystallization buffer, stepwise over a few days for equilibration, and flash-frozen in liquid nitrogen. The detailed protocols used to obtain PKD-CBM44 crystals are described elsewhere (31). Initial PKD-CBM44 data sets were collected on a home source with CuKα x-ray radiation from an Enraf-Nonius rotating anode generator operated at 5 kilowatts, with a MAR research image-plate detector. Native (at λ = 1.2915 Å) and SeMet (at the selenium-edge for MAD) data sets were collected on the beamline ID14-EH4 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using a Quantum 4 charge-coupled device detector (ADSC) with the crystals cooled at 100 K using a cryostream (Oxford Cryosystems Ltd.). Crystals belong to the tetragonal space group. The Matthews coefficient (Vₘ/ₚ = 3.6 Å³/Da) indicated the presence of one molecule in the asymmetric unit and a solvent content of 65% (32). Crystals of CBM30 native protein were grown with an equal volume (1 μl) of protein (66 mg/ml in water) and reservoir solution (0.8 M sodium and potassium tartrate in 0.1 M NaHepes, pH 7.5). A CBM30 crystal, which grew over a period of 3–4 days, was soaked for a few seconds in a modified crystallization solution containing 30% (v/v) glycerol and flash-cooled in a nitrogen stream at 100 K for data collection. Preliminary crystal characterization was performed in-house, and the diffraction experiments showed that CBM30 crystals belonged to the primitive
Structure and Function of CBM30 and CBM44

FIGURE 1. Molecular architecture of CiCel9D-Cel44A and its truncated derivatives. The modules encoded by the defined recombinant plasmids are indicated. The gray and black boxes represent the signal peptide and the dockerin, respectively. Domain boundaries are specified below the modular representation of full-length CiCel9D-Cel44A.

orhtorhombic space group with unit cell parameters a = 66.6 Å, b = 85.5 Å, and c = 88.9 Å. A complete data set of a single crystal was collected at beamline ID23-1 at the ESRF, using a marCCD detector. At a fixed wavelength of 1.072 Å, the crystal diffracted beyond the 2.27-Å resolution. Systematic absence of the odd reflections indicates that the crystal belonged to the P212121 space group. A Matthews coefficient of 3 Å³/Da suggested the presence of two molecules in the asymmetric unit, with a solvent content of ~60% (32). All data sets were processed using the programs MOSFLM (33) and SCALA (34) from the CCP4 suite (35), and the statistics are shown in Table 1.

Phasing, Model Building, and Refinement—The location of four of the five selenium sites in PKD-CBM44, phasing, and density modification were performed using the SHELXD/E package (36). SHELXD gave a correlation coefficient of 40%, and SHELXE confirmed the space group to be unambiguously P4_2_2_2. The selenium sites were refined with SOLVE (37), and the resulting phases fed into RESOLVE to give a partial (~50%) structural model. Automated model building using ARP/wARP (38) and the RESOLVE partial structure and the high energy remote SeMet data generated a model with Rwork = 29.2%. This initial model consisted of 243 residues from a total of 260 amino acid residues. This model was used to fit into the 2.17-Å native data after rigid-body refinement in REFMAC5 (39), treating each domain as a rigid body, with Rwork = 42.9%. Iterative model building with O (40), together with restrained refinement in REFMAC5 (39), resulted in a final model with Rwork = 16.9% (Rfree = 21.6%). The four residues from the PKD-CBM44 N terminus (Met, Ala, Ser, and Val) were disordered and were not included in the final model. However, Leu250 and Gln251 in the C-terminal linker were defined, although the six residues of the His tag were not. This final model includes 250 amino acid residues, 2 calcium ions, 12 ethylene glycol molecules, and 380 water molecules. The final refined native model was used to complete the refinement of the 2.1-Å SeMet data, after rigid-body fitting, using O (40) and REFMAC5 (39). The final SeMet model included 250 amino acid residues, 2 calcium ions, 3 ethylene glycol molecules, and 381 water molecules. It is similar to the native form, with significant deviations (root mean square deviation of >0.4 Å) in two loop regions (residues 47–55 and 187–194).

In relation to CBM30, subsequent to the generation of protein crystals, a structural model of the CBM was deposited in the Protein Data Bank under the accession code 1wmx. Therefore, this structure served as a model for the Patterson search methods, using program MOLREP (41) from the CCP4 suite (35). Twenty cycles of rigid-body refinement of the molecular replacement solution, using program REFMAC5 (39), gave an overall Rwork = 36.3% and Rfree = 36.9% (for a free set of 10% of the reflections). Iterative cycles of model building and restrained refinement brought the Rwork and Rfree to the final values of 21.7 and 26.4%, respectively. The final model contains 346 amino acid residues, belonging to 2 polypeptide chains, and 218 water molecules. All amino acid residues are in the allowed regions of the Ramachandran plot. The refinement statistics are summarized in Table 1.

RESULTS AND DISCUSSION

A Novel CBM Family—A previous study showed that Cell from C. thermocellum (designated CiCel9D-Cel44A in this study to reflect the modern nomenclature of glycoside hydrolases (42)) contains GH9 and GH44 catalytic modules in addition to an internal dockerin that targets the enzyme to the clostridial cellulosome (19), a multienzyme extracellular cellulose-hemicellulose complex, an N-terminal family 30 CBM (CBM30), and a C-terminal domain of unknown function (Fig. 1). CBM30, which displays affinity for β-1,4-glucopolymers (43), plays a pivotal role in the function of GH9, a typical processive endoglucanase, whereas GH44 was assigned as displaying endo-xylanase activity (19). The first 90 amino acids of the 250-residue C-terminal region of CiCel9D-Cel44A comprises a PKD module (the nature of this module is discussed below), whereas the remaining 160 amino acids display no extensive sequence identity with other modules present in glycoside hydrolases. The data presented in this study show that this C-terminal region of CiCel9D-Cel44A binds to plant structural polysaccharides (see below) and is thus classified as a novel CBM (designated CBM44), representing the founder member of CBM family 44. Alignment of CBM44 with CBM families 4, 6, 9, 16, 17, 22, 27, 29, 33, 35, and 37 revealed the presence of three very short conserved motifs at the N and C termini of these domains (Fig. 2) which, in CBM44 (see below) and in several other CBM families, play a key role in the coordination of a structural calcium. This observation provides further support for the grouping of CBM families that display a β-sandwich fold into a super-
family (2), analogous to the clan classification of glycoside hydrolases. Although the superfamily reflects an ancestral link between these CBM families, it provides no significant functional or mechanistic information, as ligand specificity and even the identity of the ligand-binding residues can vary within a family.

Ligand Specificity of CBM30 and CBM44—CBM44 and PKD-CBM44 were purified to electrophoretic homogeneity, and their biochemical properties were evaluated. AGE revealed that both proteins displayed significant affinity for xyloglucan, barley glucan, lichenan, HEC, and konjac glucomannan, although exhibiting a significantly lower $K_a$ for oats spelt xylan (Fig. 3 and Table 2). No binding to laminarin, curdlan, carob galactomannan, potato galactan, pullulan, or pustulan was detected, suggesting that the CBM targets $\alpha$-1,4-glucose polymers, consistent with the capacity of both proteins to interact with the insoluble cellulose preparations Avicel (Fig. 3) and acid-swollen cellulose (data not shown). The affinities of CBM44 and PKD-CBM44 for the various ligands were similar, suggesting that the PKD domain does not contribute to carbohydrate recognition. CBM44 is thus a type B CBM, binding individual $\alpha$-1,4-glucan chains both in soluble and insoluble polysaccharides. AGE revealed that CBM44 displays a similar ligand specificity profile to CBM30, although the affinity constants are about 10–15-fold lower for the N-terminal module (Table 2).

Figure 2. Alignment of CBM44 with CBMs from other related families. Sequences used in the alignment are labeled with the accession numbers indicating their respective CBM family. The three conserved motifs responsible for calcium binding in CBMs from the $\beta$-sandwich superfamily are boxed.
CBMs to soluble saccharides, and the molecular basis for these thermodynamic changes has been widely discussed (11, 16, 18, 45). The observation that cellulotetraose and the xyloglucan heptasaccharide XXXG (the most common repeating unit of xyloglucan; where X is a glucose-decorated α-1,6 with xylose and G is unsubstituted glucose) bind to CBM44 with similar affinities is entirely consistent with the polysaccharide data, indicating that the side chains in xyloglucan do not make steric clashes with the CBM but also do not contribute to binding and thus do not represent specificity determinants.

The observation that CBM30 and CBM44 bind to xyloglucan provides the first evidence that CBMs are able to accommodate the side chains of this decorated glucan. The specificity of the two CBMs is entirely consistent with the substrate specificity of the appended GH44 (see below). Because previous studies on CBMs have not questioned their affinity for xyloglucan, we explored whether the capacity to bind this polysaccharide is a generic feature of CBMs that recognize single β-glucan chains. We used ITC to determine the affinity of CjCBM4-1, CjCBM11, CcCBM17, BspCBM28, and PeCBM29-2 (which all bind to β-linked glucans) for xyloglucan. The data show (see Table 4) that all five modules recognize xyloglucan. It is likely, however, that the xyloglucan side chains make some steric clashes with the protein surface of CjCBM11, CcCBM17, and BspCBM28 because these modules bind to the decorated glucan less tightly than cellohexaose, which fully occupies the respective ligand-binding sites. The similar affinity of xyloglucan and cellohexaose for CjCBM4-1 indicates that the side chains of the decorated glucan do not make significant steric clashes with the protein. The modest elevation in affinity of PeCBM29-2 for xyloglucan, compared with cellohexaose, may indicate that the xylose side chains are making a positive contribution to binding. It should be noted, however, that PeCBM29-2 binds, through cooperative interactions, more tightly to polysaccharides than the corresponding oligosaccharides (which occupy the complete binding site), and thus the 3-fold increase in affinity for xyloglucan over cellohexaose may not be related to the capacity of the protein to recognize the xylose side chains (16).

Inspection of the three-dimensional crystal structures of CjCBM4-1 (46), CcCBM17 (18), and PeCBM29-2 (16) in complex with cello-oligosaccharides provides some insight into the mechanism by which these proteins bind xyloglucan. The most common repeating unit in xyloglucan is XXXG. It is likely that the undecorated glucose binds to subsite 2 in CjCBM4-1 and subsite 3 in CcCBM17 and PeCBM29-2, because at these locations the exocyclic O-6 of the hexose sugar is pointing directly at the protein surface and thus a xylose side chain would make a steric clash with the CBM preventing ligand binding. In CcCBM17, the other three subsites could all accommodate a glucose substituted at O-6 with a xylose. The protein, however, displays maximal binding to cellohexaose (~50-fold tighter than cellotetraose). Thus, the steric clashes that reduce affinity for xyloglucan (Table 4) are likely to be in subsites distal to those identified in the crystal structure of the CBM in complex with cellooctaose. In the CjCBM4-1-cellooctaose complex the exocyclic O-6 of Glc4 (in addition to Glc2) appears to point toward the protein surface, and thus decoration of this sugar with xylose would cause a significant steric clash, which is contrary to the biochemical data that show a similar affinity of CjCBM4-1 for xyloglucan and cellohexaose (Table 4). Glc4 (and Glc5), however, makes no interactions with CBM44, and thus it is highly likely that this sugar is able to rotate such that O-6 is pointing into solvent, without causing a significant loss in affinity. In PeCBM29-2 Glc1, in addition to Glc2, would clash with the protein if it is decorated with xylose at O-6. As Glc1 makes a single hydrogen bond with the CBM, the loss of this interaction, which would be required in a rotation of the sugar such that O-6 points into solvent, is likely to incur only a modest energetic penalty. It should be noted, however, that the determination of the precise conformation of xyloglucan when bound to CBMs requires the resolution of the crystal structure of these proteins in complex with oligosaccharides derived from the polysaccharide.
FIGURE 4. Representative ITC data of CBM44 binding to xyloglucan and cellohexaose. The upper parts of each panel show the raw binding heats, and the lower parts show the integrated binding heats minus the dilution control heats fitted to a single-site binding model. In the xyloglucan titration, the ligand is at 2 mg/ml and CBM44 at 75 μM. In the cellohexaose titration, the ligand was at 0.5 mM and CBM44 at 40 μM.

### TABLE 3
Affinity of CBM44 and CBM30 for oligo- and polysaccharides as determined by ITC

| Protein  | Ligand          | \( K_s \times 10^4 \) | \( \Delta G \)  | \( \Delta H \)  | \( T\Delta S \) | \( n^a \) |
|----------|-----------------|------------------------|-----------------|-----------------|-----------------|--------|
| CBM44    | Cellohexaose    | 7.2 ± 7.2              | -8.0 ± 0.1      | -15.9 ± 0.3     | -7.9 ± 0.4      | 1.1 ± 0.1 |
| CBM30    | Cellohexaose    | 6.4 ± 0.8              | -6.2 ± 0.1      | -8.0 ± 0.5      | -1.8 ± 0.2      | 1.2 ± 0.1 |
| CBM44    | Cellobiotaase   | 6.6 ± 1.3              | -6.6 ± 0.1      | -14.5 ± 0.5     | -7.9 ± 0.6      | 1.0 ± 0.0 |
| CBM30    | Cellobiotaase   | 1.2 ± 0.8              | -5.3 ± 0.3      | -6.9 ± 0.5      | -1.7 ± 0.5      | 1.3 ± 0.1 |
| CBM44    | Xyloglucan      | 8.1 ± 9.8              | -8.1 ± 0.1      | -16.3 ± 0.6     | -8.2 ± 0.7      | 1.0 ± 0.0 |
| CBM30    | Xyloglucan      | 7.2 ± 1.4              | -6.6 ± 0.1      | -10.4 ± 0.3     | -3.8 ± 0.2      | 1.0 ± 0.0 |
| CBM44    | HEC             | 12.2 ± 3.3             | -6.9 ± 0.2      | -12.5 ± 0.5     | -5.6 ± 0.7      | 1.0 ± 0.0 |
| CBM30    | HEC             | 4.5 ± 0.5              | -6.3 ± 0.1      | -10.0 ± 0.2     | -3.7 ± 0.3      | 1.0 ± 0.0 |
| CBM44    | β-Glucan        | 22.5 ± 3.6             | -7.3 ± 0.1      | -17.7 ± 0.6     | -10.4 ± 0.7     | 1.0 ± 0.0 |
| CBM30    | β-Glucan        | 2.8 ± 0.3              | -6.1 ± 0.1      | -11.2 ± 0.2     | -5.1 ± 0.3      | 1.0 ± 0.0 |
| CBM44    | Lichenan        | 12.3 ± 2.8             | -6.9 ± 0.1      | -22.6 ± 1.3     | -15.7 ± 1.4     | 1.0 ± 0.0 |
| CBM30    | Lichenan        | 3.6 ± 0.4              | -6.2 ± 0.1      | -11.5 ± 0.4     | -5.3 ± 0.4      | 1.0 ± 0.0 |
| CBM44    | Glucomannan     | 9.0 ± 2.0              | -6.7 ± 0.1      | -15.9 ± 0.8     | -9.2 ± 0.9      | 1.0 ± 0.0 |
| CBM30    | Glucomannan     | ~0.4                   |                |                 |                 |        |

\(^a^\) Number of binding sites on the protein.
\(^b^\) Data are from Arai et al. (43).
\(^c^\) Value is an estimate because affinity was too low to obtain accurate value.

### TABLE 4
Affinity of glucan-binding CBMs for xyloglucan and cellohexaose as determined by ITC

| Protein  | Cellohexaose | Xyloglucan |
|----------|--------------|------------|
| CBM4-1   | 8.7          | 4.4        |
| CtCBM11  | 9.7          | 0.6        |
| CcCBM17  | 8.7          | 1.7        |
| BspCBM28 | 4.8          | 0.4        |
| PeCBM29-2| 1.8          | 5.9        |

CBM44 Potentiates Cellulase Activity of Cel44A through a Targeting Effect—To evaluate the role of CBM44 in the function of Cel9D-Cel44A, truncated derivatives of the enzyme comprising the GH44 catalytic module (Cel44A) and Cel44A fused to CBM44 (Cel44A-PKD-CBM44) were expressed in E. coli and purified to electrophoretic homogeneity. The temperature optimum of Cel44A and Cel44A-PKD-CBM44 is ~70 °C with maximal activity at pH 5.0 (data not shown). The biochemical properties of truncated derivatives of Cel9D-Cel44A indicate that Cel44A displays mixed endo-β-1,4-glucanase/xylanase activity, hydrolyzing cellulosic substrates such as carboxymethylcellulose and mixed linked β-1,4-β-1,3-glucans, while exhibiting reduced but still significant activity against oat spelt xylan (Table 5). The enzyme did not hydrolyze the laminarin, pustulan, galactan, or galactomannan. Most interestingly, the enzyme displays significant activity against glucomannan, and its activity against the branched β-1,4-glucose polymer xyloglucan is comparable with its capacity to hydrolyze CMC or lichenan (Table 5). Cel44A is a GH44 enzyme, and glycoside hydrolases...
in this family display predominantly endo-β-1,4-glucanase or xyloglucanase activity. (www.afmb.cnrs-mrs.fr/CAZY). The data reported here show that Cel44A is indeed a cellulase/xyloglucanase that displays some xylanase activity and not a xylanase (19), as reported previously (using a qualitative assay on xylan-containing agar plates and staining with Congo Red).

A proposed role for CBMs is to maintain the proximity of the cognate catalytic domains within the context of a complex macromolecular structure such as the plant cell wall (2). This so-called targeting function could orchestrate the distribution of a large repertoire of enzymes to the different polysaccharides located in plant cell walls. Therefore, in CtCel9D-Cel44A, CBM44 would direct Cel44A to its target substrates, the branched and unbranched β-1,4-glucose cell wall polymers, in the presence of a range of other polysaccharides. To test this hypothesis, the catalytic efficiency of both Cel44A and Cel44A-PKD-CBM44 against xylglucan and CMC was evaluated in the presence and absence of polysaccharides that do not act as substrates for this enzyme. The data presented in Table 5 suggest that CBM44 "rescues" the activity of Cel44A in the presence of the polysaccharides laminarin and pustulan which, although not acting as substrate, can presumably make nonproductive interactions with the active site of the enzyme. Therefore, we propose that in the context of the plant cell wall, CBM44 potentiates the activity of Cel44A by targeting the enzyme to the regions of this macromolecular structure where Cel44A substrates are located.

Crystal Structure of CBM44—The three-dimensional structure of PKD-CBM44 was solved using MAD methods with a SeMet derivative crystal. The final model of PKD-CBM44 contains two calcium ions (Ca1 and Ca2). The structure reveals the presence of two separated domains, PKD and CBM44, both adopting a β-sandwich fold (Fig. 5a). Ca1, located in the PKD module, has an octahedral coordination; the side chain O atoms of Asn\(^{101}\) (O-κ1), Asp\(^{103}\) (bidentate coordination from O-ε1 and O-κ2), Asp\(^{105}\) (O-ε1), Asp\(^{107}\) (O-ε1), and the main chain O atom of Gly\(^{113}\) interact with the metal ion with a single water molecule completing the coordination sphere (Fig. 5b). A second calcium-binding site (Ca2) was identified in CBM44. Ca2 also shows octahedral coordination and is bound to residues Asn\(^{120}\), Lys\(^{123}\), and Arg\(^{125}\) (main chain O atoms), Asp\(^{126}\) (O-ε1), Glu\(^{103}\) (O-ε1), and Asp\(^{125}\) (bidentate coordination from O-ε1 and O-κ2). The distance between the ligands and the calcium ions varies from 2.3 to 2.6 Å (Fig. 5b). In general, β-sandwich CBMs have at least one calcium-bound metal, which is often solvent-inaccessible and is suggested to stabilize the protein fold (2). The importance of calcium in maintaining the structural integrity of a family 4 CBM was demonstrated by showing that removal of the metal from the xylan-binding module reduced its melting temperature by 23 °C, although this had no effect on ligand binding (47). It has been shown recently, however, that in a family 35 and a family 36 CBM calcium plays a direct role in ligand recognition (48, 49). The calcium-binding sites in the PKD domain and CBM44 are buried inside the protein structures supporting, in both cases, a structural role for the metal. It is interesting to note that CBM44 Glu\(^{103}\) and Asp\(^{125}\), which play a key role in Ca2 coordination, are residues belonging to the first and third amino acid consensus motifs, respectively, and are present in several members of the β-sandwich CBM superfamily. Furthermore, all the other residues belonging to these motifs are in the vicinity of the calcium ion, including the structural motif that belongs to the second motif and is highly conserved in the β-sandwich superfamily. Together, the data suggest that these three motifs fulfill a structural role of extreme importance for the CBM architecture by creating the required topological environment for calcium coordination. Therefore, we predict that the primary sequence of all β-sandwich CBMs containing a structural calcium ion that is on the opposite face of the ligand-binding site should contain these three consensus motifs.

The PKD domain is built from two β-sheets with three and four β-strands packed face to face and displays features typical of a β-sandwich (Fig. 5a). The four-stranded β-sheet has a slightly concave topology. The number and arrangement of these strands are identical to several members of the immunoglobulin and fibronectin type III superfamilies, and thus the PKD domain can be defined as displaying an immunoglobulin-like β-sandwich fold (50). The core of the PKD β-barrel is highly hydrophobic, including four phenylalanines, three tyrosines, and one tryptophan. All these residues are conserved in the PKD domain family (data not shown). PKD domains are so-called because 16 copies of this module were originally identified in the extracellular segment of polycystin-1, a large cell surface glycoprotein encoded by the pck1 gene, which is mutated in autosomal dominant polycystic kidney disease (50). One or more copies of the PKD domain are also found in several other extracellular proteins from higher organisms, eubacteria and archaeabacteria. PKD domains are present in 39 glycoside hydrolases (primarily GH18 chitinas), although their role in the function of these enzymes has not been extensively investigated. It is interesting to note, however, that a recent study showed that the PKD domain in CtCel9D-Cel44A does not modulate the function of CBM44 when binding to soluble and insoluble polysaccharides nor does it bind carbohydrates per se. The sugar-binding Trp\(^{50}\) of the chitin, binding PKD, described above, is not conserved in CtCel9D-Cel44A PKD, whereas the equivalent residue to Trp\(^{67}\) (plays a role in sugar recognition in the chitinase PKD) in the cestroidal module (Trp\(^{68}\)) is buried and therefore belongs to the hydrophobic core of the protein. It is possible that this domain functions as a nonflexible spacer domain in CtCel9D-Cel44A, although the role of these modules in glycoside hydrolases is clearly variable. The PKD domains of polycystin-1, however, are involved in inter-molecular interactions, playing an important role in intercellular adhesion (52). The PKD-CBM44 structure forms a dimer related by a crystallographic 2-fold. The four-stranded β-sheet of
FIGURE 5. The three-dimensional structure and the hydrophobic platform of PKD-CBM44. The overall structure of the PKD-CBM44 highlighting the secondary structural elements, with the β-strands of each β-sheet colored the same as is shown in a. The Trps in the binding cleft are depicted as sticks, and the calcium atoms depicted as red spheres. The hydrogen bond between the OH of Ser92 and carbonyl O of Thr94 in the linker region is shown in cyan. Note β-strand 2 of PKD has a kink at residues 29/30. b, comparison of the calcium-binding sites Ca1 in the PKD and Ca2 in the CBM44 domains and their corresponding coordinating amino acid residues. The electron density was contoured at 2σ. c depicts a top down view of the CBM44 binding cleft showing the amino acids (as sticks) that are in the vicinity of the three tryptophan residues (as balls and sticks) that comprise the hydrophobic platform. All the ribbon figures in Figs. 5, 6, and 8 were prepared using MOLSCRIPT (58) and RASTER3D (59) and the electron density figures with TURBO-FRODO (60).
the PKD domain interacts with CBM44 via four hydrogen bonds between the parallel strand 7 of the PKD domain (residues 81–90) and strand 13 of the CBM44 (residues 190–196). Therefore, it is possible that PKD modules participate in specific protein-protein interactions within the cellulosome, which may be important in orchestrating the binding of defined repertoires of plant cell wall hydrolases in each individual multienzyme complex.

Residues Ser22 to Gly65 constitute the interdomain linker sequence connecting the PKD domain to CBM44. Most interestingly, Ser22 hydrogen bonds to the main chain O of Thr11, suggesting that, at least under some conditions, the linker sequence does not possess significant flexibility (Fig. 5a). The CBM44 structure (Fig. 5a) reveals a classic distorted β-jelly roll fold consisting of two five-stranded anti-parallel β-sheets, which form a convex (β-strands 9, 12, 14, 15, and 17) and concave face (β-strands 10, 11, 13, 16, and 18). The core of the β-barrel is highly hydrophobic and includes five phenylalanine and three tryptophan residues. As in other CBMs that bind individual polymer chains, the concave side of CBM44 forms a cleft. In CBM44, this surface depression is defined by polypeptide stretches Thr111–Thr125, Gly145–Leu159, Ala177–His185, Val192–Met203, and Gln215–Thr226 and strongly resembles the equivalent grooves observed in other type B CBMs. This cleft is decorated with the side chains of Thr111, Ser113, Thr115, Glu145, Thr146, Glu148, Lys150, Asp152, Gln179, Tyr181, Met183, His185, Trp189, Trp194, Ser196, Thr199, Gln227, Gln231, and Gln235 (Fig. 5c). DALI structural similarity searches reveal that CBM44 is most similar to CBMs of families 22 (1dyo), 15 (1gny), 27 (1of3), 11 (1v0a), 4 (1gui), and 29 (1w8u, 1ulo), 22 (1dyo), 15 (1gny), 27 (1of3), 11 (1v0a), 4 (1gui and 1ulo), 29 (1w8u, 1ulo), and 17 (1j83) with more than 129 matching Cα positions and a root mean square deviation of less than 3.1 Å. All these proteins are members of the β-sandwich CBM superfamily and, apart from CBM15 and CBM29, which operate in nature at lower temperatures than most of the other modules, contain a structural calcium ion in a similar position to CBM44.

**Crystal Structure of CBM30**—Structure solution of CBM30, to a resolution of 2.27 Å, revealed two copies (chains A and B) of the polypeptide chain in the asymmetric unit of the crystal, related by a noncrystallographic 2-fold axis. Chain A of the CBM30 model consists of 174 amino acids, whereas the purified protein contains 207 amino acids, including the C-terminal His6 tag (Fig. 6a). The N-terminal stretch from residue Ser1A to Lys12A, Gln46A, Ser66A, and the C-terminal region extending from Glu121A to His203A, which includes the His6 tag, were disordered and thus were not included in the final model. The model for chain B was identical to chain A, except for the inclusion of one additional residue Lys1B. The model includes 218 water molecules. CBM30 adopts a B-sandwich fold in which the two β-sheets each contain five β-strands (Fig. 6a). The two β-sheets are connected mainly by loops, although some of these linking regions, Val44–Thr49, is a β-strand. The first β-sheet (β-sheet 1) includes β-strands 1 (Val1–Phe14), 4 (Pro4–Val8), 11 (Phe173–Asp179), 9 (Gln139–Pro140), and 6 (Tyr91–Gly98), whereas the second β-sheet (β-sheet 2) consists of strands 2 (Gly28–Ser29), 5 (Trp59–Thr63), 7 (Val106–Asp114), 8 (Asp123–Val127), and 9 (Val160 and Ser166). Each contains a hydrophobic core, which comprises 11 leucine, 8 isoleucine, and 8 phenylalanine residues. The contacts between the two monomers (water-mediated and interactions between the loop Lys112–Ile122 in both molecule A and B) is not biologically relevant because the protein is monomeric in solution (data not shown).

As observed in other type B CBMs, the concave side of the β-sandwich forms a cleft that is defined by β-sheet 2. Decorating the CBM44 groove are the residues Trp7, Trp68, Ile70, Leu72, Trp79, Asn76, Arg110, Lys112, Glu121, Asp123, Thr125, Ser166, and Arg168 (Fig. 6b). The aromatic are aligned at the edge of the cleft and comprise a hydrophobic platform.

Structural comparison of CBM30 with the 1wmx model, which was used to solve the phase problem by molecular replacement methods, shows no significant differences in the overall structure. All β-sheets and loops are conserved, and the structural differences are restricted to some side chains. The 1wmx structure (deposited in the Protein Data Bank) includes a stretch of 16 residues in the C-terminal region of only one of the monomers (from residue Glu180B to Pro196B). This sequence lies in between the interface of a symmetry-related dimer. Because of disorder, the corresponding region (from residue Glu180B to Pro203B) is not visible in the structure of CBM30 reported here. A DALI data base search for homologous macromolecules reveals structural homology of CBM30 to several members of the β-sandwich superfamily, especially CBM11 (1v0a), CBM29 (1w8u), CBM44 (1gui), CBM27 (1pmh), CBM15 (1gny), CBM22 (1dyo), CBM6 (1uxz), and CBM36 (1ux7). The root mean square deviations of these structures, when superimposed on CBM30, vary in the range of 2.3–3.7 Å. Most interestingly, although a member of the β-sandwich superfamily of CBMs, CBM30, like CBM15 and CBM29, does not contain any calcium ion in its structure (Fig. 6a).

Thus, although bound calcium is a common feature of thermostable members of this superfamily, it is not a requirement for stability at elevated temperatures. The lack of bound calcium explains why CBM30 does not contain the three consensus amino acid motifs of the superfamily that coordinate the metal.

**Probing the Location of the Ligand-binding Residues in CBM30 and CBM44**—Inspection of the surface clefts of CBM30 and CBM44 (Figs. 5 and 6) reveals the presence of three aromatic residues on the edge of both putative binding sites. Most intriguingly, of these residues only Trp27 and Trp68 are invariant in family CBM30 members that contain the complete sequence (two CBM30 sequences lack the region that includes these two residues; see below), whereas Trp79 is conserved in only two of the seven sequences that comprise this family (data not shown), suggesting differential binding mechanisms between members of CBM30. To probe the importance of the solvent-exposed residues Trp27, Trp68, and Trp79 in CBM30, the mutant proteins W27A, W68A, and W78A were produced, and their biochemical properties were compared with wild-type CBM30. The data (Fig. 7a) show that W27A, W68A, and W78A display no significant affinity for decorated and undecorated β-glucans or glucomannan, whereas the control protein W176A, with a mutation on a surface Trp that is not located on the putative cleft, exhibits similar ligand binding properties to the wild-type protein. By contrast, W78A displays reduced, but still significant, affinity for xyloglucan and barley β-glucan (~10-fold lower than wild-type CBM30) and no detectable binding to glucomannan (although a 10-fold reduction in the Km for glucomannan is likely to result in no retardation of electrophoretic migration in gels containing 0.1% (w/v) of the polysaccharide). Taken together, these results confirm that the groove located on the concave surface of CBM30 comprises the ligand-binding site, and the three aromatic residues located on the edge of the cleft are involved in ligand recognition.

In general, removal of the ligand-binding aromatic residues in CBMs completely abolishes polysaccharide binding (17, 53, 54). In CBM30, however, it is apparent that Trp79 plays a less important role in ligand binding than the other two aromatic residues, which is entirely consistent with the observation that although Trp27 and Trp68 are invariant (except for two proteins that contain N-terminal truncations) in family 30 CBMs, Trp78 is not highly conserved. In CBM29 Tyr65, located in a similar position to Trp78, at one end of the binding site, was less important in ligand recognition than the other two tryptophan residues in the cleft (17) (Fig. 8). It was suggested that the tyrosine makes less extensive hydrophobic interactions with the sugar, and we propose a similar
explanation for the retention of some ligand recognition in the W78A30 mutant. Most interestingly, two members of CBM30 family do not contain the N-terminal 80-residue sequence where the identified binding residues are located, and it is possible that these proteins do not retain a carbohydrate-binding function.

In CBM44 the loss of the central aromatic residue in the binding site, Trp194, completely abrogates ligand recognition, confirming that the groove located on the concave surface of the protein comprises the ligand-binding site (Fig. 6c and Fig. 7b). In contrast to W194A44, removal of the tryptophan residues that flank Trp194 (Trp189 and Trp198) caused a relatively modest decrease in affinity (Table 6 and Fig. 6c and Fig. 7b). It is possible that Trp194 makes a stronger hydrophobic interaction with the glucan chain than the flanking trytophans, although the dominance of this amino acid may also reflect an additional hydrogen bond between the indole nitrogen of Trp194 and the polysaccharide. Indeed, the central aromatic amino acid, Trp27, in the ligand binding site of the CBM2a of C. fimii Cel6A also plays a more important role in cellulose binding than the flanking trytophans. McLean et al. (55) pro-

**FIGURE 6. The three-dimensional structure and the hydrophobic platform of CBM30.** a, displays the overall structure of CBM30 monomer looking side-on down the binding cleft, highlighting the secondary structural elements, with the β-strands of each β-sheet colored the same. The aromatic residues involved in ligand binding are shown in ball and stick. b, displays the binding cleft of CBM30 highlighting the amino acids, depicted as sticks, that are in the vicinity of the three tryptophan residues, depicted as ball and stick.
posed that an additional hydrogen bond between the pyrrolic amine of this residue and cellulose explains why Trp54 plays a dominant role in ligand binding.

The data described above clearly show that the hydrophobic platform provided by the three tryptophan residues in both CBM30 and CBM44 play a pivotal role in ligand binding (Fig. 8). The orientation of these residues in CBM44 is such that it will bind a slightly twisted glucan chain, consistent with the conformation adopted by cello-oligosaccharides in free solution and when bound to a variety of CBMs (16, 18, 46). The length of the hydrophobic platform in CBM44 is ~24 Å, and the position of the aromatic residues would enable binding to sugars n, n + 2, and n + 4, which indicates that the protein can accommodate cello-pentaose. Therefore, it is interesting to note that CBM44 displays ~10-fold higher affinity for cellohexaose than cellopentaose. It is possible that residues at either end of the hydrophobic platform of CBM44 interact with cellohexaose, although mutagenesis studies show that Asp188, Thr202, and Met203, which are distal to Trp189 and Trp198 (Fig. 5c), do not contribute to ligand binding (see Fig. 7b and Table 6, and see below). Similarly, CBM44-1 from C. fimi Cel9B displays 250-fold higher affinity for cellotetraose than cellobiose (56), even though the protein only interacts with three consecutive glucose residues (46). It is possible that the more extended interchain hydrogen bonding network afforded by

---

**Table 6**

| Protein | $K_a \times 10^4$ | $\Delta G$ kcal mol$^{-1}$ | $\Delta H$ kcal mol$^{-1}$ | $T\Delta S$ kcal mol$^{-1}$ |
|---------|------------------|--------------------------|--------------------------|--------------------------|
| Wild type | 81.6 ± 9.8 | -16.3 ± 0.6 | -8.2 ± 0.7 |          |
| Q179A$_{44}$ | 21.2 ± 0.4 | -11.8 ± 0.5 | -4.5 ± 0.6 |          |
| M183A$_{44}$ | 22.7 ± 1.7 | -10.4 ± 0.1 | -3.1 ± 0.1 |          |
| D188A$_{44}$ | 10.6 ± 8.9 | -16.7 ± 0.6 | -8.5 ± 0.7 |          |
| W189A$_{44}$ | 3.9 ± 0.6 | -7.4 ± 0.4 | -1.2 ± 0.6 |          |
| V192A$_{44}$ | 71.6 ± 7.6 | -13.9 ± 0.5 | -5.9 ± 0.5 |          |
| W194A$_{44}$ | NB$^a$ | | | |
| S196A$_{44}$ | 49.0 ± 5.8 | -11.2 ± 0.3 | -3.4 ± 0.4 |          |
| W198A$_{44}$ | 2.2 ± 0.2 | -6.8 ± 0.5 | -0.9 ± 0.6 |          |
| T202A$_{44}$ | 77.0 ± 7.3 | -14.1 ± 0.2 | -6.1 ± 0.3 |          |
| M203A$_{44}$ | 62.6 ± 6.4 | -14.1 ± 0.4 | -6.2 ± 0.5 |          |
| Q227A$_{44}$ | 29.3 ± 2.5 | -13.3 ± 0.3 | -5.9 ± 0.3 |          |
| Q231A$_{44}$ | 12.1 ± 0.7 | -11.0 ± 0.6 | -4.1 ± 0.6 |          |

$^a$ No binding was detected.
these longer ligands stabilizes the conformation adopted by the oligosaccharide in the binding cleft. Alternatively, if the protein does not interact with O-1 of the reducing end glucose, the flexible anomeric configuration adopted by this sugar may reduce binding affinity, and thus these CBMs bind optimally to internal regions of glucan chains.

In CBM30 the orientation of the aromatic residues is different than CBM44. It is likely that Trp65, Trp92, and Trp78 will bind the same face of the sugars n, n + 1, and n + 3 in an oligosaccharide chain that adopts a slightly twisted conformation. Most interestingly, the module binds maximally to cellohexaose, whereas the tetrasaccharide is sufficient to occupy the complete hydrophobic platform. The possible mechanism for the tighter binding of ligands that extend beyond the hydrophobic platform may be similar to CjCBM4-1, described above.

A mutagenesis strategy was employed to probe the role in ligand recognition of residues in CBM44 that are solvent-exposed and in close proximity to the hydrophobic platform, which plays a key role in ligand binding, as shown above (Fig. 5e). In general, exchanging these amino acids for alanine causes only a very modest reduction in affinity, as judged by AGE (Fig. 7b) and ITC (Table 6). Furthermore, AGE showed that the influence of the nine mutants was similar for the five ligands tested, xylolucan, laminarin, glucomannan, HEC, and barley β-glucan (data not shown), indicating that these residues do not interact with specific ligands. It is particularly surprising that the affinities of the mutant proteins Q179A, S196A, M183A, and Q227A are not severely affected because the mutated amino acids are likely to interact with sugars n + 1 and n + 3, which are positioned between the three aromatic residues. Only the loss of Gln311 causes a significant decrease in affinity (~7-fold), a residue that is likely to hydrogen-bond to the sugar that stacks against the critical central aromatic amino acid Trp194. Although polar residues do not play an important role in ligand recognition in type A CBMs (55) that bind to crystalline ligands, they do contribute to the binding of type B CBMs to their target carbohydrates (53, 54, 57). As such CBM44 is an atypical type B CBM. Therefore, it is likely that the protein does not make many direct hydrogen bonds with its ligands and that glucan recognition is mediated primarily through the three aromatic residues in the binding cleft. This may explain why the side chains in xylolucan do not adversely affect the binding of the polysaccharide to CBM44, as there would be no steric clashes between the xylose decorations and amino acids that make direct hydrogen bonds with the glucan backbone.

CONCLUSIONS

The plant cell wall is an intricate macromolecular structure formed by a complex repertoire of recalcitrant polysaccharides in which β-1,4-glucose polymers predominate. Plasticity is a remarkable property of cellulases that enables the same enzyme to cleave β-1,4-glucosidic bonds in a range of polysaccharides (e.g. cellulose, xylolucan, glucomannan, and mixed linked β-1,4-β-1,3-glucans). Here we show that Cel44A of CjCel9D-Cel44A is a typical endoglucanase that is capable of cleaving a variety of glucan-based plant cell wall polysaccharides such as cellulose, β-glucans, xylolucans, and glucomannans, while retaining significant activity against xylan. It is shown that the role of CBM44, previously a domain of unknown function, is associated with the targeting of Cel44A to its substrates, which are embedded in a complex polysaccharide matrix. To mediate efficient targeting, ligand recognition by CBM44 needs to mirror the substrate specificity of Cel44A, and we show here that this promiscuity in carbohydrate recognition is intrinsic to CBM44. The properties of CBM44 are remarkably similar to the N-terminal CBM30, although binding affinities are lower for the latter CBM. As typical of type B CBMs from the β-sandwich superfamily, the crystal structures of CBM30 and CBM44 suggest that the shallow groove identified in the protein structure constitutes the ligand-binding site. Site-directed mutagenesis studies revealed that the aromatic residues that comprise the hydrophobic platform in the respective clefts of CBM30 and CBM44 play a pivotal role in ligand binding. It is suggested that the location of the three binding residues on the edge of the cleft may enable CBM30 and CBM44 to bind unsubstituted β-1,4-glucan polymers, in addition to highly branched hemicelluloses such as xylo-glucans. The topology of the binding site and the paucity of polar residues involved in ligand recognition, at least in CBM44, indicate that the decorations evident in xyloglucans will be solvent-exposed, contributing to the plasticity in ligand specificity displayed by CBM30 and CBM44. Whether the true biological function of CBMs that recognize single β-1,4-glucan chains is to target xylolucan rather than the amorphous regions of cellulose is currently unclear. Finally, this work illuminates the importance of CjCel9D-Cel44A, a major cellulolysosomal enzyme, in the function of this highly efficient multienzyme complex. In a concerted action CBM44 and Cel44A target the hydrolysis of hemicellulosic polysaccharides (β-glucans, xyloglucans, and glucomannans) that are in intimate contact with celldiolic microfibrils. Once these polysaccharides are removed, both catalytic domains, Cel9D acting as an exocellulase on crystalline cellulose and Cel44A functioning as an endoglucanase on the amorphous sections of the polysaccharide, act synergistically to degrade the most abundant and recalcitrant plant cell wall polysaccharide, cellulose.

Acknowledgment—We thank the European Synchrotron Radiation Facility for provision of data collection facilities.

REFERENCES

1. Coughlan, M. P. (1985) Biotechnol. & Genet. Eng. Rev. 3, 39–109
2. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochim. J. 382, 769–781
3. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. F., and Gilbert, H. J. (1998) Biochim. J. 331, 775–781
4. Gill, J., Rixon, J. E., Bolam, D. N., McQueen-Mason, S., Simpson, P. J., Williamson, M. P., Hazlewood, G. F., and Gilbert, H. J. (1999) Biochim. J. 342, 473–480
5. Coutinho, P. M., and Henrissat, B. (1999) in Recent Advances in Carbohydrate Bio-engineering (Gilbert, H. D., Davies, G., Henrissat, B., and Svensson, B., eds) pp. 3–12, Royal Society of Chemistry, Cambridge, UK.
6. Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) Biochemistry 28, 7241–7257
7. Raghothama, S., Simpson, P. J., Szabo, L., Nagy, T., Gilbert, H. J., and Williamson, M. P. (2000) Biochemistry 39, 978–984
8. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) EMBO J. 15, 5739–5751
9. Xu, G. Y., Ong, E., Golkes, N. R., Kilburn, D. G., Muhendram, D. R., Harris-Brandts, M., Carver, J. P., Kari, L. E., and Harvey, T. S. (1995) Biochemistry 34, 6993–7009
10. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) Biochemistry 39, 5013–5021
11. Boraston, A. B., Revett, T. J., Boraston, C. M., Nurizzo, D., and Davies, G. J. (2003) Structure (Camb.) 11, 665–675
12. Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001) J. Biol. Chem. 276, 49061–49065
13. Pires, V. M., Henshaw, J. L., Prates, J. A., Bolam, D. N., Ferreira, L. M., Fontes, C. M., Henrissat, B., Planas, A., Gilbert, H. J., and Czajek, M. (2004) J. Biol. Chem. 279, 21560–21568
14. Czajek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L., Black, G. W., Henrissat, B., and Gilbert, H. J. (2001) J. Biol. Chem. 276, 48580–48587
15. Bueren, A. L., Morland, C., Gilbert, H. J., and Boraston, A. B. (2005) J. Biol. Chem. 280, 530–537
16. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabo, L., McKie, V. A., Gilbert, H. J., and Davies, G. J. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 14077–14082
17. Flint, J., Bolam, D. N., Nurizzo, D., Taylor, E. J., Williamson, M. P., Walters, C., Davies, G. J., and Gilbert, H. J. (2005) J. Biol. Chem. 280, 23718–23726
