The Crystal Structure of the Family 6 Carbohydrate Binding Module from *Cellvibrio mixtus* Endoglucanase 5A in Complex with Oligosaccharides Reveals Two Distinct Binding Sites with Different Ligand Specificities*

Virginia M. R. Pires‡, Joanna L. Henshaw§, José A. M. Prates‡, David N. Bolam¶, Luis M. A. Ferreira†, Carlos M. G. A. Fontes‡, Bernard Henrissat§, Antoni Planas‡, Harry J. Gilbert§, and Mirjam Czjzek**

Received for publication, February 13, 2004, and in revised form, March 5, 2004
Published, JBC Papers in Press, March 8, 2004, DOI 10.1074/jbc.M401599200

Glycoside hydrolases that release fixed carbon from the plant cell wall are of considerable biological and industrial importance. These hydrolases contain non-catalytic carbohydrate binding modules (CBMs) that, by bringing the appended catalytic domain into intimate association with its insoluble substrate, greatly potentiate catalysis. Family 6 CBMs (CBM6) are highly unusual because they contain two distinct clefts (cleft A and cleft B) that potentially can function as binding sites. Henshaw et al. (Henshaw, J., Bolam, D. N., Pires, V. M. R., Czjzek, M., Henrissat, B., Ferreira, L. M. A., Fontes, C. M. G. A., and Gilbert, H. J. (2003) J. Biol. Chem. 279, 21552–21559) show that *Cm*CBM6 contains two binding sites that display both similarities and differences in their ligand specificity. Here we report the crystal structure of *Cm*CBM6 in complex with a variety of ligands that reveals the structural basis for the ligand specificity displayed by this protein. In cleft A the two faces of the terminal sugars of β-linked oligosaccharides stack against Trp-92 and Tyr-33, whereas the rest of the binding cleft is blocked by Glu-20 and Thr-23, residues that are not present in CBM6 proteins that bind to the internal regions of polysaccharides in cleft A. Cleft B is solvent-exposed and, therefore, able to bind ligands because the loop, which occludes this region in other CBM6 proteins, is much shorter and flexible (lacks a conserved proline) in *Cm*CBM6. Subsites 2 and 3 of cleft B accommodate cellobiose (Glc-β,1,4-Glc), subsite 4 will bind only to a β-1,3-linked glucose, whereas subsite 1 can interact with either a β-1,3- or β-1,4-linked glucose. These different specificities of the subsites explain how cleft B can accommodate β-1,4-β-1,3- or β-1,3-β-1,4-linked gluco-configured ligands.

The specific recognition of carbohydrates by proteins is central to many biological processes including cell signaling, host-pathogen interactions, and the microbial recycling of carbon from the plant cell wall (1). The microbial enzymes involved in plant cell wall degradation frequently display a modular structure in which non-catalytic carbohydrate binding modules (CBMs) are linked to the catalytic module. The role of the CBMs is mainly to target the enzymes to specific structural polysaccharides and enhance catalytic efficiency by increasing the effective concentration of the enzyme on the surface of these insoluble substrates (2, 3). CBMs are grouped into sequenced families, which may be found as part of a continuously updated carbohydrate-active enzyme data base at afmb.cnrs-mrs.fr/CAZY. The structures of CBMs from 22 of the 34 known CBM families have been determined, and the majority of these modules have a “jelly roll” fold composed almost exclusively of β-strands (Ref. 2; see also afmb.cnrs-mrs.fr/CAZY).

The specificity of CBMs varies between families. All characterized proteins in CBM1–5, and -10 bind to crystalline cellulose (4, 5), whereas other families contain modules that bind to a variety of different ligands, exemplified by CBM4, which contains proteins that recognize cellulose, xylan, mannan, and laminarin (6–9). The CBM6 family contains modules from ~35 enzymes (10) that display a diversity of substrate specificities. CBM6 proteins that bind to cellulose, xylan, mixed (β-1,3/1,4)-glucan and β-1,3-glucan have been described, and CBM6 sequences are also present in a α-1,6-mannanase (Swiss-Prot accession number Q9Z4P9) and several a-agarases (GenPept accession number AAF26838). The three-dimensional structure of two CBM6 modules from *Clostridium thermocellum* xylanase 10A (CtCBM6) (11) and a putative *Clostridium stercorarius* xylanase (*Cs*CBM6) (12) have been determined. In contrast to all other CBM families, the CBM6 modules contain two clefts that could potentially function as ligand binding sites. Cleft B is located on the concave surface of one β-sheet, whereas cleft A is found in the loop region connecting the inner and outer β-sheets of the jelly roll fold. Cleft B is in a similar location to the ligand binding sites of CBMs from several other families that include CBM22, CBM4, CBM29, and CBM17 (13–16), whereas cleft A is positioned in the loops connecting the two β-sheets and resembles the sugar binding sites of lectins, as has been discussed by Boraston et al. (12). NMR and

---

* 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.

† The abbreviated use is: CBM, carbohydrate binding module.

** To whom correspondence should be addressed. Tel.: 33-491-164-513; Fax: 33-491-164-536; E-mail: czjzek@afmb.cnrs-mrs.fr.
mutagenesis studies on CtCBM6 and the resolution of the structure of CsCBM6 in complex with xylooligosaccharide revealed that in both proteins the xylo-configured ligands bind exclusively to cleft A; however, it was suggested that cleft B may play a role in carbohydrate recognition in family 6 proteins that display specificities distinct from CtCBM6 and CsCBM6. To address this issue we have analyzed the structural basis for the ligand specificity displayed by the C-terminal CBM6 (CmCBM6-2) of endoglucanase 5A from Cellulovibrio mixtus. The accompanying paper (29) shows that CmCBM6-2 can accommodate β1,4- and β1,3-linked glucans in cleft A and B, and indeed, the two binding sites are required to bind insoluble cellulose. Only cleft B, however, interacts with β(1,4)(1,3)-mixed linked glucans such as lichenan and barley β-glucan, supporting the hypothesis that variation in ligand specificity within the CBM6 family reflects the functionality of the two potential carbohydrate binding clefts. Here we define the structural determinants for the distinct ligand specificity displayed by cleft A and cleft B in CmCBM6-2. Furthermore, by solving the structure of CtCBM6 with a xylooligosaccharide that occupies the complete binding site, we provide insight into the mechanism by which cleft A in family 6 CBMs can distinguish between glucose- and xylo-configured ligands.

**EXPERIMENTAL PROCEDURES**

**Crystallography**—Crystals of CmCBM6-2 (40 mg/ml) were grown from hanging drops in 12% (w/v) polyethylene glycol 6000 and 2.2 M NaCl. They belonged to space group P2₁2₁2₁, with unit cell parameters a = 63.11 Å, b = 66.96 Å, and c = 85.39 Å, suggesting a Vᵥ of 3.3 Å³/Đa with 2 molecules in the asymmetric unit. Co-crystallization of CmCBM6-2 with xylotetraose, cellobiose, and a mixed β(1,4)(1,3)-linked glucan (Glc-4Glc-3Glc-4Glc-1-Ome, see Fig. 1; in all cases the ligand concentrations were ~2 mM) yielded crystals in 11% polyethylene glycol 8000 and 2.0 M NaCl, whereas the complexes of CmCBM6-2 with two mixed β(1,3)(1,4)-linked glucans (Glc-3Glc-4Glc-3Glc and Glc-4Glc-3Glc-4Glc-1-Ome; Fig. 1), cellobiose and cellotriose, were obtained by soaking experiments. In the latter case native crystals were transferred into 20-μl drops of mother liquor containing 10% glycerol and ~2–3 mM of oligosaccharides. After 20–30 min of soaking the crystals were mounted in a cryo loop and frozen in a 100 K nitrogen stream. All crystals except those with Glc-4Glc-3Glc-4Glc-1-Ome belonged to the same space group as the native crystals and had approximately the same unit cell parameters (Table I). The co-crystals of CmCBM6-2 with Glc-4Glc-3Glc-4Glc-1-Ome belonged to space group P2₁2₁2₁, with unit cell parameters a = 63.19 Å, b = 37.93 Å, and c = 102.41 Å, leading to a Vᵥ value of 2.2 Å³/Đa, with one molecule in the asymmetric unit. Co-crystals of CmCBM6 (10 mg/ml) with xylotetraose were obtained in hanging drops made of 2 μl of a protein/xylotetraose mixture (~2 mM oligosaccharide) and 2 μl of reservoir solution containing 20% polyethylene glycol 1000, 0.2 M sodium acetate, and 0.1 M acetate buffer at pH 4.6. The crystals grew after 10 months at 4 °C. Here, the crystals belonged to space group P2₁2₁2₁, with unit cell parameters a = 27.43 Å, b = 51.68 Å, and c = 38.13 Å, and β = 102.78°. The Vᵥ value was calculated to be 1.9 Å³/Đa with 1 molecule in the asymmetric unit.

**Data Collection, Structure Determination, and Refinement**—All crystals were transferred to a cryoprotectant solution comprised of the growth buffer supplemented with 15% (v/v) glycerol. Subsequently the crystals, mounted in cryo loops, were flash-frozen in a cold nitrogen stream at 100 K. The different complex data sets were collected at beamlines ID29, ID14-EH1, and ID14-EH3 at the European Synchrotron Radiation Facilities, Grenoble, France. The details of data collection statistics are resumed in Table I. The data were processed and reduced with MOSFLM (17), and all further computing used the CCP4 (18) suite unless otherwise stated. The crystal structure of CmCBM6-2 was solved by the molecular replacement method using the program AMoRe (19) and the coordinates of the CmCBM6 (Protein Data Bank code 1nae) as the molecular starting model. Two solutions, corresponding to the two molecules in the asymmetric unit, gave a correlation coefficient of 22% and an R-factor of 53%. The model building was performed with ARP/wARP (20) using the complete data set at 1.40 Å. Then the structure was refined with REFMAC5 (18) to a final R-factor of 16.07% and Rₚᵤₜ factor of 17.76%. All complex structures having the same space group as the native protein were directly submitted to a first refinement cycle with REFMAC5 before manually constructing the substrate molecules using the program TURBO (21) and adding the water molecules. The same set of reflections was flagged for the calculation of the free R value for these complexes. Because the co-crystals of the CmCBM6 with xylotetraose did not have the same space group as the native crystals, molecular replacement with AMoRe was applied to position the protein molecules correctly in the new unit cell. The same strategy was applied to solve the structure of the crystals of CmCBM6-2 in complex with Glc-4Glc-3Glc-4Glc-1-Ome. All structures were refined with REFMAC5. The final refinement statistics are summarized in Table I. All coordinates of the different structures have been deposited with the Protein Data Bank with accession codes as follows: CmCBM6 with xylotetraose, 1uxx; native CmCBM6-2, 1uxz; CmCBM6-2 with cellobiose/cellotriose, 1uxy1uyv; CmCBM6-2 with Glc-4Glc-3Glc-4Glc-1-Ome, 1uxz; CmCBM6-2 with Glc-4Glc-3Glc-4Glc-1-Ome, 1uy0; CmCBM6-2 with xylotetraose, 1uyz.

**RESULTS AND DISCUSSION**

**Structure of CmCBM6-2**—The unliganded structure of CmCBM6-2 was solved by molecular replacement using the x-ray structure of CsCBM6-3 (RCSB Protein Data Bank code 1nae) and refined to a resolution of 1.40 Å. The final model of CmCBM6-2 comprised 2 copies of residues 1–133, 2 metal ions, and 390 water molecules. Both metal ions are common to CmCBM6, whereas CsCBM6-3 was reported to contain only one metal ion (12). In CmCBM6-2 the ions were modeled as Ca²⁺ and were both hepta-coordinated. In each case a single water molecule completes the coordination. The structure of CmCBM6-2 adopts a classical β-jelly roll fold, consisting predom-
The structure of CtCBM6 in complex with xylopyranose is very similar to the two xylan binding family 6 CBMs previously described, with a root mean square deviation value of 0.9 Å over 123 matched Cα atoms with CbCBM5-3 and a value of 1.1 Å over 125 matched Cα atoms with CbCBM6-3.

Interestingly, CmCBM6-2 has an exposed tryptophan residue in the region designated cleft B (Fig. 2), which is located in a shallow groove formed by one of the β-sheets, analogous to the location of the binding clefts of CBM4, CBM15, CBM17, CBM22, and CBM29 (13–16, 22). Data presented below and in the accompanying paper (29) demonstrate that cleft B does indeed comprise a ligand binding site in CmCBM6-2 and that Trp-39 plays a pivotal role in carbohydrate recognition. By contrast, cleft B in both CbCBM6-3 and CbCBM6-3 is occluded by solvent from a proline residue in the neighboring surface loop and, thus, is not accessible to oligosaccharides or polysaccharide ligands. Furthermore, CmCBM6-3 does not contain an aromatic residue in the equivalent position to Trp-39, which would further limit the capacity of cleft B in the C. stercorarium protein to bind ligand.

The Complex of CtCBM6 with Xylopyranose—Previously, we have shown that CtCBM6 binds with high affinity to xylans (unsubstituted and highly decorated) and xylooligosaccharides displaying the highest affinity for ligands with a degree of polymerization ≥5, suggesting the presence of five sugar binding sites (11).

The structure of CtCBM6 in complex with xylopyranose was determined to a 1.6-Å resolution (see “Experimental Procedures” for details). Electron density for all five sugar units in xylopyranose was well defined, allowing unambiguous modeling of the pentasaccharide bound to cleft A (Fig. 3A). The direction of the sugar in the binding cleft was resolved by assessing the temperature factors and the level of electron density of the sugar C5 and O5 atoms and the hydrogen bonding of O5. Xylopyranose binds in an almost 3-fold helical orientation characterized by the internal hydrogen bond from O3 to O5α′. The chair ring planes (defined by C2, C3, C5, and O5) of the three sugars Xyl-2–4 (Xyl1 defines the reducing end of the ligand, and Xyl5 defines the non-reducing end) lie at angles of 89 and 84°, respectively. Xylan thus binds in its favored 3-fold helical conformation previously determined by x-ray fiber diffraction analysis of xylan (23). Such a conformation has previously been proposed for xylohexaose bound to CBM2b-1 from "Cellulomonas fimi" Xyn11A (24), and observed when the ligand is bound to CBM15 from "Cellulobryon japonicus" Xyn10C (22).

Most CBMs determined to date display either a “plateau” of hydrophobic residues (called type A CBMs) that bind to crystalline surfaces of cellulose (25) or extended grooves (type B CBMs) stretching across and perpendicular to one of the two β-sheets that form the sandwich (22, 26). In contrast, the binding site both in CtCBM6 (11) and CbCBM6-3 (12) is unusual; it is not located in the central groove, which contains the ligand binding site in most of the CBMs studied to date but instead is on one edge of the sandwich. This binding site, designated “cleft A,” from CbCBM6-3 binds to oligosaccharides derived from xylan and cellulose with similar affinities (12), whereas in CtCBM6 cleft A accommodates primarily xylan and xylooligosaccharides (11). The binding site (cleft A) is formed by two loops, residues 28–34 and 90–97, which contain two highly conserved aromatic residues, Tyr-33 and Trp-92, that stack against the α and β face of the central xylose of the pentasaccharide located in subsite 3, whereas Asn-120 forms a tight hydrogen bond (2.6 Å) with O3 of this sugar. It is interesting to note that although two aromatic residues are also involved in ligand binding in the xylan binding CBMs from family 2a and 15, these residues are in a perpendicular orientation with respect to each other and, by stacking against xylose residues n and n + 2, play a key role in defining a binding site topology that can accommodate ligands with a 3-fold helical conformation (22, 24). Further subsites are formed in CtCBM6 by Asp-64 and Thr-65, which bind to the reducing end of the ligand at subsite 1, Pro-118 is at subsite 2, Ile-23 and Asn-93 form a tight hydrogen bond (2.6 Å) from Gly-24 to O3 of this sugar. It is interesting to note that although two aromatic residues are also involved in ligand binding in the xylan binding CBMs from family 2a and 15, these residues are in a perpendicular orientation with respect to each other and, by stacking against xylose residues n and n + 2, play a key role in defining a binding site topology that can accommodate ligands with a 3-fold helical conformation (22, 24). Further subsites are formed in CtCBM6 by Asp-64 and Thr-65, which bind to the reducing end of the ligand at subsite 1, Pro-118 is at subsite 2, Ile-23 and Asn-93 form a tight hydrogen bond (2.6 Å) from Gly-24 to O3 of this sugar.
Thus, although the movement of Ser-26 is optimal for binding unsubstituted xylooligosaccharides, this amino acid may need to be more distant from the binding cleft to accommodate xylans that are decorated at O2 and/or O3.

Previous studies show that CtCBM6 displays similar affinities for highly decorated and unsubstituted xylans (11), and the crystal structure of the protein-xylopentaose complex provides insights into the mechanism by which these side chains can be accommodated. Also xylose at subsite 3 cannot be substituted because both O2 and O3 point toward the surface of the protein, residues on both sides of this central sugar unit may be substituted. In subsite 2, both O2 and O3 point into the solvent and, thus, can be decorated, whereas at subsite 4 O2 is solvent-exposed and although O3 hydrogen bonds to ND2 of Asn-93 (3.1 Å), this hydroxyl group can still be decorated because it is perpendicular to the plane of the sugar ring. At subsite 1 both Thr-65 and Asp-64 make tight hydrogen bonds with O2 (2.5 and 2.7 Å, respectively) of the xylose unit, whereas decoration at O3 would also make a steric clash with the protein. At subsite 5 both O2 and O3 point into solvent, and thus, both hydroxyls can be substituted.

Comparison of the CtCBM6-xylopentaose complex with that of CsCBM6-3 bound to xylotriose at subsites 2, 3, and 4 reveals a similar arrangement of residues and surface loops around these three central subsites (Fig. 4). By contrast, residues involved in ligand binding at subsites 1 and 5 in the CtCBM6-xylopentaose complex have no counterpart in CsCBM6-3. In the region forming subsite 1 of CtCBM6, there is a 1-residue insert in the equivalent region of CsCBM6-3, and the backbone positions of the residues cannot be superimposed. As a consequence, Asp-64 and Thr-65 have no equivalent residues in CsCBM6-3; however, the ND2 of Asn-85 in the C. stercorarium protein is predicted to make a hydrogen bond (2.9 Å) with O3 of the xylose at subsite 1. A similar situation is observed in the region forming subsite 5. Here, the positions of the loops formed by residues 22–32 in CtCBM6-3 and 44–54 in

---

**TABLE II**

Hydrogen bonds and close contacts of xylopentaose in cleft A of CtCBM6

| Subsite | Residue | Atom | Distance (Å) |
|---------|---------|------|--------------|
| T65     | O       | O1   | 3.0          |
|         |         | O2   | 2.5          |
| T65     | O       | N    | 3.1          |
| D64     | O       | O2   | 3.5          |
|         |         | O3   | 3.6          |
| D64     | O       | Od1  | 2.7          |
|         |         | O3   | 3.9          |
| E36     | O       | Ow   | 2.8          |
| Y34     | N       | O3   | 2.9          |
| N120    | O       | O2   | 2.6          |
| N120    | O       | O3   | 3.7          |
|         |         | O4   | 3.3          |
| N120    | N       | Od1  | 2.6          |
| N93     | N       | Od2  | 3.7          |
|         |         | O4   | 3.3          |
| N93     | N       | O5   | 3.6          |
| N93     | N       | O5   | 3.5          |
| S26     | N       | O4   | 3.8          |
| T25     | O       | O3   | 3.6          |
| G24     | O       | O3   | 3.5          |
|         |         | O2   | 2.9          |

---

**Fig. 2.** Ribbon representation of CmCBM6-2 (left), CtCBM6 (middle), and CsCBM6-3, highlighting the region of cleft B for the three proteins.

**Fig. 3.** A, observed electron density map for xylopentaose bound in cleft A of CtCBM6. The map is a maximum-likelihood-weighted 2

\[ F_{o} - F_{c} \] electron density map contoured at a 1 σ level, corresponding to 1.6 e Å\(^{-3}\). B, the xylopentaose bound to CtCBM6, represented as sticks. The residues interacting with the ligand are highlighted. C, schematic representation of the protein/ligand hydrogen bonds and stacking interactions within the five binding subsites.
CsCBM6-3 are different. CsCBM6-3, however, contains several residues that are likely to form a functional subsite 5; Pro-48 is predicted to stack again the xylopyranose ring, whereas Asp-113 OD1 and the backbone carbonyl of Ser-46 will make hydrogen bonds with O5 and O2, respectively. Thus, both CsCBM6-3 and CtCBM6 are likely to contain five subsites, although the structures of the distal subsites in the two proteins are very different.

Three-dimensional Structures of CmCBM6-2 Oligosaccharide Complexes—The ligand binding properties of CmCBM6-2 are described in the accompanying paper (29). Briefly, this CBM6 binds to cellobiose, the β1,4-β1,3-mixed linked glucans lichenan, and barley β-glucan, insoluble forms of cellulose and displays weak affinity for the β1,3-glucan laminarin. Additionally, mutagenesis studies show that both cleft A and B can accommodate laminarin and cellulooligosaccharides, whereas only cleft B binds to glucans with mixed β1,4-β1,3 linkages such as lichenan and barley β-glucan. It also appears that the binding of CmCBM6-2 to insoluble cellulose involves synergistic interactions between cleft A and cleft B (see the accompanying paper, Ref. 29).

We were able to obtain complexes of CmCBM6-2 with five different ligands, cellobiose, cellotriose, xylotetraose, and two β1,4-β1,3-mixed linked glucans (Glc-4Glc-3Glc-4Glc-OMe and Glc-3Glc-4Glc-3Glc; see Fig. 1). However, no crystals of complexes with pure β1,3-linked glucans were obtained either by soaking or co-crystallization. The lack of complex formation for this substrate by soaking experiments might be due to crystal packing constraints since cleft A is in close contact to a neighboring molecule. The much lower affinity displayed by the two binding clefts for the β1,3-linked glucans might explain the lack of complex formation in the co-crystallization trials. The asymmetric unit for the cellulooligosaccharide complexes contains two protein molecules, so there are four putative ligand binding sites in the asymmetric unit (cleft A and B in both CmCBM6 molecules). In both the cellobiose and cellotriose complexes three of these sites are occupied, cleft A and B of the first molecule and only cleft B of the second molecule. Hereafter, ligand binding described for cleft A refers to one protein molecule in the asymmetric unit, whereas binding to cleft B relates to both copies of CmCBM6 in the asymmetric unit unless otherwise specified.

Three-dimensional Structure of CmCBM6-2 in Complex with Cellubiose and Cellotriose at Cleft A—In crystals of CmCBM6-2 soaked with cellobiose or cellotriose electron density compatible with the disaccharide was evident in subsites 3 and 4 (using the nomenclature for the subsites in CtCBM6 cleft A) of cleft A. The positions of C6 and O6 are undefined, which likely reflects ligand binding in both orientations. Electron density is only clearly defined for one sugar unit, the second sugar is disordered for both the di- and trisaccharide ligands, and no density is evident for the third sugar of cellotriose. The lack of electron density for a third sugar in the cellotriose complex suggests that one of the glucose units in this ligand is completely disordered, and thus, does not form stable interactions with cleft A. This is entirely consistent with the data of Henshaw et al. (29), which show that cleft A contains only two sugar binding sites. The hydrogen bonds formed between protein residues and the hydroxyl groups of the sugar units as well as close contacts are summarized in Table III. Binding of cellubiose and cellotriose in cleft A occurs via stacking interactions of a terminal sugar unit between the two aromatic residues, Tyr-33 and Trp-92, at subsite 3, although the orientation of the bound ligand is unclear. In both orientations, three OH groups are stabilized by residues Glu-20 and Asn-120 (Fig. 5 and Table III). For complexes in which the non-reducing end of the ligand is bound at subsite 3, O2, O3, and O4 interact with these two amino acids, whereas O1, O2, and O3 hydrogen bond to Glu-20 and Asn-120 when Glc1 (reducing end sugar) is sandwiched between the two aromatic residues in cleft A. The second glucose molecule interacts only with the hydroxyl group of Tyr-33 via either O2 or O3, depending on the orientation of the ligand, although the CBM does make indirect water-mediated hydrogen bonds with this sugar residue.

Three-dimensional Structure of CmCBM6-2 in Complex with Cellubiose and Cellotriose at Cleft B—At cleft B the “walls” of the main binding site, designated subsite 3, are formed by Trp-39 on one side and a short loop containing two adjacent glycine residues (residues 73 to 77) on the opposite side. Glc1, Glc2, and Glc3 of cellotriose are located in subsites 1, 2, and 3, respectively (Fig. 6). O4 and O6 of Glc1 make tight hydrogen bonds with the side chain of Glu-73 and to the main-chain atoms of the loop consisting of residues 73–77. The interaction of Glu-73 with the C6OH group explains why cleft B is able to bind gluco- but not xylo-configured ligands. In subsite 2, hydrogen bonds are formed between Glc2 and both Ser-41 and Gln-113, whereas Glc1 is stabilized in subsite 1 through an interaction with Gln-13 (Fig. 6 and Table III). Cellubiose, which is located in subsites 3 and 2, makes the same interactions with the CBM as Glc3 and Glc2 of cellotriose. The structural identification of three subsites for cellulooligosaccharides agrees with the biochemical data showing that cleft B displays maximum affinity for β1,4-linked glucans with a degree of polymerization of ≥3 (29). Unlike cleft A, the electron density for both C6 and O6 in cellubiose and cellotriose are clear, indicating that these ligands are located in a single orientation in cleft B. Indeed the orientation of cellulose chains in...
CBM binding sites is a matter of some debate. The solution structure of a Cellulomonas CBM4 suggests that cellooligosaccharides can bind in both orientations (27), whereas the crystal structure of the same protein in complex with cellopentaose (15) and a CBM17 module bound to cellohexaose indicate that these ligands interact with the two CBMs in a single orientation (16). Uniquely, CmCBM6 appears to accommodate cellooligosaccharides uni- or bi-directionally, dependent on the binding site occupied.

Three-dimensional Structure of CmCBM6-2 in Complex with Xylotetraose—Interestingly, we were able to obtain a complex of CmCBM6-2 with xylotetraose even though the protein was shown to bind very weakly to xylo-configured ligands at cleft A but not at cleft B (29). The complex is formed by the two adjacent CmCBM6-2 molecules, present in the asymmetric unit. All four sugar units are defined by electron density with Xy1 bound in cleft A, whereas Xy2–4 are positioned along the surface of a neighboring CBM6 molecule in a region that does not comprise either cleft A or B. Only very few hydrogen bonds are formed between this second CBM6 molecule and the three xylose units that lie parallel to the surface of the protein. The xylose residue in cleft A is sandwiched between the two aromatic residues, Tyr-33 and Trp-92, as observed for all other ligands bound at subsite 3; however, the sugar is orientated at 90° to the xylose bound in the equivalent position in CtCBM6. The O4 of Xy1 points into solvent, and thus, the other xylose molecules in the tetrasaccharide interact with a different CBM6 molecule (Fig. 4; Fig. 5C and Table III). Although the interaction of Xy1 is consistent with the observation that cleft A interacts weakly with xyloooligosaccharides, we assume that the complex of CmCBM6-2 with xylotetraose is due to opportunistic binding in the particular conditions of co-crystallization.

Three-dimensional Structure of CmCBM6-2 in Complex with β-1,4-β-1,3-Linked Glucans—The crystal structure of CmCBM6-2 in complex with the mixed β1,4-β1,3-linked ligand, Glc-4Glc-3Glc-4Glc-O-methyl, revealed that Glc4 binds in cleft A, whereas Glc1-Glc3 is located in cleft B of a second, crystallographically related CBM6 molecule (Fig. 7A). This reflects the capacity of cleft B to bind β1,4-β1,3-linked glucans and cleft A to accommodate terminal glucose molecules in subsite 3; however, the stabilization of the trimolecular complex by ligand binding to 2 protein molecules most probably is an artifact of crystallization because cleft A does not contribute to the binding of mixed-linked glucans. Interestingly, Glc4 is bound in cleft A in a manner similar to Xy1 of xylotetraose, with O1, O2, and O3 interacting with Glu-20 and Asn-121 (molecule A in Fig. 7A). The β-1,4-linked Glc1 and Glc2 residues are bound in

### Table III

| Cleft A subsite 3, celllobiose | Glc-4Glc-3Glc-4Glc-OMe | Glc-3Glc-4Glc-3Glc | Xylotetraose |
|-------------------------------|------------------------|-------------------|-------------|
| N121 Od1                      | O1 (O4) 3.5 4.0        | O4 3.9            |             |
| N121 Nd2                      | O2 (O3) 2.3 2.5        | O3 2.8            |             |
| E20 Oe1                      | O1 (O4) 2.7 3.3        | O4 2.5            | O1 2.3      |
| E20 Oe2                      | O2 (O3) 3.6 3.7        | O3 3.4            |             |
| Y33 N                       | O3 3.3                | O2 2.9            | O2 3.8      |
|                             | O2 (O3) 2.7 2.5        | O3 2.4            | O2 3.1      |
|                             | O2 (O3) 3.2 3.5        | O3 4.0            | O1 2.8      |
|                             | O2 (O3) 3.8 3.5        | O5 3.6            |             |
| Cleft A subsite 2            |                        |                   |             |
| Y33 Oh                       |                        |                   |             |
| Cleft B subsite 4, celllobiose (cellotriose) | O2 (O3) 2.6 3.6 | O5 2.8 |             |
|                             |                        |                   |             |
| Cleft B subsite 3            |                        |                   |             |
| K114 N                       | O5 3.2                | O5 3.2            |             |
| E73 Oe2                      | O6 3.5                | O6 3.1            |             |
| E74 O                        | O6 3.5                | O6 3.1            |             |
| G78 N                        | O6 3.5                | O6 3.1            |             |
| G77 N                        |                        |                   |             |
| Cleft B subsite 2            |                        |                   |             |
| G78 N                        | O3 3.6                | O3 3.4            |             |
|                             | O4 4.0                | O4 3.7            |             |
|                             | O3 3.3                | O3 3.4            |             |
|                             | O1 2.5                | O1 3.2            | O5 2.8      |
|                             | O2 (O3) 2.7           | O2 2.9            |             |
|                             | O1 3.0 (3.6)          | O1 3.5            |             |
|                             | O5 2.9 (2.9)          |                   |             |
| Cleft B subsite 1            |                        |                   |             |
| Q13 Oe1                      |                        |                   |             |
|                             | O4 - (3.3)            |                   |             |
| Q13 Ne2                      |                        |                   |             |
|                             | O6 - (4.9)            |                   |             |

Values in italics are for the non-reducing end of cellotriose bound in subsite 3 of cleft A.

Values in parentheses are for the complex with cellotriose bound in cleft B.
subsites 2 and 3 of cleft B, respectively, in a pattern identical to Glc2 and Glc3 of cellobiose (Fig. 6, A and B, and 7A). Subsite 4 is occupied by β-1,3-linked Glc3, whereas subsite 1 is unoccupied. In subsite 4, three hydrogen bonds stabilize the glucose unit, Lys-114 with O6 and O5, whereas the carbonyl of Gly-37 interacts with O6. Subsite 4 is positioned relative to the two central subsites 2 and 3 such that only a β-1,3-linked glucose can interact with this distal subsite. Furthermore, tight hydrogen bonds between subsite 3 and O6 and O4 prevents a β-1,4-linked glucose moiety binding at subsite 4 (Fig. 7B).

In the last complex of CmCBM6-2 with the mixed β1,4-β1,3-linked oligosaccharide Glc-3Glc-4Glc-3Glc the ligand is bound only in clefts B. In common with all other oligosaccharides located in cleft B the primary interaction between the protein and ligand is between subsites 2 and 3 and 2 β-1,4-linked glucose moieties depicted in Fig. 6 and Table III. In one of the two molecules in the asymmetric unit only β-1,4-linked Glc2 and Glc3 are defined by electron density, whereas in the second CmCBM6-2 molecule all four binding subsites in cleft B are occupied with glucose units. Thus, subsite 1 can accommodate a β-1,3- or β-1,4-linked glucose molecule; this latter, therefore, represents the maximum binding that can be obtained in cleft B with mixed β-1,4–1,3-linked glucans.

Mechanisms of Ligand Recognition—The binding mode in cleft A is clearly different in CBM6 modules that display distinct specificities. The examples described here reflect the two extremes, CbCBM6, which is able to bind internal regions of xylan chains as reflected by its complex with xylopentaose (Fig. 3B), and CmCBM6-2, which binds to the terminal residues of both xylo- and gluco-configured oligosaccharides (Fig. 7A). In all CBM6 structures characterized to date, the conserved residues Asn-120, Trp-92, and Tyr-33 (numbers are from CmCBM6-2)
play a pivotal role in ligand binding. These residues define the two subsites, 3 and 4, that are invariant in the three CBM6 proteins. CmCBM6-2 cleft A appears to bind only at subsites 3 and 4, whereas Boraston et al. (12) define 3 subsites in CsCBM6-3 and the complex of CtCBM6 with xylopaentose identifies 5 subsites. Overlaying the structure of CsCBM6-3 and CtCBM6 indicates that the C. stercorarium module also contains five subsites, although the structure of the distal subsites, 1 and 5, are very different in the two proteins. One of the key differences between CsCBM6-3 and CtCBM6 is that the C. thermocellum protein binds xylohexaose 100-fold tighter than cellohexaose, whereas the affinity of the C. stercorarium module for these two ligands is similar (xylohexaose binds 100-fold tighter than cellohexaose). The complex of CtCBM6 with xylopaentose provides some insight into the mechanism by which this protein is able to discriminate between xylo- and gluco-configured ligands. Specificity for xylose-containing polymers is conferred at subsites 4 and 5 where Ile23 and Ser-26, respectively would make steric clashes with the C6OH of glucose. The capacity of CsCBM6-3 to bind cellohexaose more tightly than CtCBM6 may be due to differences in subsite 5, where the lack of an equivalent residue to Ser-26 in the C. stercorarium protein would facilitate the accommodation of the C6OH moiety of glucose. The accommodation of glucose in subsite 4 of CsCBM6-3 is less clear because Phe-45 (equivalent to CtCBM6 Ile23) would likely clash with the C6 group. It is possible that the cavity behind the aromatic side chain could provide sufficient space for Phe-45 to undergo a conformational change upon cellohexaose binding. Indeed, the 200,000-fold variation in the capacity of the -1 subsite of GH10 xylanases to accommodate glucose is dependent upon the conformational freedom of an invariant tryptophan at the active site; enzymes that contain a cavity behind the indole ring display the highest activity against gluco-configured substrates (28).

Inspection of the crystal structure of CmCBM6 provides some insight into its specificity for terminal sugar residues. The surface loop from Gln-18 to Tyr-33 adopts a slightly different conformation to the equivalent region of CtCBM6 and constrains the binding cleft at subsite 5. Indeed, overlaying the liganded CtCBM6 structure with CmCBM6 indicates that

Fig. 6. Schematic representation of ligands bound to cleft B of Cm-CBM6-2. A, a cellotriose molecule bound in subsites 1–3 of cleft B. B, the tetrasaccharide Glc-4Glc-3Glc-4Glc-OMe bound to subsites 4, 3, and 2 and the reducing end bound to cleft A of a symmetry-related CmCBM6-2 molecule. C, the tetrasaccharide Glc-3Glc-4Glc-3Glc bound to subsites 4, 3, 2, and 1 of binding cleft B.
Thr-23 in the *Cellvibrio* protein will clash with O4 of the xylose at subsite 5. Furthermore, *CmCBM6* Glu-20, the equivalent residue to Phe-45 and Ile-23 in *CcCBM6*-3 and *CcCBM6*, respectively, while playing a role in ligand binding at subsite 3, is predicted to clash with a sugar residue that occupies subsite 4.

*CmCBM6-2* and *CcCBM6-3* (12) are CBM6s for which binding has been observed in cleft B. From sequence alignments (see alignment in Boraston et al. (15) for example), however, one can conclude that several other CBM6 proteins will also have an equivalent ligand binding cleft. Our structure identifies a maximum of 4 binding sites in cleft B of *CmCBM6*-2, with subsites 2 and 3 comprising the major binding sites. Discrimination between *gluco-* and *xylo-* configured ligands occurs only at subsite 2, where the C6OH of glucose makes strong hydrogen bonds with Glu-73 and the backbone carbonyl of Glu-74, explaining why the protein does not bind xylan at cleft B. Supplementary binding is possible at subsites 1 and 4. At these distal binding sites two highly flexible residues, Lys-14 in subsite 4 and Gln-13 in subsite 1, undergo conformational changes so that the side chains can interact with glucose molecules located at these subsites. It is noteworthy that discrimination between β-1,4-linked glucans (cellulose) and mixed β-1,3–1,4-linked glucans occurs at subsite 4, where only a β-1,3-linked glucose can be accommodated, whereas subsite 1 can interact with a sugar that is linked either β-1,3 or β-1,4. Thus, the order of the β-1,3–1,4 linkages in glucans is important in defining affinity because only Glc-β-1,3-Glc-β-1,4-Glc-β-1,3-Glc and Glc-β-1,3-Glc-β1,4-Glc-β-1,4-Glc will occupy all four subsites.

Acknowledgments—We thank the staff of the European Synchrotron Radiation sources (Grenoble, France) for provision of beamtime and for technical assistance at the beamlines ID29 and ID-14 EH1 and EH3.

REFERENCES
1. Brett, C. T., and Waldren, K. (1996) *Physiology and Biochemistry of Plant Cell Walls*. Chapman & Hall, London
2. Gilbert, H. J., Bolam, D. N., Szabo, L., Xie, H., Williamson, M. P., Simpson, P. J., Jamal, S., Boraston, A. B., Kilburn, D. G., and Warren, R. A. (2001) in *Carbohydrate Bioengineering: Interdisciplinary Approaches* (Teeri, T. T., Svensson, B., Gilbert, H. J., and Feizi, T., eds) Royal Society of Chemistry, Cambridge, UK
3. Gill, J., Dixon, 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
4. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T. T. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, 484–489
5. Hogg, D., Pell, G., Dupree, P., Goubet, F., Martin-Orue, S. M., Armand, S., and Gilbert, H. J. (2003) *Biochem. J.* 371, 1027–1043
6. Boraston, A. B., Warren, R. A., and Kilburn, D. G. (2001) *Biochemistry* 40, 14679–14686
7. Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxa, E., Raghothama, S., Simpson, P. J., Gilbert, H. J., Williamson, M. P., and Holst, O. (2000) *Biochem. J.* 345, 53–60
8. Tomme, P., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (1996) *Biochemistry* 35, 13885–13894
9. Zverlov, V. V., Volkov, I. Y., Velikodvorskaya, G. A., and Schwarz, W. H. (2001) *Microbiology* 147, 621–629
10. Henriassat, B., and Davies, G. J. (2000) *Plant Physiol.* 124, 1515–1519
11. Czajek, M., Bolam, D. N., Mosbah, A., Allouche, J., Fontes, C. M., Ferreira, L. M., Borne, O., Zamboni, V., Darbin, H., Smith, N. L., Black, G. W., Henriassat, B., and Gilbert, H. J. (2001) *J. Biol. Chem.* 276, 48580–48587
12. Boraston, A. B., Notenboom, V., Warren, R. A., Kilburn, D. G., Rose, D. R., and Davies, G. G. (2003) *J. Mol. Biol.* 327, 659–669
13. 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
14. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabo, L., McKay, V. A., Gilbert, H. J., and Davies, G. G. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 14077–14082
15. Boraston, A. B., Nurizzo, D., Notenboom, V., Dueros, V., Rose, D. R., Kilburn, D. G., and Davies, G. G. (2002) *J. Mol. Biol.* 319, 1143–1156
16. Notenboom, V., Boraston, A. B., Chiu, P., Freeeole, A., Kilburn, D. G., and Rose, D. R. (2001) *J. Mol. Biol.* 314, 797–806
17. Powell, H. R. (1999) *Acta Crystallogr. Sect. D 55*, 1690–1695
18. Collaborative Computational Project No. 4 (1994) *Acta Crystallogr. Sect. D 50*, 760–763
19. Navaza, J. (1994) *Acta Crystallogr. Sect. A 50*, 157–163
20. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) *Acta Crystallogr. Sect. D 53*, 448–455
21. Roussel, A., and Cambillau, C. (1991) in *Silicon Graphics Geometry Partners Directory 88*, Mountain View, CA
22. Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. G. (2001) *J. Biol. Chem.* 276, 49061–49065
23. Atkins, E. D. T. (1992) in *Xylan and Xylanases: Progress in Biotechnology* (Visser, J., Balden, G., van Kusters, S., and Voragen, A. G. L., eds) Vol. 7, pp. 39–50, Elsevier Science Publishers B. V., Amsterdam
24. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000) *J. Biol. Chem.* 275, 41157–41142
25. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoahm, Y., and Steitz, T. A. (1996) *EMBO J.* 15, 5739–5751
26. Brun, E., Johnson, P. E., Creagh, A. L., Tomme, P., Webster, P., Haynes, C. A., and McNab, L. P. (2000) *Biochemistry* 39, 2445–2458
27. Johnson, P. E., Brun, E., MacKenzie, I. F., Withers, S. G., and McNab, L. P. (1999) *J. Mol. Biol.* 287, 609–625
28. Pell, G., Szabo, L., Charnock, S. J., Xie, H., Glover, T. M., Davies, G. J., and Gilbert, H. J. (2004) *J. Biol. Chem.* 279, 11777–11786
29. Henrissat, B., Bolam, D. N., Pires, V. M. R., Czajek, M., Henriassat, B., Ferreira, L. M. A., Fontes, C. M. G. A., and Gilbert, H. J. (2003) *J. Biol. Chem.* 278, 21552–21559

Fig. 7. Ribbon representation of mixed β-1,3–1,4-linked glucan tetrasaccharides bound to *CmCBM6*-2. A, the tetrasaccharide Glc-4Glc-3Glc-2Glc-OMe is bound to two neighboring *CmCBM6*-2 molecules. The glucose unit at the reducing end is bound to cleft A of one molecule (Mol A), whereas the following glucose units are bound to a second *CmCBM6*-2 molecule (Mol B). B, the tetrasaccharide Glc-3Glc-4Glc-3Glc is bound to subsites 4, 3, 2, and 1 of binding cleft B. Residues important for binding are highlighted.
The Crystal Structure of the Family 6 Carbohydrate Binding Module from *Cellvibrio mixtus* Endoglucanase 5A in Complex with Oligosaccharides Reveals Two Distinct Binding Sites with Different Ligand Specificities

Virgínia M. R. Pires, Joanna L. Henshaw, José A. M. Prates, David N. Bolam, Luís M. A. Ferreira, Carlos M. G. A. Fontes, Bernard Henrissat, Antoni Planas, Harry J. Gilbert and Mirjam Czjzek

*J. Biol. Chem.* 2004, 279:21560-21568.
doi: 10.1074/jbc.M401599200 originally published online March 8, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M401599200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 26 references, 9 of which can be accessed free at http://www.jbc.org/content/279/20/21560.full.html#ref-list-1