The family 11 Carbohydrate-Binding Module of Clostridium thermocellum Lic26A-Cel5E accommodates β-1,4 and β-1,3-1,4-mixed linked glucans at a single binding site

Ana L. Carvalho‡*, Arun Goyal§*, José A.M. Prates§, David N. Bolam†, Harry J. Gilbert†, Virginia M.R. Pires§, Luís M.A. Ferreira§, Antoni Planas¶, Maria J. Romão** and Carlos M.G.A. Fontes§**

‡ REQUIMTE / CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal
§ CIISA - Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Rua Prof. Cid dos Santos, 1300-477 Lisboa, Portugal
† School of Cell and Molecular Biosciences, University of Newcastle upon Tyne, The Agriculture Building, Newcastle upon Tyne NE1 7RU, United Kingdom
¶ Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, 08017 Barcelona, Spain

* Contributed equally to this work.
** To whom correspondence should be addressed:
Tel.: 351 212948310; Fax: 351 212948550; e-mail: mromao@dq.fct.unl.pt.
Tel.: 351 213652876; Fax: 351 213652889; e-mail: cafontes@fmv.utl.pt.

Running title: Structure and function of family 11 Carbohydrate Binding Module.

Coordinates: Coordinates and observed structure factor amplitudes have been deposited in the Protein Data Bank (accession code 1v0a).
SUMMARY

Modular glycoside hydrolases that attack recalcitrant polymers generally contain non-catalytic carbohydrate binding modules (CBMs), which play a critical role in the action of these enzymes by localising the appended catalytic domains onto the surface of insoluble polysaccharide substrates. Type B CBMs, which recognise single polysaccharide chains, display ligand specificities that are consistent with the substrates hydrolysed by the associated catalytic domains. In enzymes that contain multiple catalytic domains with distinct substrate specificities it is unclear how these different activities influence the evolution of the ligand recognition profile of the appended CBM. To address this issue we have characterised the properties of a family 11 CBM (CrCBM11) in Clostridium thermocellum Lic26A-Cel5E, an enzyme that contains GH5 and GH26 catalytic domains that display β-1,4 and β-1,3-1,4-mixed-linked endoglucanase activity, respectively. Here we show that CrCBM11 binds to both β-1,4 and β-1,3-1,4-mixed linked glucans, displaying $K_a$ values of $1.9 \times 10^5$ M$^{-1}$, $4.4 \times 10^4$ M$^{-1}$ and $2 \times 10^3$ M$^{-1}$ for Glc-β1,4-Glc-β1,4-Glc-β1,3-Glc, Glc-β1,4-Glc-β1,4-Glc-β1,4-Glc and Glc-β1,3-Glc-β1,4-Glc-β1,3-Glc, respectively, demonstrating that CBMs can display a preference for mixed linked glucans. To determine whether these ligands are accommodated in the same or diverse sites in CrCBM11 the crystal structure of the protein was solved to a resolution of 1.98 Å. The protein displays a β-sandwich with a concave side that forms a potential binding cleft. Site-directed mutagenesis revealed that Tyr22, Tyr53 and Tyr129, located in the putative binding cleft, play a central role in the recognition of all the ligands recognised by the protein. We propose, therefore, that CrCBM11 contains a single ligand binding site that displays affinity for both β-1,4 and β-1,3-1,4-mixed linked glucans.
INTRODUCTION

The major plant cell wall polysaccharides, cellulose and hemicellulose, are hydrolysed into soluble sugars by a consortium of microbial enzymes. In nature, these enzyme systems play a major role in recycling plant cell wall fixed carbon and are therefore of considerable biological and biotechnological importance. In general, glycoside hydrolases (GH) that catalyse plant cell wall degradation contain non-catalytic carbohydrate binding modules (CBMs) that interact with polysaccharides (1). Based on primary structure similarity, CBMs have been grouped into 39 different families ((2) http://afmb.cnrs-mrs.fr/CAZY). By mediating an intimate and prolonged association of the enzyme with its target substrate, CBMs enhance the activity of the catalytic module against insoluble polysaccharides (3,4). Therefore, CBMs play a major role in potentiating the capacity of cellulases and hemicellulases to degrade plant cell wall polysaccharides.

Structural studies have revealed that the topology of the ligand-binding site of CBMs varies. CBMs that interact with the flat surfaces of crystalline polysaccharides contain a planar hydrophobic carbohydrate binding site and are referred to as Type A CBMs (5-8). In contrast, CBMs that bind to single polysaccharide chains accommodate these ligands in extended clefts and are defined as Type B CBMs (9-12). Type B CBMs have been described that bind to a diversity of ligands, with some modules displaying plasticity in their capacity to accommodate heterogeneity in the sugar backbone both in terms of the identity of the saccharides and the nature of the linkage. Thus, polar residues in CBM29 are able to hydrogen bond to the axial O2 in mannose and the equatorial O2 in glucose enabling the protein to bind cellulose, glucomannan and mannan (13). Some CBMs from families 4 and 6 display affinity for the β-1,3-1,4- mixed linked glucans (barley β-glucan and lichenan) and β-1,4 glucans such as amorphous cellulose and cellohexaose (14,15). It is currently unclear
whether these proteins interact exclusively with the β-1,4 regions of the mixed linked polysaccharide or display specificity for sequences of glucose units that are linked by a mixture of β-1,3 and β-1,4 glycosidic bonds. In general the ligand specificity of Type B CBMs reflects the substrate hydrolysed by the associated catalytic modules (1). In enzymes that contain multiple catalytic domains with distinct substrate specificities, however, it is unclear how these different activities influence the evolution of the ligand recognition profile of the appended CBM.

To investigate whether CBMs can display enhanced affinity for mixed linked glucans and to interrogate the evolutionary pressure exerted on the ligand specificity of CBMs in enzymes that contain distinct catalytic modules, we have studied the structure and function of the family 11 CBM (CtCBM11) in Clostridium thermocellum Lic26A-Cel5E (16). This enzyme contains GH5 and GH26 catalytic domains that display β-1,4 and β-1,3-1,4-mixed-linked endoglucanase activity, respectively. Here we show that CtCBM11 does indeed display a preference for specific β-1,3-1,4-mixed linked glucans although the protein is able to bind to β-1,4 glucose polymers. The 3D structure of CtCBM11 in harness with mutagenesis studies reveals that the protein contains a single ligand binding cleft that can accommodate both β-1,3-1,4- and β-1,4- linked glucans.

**EXPERIMENTAL PROCEDURES**

*Protein Expression and Purification*

To express CtCBM11 in Escherichia coli, the region of the Lic26A-Cel5A gene (lic26A-cel5A) encoding the internal family 11 CBM was amplified from C. thermocellum YS genomic DNA using the thermostable DNA polymerase pFU Turbo (Stratagene). The
primers used, 5´-CTCGCTAGCGCTGCTGGTGAAAAATG-3´ and 5´-CACCTCGAGACACCAATCAGCTTGAT-3´, incorporated NheI and XhoI restriction sites, which are depicted in bold. The PCR product was cloned into pGEM T-easy (Promega) and sequenced to ensure that no mutations had occurred during the polymerase chain reaction. The recombinant pGEM T-easy derivative was digested with NheI and XhoI and the excised CtCBM11 encoding gene was cloned into the similarly restricted expression vector pET21a to generate pAG1. CtCBM11 encoded by pAG1 contains a C-terminal His$_6$-tag. E. coli BL21 harbouring pAG1 was cultured in LB containing 100 µg/ml ampicillin at 37 °C to mid-exponential phase (A$_{550}$ 0.6) at which point isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and the cultures were incubated for a further 5 hours. Cells were collected by centrifugation and the cell pellet was resuspended in a 50 mM sodium Hepes buffer, pH 7.5, containing 1 M NaCl and 10 mM imidazole. For biochemical assays, recombinant CtCBM11 was purified by immobilized metal ion affinity chromatography as described previously (17).

Source of sugars used

All soluble polysaccharides were purchased from Megazyme International (Bray County Wicklow, Ireland), except oat spelt xylan and hydroxyethylcellulose, which were obtained from Sigma. Cellobioisaccharides were from Seikagaku Corp. (Japan). Avicel (PH101) was obtained from Serva while acid-swollen cellulose was prepared as described previously (18). The mixed linked glucooligosaccharides Glc-β1,4-Glc-β1,4-Glc-β1,3-Glc and Glc-β1,3-Glc-β1,4-Glc-β1,3-Glc were synthesised by the method of Faijes et al. (19).
**Affinity gel electrophoresis**

The affinity of CtCBM11 for a range of soluble polysaccharides was determined by affinity gel electrophoresis (AGE). The method was essentially as described by Tomme et al. (20) using the polysaccharide ligands at a concentration of 0.1 % (w/v). Electrophoresis was carried out for 4 h in native polyacrylamide gels containing 10% (w/v) acrylamide. The non-binding negative control, was bovine serum albumin. Quantitative assessment of binding was carried out as described previously (21).

**Binding to insoluble polysaccharides**

Qualitative assessment of CtCBM11 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 % (v/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 4 °C with gentle shaking, after which time the insoluble ligand was precipitated by centrifugation at 13000 x 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 analysed by SDS-PAGE using a 14 % acrylamide gel. Controls containing protein but no polysaccharide were performed in parallel to ensure no precipitation occurred.

**Isothermal titration calorimetry (ITC)**

ITC measurements were made at 25 °C for all ligands except lichenan, which was also measured at 60 °C in order to determine the ΔC_p, following standard procedures (22) using a Microcal Omega titration calorimeter. The proteins were dialyzed extensively against 50 mM
sodium phosphate buffer, pH 7.0, and the ligand was dissolved in the same buffer to minimize heats of dilution. During a titration experiment the protein sample stirred at 300 rpm in a 1.4331ml reaction cell was injected with 30 successive 10µl aliquots of polysaccharide (1.5-5 mg ml\(^{-1}\)) or oligosaccharide (1-10 mM) ligand at 200s intervals. The molar concentration of \(Ct\)CBM11 binding sites present in the polysaccharide ligands was determined as described previously (11). Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single-site binding model (Microcal Origin, Version 5.0). The fitted data yield the association constant \((K_a)\), number of binding sites on the protein \((n)\), and the enthalpy of binding \((\Delta H)\). Other thermodynamic parameters were calculated using the standard thermodynamic equation \(-RT \ln K_a = \Delta G = \Delta H - T \Delta S\). Titrations were carried out in triplicate for most ligands, and the errors are the S.D. of the mean of these replicates.

**Site-directed mutagenesis**

Mutants of \(Ct\)CBM11 were generated using the PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer’s instructions. The primers used to generate these mutants were as follows: TGGGGTTTCAGCCTCCGGTGAAAGGTGC and TGCACCTTCACCAGGGCTGAACCCCA, Y22A; GGGACAACGCGACGCGCTGGGGGAACAGTATAC and GTATACTGTCCCAGCGGCGCTGGGTGTCCC, Y53A; AGACTTGATGCTCAGCCGTGGGACAG and CTGTCCAGGCGGCTGAGCATCAAGTCT, Y129A; CAATTCACTTCATGGCTGCCAACAATGTCTG and CGACTTTGATTGCAGCCATGAAGTGAATTG, Y152A. The mutated DNA sequences were sequenced by MWG (Germany) to ensure that only the appropriate mutations had been incorporated into the nucleic acid.
**Production and purification of seleno-L-methionine containing CtCBM11**

The methionine auxotroph *E. coli* B834 (DE3), transformed with pAG1, was cultured at 37°C in one litre of culture medium as described by Charnock *et al.* (9). Expression of CtCBM11 was induced by the addition of 1 mM IPTG when cells were at mid-exponential growth phase. Cells were incubated at 37°C for a further 16 h, after which time were collected and the recombinant protein purified by affinity chromatography as described above. For crystallization trials, gel filtration was included as a further purification step. The enzyme was buffer exchanged, using a PD-10 Sephadex G-25M gel filtration columns (Amersham Biosciences), into 50 mM Hepes buffer, pH 7.5, containing 200 mM NaCl (Buffer A), concentrated to 20 mg/ml with Amicon 10 kDa molecular-weight centrifugation membranes, and subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (Amersham Biosciences) with protein eluted at 1 ml/min in Buffer A. Purified enzyme was concentrated, as described before, washed three times with 5 mM DTT using the same centrifugal membranes, and the final protein concentration was adjusted to 50 mg/ml.

**Crystallization and data collection**

Crystals of seleno-L-methionine-containing protein were grown by vapor-phase diffusion using the hanging drop method with an equal volume (1 µl) of protein (50 mg/ml in water) and reservoir solution (20-24 % PEG 4000, 0.1 M sodium acetate pH 4.6 and 0.2 M ammonium sulfate). Crystals grew over a period of 3 days and were harvested in a buffer containing 30-35% PEG 4000 and 0.2 M ammonium sulphate. The crystals were flash-frozen in liquid nitrogen using 30% (v/v) glycerol as cryoprotectant added to the harvesting solution. A three-wavelength MAD data set of one seleno-methionine derivative crystal was collected on beamline BM30 at the European Synchrotron Radiation Facility (Grenoble, France). The
crystal was maintained at 110 K by using a nitrogen cryostream. The CtCBM11 crystal belonged to space group P2_12_12 with cell constants \( a = 75.1 \text{ Å}, b = 50.9 \text{ Å} \) and \( c = 40.9 \text{ Å} \), which corresponds to a calculated Matthews coefficient of \( 2.0 \text{ Å}^3 \text{Da}^{-1} \) and a solvent content of 36.5%. Data were processed with the HKL package (23), and the statistics are shown in Table 1.

**Phasing, model building and refinement**

Four selenium sites were located and refined with SOLVE (24), using the peak (0.9810 Å) and the edge (0.9813 Å) data sets, which showed a high correlation of anomalous differences. This resulted in phases to 1.98 Å with a figure of merit of 0.62, which were further improved by solvent flattening with DM (25). Automated model building using warpNtrace (26) generated a model with \( R = 37.4\% \). This initial model consisted of 163 of a total of 178 amino-acid residues and did not include the CtCBM11 primary sequence. Iterative model building with TURBO (27), together with refinement in REFMAC5 (28) and incorporation of the CtCBM11 primary sequence, resulted in a final model with \( R = 19.3\% \) (\( R_{\text{free}} = 23.2\% \)). This final model includes 170 amino-acid residues (five of which are seleno-methionines), two calcium ions, two sulphate, and 182 water molecules. Residues Asp 79, Gly 80 and Ser 81, belonging to a disordered loop, were not included in the final model, as well as Se-Met 1 and Ala 2 of the N-terminus and 3 histidine residues of the C-terminus, part of the His tag.

**RESULTS AND DISCUSSION**

**Ligand specificity of CtCBM11**

The family 11 CBM of endoglucanase F of *Fibrobacter succinogenes* S85 was previously shown to bind Avicel (29). It was based on this observation that members of this family were classified as CBMs. To characterize the ligand binding specificity of CtCBM11, the protein
module was expressed as a discrete entity in *E. coli* and purified to electrophoretic homogeneity. The affinity of CtCBM11 for barley β-glucan, lichenan, glucomannan, hydroxyethyl cellulose and oat spelt xylan was determined by quantitative AGE. The data showed that CtCBM11 displays highest affinity for β-1,3-1,4-glucons, while exhibiting significantly weaker binding to hydroxyethyl cellulose, glucomannan and oat spelt xylan (see Table 3). CtCBM11 did not associate with arabinan, galactomannan, laminarin, rhamnogalacturan, glucuronoxylan or rye-arabinoxylan.

To explore the ligand specificity of CtCBM11 in more detail, the binding of the protein to polysaccharides and oligosaccharides was evaluated using ITC. Example titrations, displayed in Figure 1, reveal sigmoidal titration curves for β-1,3-1,4 and β-1,4-glucooligosaccharides as well as the mixed linked polysaccharides barley β-glucan and lichenan, enabling the $K_a$, $\Delta H$ and stoichiometry of binding to be accurately determined for these ligands (Table 2). The data show that CtCBM11 displays similar affinity for cellohexaose, cellopentaose and cellotetraose. Although the protein retains significant affinity for cellotriose ($K_a \ 1.6 \times 10^4$), binding to cellobiose is too weak to accurately quantify by ITC ($\sim 1.3 \times 10^3$). These data suggest that CtCBM11 contains at least 4 sugar binding sites, which is consistent with the ligands accommodated by other Type B CBMs. Interestingly, the protein displays approximately 4-fold higher affinity for the two β-1,3-1,4-mixed linked glucan polymers indicating that the CBM displays a preference for a β-1,3-linked glucose in at least one sugar binding site. To investigate this bond preference in more detail the capacity of CtCBM11 to bind a range of oligosaccharides was assessed. The protein displayed no significant affinity for laminohexaose, a β-1,3-linked oligosaccharide of glucose, indicating that not all the sugar binding sites in the CBM can accommodate β-1,3-linked glucose residues (data not shown). CtCBM11, however, bound to the mixed linkage tetraoligosaccharide Glc-β1,4-Glc-β1,4-
Glc-β1,3-Glc approximately four times more tightly than cellotetraose, supporting the view that the protein displays a preference for a β-1,3-linked glucose in at least one subsite. A second mixed linkage glucotetra-oligosaccharide, Glc-β1,3-Glc-β1,4-Glc-β1,3-Glc, bound only weakly to CtCBM11 ($K_a \sim 2 \times 10^3$; affinity too low to accurately quantify by ITC) suggesting that the protein may only be able to accommodate a single β-1,3-linked glucose. The ligand specificity of CtCBM11 is consistent with the observation that natural glucose polymers containing mixed linkages of β1,4 and β1,3-glycosidic bonds, do not have the alternating substructure of Glc-β1,3-Glc-β1,4-. Indeed lichenan and cereal β-glucans contain approximately 2.3-3 β1,4 bonds for every β1,3 linkage, with fine structures comprising Glc-β1,4-Glc-β1,3-Glc, Glc-β1,4-Glc-β1,3-Glc and even Glc-β1,4-Glc-β1,4-Glc-β1,4-Glc-β1,3-Glc that are linked by β1,4 bonds. Thus the lack of specificity for alternating Glc-β1,3-Glc-β1,4- linkages is consistent with their absence from the target polysaccharides, of mixed linked β-glucans. The binding of CtCBM11 to its ligands is associated with negative enthalpy and entropy. The thermodynamics of the interaction of CtCBM11 with glucose polymers is typical of the binding of proteins to soluble saccharides, which is invariably enthalpically driven, with an unfavorable entropic contribution. The heat capacity of CtCBM11 for lichenan revealed a value of -74 cal mol$^{-1}$ K$^{-1}$. Negative values for ΔCp are often associated with the burial of significant apolar surfaces upon ligand binding. This is entirely consistent with the important role aromatic amino acids play in the interaction of CtCBM11 with all its ligands (Table 3 and see below).

The ligand specificity of CtCBM11 mirrors the substrate specificity of the cognate enzyme Lic26A-Cel5E, which displays both β-1,3-1,4-mixed linked and β-1,4-linked endoglucanase activity, mediated by the Lic26A and Cel5E catalytic modules, respectively (manuscript in preparation). It should be appreciated, however, that Lic26A-Cel5E is a component of a large
*C. thermocellum* extracellular multiprotein complex, termed the “cellulosome”, that contains numerous hydrolytic enzymes including several additional endo-β-1,4-glucanases and lichenases (30). Therefore, interpreting the properties of CtCBM11 exclusively within the context of the activities displayed by Lic26A-Cel5E may be naive, as the ligands accommodated by the module may potentiate glycosidic bond cleavage by other enzymes within the complex that display substrate specificities similar to Lic26A and Cel5E. It should be noted, however, that in modular enzymes within the cellulosome the ligand specificities of CBMs are similar to the substrate specificity of the appended catalytic domains. Thus, the ligands recognised by CBMs in *C. thermocellum* enzymes are not influenced *per se* by their location within the bacterium’s multicellulase/hemicellulase complex, suggesting that the specificity of these modules is determined by the substrates hydrolysed by the cognate enzyme.

**Crystal structure of CtCBM11 and its similarity with other CBMs**

The structure of the CtCBM11 was solved using MAD methods with a seleno-methionine derivative crystal. Experimental phases were obtained to a resolution of 1.98 Å, which generated electron density maps of sufficiently high quality for automatic interpretation. Statistics of the diffraction data and the final model are presented in Table 1. The CtCBM11 model consists of 170 amino-acid residues from a total of 178 (Figure 2). Residues Se-Met 1, Ala 2, Asp 79, Gly 80, Ser 81, His 176, His 177 and His 178 were excluded from the final model due to disorder. The model includes 182 water molecules, two sulphate molecules and two calcium ions (Ca1 and Ca2). The CtCBM11 structure reveals a classical distorted β-jelly roll fold consisting of two six-stranded anti-parallel β-sheets, which form a convex side (β-strands 1, 3, 4, 6, 9 and 12) and a concave side (β-strands 2, 5, 7, 8, 10 and 11). Residues Phe 120, Ser 121 and Ser 122 define a 3<sup>10</sup>-helix. The core of the β-barrel is highly hydrophobic.
and includes seven phenylalanine and six tryptophan residues. The coordination of the two calcium ions is illustrated in Figure 3. Ca1 is coordinated by side chain O atoms of Glu 91 (bidentate coordination from O\(_{\varepsilon 1}\) and O\(_{\varepsilon 2}\)), Glu 101 (O\(_{\varepsilon 1}\)), Asp 135 (O\(_{\delta 1}\) and O\(_{\delta 2}\)), Ser 137 (O\(_{\gamma}\)), Asp 141 (O\(_{\delta 2}\)) and the main chain O atom of Thr 139. Although calcium was not included in purification and crystallization, a second binding site (Ca2) was identified in \(\text{CtCBM11}\). Ca2 also shows octahedral coordination and is bound to residues Asp 12, Thr 38 and Asn 40 (main chain O atoms), Glu 14 (O\(_{\varepsilon 1}\)) and Asp 163 (O\(_{\delta 1}\) and O\(_{\delta 2}\)). One water molecule completes the Ca2 coordination sphere. The distances between the ligands and the calcium ions varied from 2.3 to 2.6 Å.

Similar to \(\text{CtCBM11}\), representatives of CBM families 4, 6, 9 and 22 also bind calcium ions. Usually these ions are solvent inaccessible and are suggested to stabilize the protein fold (31). The importance of calcium in maintaining the structural integrity of a family 4 CBM, which also contains two binding sites for the metal ion, was demonstrated by showing that removal of the ions from the xylan-binding module reduced the melting temperature by 23 ºC (32). Recently, it was shown that in CBMs from families 35 and 36, calcium plays a direct role in ligand binding (22, 33). The location of the two calcium-binding sites in \(\text{CtCBM11}\), however, is distant from the carbohydrate-binding cleft, supporting a structural role for the metal.

As in other CBMs that bind individual polymer chains, the concave side of \(\text{CtCBM11}\) forms a cleft. In \(\text{CtCBM11}\), this surface depression is defined by polypeptide stretches Gly 20 – Glu 25, Asp 51 – Ser 59, Glu 84 – Glu 91, Gly 98 – Ile 107, Phe 123 – Gly 133 and Asp 146 – Asn 154, and strongly resembles the equivalent grooves observed in CBM4 (34), CBM6 (35), CBM15 (11), CBM17 (31), CBM22 (9,36), CBM27 (10) and CBM29 (13). The side chains of Tyr 22, Tyr 53, Tyr 129, Tyr 152, Arg 86, Arg 126, His 102, His 149, Asp 51, Asp 99, Asp
Asp 146, Ser 59, Ser 106, Ser 147 and seleno-methionines 88 and 151 decorate the cleft of CtCBM11. As a result of space group constraints, the entrance to this cleft is partially occupied by the C-terminal residues Glu 172*, His 173* and His 174* of a symmetry-related molecule (Figure 2b). Residue Glu 172*, belonging to the His-tag, forms direct hydrogen contacts with residues Arg 126 and Tyr 152. Furthermore, the side chain of Glu 172* lies between the aromatic rings of Tyr 53 and Tyr 129, which are nearly coplanar and approximately 7.5 Å apart. Other contacts are mediated by bridging water molecules. Residue Ser 3*, from the N-terminus of the same symmetry-related molecule, makes a direct hydrogen bond contact with an O\(_\varepsilon\) of Glu 25, one of the residues that narrows the entrance to the cleft. The narrow cleft appears to be the likely carbohydrate-binding site in the protein, with residues Tyr 22, Tyr 53, Tyr 129 and Tyr 152 playing a key role in ligand recognition.

A search for proteins structurally homologous to CtCBM11 was done using the DALI database (37). The results show that CtCBM11 is topologically similar to several carbohydrate binding domains from different organisms. These include the crystallographic structures of family 29 CBM from Piromyces equi NCP1 (PDB entry 1gwk), the N-terminal domain of chondroitin sulphate ABC lyase I from Proteus vulgaris (PDB entry 1hn0), the family 4 CBM from Thermotoga maritima Lam16A (PDB entry 1gui), the family 22 CBM from C. thermocellum Xyn10B (PDB entry 1dyo), the family 15 CBM from Cellvibrio japonicus Xyn10C (PDB entry 1gny), the family 27 CBM from T. maritima Man5A (PDB entry 1of3), the ligand-binding domain of the EPHB2 receptor tyrosine kinase from Mus musculus (PDB entry 1nuk), the family 17 CBM from Clostridium cellulovorans Cel5A (PDB entry 1j83) and the family 6 CBM from Cellvibrio mixtus Cel5A (PDB entry 1uxz). They all share a structurally conserved β-barrel fold and the main differences are restricted to solvent-exposed loops. The r.m.s. deviations of these structures when superimposed on
CtCBM11 vary in the range of 2.2 to 3.5 Å. A single calcium binding site is reported in the structures of CBM4 and CBM27 from *T. maritima*, CBM6 and CBM22 from *C. thermocellum* and CBM17 from *C. cellulovorans*, and this corresponds to position Ca2 in CtCBM11, although the coordinating ligands differ. In all of these related structures, the proposed carbohydrate-binding cleft is also conserved and results from the distortion of the β-barrel. CtCBM11 shows a pronounced cleft, whose entrance is narrowed by loops Lys 92 – Glu 101 and Arg 125 – Thr 139. In the structures of CBM6 and CBM17 this narrowing of the cleft is comparatively more pronounced than in CtCBM11, also due to loop structures.

It is now clear that the β-sandwich fold presented by CtCBM11 is the most common among CBMs and that families 2, 3, 4, 6, 9, 15, 17, 22, 27, 28 and 29, together with the CBM family 11 module described here, constitute a fold superfamily as first suggested by Sunna *et al.* (38). There is a strong possibility that all these CBM families might share a common evolutionary origin. Considering that in evolutionary time catalytic modules might have appeared before the non-catalytic CBMs, the glycoside hydrolases are strong candidates for the common ancestor of the β-sandwich super-family of CBMs. A very subtle change in the catalytic domain, such as the mutation of the nucleophile in a GH, could directly generate a CBM. Indeed it is established that very subtle amino acid substitutions can dramatically alter the ligand specificity of CBMs and therefore the accumulation of those mutations could have contributed for the evolution of the ligand specificities required by the various enzymes (39). CtCBM11 shows significant structural homology to several glycoside hydrolases including endo-β-1,4-glucanase from *Streptomyces lividans* (PDB entry 1nlr), bovine lysosomal α-mannosidase (PDB entry 1o7d), domain II of 4-α-glucanotransferase from *Thermococcus litoralis* (PDB entry 1k1w), the jelly-roll like domain of phosphomanose isomerase from *Candida albicans* (PDB entry 1pmi), human β-glucuronidase (PDB entry 1bhg) and 1,4-β-D-
glucanelllobiohydrolase I from *Trichoderma reesei* (PDB entry 1cel). In the endo-glucanase structure from *S. lividans* (40), a modified glucose trimer (2-deoxy-2-fluorocellotrioside) was identified in the carbohydrate-binding cleft (PDB entry 2nlr), covalently linked to a glutamate residue, forming a glycosyl-enzyme intermediate. Two asparagines, one histidine and one tryptophan residues are H-bonded to the trisaccharide molecule.

**Identification of CtCBM11 binding cleft**

As described above, inspection of the surface of the cleft of CtCBM11 reveals the presence of several aromatic residues, which are conserved in the other three CBM11 members (Figure 4). Of these amino acids only Tyr 129 is invariant in all four sequences. Tyr 22, Tyr 53, and Tyr 152 are conserved in at least two out of the four sequences available in this family. Lack of conservation of key residues involved in ligand binding is not completely surprising considering the plasticity of specificities often observed in Type B CBM families. However, it is clear that all the identified residues are solvent exposed in CtCBM11 and could interact with the target ligands via hydrophobic stacking interactions. To evaluate the importance of residues Tyr 22, Tyr 53, Tyr 129 and Tyr 152 in ligand binding, the mutant proteins Y22A, Y53A, Y129A and Y152A were produced and their biochemical properties compared to wild type CtCBM11. The data (Table 3 and Figure 5) show that Y22A displayed no detectable affinity for the ligands analysed, while Y53A and Y129A exhibited significantly reduced binding for these polysaccharides. The similar proportional reduction in the affinities of the various mutants against all the ligands, compared to wild type CtCBM11, suggests that these three residues play an important role in the recognition of glucomannan, xylan, β-1,4- and β-1,4-β-1,3-linked glucans. By contrast, Y152A displays similar affinity to wild type CtCBM11 for these soluble ligands indicating that Tyr 152 does not contribute to ligand recognition. Qualitative binding assays were also carried out to probe the capacity of CtCBM11 to interact
with insoluble polysaccharides. The data, displayed in Figure 6, demonstrate that the family 11 CBM of Lic26A-Cel5A is able to interact with Avicel and ASC. The mutants Y22A, Y53A and Y129A display dramatically reduced binding to the insoluble ligands, suggesting that cellulose binds to the CBM11 within the extended cleft. The topography of the cleft precludes binding to the planar face of crystalline regions of cellulose, and thus the CBM is more likely to interact with single cellulose chains in the unstructured regions of the two forms of cellulose. The mutation of Tyr 152 to alanine had no apparent effect on binding to the insoluble ligands. Taken together, these results suggest that the groove located in CtCBM11 is the carbohydrate-binding cleft and that residues Tyr 22, Tyr 53 and Tyr 129 play a major role in the interaction of CtCBM11 with its ligands. Structural studies of family 6 CBMs (CBM6), which are closely related to the CBM11 family, demonstrate that these proteins contain two topologically and spatially different clefts that display different ligand specificities (35). Recently, one of the two binding clefts (cleft B) in CmCBM6 from Cellvibrio mixtus endoglucanase 5A was shown to interact with β-1,3-1,4-linked and β-1,4-linked glucans with similar affinities, binding only weakly to β-1,3-glucan (41). The crystal structure of CmCBM6 showed that in cleft B, while only a β-1,4 glucose can be accommodated at two sugar binding subsites, the third subsite can interact with either β-1,4 or β-1,3-linked glucose molecule, while the fourth subsite displays exclusive specificity for a β-1,3-linked glucose (35). Analysis of the CtCBM11 structure, together with the mutagenesis data, indicates that unlike CBM6, the family 11 module contains only a single ligand binding site, in common with the majority of CBMs characterised to date. In addition, the preference displayed by CtCBM11 may reflect a general feature of CBMs that recognise β-1,3-1,4-mixed linked glucans. Thus, the capacity of the Clostridium CBM to bind β-1,4 and β-1,4-β-1,3-mixed ligands reflects plasticity at a single binding site rather than the presence of multiple carbohydrate binding regions.
Conclusions

This report reveals that CtCBM11 binds preferentially to β-1,3-1,4 glucans while displaying considerable affinity for β-1,4 linked glucose polymers and no affinity for β-1,3 glucans. This ligand specificity reflects the substrate specificity of the associated GH26 and GH5 catalytic modules that act specifically on β-1,3-1,4 or β-1,4 glucans, respectively (unpublished data). The crystal structure of CtCBM11 combined with site-directed mutagenesis experiments show that the deep groove identified in the protein structure constitutes the ligand-binding site and that residues Tyr 22, Tyr 53 and Tyr 129 play a central role in carbohydrate recognition. Tyr 53 and Tyr 129 are located directly opposite to each other in CtCBM11 cleft and are conserved in all CBM11 members, and thus are likely to stack against the two faces of a single sugar unit. Tyr 22, however, is only present in two of the four known CBM11 proteins. Whilst it is possible that this lack of conservation reflects differences in ligand specificity in other CBM11 modules, it is interesting that mutation of Tyr 22 caused a more pronounced reduction in the affinity for polysaccharides than Tyr 53 and Tyr 129. The data presented in this report demonstrate that the different substrate specificities displayed by the two catalytic domains in Lic26A-Cel5E both influence the ligand specificity of the appended CBM11. We also provide evidence that CBMs can display specificity for stretches of glucose residues that are linked by a specific combination of β-1,4 and β-1,3 glycosidic bonds.

Acknowledgments

The authors would like to thank the beamline scientists at BM30, ESRF Grenoble, for assistance during data collection.
REFERENCES


Figure legends

Figure 1. Isothermal titration calorimetry of wild-type CtCBM11 with oligo- and polysaccharides.

The upper parts of each panel show the raw heats of binding, whereas the lower parts are the integrated heats fit to a one-site binding model. Heats of dilution were measured independently and found to be negligible. G4G4G3G is the mixed linkage tetraooligosaccharide Glc-β-1,4-Glc-β-1,4-Glc-β-1,3-Glc.

Figure 2. Ribbon representation of the three-dimensional structure of CtCBM11.

a, Stereo representation of the overall structure of CtCBM11. The CtCBM11 structure consists of a distorted β-jelly roll fold composed of two six-stranded antiparallel β-sheets, which form a convex side (β-strands drawn in dark blue) and a concave side (β-strands drawn in light blue). The two calcium ions are indicated as green spheres. The concave side of CtCBM11 forms a cleft. Residues inside the putative binding cleft are shown as ball-and-stick models. Tyrosine residues are represented in red, arginines in yellow, histidines in pink and aspartates in orange. The picture was drawn with program MOLMOL (42). b, Stereo view of the CtCBM11 cleft, in the same orientation as in a, occupied by the C-terminus residues of a symmetry-related molecule. The residues are labelled in red and represented in ball-and-stick. The 2mFo-DFc electron density map around the molecule is shown in dark blue and contoured at 1.1σ. Residues inside the cleft (colour code) are shown as ball-and-stick models and labelled blue. The calcium-binding site, Ca1, is represented as a green sphere. The picture was produced with program TURBO-FRODO (27).
**Figure 3. The Ca$^{2+}$-binding sites of CtCBM11.**

Stereo view of the Ca$^{2+}$-coordination in CtCBM11 superimposed in the 2mFo-DFc electron density map, contoured at 1.3σ. The residues involved in calcium binding are represented as stick models and labelled black. The polypeptide chain atoms are represented in colour code and the Ca$^{2+}$ ions are shown as orange spheres. The pictures were produced with program TURBO-FRODO (27).

**Figure 4. Alignment and secondary structure of CBM11 members.**

CBM11 aligned primary sequences are from *Clostridium thermocellum* (CtCBM11), *Clostridium cellulolyticum* (CcCBM11), *Streptomyces avermitilis* (SaCBM11) and *Fibrobacter succinogenes* (FsCBM11). Identity to CtCBM11 is indicated with grey boxes. Residue numbers refer to the sequence of CtCBM11, which was purified and crystallized. Residues in CtCBM11 that have side chains inside the putative binding cleft are marked with dark grey boxes. The depicted secondary structure corresponds to CtCBM11. The sequence alignment was calculated with program CLUSTALW (43) and the picture was produced with program ALSCRIPT (44).

**Figure 5. Quantitative AGE with barley β-glucan as the ligand.**

Bovine serum albumin (lane S), CtCBM11 (lane 1), Y53A (lane 2), Y152A (lane 3), Y22A (lane 4) and Y129A (lane 5) were subjected to nondenaturing electrophoresis in gels containing a range of different concentrations of barley β-glucan. The percentage concentrations of the soluble polysaccharide in the example gels shown were: 0% (a), 0.001% (b), 0.002% (c), 0.004% (d), 0.006% (e), 0.008% (f), 0.01% (g) and 0.02% (h).
Figure 6. Binding of CtCBM11 and its mutants to insoluble polysaccharides as assessed by SDS-PAGE.

Protein was mixed with Avicel and acid swollen cellulose (ASC) and incubated for 2 h at 4°C. Unbound protein was removed after centrifugation (lane -) and the pellet was washed for three times in 50 mM Tris.HCl buffer, pH 7.5, containing 0.05 % (v/v) Tween 20 and 5 mM CaCl₂. Bound protein was eluted by boiling in 10% (w/v) SDS (lane +). The positions of the 30, 20 and 14 kDa markers are indicated in lane m.
Table I
X-ray data collection and structure refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Peak</th>
<th>Edge</th>
<th>Remote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength, Å</td>
<td>0.9810</td>
<td>0.9813</td>
<td>0.9763</td>
</tr>
<tr>
<td>Resolution of data (outer shell), Å</td>
<td>25.00 – 1.98 (2.02 – 1.98)</td>
<td>25.00 – 1.98 (2.02 – 1.98)</td>
<td>25.00 – 1.97 (2.01 – 1.97)</td>
</tr>
<tr>
<td>$R_{	ext{sym}}$ (outer shell), % $^a$</td>
<td>4.9 (8.7)</td>
<td>5.7 (9.8)</td>
<td>5.1 (9.9)</td>
</tr>
<tr>
<td>Mean I/$\sigma$(I)</td>
<td>30.3 (8.4)</td>
<td>21.2 (5.6)</td>
<td>20.9 (4.6)</td>
</tr>
<tr>
<td>Completeness (outer shell), %</td>
<td>99.5 (99.3)</td>
<td>99.4 (94.1)</td>
<td>99.0 (92.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>1.8</td>
<td>1.8</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Structure refinement

| No. protein atoms               | 1293            |
| No. solvent waters              | 182             |
| Resolution used in refinement, Å | 19.00 – 1.98    |
| Reflections                     | 10598           |
| $R_{	ext{work}} / R_{\text{free}}$ (%) $^b$ | 19.3 / 23.2 |
| Rms deviation 1-2 bonds (Å)     | 0.005           |
| Rms deviation 1-3 bonds (degrees) | 0.829         |
| Rms deviation chiral volume (Å³) | 0.057          |
| Avg B factors (Å²)              |                 |
| main-chain                      | 16.1            |
| side-chain                      | 16.3            |
| Ca$^{2+}$ (1)                   | 11.9            |
| Ca$^{2+}$ (2)                   | 15.4            |
| SO$_4^{2-}$ (1)                 | 35.8            |
| SO$_4^{2-}$ (2)                 | 44.0            |
| water molecules (182)           | 28.9            |

$^a$ $R_{\text{sym}} = \Sigma |I-<I>| / \Sigma <I>$, where $I$ is the measured intensity of each reflection and $<I>$ is the intensity averaged from multiple observations of symmetry-related reflections.

$^b$ $R_{\text{work}} = \Sigma |F_{\text{calc}}| - |F_{\text{obs}}| / \Sigma |F_{\text{obs}}| \times 100$, where $F_{\text{calc}}$ and $F_{\text{obs}}$ are the calculated and observed structure factor amplitudes, respectively ($R_{\text{free}}$ is calculated for a randomly chosen 5% of the reflections.)
Table II
Quantitative assessment of CtCBM11 binding to oligo- and polysaccharides as determined by ITC

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Temperature (K)</th>
<th>$K_a \times 10^4$ (M$^{-1}$)</th>
<th>$\Delta G$ (kcal mol$^{-1}$)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$T\Delta S$ (kcal mol$^{-1}$)</th>
<th>$n^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lichenan</td>
<td>298.15</td>
<td>30.1 (±0.4)$^b$</td>
<td>-7.5 (±0.1)</td>
<td>-10.4 (±0.2)</td>
<td>-2.9 (±0.2)</td>
<td>1.0 (±0.0)</td>
</tr>
<tr>
<td>Lichenan</td>
<td>333.15</td>
<td>5.3 (±0.1)</td>
<td>-7.2 (±0.0)</td>
<td>-13.0 (±0.0)</td>
<td>-5.8 (±0.0)</td>
<td>1.1 (±0.0)</td>
</tr>
<tr>
<td>β-glucan</td>
<td>298.15</td>
<td>27.1 (±0.5)</td>
<td>-7.4 (±0.1)</td>
<td>-11.2 (±0.3)</td>
<td>-3.8 (±0.2)</td>
<td>1.0 (±0.0)</td>
</tr>
<tr>
<td>Cellohexaose</td>
<td>298.15</td>
<td>7.8 (±0.1)</td>
<td>-6.6 (±0.0)</td>
<td>-9.5 (±0.2)</td>
<td>-2.9 (±0.2)</td>
<td>0.8 (±0.0)</td>
</tr>
<tr>
<td>Cellopentaose</td>
<td>298.15</td>
<td>5.9 (±0.3)</td>
<td>-6.5 (±0.0)</td>
<td>-8.7 (±0.3)</td>
<td>-2.2 (±0.3)</td>
<td>0.9 (±0.0)</td>
</tr>
<tr>
<td>Cellotetraose</td>
<td>298.15</td>
<td>4.4 (±0.8)</td>
<td>-6.3 (±0.1)</td>
<td>-9.8 (±0.1)</td>
<td>-3.5 (±0.1)</td>
<td>1.0 (±0.0)</td>
</tr>
<tr>
<td>G4G4G3G$^c$</td>
<td>298.15</td>
<td>19.2 (±1.5)</td>
<td>-7.2 (±0.1)</td>
<td>-10.2 (±0.1)</td>
<td>-3.0 (±0.1)</td>
<td>1.1 (±0.0)</td>
</tr>
</tbody>
</table>

$^a$ $n$ is the number of binding sites on the protein.
$^b$ The values in parenthesis are the standard deviations of replicate titrations.
$^c$ Mixed linkage glucotetraoligosaccharide: Glc-β-1,4-Glc-β-1,4-Glc-β-1,3-Glc.
Table III
Binding of wild-type CtCBM11 and its mutant derivatives to soluble polysaccharides quantified by AGE

<table>
<thead>
<tr>
<th>Ligand</th>
<th>( K_a (c^{-1})^a )</th>
<th>Wild-type</th>
<th>Y22A</th>
<th>Y53A</th>
<th>Y129A</th>
<th>Y152A</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )-Glucan</td>
<td>1194,9</td>
<td>15,5</td>
<td>74,2</td>
<td>56,2</td>
<td>1080,4</td>
<td></td>
</tr>
<tr>
<td>Lichenan</td>
<td>701,6</td>
<td>9,7</td>
<td>89,1</td>
<td>68,1</td>
<td>690,1</td>
<td></td>
</tr>
<tr>
<td>HEC</td>
<td>24,4</td>
<td>NB(^b)</td>
<td>NB</td>
<td>NB</td>
<td>53,6</td>
<td></td>
</tr>
<tr>
<td>Glucomannan</td>
<td>25,9</td>
<td>NB</td>
<td>NB</td>
<td>NB</td>
<td>20,4</td>
<td></td>
</tr>
<tr>
<td>Oat spelt xylan</td>
<td>17,5</td>
<td>NB</td>
<td>NB</td>
<td>NB</td>
<td>nd(^c)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)ac, ligand concentration, % (w/v).
\(^b\)NB, \( K_a \) below 2.
\(^c\)nd, not determined.