The Location of the Ligand-binding Site of Carbohydrate-binding Modules That Have Evolved from a Common Sequence Is Not Conserved*

Received for publication, September 21, 2001, and in revised form, October 18, 2001
Published, JBC Papers in Press, October 22, 2001, DOI 10.1074/jbc.M109142200

Mirjam Czjzek‡§, David N. Bolam¶¶, Amor Mosbah¶¶¶, Julie Allouch‡‡, Carlos M. G. A. Fontes§§, Luis M. A. Ferreira§§, Olivier Bornet‡, Hervé Darbon‡, Nicola L. Smith¶¶¶, Gary W. Black¶¶¶, Bernard Henriassat‡, and Harry J. Gilbert‡

From the ¶Laboratoire d’Architecture et de Fonction des Macromolécules Biologiques, IBSM, CNRS Marseille and University Aix-Marseille I & II, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France, ‡Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, United Kingdom, §§CIISA-Faculdade de Medicina Veterinaria, Rua Gomes Freire, 1199 Lisboa Codex, Portugal, and ¶¶¶School of Applied and Molecular Sciences, University of Northumbria at Newcastle, Newcastle upon Tyne NE1 8ST, United Kingdom

Polysaccharide-degrading enzymes are generally modular proteins that contain non-catalytic carbohydrate-binding modules (CBMs), which potentiate the activity of the catalytic module. CBMs have been grouped into sequence-based families, and three-dimensional structural data are available for half of these families. *Clostridium thermocellum* xylanase 11A is a modular enzyme that contains a CBM from family 6 (CBM6), for which no structural data are available. We have determined the crystal structure of this module to a resolution of 2.1 Å. The protein is a β-sandwich that contains two potential ligand-binding clefts designated cleft A and B. The CBM interacts primarily with xylan, and NMR spectroscopy coupled with site-directed mutagenesis identified cleft A, containing Trp-92, Tyr-34, and Asn-120, as the ligand-binding site. The overall fold of CBM6 is similar to proteins in CBM families 4 and 22, although surprisingly the ligand-binding site in CBM4 and CBM22 is equivalent to cleft B in CBM6. These structural data define a superfamily of CBMs, comprising CBM4, CBM6, and CBM22, and demonstrate that, although CBMs have evolved from a relatively small number of ancestors, the structural elements involved in ligand recognition have been assembled at different locations on the ancestral scaffold.

Microbial plant cell wall hydrolases, in general, have a modular structure in which the catalytic modules are attached, via linker sequences to non-catalytic modules that bind polysaccharides (1). These carbohydrate-binding modules (CBMs)1 maximize catalytic activity by mediating prolonged and intimate contact between the substrate and the enzyme (2). The ligand specificity of CBMs generally reflects the target substrate of the catalytic module of the enzyme. Thus, cellulose, xylan, mannann, and β-1,3-glucan-binding CBMs are located in cellulases (3, 4), xylanases (5), mannanases (6), and lichenases (7), respectively, although it should be emphasized that numerous enzymes that do not cleave cellulose contain CBMs that recognize the crystalline form of this polysaccharide (8, 9).

CBMs have been classified into more than 28 families based on primary structure similarity (afmb.cnrs-mrs.fr/~pedro/CAZY/cbm.html; see Ref. 10). Ligand specificity can vary both between and within families, demonstrating considerable diversity in polysaccharide recognition in these protein modules. For example, the CBM2 family contains proteins that bind to xylan and crystalline cellulose (2, 5), whereas xylan, single cellulose chains, and β-1,3-glucan are recognized by different members of CBM4 (4, 7, 11). The three-dimensional structure (see afmb.cnrs-mrs.fr/~pedro/CAZY/cbm.html) of all CBMs characterized to date consist primarily of β-strands. Those that bind to crystalline cellulose have a flat ligand-binding surface (3, 12), whereas CBMs that interact with single polysaccharide chains contain clefts that accommodate the target carbohydrate (13–15). Ligand binding in CBMs is dominated by hydrophobic interactions between the sugar rings and aromatic residues on the surface of the protein-binding site (see Ref. 1 for review), and a recent study has shown that the orientation of these aromatic amino acids can play an important role in ligand specificity (16).

Members of CBM family 6 are found in a variety of enzymes that include α-agarases, α-1,6-mannanases, xylanases, acetylxylan esterases, celloextrinases, β-1,3-glucanases, and cellulases (afmb.cnrs-mrs.fr/~pedro/CAZY/db.html). Currently there is no three-dimensional structural data for any CBM6 module, and there is no information on which amino acids play a key role in ligand binding. To investigate the structural basis for the capacity of family 6 CBMs to bind to polysaccharides, we have used a combination of x-ray crystallography, NMR spectroscopy, and biochemical studies to define the ligand-binding site in CBM6 from *Clostridium thermocellum* xylanase 11A (Xyn11A, formerly XynU). This modular xylanase consists of

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

† To whom correspondence should be addressed. Tel.: 33 491 164 513; Fax: 33 491 164 536; E-mail: czjzek@afmb.cnrs-mrs.fr.

¶ Contributed equally to this work.

** Recipient of a Ph.D. fellowship from the CNRS.

‡‡ Recipient of a Ph.D. fellowship by the region Provence, Alpes, Côte-d’Azur (France) and Goémar Industries (St. Malo, France).

†§§ Recipient of a Ph.D. fellowship by the region Provence, Alpes, Côte-d’Azur (France) and Goémar Industries (St. Malo, France).

1 The abbreviations used are: CBM, carbohydrate-binding module; ITC, isothermal titration calorimetry; AGE, affinity gel electrophoresis; ESRF, European Synchrotron Radiation Facility; MAD, multiple wavelength anomalous dispersion; TLS, translational, librational and screw motions; dp, degree of polymerization; COSY, correlated spectroscopy; TOCSY, total correlation spectroscopy; HSQC, heteronuclear single quantum spectroscopy; ORF, open reading frame.

48580 This paper is available on line at http://www.jbc.org
an N-terminal catalytic module that belongs to glycoside hydrolase family 11, an internal dockerin module (17), and a C-terminal CBM6 that binds primarily to insoluble xylan but can also interact with insoluble cellulose (18). Here we report the crystal structure of the first representative of CBM family 6, Xyn11A CBM6. This has facilitated the definition of a novel CBM superfamily that encompasses CBM6, CBM4, and CBM22. Surprisingly, the location of the ligand-binding site in CBM6 is clearly different than that of CBM4 and CBM22.

**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—The CBM6 from Xyn11A was produced and purified as described earlier (18). Mutants of the protein were generated by the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions. To investigate the ligand-binding site of the protein, the mutants Y34A and N120A were generated. To assist in solving the crystal structure of the protein, four methionine mutants (Y40M, R72M, W92M, and Y112M) containing 30% polyethylene glycol 4000, 100 mM sodium citrate buffer, pH 5.0 and reservoir solution, containing 30% polyethylene glycol 4000, 100 mM sodium citrate buffer, pH 5.5 and 100 mM (NH₄)₂SO₄ were mixed. Needle-like crystals grew from microseeds after several weeks at 4 °C. The seleno-methionine-substituted protein crystals for two of the four mutants, namely R72M and Y112M, grew under the same conditions, except that the reservoir solution additionally contained a 10 mM tris(2-carboxyethyl)phosphine hydrochloride solution. A 10% glycerol solution was added to the mother liquor in both cases, prior to freezing the crystals. A rayon fiber loop was used to transfer the crystals into a liquid nitrogen cold stream. Preliminary x-ray diffraction analysis revealed that they belong to space group P6₁ 22 or P6 5 22 with unit cell dimensions a = 60.12 Å and b = 60.12 Å and c = 157.43 Å and with one molecule in the asymmetric unit. The crystals of the R72M seleno-methionine containing CBM6 mutant were larger in size and diffracted to higher resolution and were therefore chosen for the MAD experiment.

**MAD X-ray Diffraction Data Collection and Phasing**—A three-wavelength MAD experiment was conducted on beamline ID14-4 at the European Synchrotron Radiation Facility (ESRF) at Grenoble, France, using an ADSC Quantum-4 CCD detector. A single crystal of seleno-methionine containing CBM6 was flash-frozen in the laboratory and then preserved in liquid nitrogen. The crystal was transported to the ESRF, where it was remounted on the single axis goniometer. The wavelengths for the MAD experiment were chosen by scanning through the absorption edge of the R72M seleno-methionine containing CBM6 crystal. Data were collected at three wavelengths to maximize the anomalous signal as follows: at minimum f’, maximum f’, and a reference wavelength at an energy above the absorption edge (Table I). After indexing and a initial diffraction image using the program package HKL2000 (20), the program STRATEGY (21) was used to determine the optimal f range to collect complete data using a minimal oscillation sweep. A total of 60 images with 1° oscillation was collected at each of the three wavelengths. The data were processed with DENZO as part of the HKL2000 package and scaled with SCALA, part of the CCP4 suite of programs (22).

Because only one selenium was present in the protein, the Patterson maps were deconvoluted by hand, and the site was confirmed by the SOLVE (23) automated Patterson search. The Selenium position was refined, and the phase calculations were performed using the program SHARP (24) in both space groups and in the resolution range between 32 and 3.0 Å. Subsequently, a cycle of phase improvement was applied using the program DM (25). The correct enantiomorph (P6₃ 22) was determined by the quality of the resulting electron density maps.

### Table I

| Wavelength | e1 (peak) | e2 (reflection) | e3 (remote) | Native |
|------------|-----------|----------------|-------------|--------|
| 0.9793 Å |
| 0.9795 Å |
| 0.9465 Å |
| 0.933 Å |
| 2.6 Å |
| 2.6 Å |
| 2.88 Å |
| 2.0 Å |
| 0.065 |
| 0.067 |
| 0.056 |
| 0.055 |
| 5530 |
| 5210 |
| 4163 |
| 11911 |
| 98.2 (96.9) |
| 99.0 (96.4) |
| 99.0 (96.4) |
| 99.3 (98.2) |
| 4.0 |
| 5.0 |
| 5.0 |
| 6.2 |
| 10.0 (5.6) |
| 9.5 (6.4) |
| 10.8 (8.3) |
| 8.7 (1.2) |
| 0.034 |
| 0.028 |
| 0.030 |
| 97.3 (95.1) |
| 97.6 (97.1) |
| 98.5 (96.5) |

Refinement statistics:

- Resolution range (Å): 32.0–2.1
- R_cryst factor (%): 20.0 (11375)
- R_free factor (%): 21.6 (571)
- r.m.s.d. deviation from ideal geometry: 0.009
- Bond angles (°): 1.55
- Estimated coordinate error (Å): 0.18
- Average B factor for all atoms (Å²): 18.57

---

**Affinity Gel Electrophoresis (AGE)**—The capacity of CBM6 and its derivatives to bind to a range of soluble plant structural polysaccharides was evaluated by AGE. The method was carried out as described by Charnock et al. (13) except that gels were subjected to electrophoresis for 3 h. AGE was also carried out in the presence of a range of oligosaccharide xylan concentrations (0.1 to 0.005 mg ml⁻¹) and affinity constants for the interaction of native and mutant forms of CBM6 with ligand determined using equations derived from Takeo (19).
**Structure and Function of CBM6**

**RESULTS**

**Ligand Binding Studies**—To evaluate the binding specificity of CBM6 in more detail, the protein was subjected to affinity gel electrophoresis. The data (not shown) revealed that the protein binds to oat spelt, birchwood, rye, and wheat xylans, which vary in their degree of substitution, interacts weakly with barley β-glucan and soluble forms of cellulose (hydroxyethylcellulose and methylcellulose), but does not associate with sugar beet arabinan, potato, and pectic β-1,4-galactan, lime and apple pectin, rhamnogalacturonan, locust bean and carob galactomannan, or carboxymethyl β-1,3-glucan.

ITC was used to quantify the affinity of CBM6 for xylan, xylooligosaccharides, and cellohexaose. The full data are presented in Table II. The results show that the protein binds to xylooligosaccharides with a dp of 2 or more, with the affinity increasing up to xylotenase. The association constant of the CBM for xylotenase is ~50 and 100 times higher than for xylobiose and cellohexaose, respectively. The affinity of the protein for highly substituted arabinoxylans (wheat and rye) or poorly substituted xylans (oat spelt and birchwood xylan) is broadly similar. These data suggest that the protein has five sugar-binding sites and can accommodate highly decorated xylans. The stoichiometry of binding for all oligosaccharides was 1:1, indicating that there is only one ligand-binding site per protein molecule. This is consistent with the other CBMs such as CBM4 (4), CBM9 (36), and CBM22 (13), which also contain single carbohydrate-binding sites, but is in contrast to the starch-binding modules that can interact with two amylose (4, 13, 36), whereas favorable entropic ligand interactions are similar to other proteins that bind soluble carbohydrates (4, 13, 36), whereas favorable entropic forces dominate the binding of CBM2a proteins to crystalline cellulose (38).

**Crystal Structure of CBM6**—The crystals of CBM6 belong to the space group P6121 with one molecule in the asymmetric unit. The phases determined by the MAD method (figure of merit 0.51) with subsequent phase improvement using DM gave a map with a mean figure of merit of 0.67. The final model of CBM6 consists of residues 4–129 (with 4 residues displaying two alternate conformers), 2 ions, and 140 water molecules. The coordinates have been deposited at the Protein Data Bank (code 1gmm). The overall mean temperature factor value for the CBM6 molecule is 18.57 Å²; however, TLS refinement of recorded on a 500-MHz DRX Bruker spectrometer equipped with a 5-mm triple resonance HCN probe with self-shielded triple axis gradi- ents. DQF-COSY, Clean TOCSY, and NOESY spectra were acquired in a phase-sensitive mode using States-TTPI. Heteronuclear NMR experiments were performed at 307 K on 0.5 mM uniformly 15N-labeled protein in 20 mM deuterated sodium acetate buffer at pH 5.0 containing 10% (v/v) D2O, and on 0.3 mM uniformly 13C-15N-labeled protein under the same conditions. Two-dimensional HSQC, two-dimensional HSQC-NOESY, and two-dimensional HSQC-TOCSY as well as three-dimen- sional NOESY-HSQC, three-dimensional TOCSY-HSQC, and three-di- mensional HSQC-NOESY-HSQC were acquired using the Fast HSQC scheme described by Mori et al. (29). A three-dimensional HNCA (30), HNCO (31), HCCH-TOCSY (32), CBCAcoNH (33), and CBCANH (34) spectra were recorded with the 13C-15N-labeled protein. Spectral anal- ysis was performed using the XEASY software, following the standard strategy described by Wuthrich (35) combined with heteronuclear strategy.

**NMR Ligand Titration**—To identify the residues of CBM6 involved in binding xylohexaose, five 1H-15N HSQC spectra were acquired. A first HSQC spectrum was recorded on CBM6 alone at 0.5 mM. Four further 1H-13N HSQC spectra were then recorded with 0.2, 0.6, 0.9, and 1.2 mM xylohexaose, respectively.

---

**Model Building and Refinement**—The electron density map at 3.0 Å, calculated after DM in space group P6121, displayed extensive well defined regions revealing continuous stretches of main chain density, with unambiguous density for carbonyl oxygen atoms and side chains. A model comprising 120 amino acids was built from the initial map with the graphics program TURBO-FRODO (26). This model was refined using the CNS program (27).

Data to 2.0 Å resolution were collected on a native CBM6 crystal on beamline ID14-2 (λ = 0.933 Å) at the ESRF Grenoble, France (Table I). Data were processed and reduced as described above. The refinement of the 3.0 Å model was continued using these data at 2.1 Å resolution. Water molecules and alternative conformers were added, and TLS refinement, using the TLS option in REFMAC (28), was performed as deemed appropriate from the behavior of the cross-validation (R_fre) subset of reflections (5%). The final electron density map in the region of Tyr-34 and Trp-92 is shown in Fig. 1.

**NMR Spectroscopy**—Homonuclear NMR experiments were performed on a 2.0 mM unlabeled protein sample at 307 K, unless other- wise indicated, in 90% H2O and 10% (v/v) D2O containing 20 mM deuterated sodium acetate buffer at pH 5.0. All NMR spectra were
selected groups of residues showed that the particular loop region including residues 22–33 displayed highly anisotropic librational motion. Refinement statistics are given in Table 1.

The structure of CBM6 forms a classic lectin-like β-jelly roll (Fig. 2), predominately consisting of five antiparallel β-strands on one face and four antiparallel β-strands on the other face. This fold displays most similarity to the Bacillus thuringiensis insecticidal toxin (Protein Data Bank code 1cy) and a tyrosine kinase receptor domain (Protein Data Bank code 1nuk) but also to CBM22 (Protein Data Bank code 1dyo), CBM4 (Protein Data Bank code 1ulo), and a Grifonia simplicifolia lectin (Protein Data Bank code 1led). The respective root mean square deviations are 2.1 Å (115 matched Ca atoms), 2.7 Å (117 matched Ca atoms), 3.5 Å (117 matched Ca atoms), 3.2 Å (108 matched Ca atoms), and 3.4 Å (105 matched Ca atoms) when calculated using DALI (39).

CBM6 displays a small shallow cleft, lined by two aromatic residues Trp-92 and Tyr-34 (Fig. 2, left A). However, when the fold is compared with CBM22 and CBM4, a second putative ligand-binding site becomes apparent on the concave face of the jelly roll (Fig. 2, left B). This latter cleft strongly resembles those observed in CBM22 (13) and CBM4 (14), and two aromatic residues are equally located close to the surface, namely Tyr-40 and Tyr-112, the aromatic nature of which are relatively conserved in CBM family 6 (Tyr-40 is replaced by a tryptophan in a number of family members; Fig. 3). Surprisingly, a proline residue of a neighboring loop (residues 73–79) covers up this groove, making the surface aromatic residues inaccessible for xylooligosaccharides.

Two ion sites were identified in the CBM6 crystal structure. The first site displays a typical hepta-coordination geometry to accommodate a calcium ion and links the side chains of Glu-8 and Glu-10 (bidentate coordination) of the N-terminal region to the side chain of Asp-122 of the C-terminal stretch. Two main chain carbonyl groups of Ser-30 and Glu-8 complete the coordination sphere, and three water molecules are in close contact (Fig. 1b). This ion can be considered as a structural calcium ion. It is located at an equivalent position as the structural calcium ion found in CBM22 (13), bridging the N- and C-terminal regions, although neither the residues that interact with the ion nor the number of different ligation types are identical. Structural calcium ions often play a key role in protein stability (40), and recent data have shown that removal of the divalent ion from CBM6 greatly increased the susceptibility of the protein to proteolytic degradation.

The coordination of the second ion involves two main chain carbonyl groups (Ile-35 O and Val-119 O) and three water molecules. Most likely this site is occupied by a metal ion, although the biological significance of this bound metal ion is unclear.

**Fig. 2. Ribbon and surface representations of CBM6.** a, view showing the ligand-binding cleft (left A) formed by the loops between the two β-sheets of the sandwich fold. b, surface representation of CBM6 in the same orientation as in a. The shallow binding cleft on top of the globular molecule is formed by Tyr-34 and Trp-92. c, ribbon representation of CBM6 in a perpendicular view with respect to a, showing the second possible cleft (left B), which is obstructed by a short loop in CBM6, situated on the concave face of the β-sheet sandwich. a and c were produced with Molscript (45) and Raster3D (46), whereas b was produced using GRASP (47).
FIG. 3. Sequence alignment and phylogenetic tree of the modules CBM family 6. a, sequence alignment of selected sequences. The numbering corresponds to the sequence of CBM6 from *C. thermocellum* Xyn11A (formerly XynU). The secondary structure elements (exclusively $\beta$-strands) are marked below the sequences in blue when part of the structural jelly roll fold and in green for two $\beta$-strands that are not part of the $\beta$-sandwich. The three residues, Tyr-34, Trp-92, and Asn-120, which have been shown to be involved in ligand binding, are marked by red arrowheads above the sequences. The pink circles above the sequences mark the position of two aromatic residues situated in cleft B. The flexible loop region, adjacent to the ligand-binding site cleft A, is indicated above the sequences. The alignment was performed with ClustalW, and the color coding of homology (yellow boxes, cut off of 5; red boxes, cut off of 9; the maximum value for strictly conserved residues is 10) was calculated with an ALSCRIPT option. The figure was produced with ALSCRIPT (41).

b, phylogenetic tree for all known members of family CBM6, based on the alignment produced with ClustalW. The red numbers above the branches indicate the glycoside hydrolase (GH) family of the catalytic domain to which the corresponding CBM6 is attached, as reported by afmb.cnrms.fr/ pedro/CAZY/db.html (see Ref. 10). From top to bottom: xylanase U (Xyn11A) from *C. thermocellum* (Swiss-Prot accession number O52780); xylanase V from *C. thermocellum* (Swiss-Prot accession number O52779); xylanase Z from *C. thermocellum* (Swiss-Prot accession number P10478); m-1, m-2, and m-3 of xylanase A from *C. stercorarium* strain NCIB11754 (GenBank accession number AF417638); ORF SC2H12.11c from *Streptomyces coelicolor* (GenBank accession number AL359215); cellobiohydrolase from *Microbispora bispora* (Swiss-Prot accession number Q99566); ORF BH1908 from *Bacillus halodurans* (GenBank accession number AL359215).
multiple amide proton chemical shifts. This is exemplified by the arrows in Fig. 4a. The complete assignment of intra-residue HN-H, and HN-N and Cα-N cross-peaks could not be achieved because of the multiple conformations that induced many ambiguities. We were, however, able to assign some regions of the sequence (residues 6–12, 21–26, 28–40, 67–70, 91–96, and 112–117). Fig. 4a provides an example of the sequential assignment encompassing residues 28–40. Fig. 4b illustrates the loop formed by these residues, adjacent to cleft A.

**CBM6 Titration Experiments**—The 1H–15N HSQC spectra of CBM6 titrated with increasing amounts of xylohexaose show that the protein and its ligand are in fast exchange between the free and bound states. This allowed us to identify the backbone CBM6 amide groups that are affected by oligosaccharide binding (Fig. 4c).

The chemical shift variations of the N amide and the HN proton of the residues that are close to the xylohexaose-binding site are listed in Table III. Among the 23 cross-peaks that are shifted upon ligand binding, only 9 were sequentially identified using the conventional method (see “Experimental Procedures”). The remaining 14 cross-peaks fortunately corresponded to known spin systems, although they were not sequentially assigned. This problem was solved by first determining the type of amino acid corresponding to a given spin system by virtue of its 13Cα and/or 13Cβ chemical shift value and second by looking for sequential connectivity on HNCA/NHeCoA and CBCANH/CBCACoNH groups of spectra between the given amino acid and the preceding and following residues. Following this strategy, all 23 amino acids, whose chemical shifts were influenced by ligand binding and are thus likely to be at, or close, to the ligand-binding site, could unambiguously be assigned (Table III). These residues include Trp-92, Tyr-34, and Asn-120, located in cleft A, indicating that they are part of the ligand-binding site. None of the residues situated in cleft B are affected by the titration experiment.

**Site-directed Mutagenesis**—Amino acids located on the surface of the two clefts of CBM6 were mutated, and the ligand binding capacity of the resultant proteins was assessed. Mutations to Tyr-40, Arg-72, or Tyr-112, located in cleft B, did not result in a significant decrease in the affinity of the protein for xylan (Table IV). In contrast, the mutants W92M, Y34A, and N120A exhibited greatly reduced affinity for oat spelt xylan. The location of Trp-92, Asn-120, and Tyr-34 in cleft A indicates that this region composes the ligand-binding site of CBM6.

**DISCUSSION**

This report shows that CBM6 binds preferentially to xylan and xylooligosaccharides of at least two xylose moieties, exhibiting maximum affinity for the pentasaccharide. CBM6 is thus primarily a xylan-binding module, and by analogy it is likely that the other CBM6s, which are components of xylanases, also recognize xylan. These modules, however, are also found in glycoside hydrolases that act on a diversity of plant polysaccharides other than xylan. Thus, it is likely that, similar to the CBM families 2 and 4 (3, 4, 7, 11, 16), different members of CBM6 will recognize distinct polysaccharides. This view is supported by a recent study that showed that the N-terminal CBM6 in xylanase 11 A from Clostridium stercorarius strain NCBI11754 binds specifically to xylan, whereas the other two CBM6 modules in this protein bind cellulose- and xylooligosaccharides with similar affinity (42). The crystal structure of CBM6 combined with NMR titration and site-directed mutagenesis experiments show that the shallow groove formed by Tyr-34 and Trp-92 is part of the ligand-binding site. Surprisingly, these residues are not invariant within CBM family 6 (Fig. 3). However, this lack of conservation is plausible when we-as-
It is rather more surprising that proteins that recognize the same ligand, and have evolved from a common progenitor sequence, bind to xylan in different locations. It should be noted that CBM6 contains a surface loop that partially occludes cleft B and thus prevents access to the ligand. It is possible to note that for other members of the CBM6, the loop that extends from Leu-73 to Thr-79 is in a different conformation, making cleft B accessible to xylan and xylooligosaccharides. Support for this view comes from the sequence alignment, displayed in Fig. 3, which shows that, apart from one exception, the CBM6 modules that do not have a tyrosine in the equivalent position to Tyr-34 (cleft A) have a tryptophan in the position equivalent to Tyr-40 (situated in cleft B) and actually lack the proline residue that covers up cleft B in Xyn11A CBM6. Another intriguing possibility is that the location of the ligand-binding site in CBM6 modules varies depending on ligand specificity, a view strengthened by the lack of conservation, in family 6 members, of residues that are known to play a key role in ligand binding in Xyn11A CBM6. This suggests that the CBM6 scaffold displays a remarkable structural and functional plasticity.

Acknowledgments—We thank the staff at the ESRF for provision of data collection facilities. We are indebted to Alain Roussel and Caroline Penef for help during the MAD data collections.

| TABLE III |
|----------------|
| Chemical shifts of the residues implicated in the binding of the xylohexaose CBM6 at 300 K |
| N | HN | Co | Residue |
|---|---|---|---|
| 122.94 | 9.62 | 59.03 | Ser-18 |
| 130.18 | 9.78 | 62.15 | Ile-23 |
| 108.92 | 8.59 | 44.75 | Gly-24 |
| 112.73 | 6.48 | 59.99 | Thr-25 |
| 107.78 | 9.25 | 45.79 | Gly-28 |
| 109.31 | 7.50 | 44.44 | Gly-29 |
| 122.79 | 6.95 | 51.70 | Gly-30 |
| 121.92 | 8.94 | 56.60 | Tyr-34 |
| 129.72 | 9.41 | 59.20 | Ile-35 |
| 122.18 | 8.87 | 51.16 | Ala-60 |
| 113.86 | 8.77 | 57.61 | Ser-61 |
| 122.64 | 8.71 | 57.70 | Ser-62 |
| 126.01 | 7.79 | 51.85 | Ala-63 |
| 119.07 | 8.72 | 53.66 | Ala-64 |
| 124.95 | 8.36 | 63.89 | Thr-67 |
| 129.76 | 9.10 | 53.36 | Asn-68 |
| 112.75 | 8.6 | 60.41 | Trp-89 |
| 125.93 | 9.40 | 52.10 | Gly-90 |
| 116.94 | 9.48 | 46.01 | Gly-91 |
| 118.94 | 11.36 | 57.79 | Trp-92 |
| 133.95 | 8.61 | 51.50 | Asn-120 |
| 122.16 | 9.69 | 55.29 | Asp-122 |
| 124.62 | 9.31 | 55.95 | Tyr-123 |

**Affinity determined by quantitative AGE as described under “Experimental Procedures.”**

| TABLE IV |
|----------------|
| Affinity of wild and mutants of CBM6 for xylan |
| Variant of CBM6 | Affinity constant \( \text{Kd} \) |
|---|---|
| Wild type | 437 |
| Y34A | 52 |
| N120A | 3 |
| W92M | 8 |
| Y40M | 513 |
| R72M | 474 |
| Y112M | 520 |

**a Affinity determined by quantitative AGE as described under “ Experimental Procedures.”**
29. Mori, S., Abeygunawardana, C., Johnson, M., and Van Zijl, P. C. M. (1995) J. Magn. Reson. 108, 94–98
30. Grzesiek, S., and Bax, A. (1992) J. Magn. Reson. 96, 432–440
31. Bax, A., and Ikura, M. (1991) J. Biomol. NMR. 1, 99–104
32. Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993) J. Magn. Reson. 101, 333–337
33. Muhandiram, D. R., and Kay, L. E. (1994) J. Magn. Reson. 103, 203–216
34. Grzesiek, S., and Bax, A. (1992) J. Magn. Res. 99, 201–207
35. Wuthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley Interscience, New York
36. Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A. J., and Kilburn, D. G. (2001) Biochemistry 40, 6240–6247
37. Sorimachi, K., Jacks, A. J., Le Gal-Coeffet, M.-F., Williamson, G., Archer, D. B., and Williamson, M. P. (1997) Structure 5, 647–661
38. Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 12229–12234
39. Holm, L., and Sander, C. (1994) Nucleic Acids Res. 22, 3660–3669
40. Spurway, T. D., Morland, C., Cooper, A., Sumner, I., Hazelwood, G. P., O’Dennell, A. G., Pickersgill, R. W., and Gilbert, H. J. (1997) J. Biol. Chem. 272, 17525–17530
41. Barton, G. J. (1993) Protein Eng. 6, 37–40
42. Boraston, A. B., McLean, B. W., Chen, G., Li, A., Warren, R. A. J., and Kilburn, D. G. (2001) Mol. Microbiol., in press
43. Xie, H., Gilbert, H. J., Charnock, S. J., Davies, G. J., Williamson, M. P., Simpson, P. J., Raghothama, S., Fontes, C. M. G. A., Dias, F. M., Ferreira, L. M. A, and Bolam, D. N. (2001) Biochemistry 40, 9167–9176
44. Brun, E., Johnson, P. E., Creagh, A. L., Tomme, P., Haynes, F. W. C. A., and McIntosh, L. P. (2000) Biochemistry 39, 2445–2458
45. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
46. Meritt, E. A., and Murphy, M. E. P. (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 869–873
47. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Proteins 11, 281–296