A Novel Carbohydrate-Binding Module from Sugar Cane Soil Metagenome Featuring Unique Structural and Carbohydrate Affinity Properties

Bruna Medeiros Campos1a, Marcelo Vizona Liberato2a, Thabata Maria Alvarez2a, Letícia Maria Zanphorlin2, Gabriela Cristina Ematsu2, Hernane Barud6, Igor Polikarpov4, Roberto Ruller2, Harry J Gilbert3, Ana Carolina de Mattos Zeri3 and Fabio Marcio Squina2*

1Laboratório Nacional de Biociências (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Caixa Postal 6192, CEP 13083-970, Campinas, São Paulo, Brazil.
2Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Caixa Postal 6192, CEP 13083-970, Campinas, São Paulo, Brazil.
3Laboratório Nacional de Luz Sincrotron (LNLS), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Caixa Postal 6192, CEP 13083-970, Campinas, São Paulo, Brazil.
4Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil.
5Institute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle-upon-Tyne NE 4HH, United Kingdom.
6Centro Universitário de Araraquara-UNIARA, BioPolMat, Araraquara-SP, Brazil.

Running title: A Novel CBM with both Type A and Type B characteristics

*Correspondence to: Fabio M Squina, Laboratório Nacional de Ciência e Tecnologia do Bioetanol – CTBE/CNPEN, Caixa Postal 6170, CEP 13083-970, Campinas, São Paulo, Brazil, Tel.: +55 19 3518 3111, Fax: +55 19 3518 3104, e-mail: fabio.squina@bioetanol.org.br

a These authors contributed equally to this work

Keywords: Carbohydrate-binding protein, X-ray crystallography, metagenomics, cellulose, biofuel

ABSTRACT

Carbohydrate-binding modules (CBMs) are appended to glycoside hydrolases and contribute to the degradation of complex recalcitrant substrates such as the plant cell wall. For application in bioethanol production, novel enzymes with high catalytic activity against recalcitrant lignocellulosic material are being explored and developed. In this work, we report the functional and structural study of CBM_E1, discovered through a metagenomics approach, which is the founding member of a novel CBM family, CBMxx. CBM_E1, which is linked to an endoglucanase, displayed affinity for mixed linked β-1,4- β-1,3-glucans, xyloglucan, avicel and cellooligosaccharides. The crystal structure of CBM_E1 in complex with cellopentaose displayed a canonical beta-sandwich fold comprising two beta sheets. The planar ligand binding site, observed in a parallel orientation with the beta strands, is a typical feature of Type A CBMs, although the expected affinity for bacterial crystalline cellulose was not detected. Conversely, the binding to soluble glucans was enthalpically driven, which is typical of Type B modules. These unique properties of CBM_E1 are at the interface between Type A and Type B CBMs.

INTRODUCTION

Carbohydrate-binding modules (CBM) are functionally and structurally discrete units that are linked to a range of carbohydrate active enzymes (CAZymes), although primarily glycoside hydrolases (GH) (1). CBM-containing GHs target insoluble polysaccharides exemplified by the plant cell wall (2). The major industrial interest in these enzymes is for the production of lignocellulosic-derived ethanol, which is a promising alternative to environmentally damaging and finite fossil fuels. The study and development of GHs and their
accessory modules, primarily CBMs, are crucial to overcome accessibility issues regarding the deconstruction of plant cell wall polysaccharides into their monomeric fermentable units (3,4).

CBMs are capable of binding to different carbohydrates (5). Although non-catalytic, CBMs can increase the enzymatic efficiency of its associated catalytic module against insoluble substrates through proximity effects (6).

Currently, CBMs are classified into 80 sequence-based families according to the CAZy database (7). In addition, it is proposed that CBM classification should also be based on the structure and binding specificity of these modules: ‘surface-binding to crystalline ligands’ or Type A, endo-single chain glycan binding or Type B and ‘exo-binding’ or Type C (1). Type A CBMs bind to insoluble crystalline carbohydrates, crystalline cellulose and/or chitin, and have a planar binding site, usually composed of three aromatic amino acids, displaying little or no affinity for soluble carbohydrates. Type B CBMs on the other hand, have a binding site shaped as a cleft and bind to a wide variety of glycans such as xylans, mannans, galactans, glucans of mixed linkage and non-crystalline cellulose. Aromatic amino acids also play an important role in ligand binding in Type B CBMs (2). The binding sites of Type C CBMs display a pocket topology explaining why they recognize the non-reducing termini of glycans.

Here we report the comprehensive structural and biochemical characterization of a founding member of a new CBM family, CBM_E1. The CBM comprises the C-terminal region of a GH family 5 (GH5) endoglucanase derived from the sugar cane soil metagenomic library (4). CBM_E1 shows biophysical properties that are characteristic of Type B CBMs, but contains a planar binding site, typical of Type A CBMs. Thus CBM_E1 displays unique properties that are at the interface between Type A and Type B CBMs.

RESULTS

**CBM_E1** is the C-terminal domain of CelE1, an endoglucanase derived from sugar cane soil metagenome – Previous functional screening of a metagenomic library using carboxymethyl cellulose (CMC) as substrate identified a clone encoding a 427 amino acid polypeptide (4). The full length CelE1 (accession number KF498957) contains a putative N-terminal signal peptide (prediction performed on Signal-Blast) (8), a catalytic module belonging to glycoside hydrolase family 5 (GH5), predicted on dbCAN web server (9), a serine-rich linker sequence and a C-terminal region of unknown function, which was defined here as CBM_E1 (Figure 1; accession number KJ917170). Based on BLASTp sequence comparison, the most similar protein is the C-terminal region (unknown function) of another GH5 cellulase from *Pseudomonas sp.* ND137 (GenBank accession number BAB79288.1), presenting 34% identity to CBM_E1.

Ligand binding properties of CBM_E1 – To evaluate whether CBM_E1 binds to polysaccharides, the protein module, comprising residues 335 to 427 of full length CelE1, was expressed in soluble form in *Escherichia coli* and purified to electrophoretic homogeneity. Isothermal titration calorimetry (ITC) was used to explore binding to potential soluble ligands. The data, reported in Table 1 and Figure 2, show that CBM_E1 bound with highest affinity for barley β-glucan, with a $K_a$ of ~ $10^4$ M$^{-1}$, whereas the interaction with xyloglucan was almost 10-fold weaker and the protein displayed no recognition of birchwood xylan. With respect to oligosaccharides, CBM_E1 displayed affinity for cellohexaose (C6), with a $K_a$ of $1.2 \times 10^4$ M$^{-1}$, and cellopentaose (C5), with $K_a$ of $4.3 \times 10^3$ M$^{-1}$. The calculated thermodynamic parameters indicate that protein binding to oligosaccharides is enthalpically driven, which is typical for Type B CBMs (10). Binding to mannohexaose (M6), xylohexaose (X6), cellotetraose (C4) and XXXG (heptasaccharide derived from xyloglucan, the X stands for a glucose decorated with xylose while G indicates an undecorated glucose) was not detected.

A pull down assay was performed using Avicel or BMCC (Bacterial Microcrystalline Cellulose) to determine whether CBM_E1 binds to insoluble and crystalline forms of cellulose. As observed in Figure 3, CBM_E1 bound to Avicel (the protein was present only in the insoluble fraction), but not to BMCC (CBM_E1 appeared only in the soluble fraction). Considering that
A novel CBM with both Type A and Type B characteristics

BMCC has a high degree of crystallinity, approximately 95% (11), and Avicel contains around 40% of amorphous regions (12), the results obtained suggest that CBM_E1 does not bind to crystalline cellulose but targets disordered regions of the polysaccharide.

The Structure of CBM_E1 – The native crystal structure of CBM_E1 was solved in the absence and presence of the ligand cellopentaose at 1.5 and 1.74 Å resolution, respectively. The initial phases were found by single-wavelength anomalous dispersion method (SAD) using the anomalous scattering of gadolinium. The derivative data set presented twelve heavy atoms with clear electron densities (data not shown). The statistics of data collection and refinement are described in Table 2.

The final structures presented a monomer in the asymmetric unit and all residues were built with the exception of the vector encoded N-terminal two and four residues from ligand-complexed CBM_E1 and apo CBM_E1, respectively, due to poor electron density. The structural alignment of apo and ligand-complexed protein resulted in a root mean square deviation (RMSD) of 0.19 Å, demonstrating that they are almost identical (Figure 4), and thus ligand binding did not induce conformational changes in the protein. Reflecting the high degree of structural similarity, only the cellopentaose-CBM_E1 complex was considered further.

Characteristic of CBMs (13), CBM_E1 has a globular β-sandwich fold composed of two antiparallel β-sheets with four and five β-strands connected by loops (Figure 4). β-Sheet 1 and 2 comprised strands β2, β4, β7 and β9, and strands β1, β3, β5, β6 and β8, respectively. A striking feature of CBM_E1 is a completely solvent exposed planar surface that is oriented parallel and close to the strands β4, β7 and β9 in β-sheet 2. This planar surface is dominated by three tryptophan residues, W375, W398 and W427. The side chains of W375 and W427 are oriented parallel with β-sheet 2, while the indole ring of W398 is positioned perpendicular to this secondary structural element. The planar surface is reminiscent with the binding sites of CBMs from type A families such as CBM2 (14), CBM3 (15), CBM5 (16) and CBM10 (17,18).

The ligand-complexed structure has an electron density that clearly represents the co-crystallized cellopentaose (Figure 5E), present in two conformations orientated 180° with respect to each other, perpendicularly to the rotational axis (Figure 5). This electron density is positioned at the interface in a 2-fold rotational symmetry axis, over the planar surface of the CBM. The conformation of the bound cellopentaose adopts a perfect 2-fold screw axis in which adjacent glucose molecules are orientated 180° with respect to each other. Thus, the ligand displays the conformation adopted by cellulose chains in the crystalline polysaccharide, and not the twisted helical structures displayed by cellobiose saccharides and β-glucans in solution. It is evident that the binding sites of both molecules are packing the cellopentaose as a sandwich and each CBM_E1 is able to bind to cellopentaose in both orientations (Figure 5).

The ligand makes extensive parallel hydrophobic contacts (CH-π interactions) with W375 and W427 and hydrogen bonds with W398 and K423 (Figure 5B). In one orientation (orientation 1) W375 and W427 interact with G2 and G4, respectively (G1 is the non-reducing glucose of cellopentaose and G5 is the reducing terminal glucose), the indole nitrogen of W398 makes a hydrogen bond with the glycosidic oxygen between G2 and G3 and the N-ε of K423 makes hydrogen bonds with oxygens O2 and O3 of G2. In the opposite orientation, (orientation 2) W375 and W427 interact with G4 and G2, respectively, W398 hydrogen bonds with the glycosidic O between G3 and G4, and K423 interacts with G4. As CBM_E1 interacts only with symmetrical regions of the ligand (does not interact with O6 or the endocyclic O), it is not entirely clear which orientation is biologically significant, or, indeed, whether the module can interact with both orientations of cellopentaose in vivo. In orientation 1, W375 and W427 interact with the α-face of G2 and G4, respectively, which appear to be more hydrophobic by virtue of both axial H-5 and H-3 hydrogens and the aliphatic C5-C6 bond. H-5 “points” into the π-electron cloud of the aromatic ring, hinting at “ring-current” hydrogen-bonding possibilities. Indeed, in all CBM-ligand complexes reported to date,
A novel CBM with both Type A and Type B characteristics

aromatic residues make apolar interactions exclusively with the \( \alpha \)-face of sugar rings.

Although CBM_E1 has a typical fold of such protein modules, the CBM_E1 does not have significant structural identity with any other CBM in the PDB, based on similarity search using the PDBRefine server. The most similar structures are a collagenase from Clostridium histolyticum, PDBid 1NQ1 (19), with a Q-score of 0.42, and a DNA-binding kinetochore protein from Saccharomyces cerevisiae, PDBid 2VPV (20), with a Q-score of 0.33. The low structural and sequence similarity with known CBM families indicates that CBM_E1 is the founding member of a novel CBM family, defined as CBMxx.

Mutation of tryptophans abolishes protein-substrate interaction – To evaluate the importance of tryptophan residues (W375, W398 and W427) and K423 in substrate binding, these amino acids were mutated to alanine. The pull down assay with Avicel showed that substitution of any of the three tryptophans abolished cellulose binding (Figure 6). Although mutation of K423 reduced binding, it did not abolish interaction, as can be seen in Figure 6 where the K423A mutant is found in soluble and insoluble fractions.

For further evaluation of the mutants, intrinsic tryptophan fluorescence (ITF) experiments were carried out (Table 3 and Figure 7). The data revealed that wild type CBM_E1 bound to cellohexaose and cellopentaose (\( K_{\text{APP}} = 4.3 \times 10^3 \) and \( 2.3 \times 10^3 \) M\(^{-1} \)), respectively, but not to celloctetraose, as observed in ITC experiments (shown above). The mutant K423A interacts only with cellohexaose (\( K_{\text{APP}} = 5 \times 10^3 \) M\(^{-1} \)), and the other mutants did not interact with any of the ligands. According to pull down assays and ITF, K423 influences substrate binding but it is not essential however, all tryptophans residues are necessary for ligand recognition.

CBM_E1 is a monomer in solution – The oligomeric state of CBM_E1 in solution was accessed by dynamic light scattering and analytical size exclusion chromatography. Both techniques indicated that CBM_E1 is a monomer in solution. Analytical size exclusion chromatography suggests a protein with 7.7 kDa and 1.3 nm radius (Figure 8A), and the DLS data indicated an 11.3 kDa protein with 1.6 nm radius (Figure 8B). Both experiments presented a single peak and agreed with the radius of gyration of the monomer in the crystallographic structure (1.18 nm) and calculated molecular weight (11.8 kDa). A CBM crystallographic ligand-mediated dimer has been observed in a Type A CBM, an integrated part of an expansin from Bacillus subtilis (10). According to our data, the dimerization observed in CBM_E1 structures is probably due to crystal packing.

DISCUSSION

This report describes the characterization of CBM_E1, which is the founding member of a novel CBM class. Despite the typical beta-sandwich fold, CBM_E1 displays no structural similarity to any CBM family in the CAZy database. The 3D-structure of CBM_E1 was solved in the presence and absence of cellopentaose and the data indicated that the CBM underwent no obvious conformation change upon ligand binding.

Regarding the structural properties of CBM_E1, the first striking characteristic is the orientation of the binding-site, which is parallel to the beta-strands. From the best of our knowledge, only Type A CBMs display this characteristic (Figure 9). Type A CBMs present a beta-sandwich fold, like a chitin-binding domain from Pyrococcus furiosus (21), whose binding-site is almost parallel (with a small slope) in relation to the beta-strands. The binding-site of Type B CBMs often consists of the variable loops site (VLS), which connects the beta-strands at one end of the beta-sandwich, or comprises the concave face site (CFS), whose beta-strands are perpendicular to the ligand chain (13).

The second striking feature of CBM_E1 is the completely solvent exposed planar conformation of the binding site, which is typical of Type A CBMs (2). The ligand-binding site of CBM_E1 is composed of three tryptophans and a lysine residue, and mutation of any of these aromatic residues abolished ligand binding. In Type A CBM binding sites, the side chain of the three aromatic residues are displayed in the same plane, resulting in CH-\( \pi \) interactions with the glucose rings from the ligand. In contrast, only two tryptophan residues from CBM_E1 (W375 and W427) adopt a planar conformation with the ligand. The third tryptophan
and the lysine residue form hydrogen bonds with cellopentaose, which can help to explain the enthalpically driven interaction of CBM_E1 according to ITC data. This finding corroborates the classification of CBM_E1 as a Type B CBM, in contrast to ligand-binding Type A CBM, which is entropically driven (10).

Although CBM_E1 has a typical Type A planar binding-site, its classification as a Type B CBM was defined based on its binding properties. The CBM_E1 binds to Avicel, which is composed of 60% crystalline cellulose (12), but not to BMCC, composed of a high content of crystalline regions, around 95% (11). Consequently, CBM_E1 probably binds to the amorphous region of Avicel and is not able to bind to crystalline cellulose. CBM_E1 also displays significant binding to barley β-Glucan and xylloglucan, as well to cellohexaose and cellopentaose, which are ligands compatible with Type B CBMs (1). The inability of CBM_E1 to bind on BMCC (crystalline cellulose model) may reflect the presence of only two tryptophan residues that make CH–π interactions with cellulose, while in Type A CBMs three aromatic residues are required to bind to crystalline ligands (17,22). Indeed, mutating any of the three aromatic residues in Type A CBMs greatly reduces, and in some cases completely abolishes ligand binding (14).

In conclusion, we characterized the structure and binding properties of CBM_E1, a novel CBM derived from soil metagenomics that represents the new CBMxx family. Although it is classified as a Type B CBM, due to its affinity for soluble carbohydrates and the enthalpically driven binding, CBM_E1 presents a binding-site structure that resembles Type A. It is possible that CBM_E1 binds to regions of paracrystalline cellulose and thus targets the cognate enzyme to areas of the substrate at the interface between crystalline and amorphous structures. Indeed, within type A and type B groups, different CBMs recognize distinct regions of crystalline and amorphous forms of cellulose (23,24). Thus, the specificity of CBM_E1 may contribute to the multiple CBM targeting roles required to fully deconstruct the myriad of structures present in plant cell wall cellulose.

MATERIALS AND METHODS

Sequence analysis of CBM_E1 gene – The CBM_E1 gene nucleotide sequence was deposited in the GenBank database (accession number KJ917170). Physical and chemical parameters were predicted using the ProtParam tool (25). The sequence of amino-acid residues from the CBM_E1 gene was aligned with reference sequences from the non-redundant NCBI database using the ClustalX 1.83 program (26).

Protein production and purification – The gene encoding CBM_E1 (KJ917170) was amplified by PCR using full-length CelE1, retrieved from a sugarcane soil metagenomics library as a template (4). The cloning in pET28a is described by Ref 27. The CBM_E1 gene was also cloned in pET41a using the CBM_E1 pET28a clone as the template and the forward (5’-CTGCCGGGATCCACGCATCATGCGGTAGC-3’) and reverse primers (5’-CGCGAGCTCGAGTTACCAGTTATCGAACTTCAC-3’) which contains a BamHI and XhoI restriction site (in bold), respectively. The 282 bp product and the expression vector were digested with BamHI and XhoI restriction enzymes. The ligation mixture was transformed in Escherichia coli DH5α competent cells and cloning was verified by PCR. The final construct in pET41a encodes CBM_E1 fused to both N-terminal GST and His-tags.

Recombinant His-tag CBM_E1 was expressed in E. coli strain Origami 2 (DE3) (Novagen) and purified as described previously (27). The recombinant CBM_E1 protein with GST-tag was expressed in E. coli strain BL21 (DE3). A single colony was used to inoculate a 10 ml Luria-Bertani (LB) starter culture supplemented with kanamycin (50 mg/ml) and this was used to inoculate 1 l LB medium. The bacteria were cultured at 37 °C and 250 rpm until the OD 600nm reached 0.6, followed by induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 hours at 16 °C. The cells were harvested by centrifugation at 5,800 x g for 30 min., suspended in binding buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl) and sonicated. CBM_E1 was purified from lysed cells by immobilized metal ion affinity chromatography (IMAC) using Talon resin (Clontech). After incubation, the beads were washed once with wash...
A novel CBM with both Type A and Type B characteristics

buffer A (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM Imidazol) and the protein eluted with an elution buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 100 mM Imidazol). Purified CBM_E1 GST-tag with 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 100 mM Imidazol and CBM_E1 His-tag with 20 mM Sodium phosphate pH 7.2, 50 mM NaCl were stored at 4 °C.

Crystallization and Data Collection –

Crystallization experiments were performed manually (CBM_E1 His-tag with the tag removed), using the hanging drop vapor-diffusion method at 18 °C. The crystals of apo-CBM_E1 (9 mg/ml) were obtained in 0.1 M CAPS pH 10.5, 0.2 M Lithium sulfate, 2 M Ammonium sulfate. CBM_E1 in complex with cellopentaose (C5) (molar ratio of 1 CBM_E1: 2 C5) was crystallized at 6 mg/ml in 4M sodium formate. The crystals were soaked in a cryoprotection solution (15% glycerol and crystallization solution) and flash cooled in a stream of gaseous nitrogen at 100 K. For derivatization, CBM_E1-C5 crystal was soaked in a cryoprotection solution containing 1 M Gadolinium sulfate. The X-ray diffraction data were collected in the MX2 beamline (28) of the Brazilian Synchrotron Light Laboratory (LNLS, Campinas-SP) using a MAR 225 detector.

Collected data were processed with iMOSFLM (29) or XDS (30) and AIMLESS (31). The structure was solved using SAD method AUTOSOL (32) from PHENIX (33). A single solution for space group I23 was obtained for the derivatized data and for CBM_E1-C5. For the Apo structure, the space group I21 was found. The models were adjusted and refined using REFMAC5 (34) interspersed with model adjustment in COOT (35) to give the final model to a resolution of 1.5 Å for CBM_E1-C5. The Apo-CBM_E1 structure was solved by molecular replacement, using CBM_E1-C5 as model and PHASER MR (36). The final structures were deposited in the Protein Data Bank with the following IDs: CBM_E1/Apo PDBiD 5KLC and CBM_E1/C5 PDBiD 5KLE; CBM_E1/Gd PDBiD 5KLF.

Site-directed Mutagenesis –

Site-directed mutagenesis was carried out employing PCR-based Q5® Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer’s instruction, using CBM_E1 cloned in pET28a as the template. Trp375, Trp398, Lys423 and Trp427 were replaced by Ala. The mutated DNA sequences were sequenced to ensure that only the appropriate mutations had been incorporated into the amplified DNA. The mutant proteins W375A, W427A and K423A were expressed as 6xHis tag fusions and purified as described by Ref 27. W398A mutant protein was expressed in E. coli strain Origami 2 (DE3) (Novagen). The bacteria were cultured at 23 °C and 200 rpm for 16 h, followed by induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 6 hours at 18 °C. W398A protein purification followed the protocol described previously (27).

BMCC preparation –

Bacterial cellulose membranes were produced as described previously (37). Briefly, cultures of Gluconacetobacter hansenii (strain ATCC 23769) were incubated for 96 h at 28 °C in trays measuring 30 cm × 50 cm, using a static culture liquid medium (HS) composed of 50 g/l glucose, 4 g/l of yeast extract, 0.73 g/l of MgSO\(_4\).7H\(_2\)O, 2 g/l KH\(_2\)PO\(_4\), 20 g/l ethanol and distilled water. Then, BC membranes obtained were washed in 1 wt% aqueous NaOH at 70 °C in order to remove bacteria and then several times in water, until reaching a neutral pH.

Pull Down Assay –

The experiments were based on the protocol described previously (38), with some modifications. 10 to 20 μg of purified proteins (WT or mutants) were incubated with 200 μl of solution containing 35 mg/ml Bacterial Microcrystalline Cellulose (BMCC) or Avicel (PH-101 – Fluka Analytical), dissolved in 25 mM Ammonium acetate pH 5.0, for 20 minutes at 8 °C and under 1000 rpm agitation. The mixture was centrifuged at 14000 rpm for 15 minutes. The soluble fraction was collected, concentrated and mixed with Laemli buffer. The insoluble fraction was washed three times with 25 mM ammonium acetate pH 5.0, 1 M NaCl. After centrifugation, the pellet was resuspended in 100 μl SDS sample buffer. Soluble and insoluble fractions were analyzed by SDS-PAGE.

Isothermal titration calorimetry (ITC) -

Thermodynamic parameters of binding of the CBM_E1 with GST-tag to soluble polysaccharides and oligosaccharides were determined by ITC,
using a VP-ITC calorimeter (Microcal, Northampton, MA). Titrations consisted of 10 μl injections of 5 mM oligosaccharides or 10 mg/ml of polysaccharides in 50 mM Na-HEPES buffer, pH 8.0, into the cell containing 100 μM CBM_E1 dialyzed into the Na-HEPES buffer, at 25 °C. The recorded data were analyzed using the Microcal Origin 7.0 software to derive n, K_a and ΔH values. ΔS was calculated using the standard thermodynamic equation, \(RT\ln K_a = \Delta G = \Delta H - T\Delta S\). All soluble polysaccharides and oligosaccharides were purchased from Megazyme International (Bray, County Wicklow, Ireland).

**Intrinsic Fluorescence Emission** – The intrinsic fluorescence emission measurements were performed in a Cary Eclipse Fluorescence Spectrophotometer (Varian) using a 10 mm path-length cell with CBM_E1 wild type or mutants (10 μM) in 20 mM Sodium Phosphate pH 7.4, 50 mM NaCl buffer, at room temperature. The excitation wavelength (λ) was set to 295 nm with a bandpass of 5 nm and emission was measured from 305 to 550 nm with a bandpass of 5 nm. Titration of cellotetraose, cellopentaose or cellohexaose was performed by adding from 0 to 400 μM of the ligand to the protein. Fluorescence was monitored immediately after the ligand was added. The spectra were concentration normalized and the data were analyzed using the spectral center of mass (\(<\lambda>\)), where \(<\lambda> = \Sigma F_i / \Sigma F_i\) (40). The \(K_a\) was obtained by obtaining the center of mass (\(<\lambda>\)) data versus ligand concentration using a hyperbole model according to the equation: \(y = P1 * x / (P2 + x)\).

**Dynamic Light Scattering** – Dynamic light scattering (DLS) was measured with a Zetasizer Nano Series dynamic light scattering instrument (Malvern) at 20 °C. Sample of CBM_E1 in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2% Glycerol buffer was used immediately after size exclusion chromatography at concentration of 0.75 mg/ml. Data were analyzed with software provided by the instrument.

**Analytical Size Exclusion Chromatography** – Analytical size exclusion chromatography of CBM_E1 was performed at room temperature in a Superdex 75 10/300 GL column with an AKTA instrument (GE Healthcare). Absorbance was recorded at wavelength of 280 nm. The system was calibrated with the following globular (3 mg/ml each), compact molecules with known hydrodynamic radius: conalbumin (75 kDa, 40.4Å), ovalbumin (44 kDa, 30.5 Å), carbonic anhydrase (29 kDa, 20.1 Å), ribonuclease A (13.7 kDa, 16.4 Å) and aprotinin (6.5 kDa, 1.35 Å), from GE calibration kit. The chromatography was performed at 0.5 ml/min with 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2% Glycerol buffer. The protein was injected at 0.5 mg/ml. After the run, the elution volume (V_e) was determined for each protein and the void volume (V_o) was determined with Dextran Blue 2000.

**Acknowledgments:** We gratefully acknowledge the time provided on the MX2 beamline (LNLS - Brazilian Synchrotron Light Laboratory) and Robolab (LNBio – Brazilian Biosciences National Laboratory) at the National Center for Research in Energy and Materials (CNPEM) (Campinas, Brazil).

**Conflict of Interest:** The authors declare that they have no conflicts of interest with the contents of this article.

**Author Contribution:** Performed experiments: BMC, MVL, TMA and GCE. Analyzed data: BMC, MVL, TMA and LMZ. Contributed reagents/materials/analysis: BMC, MVL, TMA, LMZ, HB, IP, RR, ACMZ and HJG. Wrote the paper: BMC, MVL, TMA, HJG and FMS. All authors read and agreed with the submitted version of the paper.

**References**
A novel CBM with both Type A and Type B characteristics

A novel CBM with both Type A and Type B characteristics


A novel CBM with both Type A and Type B characteristics


FOOTNOTES
This work was funded by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP – Process 2013/06336-0; 2014/04105-4; 2010/11469-1; 2008/58037-9) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

The abbreviations used are: CBM, Carbohydrate-binding module; GH, glycoside hydrolases; CAZy, carbohydrate-active enzymes database; CMC, carboxymethyl cellulose; ITC, isothermal titration calorimetry; C6, cellohexaose; C5, cellopentaose; C4, cellotetraose; BMCC, bacterial microcrystalline cellulose; ITF, intrinsic tryptophan fluorescence; VLS, variable loops site; CFS, concave face site.

FIGURE LEGENDS

**FIGURE 1.** A) Domain architecture of CelE1. B) SDS-PAGE analysis of purified CBM_E1 (11.8 kDa). MW = Molecular weight marker (PageRuler Unstained Protein Ladder); E1 to E4 = Elutions from affinity chromatography. A23 to A26 = fractions from size exclusion chromatography using a Superdex 75 10/30 GL column.

**FIGURE 2.** Representative ITC data of CBM_E1 binding to soluble ligands. A) 5 mM Cellohexaose (C6); B) 5 mM Cellopentaose (C5); C) 5 mM Cellotetraose (C4); D) 1% β-Glucan; E) 1% Xyloglucan; F) 1% Xylan Birchwood. The ligand in the syringe was titrated into CBM_E1 (100 μM) in the cell. The top half of each panel shows the raw ITC heats; the bottom half displays the integrated peak areas fitted using a one single binding model by MicroCal Origin software. ITC was carried out in 50 mM Na-HEPES, pH 8.0, at 25 °C.

**FIGURE 3.** A) Pull down assay for CBM_E1 with Avicel. B) Pull down assay with BMCC. The large arrows indicate the protein bound to Avicel (insoluble fraction) and the thin arrows, the protein in the soluble fraction, indicating that CBM_E1 does not bind to BMCC. MW = Molecular weight marker (PageRuler Unstained Protein Ladder); E1 = purified CBM_E1; I1 = Insoluble fraction from the assay with 10 μg of CBM_E1; S1 = Soluble fraction from the assay with 10 μg of CBM_E1; I2 = Insoluble fraction from the assay with 20 μg of CBM_E1; S2 = Soluble fraction from the assay with 20 μg of CBM_E1.
A novel CBM with both Type A and Type B characteristics

**FIGURE 4.** Crystal structures of CBM_E1 apo and holo overlayed. Cartoon depicting the 3D structure of CBM_E1 apo (orange) and holo (cyan) showing the two β-sheets formed by 5 and 4 β-strands and the disulfide bond represented in spheres. The only significant difference between them is observed in the first three N-terminal amino acids (N).

**FIGURE 5.** Structure of CBM_E1 in complex with cellopentaose. A) Model of CBM_E1 composed of the monomer in the asymmetric unit (Cyan) and another symmetric related molecule (green). The monomers bind to cellopentaose that is located perpendicularly to the 2-fold rotational symmetry axis. Consequently, cellopentaose is presented in two orientations – orientation 1 in cyan and orientation 2 in green. B) Details of the ligand-binding site illustrating the interactions between amino acids and ligand. C) 90° rotated view and D) 45° rotated view between B and C images, revealing the parallel alignment between the amino acids composing the binding site with the ligand and the perpendicular orientation between ligand and rotational symmetry axis. E) Cellopentaose, co-crystallized with CBM_E1, present at the binding site in two conformations (about 50 % occupancy each one) orientated 180° with respect to each other (Orientation 1 in cyan and orientation 2 in green). The 2Fo – Fc electron density map contoured at 1σ confirmed the presence of the double conformation. The symmetry axis is indicated by a dashed line or the symbol †, where the axis is perpendicular to the figure.

**FIGURE 6.** Pull down assay of CBM_E1 wild type and mutants against Avicel. The thin arrows show that the wild type interacts with Avicel. The thick arrows highlight that the mutants W375A, W398A, W427A do not interact with Avicel. The mutant K423A interacts partially with the substrate, as the protein can be found both in insoluble (thin arrow) and soluble (thick arrow) fractions. MW = Molecular Weight marker; C = Control (recombinant protein); I = insoluble fraction; S = Soluble fraction. WT = wild type CBM_E1; W375 = W375A mutant; W398 = W398A mutant; W427 = W427A mutant; K423 = K423A mutant.

**FIGURE 7.** Intrinsic Tryptophan Fluorescence Assay of CBM_E1 wild type and mutant K423A. A) CBM_E1 wild type assay. The calculated 1/⟨λ⟩ signals as a function of cellohexaose concentration present a consistent blue shift. This result suggests that the interaction of CBM_E1 wild type with cellohexaose involves W375, W398 and W427. B) CBM_E1 K423 assay. The calculated 1/⟨λ⟩ signals as a function of cellohexaose concentration also present a consistent blue shift. This result suggests that mutation of the residue K423 does not affect the tryptophan region surfaces and the protein-ligand interaction.

**FIGURE 8.** Oligomeric analysis of CBM_E1. A) Analytical Gel Filtration performed with the Superdex 75 10/300 GL Chromatographic column. Run performed with calibration kit, composed of Mix A (black line) - Conalbumin (C), Carbonic Anidrase (CA), Ribonuclease A (R) and Aprotinin (Ap). Mix B (large dashes) - Ovalbumin (O), Ribonuclease A (R) and Aprotinin (Ap). Short dashes: CBM-E1 at 0.5 mg/ml. B) Dynamic Light Scattering performed with CBM_E1 at 0.75 mg/ml.

**FIGURE 9.** A) The substrate-interacting tryptophans and the beta-strands that compose the binding site of CBM_E1 have a distinct parallel orientation. B) Type A CBMs, represented here by CBM2 from *P. furiosus*, have an almost parallel (with a small slope) orientation between substrate-interacting aromatic residues and the beta-strands (B). Conversely, Type B CBMs have a perpendicular orientation between substrate and beta-strands, independent of the binding site position: C) CFS (concave face site), represented by a CBM6 from *Clostridium stercorarium* or D) VLS (variable loops site), represented by a CBM15 from *Cellvibrio japonicus*. Geometric figures represent the orientations of substrates (green), substrate-interacting aromatic residues (cyan) and beta-strands (red).
Table 1: Affinity and thermodynamic parameters of CBM_E1 binding to polysaccharides and oligosaccharides

<table>
<thead>
<tr>
<th>Ligand</th>
<th>$K_a \times 10^4$</th>
<th>$\Delta G$ (kcal mol$^{-1}$)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$T\Delta S$ (kcal mol$^{-1}$)</th>
<th>$n^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellohexaose (C6)</td>
<td>1.23 ± 0.21</td>
<td>-5.57</td>
<td>-8.24 ± 1.64</td>
<td>-2.67</td>
<td>0.995 ± 0.172</td>
</tr>
<tr>
<td>Cellopentaose (C5)</td>
<td>0.433 ± 0.06</td>
<td>-4.95</td>
<td>-5.07 ± 1.66</td>
<td>-0.12</td>
<td>0.955 ± 0.285</td>
</tr>
<tr>
<td>Cellotetraose (C4)</td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manohexaose (M6)</td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xylohexaose (X6)</td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-Glucan (BG)</td>
<td>1.37 ± 0.473</td>
<td>-5.63</td>
<td>-3.31 ± 1.49</td>
<td>2.32</td>
<td>1.00 ± 0.4</td>
</tr>
<tr>
<td>Xyloglucan (XG)</td>
<td>0.5 ± 0.1</td>
<td>-5.04</td>
<td>-5.55 ± 2.76</td>
<td>-0.51</td>
<td>1.03 ± 0.474</td>
</tr>
<tr>
<td>XXXG</td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xylan Birchwood (XB)</td>
<td>NB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a the ITC data were fitted to a single site binding model for all ligands. For polysaccharide ligands, in which the molar concentration of binding sites is unknown, the $n$ value was iteratively fitted as close as possible to one, by adjusting the molar concentration of the ligand (39).

b NB, no binding
Table 2: Data Collection and Processing Statistics

<table>
<thead>
<tr>
<th></th>
<th>CBM_E1/CS</th>
<th>CBM_E1/Apo</th>
<th>CBM_E1/Gd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.96</td>
<td>1.46</td>
<td>1.48</td>
</tr>
<tr>
<td>Space group</td>
<td>ı2,3</td>
<td>ı2,1</td>
<td>ı2,3</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>88.8</td>
<td>28.4, 26.5, 99.3</td>
<td>88.41</td>
</tr>
<tr>
<td>β (°)</td>
<td>92.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>44.4-1.5 (1.53-1.5)</td>
<td>27.62-1.74 (1.78-1.74)</td>
<td>31.26-1.8 (1.84-1.8)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>114680</td>
<td>51818</td>
<td>113830</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>18747</td>
<td>7638</td>
<td>10827</td>
</tr>
<tr>
<td>Average mosaicity</td>
<td>0.23</td>
<td>0.43</td>
<td>0.85</td>
</tr>
<tr>
<td>R_{merge}</td>
<td>0.08 (0.84)</td>
<td>0.08 (0.97)</td>
<td>0.13 (0.81)</td>
</tr>
<tr>
<td>R_{free}</td>
<td>0.05 (0.52)</td>
<td>0.05 (0.58)</td>
<td>0.06 (0.38)</td>
</tr>
<tr>
<td>I / σI</td>
<td>17.9 (2.2)</td>
<td>14.8 (1.9)</td>
<td>15.1 (4.0)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.5 (90.7)</td>
<td>98.9 (81.0)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.1 (5.6)</td>
<td>6.8 (5.5)</td>
<td>10.5 (10.3)</td>
</tr>
<tr>
<td>CC ½ (%)</td>
<td>99.9 (70.4)</td>
<td>99.8 (31.9)</td>
<td>99.7 (90.6)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R_{work} / R_{free} (%)</td>
<td>15.50/18.29</td>
<td>15.39/20.73</td>
<td></td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>713</td>
<td>676</td>
<td></td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>179</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>13.5</td>
<td>14.9</td>
<td></td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>20.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>31.1</td>
<td>29.6</td>
<td></td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.022</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.72</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Ramachandran</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>97.9</td>
<td>95.6</td>
<td></td>
</tr>
<tr>
<td>Allowed</td>
<td>2.1</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Outliers</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Apparent affinity parameters of CBM_E1 and mutants binding to oligosaccharides

<table>
<thead>
<tr>
<th>CBM_E1</th>
<th>Ligand</th>
<th>$K_{a_{app}} (M^{-1})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>Cellohexaose</td>
<td>$4.3 \times 10^3$</td>
</tr>
<tr>
<td>W375A</td>
<td>Cellohexaose</td>
<td>No binding</td>
</tr>
<tr>
<td>W378A</td>
<td>Cellohexaose</td>
<td>No binding</td>
</tr>
<tr>
<td>W427A</td>
<td>Cellohexaose</td>
<td>No binding</td>
</tr>
<tr>
<td>K423A</td>
<td>Cellohexaose</td>
<td>$5.1 \times 10^3$</td>
</tr>
<tr>
<td>WT</td>
<td>Cellopentaose</td>
<td>$2.3 \times 10^3$</td>
</tr>
<tr>
<td>W375A</td>
<td>Cellopentaose</td>
<td>No binding</td>
</tr>
<tr>
<td>W378A</td>
<td>Cellopentaose</td>
<td>No binding</td>
</tr>
<tr>
<td>W427A</td>
<td>Cellopentaose</td>
<td>No binding</td>
</tr>
<tr>
<td>K423A</td>
<td>Cellopentaose</td>
<td>No binding</td>
</tr>
<tr>
<td>WT</td>
<td>Cellotetraose</td>
<td>No binding</td>
</tr>
<tr>
<td>W375A</td>
<td>Cellotetraose</td>
<td>No binding</td>
</tr>
<tr>
<td>W378A</td>
<td>Cellotetraose</td>
<td>No binding</td>
</tr>
<tr>
<td>W427A</td>
<td>Cellotetraose</td>
<td>No binding</td>
</tr>
<tr>
<td>K423A</td>
<td>Cellotetraose</td>
<td>No binding</td>
</tr>
</tbody>
</table>
A novel CBM with both Type A and Type B characteristics

Figure 1

A novel CBM with both Type A and Type B characteristics
Figure 3

A novel CBM with both Type A and Type B characteristics.
A novel CBM with both Type A and Type B characteristics.

Figure 4
A novel CBM with both Type A and Type B characteristics

Figure 5
A novel CBM with both Type A and Type B characteristics

Figure 6
Figure 7

A novel CBM with both Type A and Type B characteristics
A novel CBM with both Type A and Type B characteristics

Figure 8

<table>
<thead>
<tr>
<th>Mean [r.m.m]</th>
<th>MW [kDa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM_F1</td>
<td>1.67</td>
</tr>
</tbody>
</table>

Size Distribution by Volume
A novel CBM with both Type A and Type B characteristics

Figure 9
A novel carbohydrate-binding module from sugar cane soil metagenome featuring unique structural and carbohydrate affinity properties
Bruna Medeia Campos, Marcelo Vizona Liberato, Thabata Maria Alvarez, Leticia Maria Zanphorlin, Gabriela Cristina Ematsu, Hernane Barud, Igor Polikarpov, Roberto Ruller, Harry J. Gilbert, Ana Carolina de Mattos Zeri and Fabio Marcio Squina

J. Biol. Chem. published online September 12, 2016

Access the most updated version of this article at doi: 10.1074/jbc.M116.744383

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