Unique properties of a *Dictyostelium discoideum* carbohydrate-binding module expand our understanding of CBM–ligand interactions

Marcelo Vizona Liberato[1,‡], Bruna Medeia Campos[2,‡], Geizecler Tomazetto[3], Lucy Isobel Crouch[4], Wanius Garcia[5], Ana Carolina de Mattos Zeri[6], David Nichol Bolam[7], and Fabio Marcio Squina[1,*]

From the [1]*Programa de Processos Tecnológicos e Ambientais, Universidade de Sorocaba (UNISO), Sorocaba, SP, Brazil;*
[2]*Laboratório Nacional de Biociências (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil;* [3]*Department of Biological and Chemical Engineering (BCE), Aarhus University, Aarhus, Denmark;* [4]*Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, United Kingdom;* [5]*Centro de Ciências Naturais e Humanas, Universidade Federal do ABC (UFABC), Santo André, São Paulo, Brazil;* [6]*Laboratório Nacional de Luz Sincrotron (LNLS), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil;* [7]*Institute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle, United Kingdom*

Edited by Gerald Hart

Deciphering how enzymes interact, modify, and recognize carbohydrates has long been a topic of interest in academic, pharmaceutical, and industrial research. Carbohydrate-binding modules (CBMs) are noncatalytic globular protein domains attached to carbohydrate-active enzymes that strengthen enzyme affinity to substrates and increase enzymatic efficiency *via* targeting and proximity effects. CBMs are considered auspicious for various biotechnological purposes in textile, food, and feed industries, representing valuable tools in basic science research and biomedicine. Here, we present the first crystallographic structure of a CBM8 family member (CBM8), DdCBM8, from the slime mold *Dictyostelium discoideum*, which was identified attached to an endo-β-1,4-glucanase (glycoside hydrolase family 9). We show that the planar carbohydrate-binding site of DdCBM8, composed of aromatic residues, is similar to type A CBMs that are specific for crystalline (multichain) polysaccharides. Accordingly, pull-down assays indicated that DdCBM8 was able to bind insoluble forms of cellulose. However, affinity gel electrophoresis demonstrated that DdCBM8 also bound to soluble (single chain) polysaccharides, especially glucomannan, similar to type B CBMs, although it had no apparent affinity for oligosaccharides. Therefore, the structural characteristics and broad specificity of DdCBM8 represent exceptions to the canonical CBM classification. In addition, mutational analysis identified specific amino acid residues involved in ligand recognition, which are conserved throughout the CBM8 family. This advancement in the structural and functional characterization of CBMs contributes to our understanding of carbohydrate-active enzymes and protein–carbohydrate interactions, pushing forward protein engineering strategies and enhancing the potential biotechnological applications of glycoside hydrolase accessory modules.

Glycoside hydrolases (GHs) are enzymes capable of breaking glycosidic bonds. They are found in all living beings and are involved in essential functions, such as cell wall modeling, defense, symbiosis, signaling, biosynthesis, and nutrient acquisition (1). These enzymes are broadly applied in the industrial production of paper, fabrics, and food and have gained visibility in biorefining processes for conversion of biomass into renewable fuels and chemicals (2–4).

GHs are composed of a catalytic domain, which is often covalently linked to one or more accessory modules that regulate their activity, such as carbohydrate-binding modules (CBMs). CBMs are discrete folding units capable of binding to different types of carbohydrates, and their main role is to mediate the interaction between the enzyme and the target substrate, leading to modifications in catalytic efficiency (5). More specifically, CBMs perform functions such as increasing substrate accessibility through disruption of the crystalline structure of cellulose (6, 7), promoting specificity (8), and complementing the substrate-binding site of catalytic domains (9). To date, CBMs have been grouped into 89 different families in the carbohydrate-active enzymes (CAZy) database based on amino acid sequence similarities (10). CBMs are also classified based on their functional properties: type A CBMs possess flat binding faces, capable of binding to crystalline polysaccharides; type B CBMs bind internally to soluble polysaccharides *via* cleft-shaped contact sites; and type C CBMs interact with the terminal regions of carbohydrates through protein pocket–shaped sites (5, 11).

CBMs are considered auspicious for various biotechnological purposes, such as modification of the physical properties of fibers and potentiating enzymatic degradation of polysaccharides (12, 13), suitable for application in textile, food, and feed industries. CBMs are valuable tools in basic science research, used on *in situ* visualizations of polysaccharides, *in vivo* expression in plant physiology studies (14), and high-throughput analysis of polysaccharides based on microarrays (15). In synthetic biology

initiatives, these protein modules are used as building blocks to construct minicellulosomes (16). In biomedicine, CBMs are employed to functionalize carbohydrate-based biomaterial and improve recombinant protein technology, promoting expression, purification, stabilization, and immobilization of heterologous proteins (17–19). In this sense, the CBMs were employed for expression at the surface of proteins of pathogens (17), such as antigenic protein fragments of severe acute respiratory syndrome coronavirus 2 (20).

The number of different sequences deposited in CAZy and the number of families have been rapidly increasing because of next-generation sequencing technologies (21). Occasionally, members from novel or poorly studied families reveal new characteristics that show uncertainties in the current classifications (22) as well as disclose opportunities to be explored for biotechnological purposes. Therefore, we selected CBM8 (DdCBM8) from the endo-β-1,4-glucanase (CelA) of *Dictyostelium discoideum* (Fig. 1). DdCBM8 is located at the C-terminal region of CelA and is connected to the catalytic GH9 domain through a threonine–glutamate–threonine–proline repeat linker (23). According to our studies, the DdCBM8 presented a planar binding site and has the ability of binding to insoluble crystalline cellulose, conferring type A CBM characteristics. However, DdCBM8 showed highest affinity for soluble polysaccharides, such as glucomannan, resembling type B CBMs. We aimed to study the DdCBM8 using biochemical and biophysical methods to determine the unique structural and functional properties to provide novel insights into CBM–ligand interactions that can be useful for protein engineering strategies, increasing the biotechnological application space of GH accessory modules.

Results

Ligand-binding properties of DdCBM8

The gene encoding DdCBM8, comprising amino acid residues 555 to 705 from full-length CelA (Fig. 1), was inserted into the pET28a(+) bacterial vector, expressed in soluble form using *Escherichia coli* as the host, and purified to electrophoretic homogeneity using affinity and size-exclusion chromatography (data not shown).

Initially, a pull-down assay was performed to determine whether DdCBM8 binds to insoluble polysaccharides. About 10 micrograms of DdCBM8 were incubated with 35 mg/ml of Avicel and bacterial microcrystalline cellulose (BMCC). After several washing steps, the protein bound to the polysaccharides was detected with SDS gel. As shown in Figure 2A, DdCBM8 was able to bind to both the polysaccharides (protein is present in the insoluble fraction). An attempt to determine the affinity constant to insoluble polysaccharides was performed using depletion isotherm assay, where different amounts of DdCBM8 were incubated with constant concentration of each ligand, and the plot of bound *versus* unbound (free) protein could reveal the affinity constants. In fact, the binding was confirmed (Fig. 3A). However, saturation was not achieved, precluding determination of the affinity of DdCBM8 for BMCC and Avicel.

The ability of DdCBM8 to bind soluble polysaccharides was assessed using an affinity gel electrophoresis (AGE). Basically, nondenaturing gels polymerized with different polysaccharides were used in electrophoresis, and the protein migration was reduced when binding occurred between protein and ligand, in comparison to a gel lacking the ligand (native). As shown in Figure 2B, DdCBM8 bound to xyloglucan, glucomannan, β-glucan, and hydroxyethyl cellulose (HEC), but not to xylan. Subsequently, intrinsic fluorescence spectroscopy was performed to monitor (Fig. 3B) and estimate the binding capacity of DdCBM8 to these polysaccharides (Fig. 3C and Table S1). The data were fit to a one-site binding model. The average molecular weight (MW) of polysaccharides was employed to determine the binding constants to these polysaccharides, although this approach may overestimate the affinity as each chain can often bind multiple CBMs. The DdCBM8 binding constant (K_a) for glucomannan was 13.3×10^4 M^{-1}, whereas for β-glucan and HEC was 3.1×10^4 M^{-1} and 0.66×10^4 M^{-1}, respectively. Affinity for xyloglucan was also detected; however, as no average MW was given by manufacturers owing to its heterogeneity, the binding affinity constant was not determined. Finally, no binding of DdCBM8 to xylan or any tested oligosaccharide, such as cellopentaose, cellohexaose, and xyloglucan heptasaccharide (X_3Glc_4 or XXXG, where X stands for a glucose decorated with xylose and G indicates an undecorated glucose) could be detected, based on fluorescence and isothermal titration calorimetry (data not shown) methods.

General structural characteristics

Native DdCBM8 crystallographic structure was determined at 1.51 Å resolution. Initial phases were determined using the single-wavelength anomalous dispersion method using anomalous scattering of iodine, as described in the Experimental procedures section. At the end of refinement, the derivative dataset (resolution of 1.8 Å) displayed 13 iodine atoms with occupancies varying from 0.42 to 0.83. The statistics for data collection and refinement are presented in Table S2.

The final crystal structure presented a monomer in the asymmetric unit, and all amino acid residues (555–705) were built with well-defined electron densities. The amino terminal end also contained three residues (552–554; Ser, His, and Met) derived from the cloning vector, pET28a(+).

DdCBM8 has a compact globular β-sandwich fold, typical of the CBMs, and composed of two antiparallel β-sheets with five and six β-strands (β1, β4, β6, β7, and β9, and strands β2, β3, β5, β8, β10, and β11) that are connected by loops and one α-helix
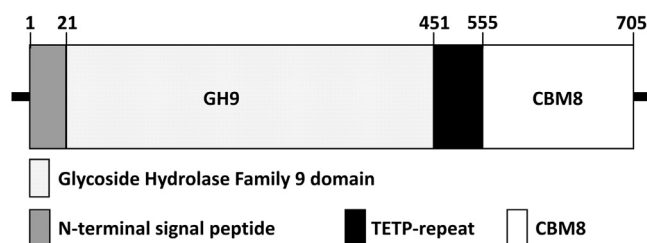

Figure 1. Schematic representation of *Dictyostelium discoideum* Cel9A primary structure (UniProt ID: P22699).
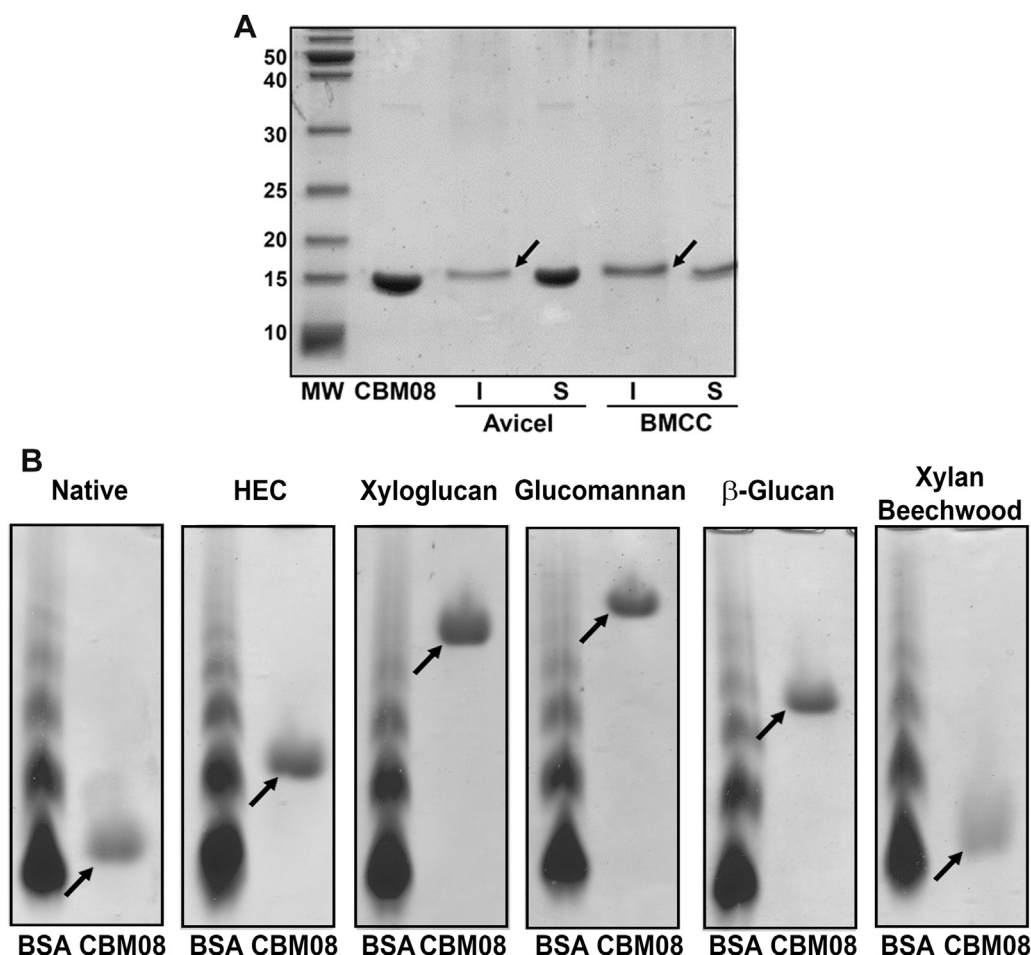
Figure 2. Qualitative assessment of binding of DdCBM8 to soluble and insoluble polysaccharides. *A*, pull-down assay with insoluble forms of cellulose (Avicel and BMCC). The *arrows* indicate the protein bound to Avicel or BMCC (insoluble fraction). *B*, affinity gel electrophoresis. The experiment demonstrated that DdCBM8 binds to β-glucan, glucomannan, xyloglucan, and HEC but not to xylan. The *arrows* indicate the band corresponding to DdCBM8. BSA was used as a control as it does not interact with carbohydrates. BMCC, bacterial microcrystalline cellulose; BSA, bovine serum albumin; CBM, carbohydrate-binding module; DdCBM8, purified DdCBM8; HEC, hydroxyethyl cellulose; I_{Avicel}, insoluble fraction from the assay with Avicel; I_{BMCC} = Insoluble fraction from the assay with BMCC; MW, molecular weight marker (PageRuler Unstained Protein Ladder); S_{Avicel}, soluble fraction from the assay with Avicel; S_{BMCC}, soluble fraction from the assay with BMCC.

(Fig. 4). A structural comparison conducted with DALI server (24) and PDBeFold server (25) revealed that DdCBM8 shared low similarity with members of other CBM families: CBM11 (Protein Data Bank [PDB] ID: 1V0A), with 2.02 Å rmsd; CBM30 (PDB ID: 1WZX), with 2.02 Å rmsd; and CBM29 (PDB ID: 1GWM), with 2.15 Å rmsd.

Characteristics of the ligand-binding site

Attempts to cocrystallize and soaking with xyloglucan heptasaccharide (XXXG) and cello-oligosaccharides (cellopentaose and cellohexaose) failed, which is consistent with the absence of affinity determined for these ligands. However, three aromatic residues exposed to solvent (W572, W574, and Y600) formed a planar surface resembling a typical ligand-binding site in type A CBMs (Fig. 5). Despite the overall low structural similarity with other CBMs, these aromatic residues were significantly aligned with the binding site of a member of the CBM29 family, CBM29-2 (PDB ID: 1GWM) (26), which was cocrystallized with cellohexaose. Both CBMs (DdCBM8 and CBM29-2) present the same three aromatic residues in similar positions, which in the

case of CBM29-2 were responsible for ligand binding through CH−π interactions. Furthermore, two other residues that are involved in ligand stabilization through hydrogen bonding in CBM29-2, share common positions: R634 and Q686 from DdCBM8 and K74 and Q116 from CBM29-2, respectively (Fig. 5A). The main difference in the binding sites is the presence of another three amino acids (E78, E83, and R112) in CBM29-2, which interact with the ligand through hydrogen bonds, but not in DdCBM8. Furthermore, DdCBM8 has three residues shorter β-strand (β7) compared with that in CBM29-2, which is responsible for creating a cleft where the ligand chain fits (Fig. 5B). The shorter β-strand found in DdCBM8 confers a planar shape for this region that probably allows crystalline polysaccharide binding.

Contribution of specific amino acid residues to ligand binding

To confirm the ligand-binding site inferred from the crystallographic structure and to evaluate the role of each amino acid in the interaction, five residues of the sites were mutated to alanine (W572, Y600, F608, R634, and Q686), and the
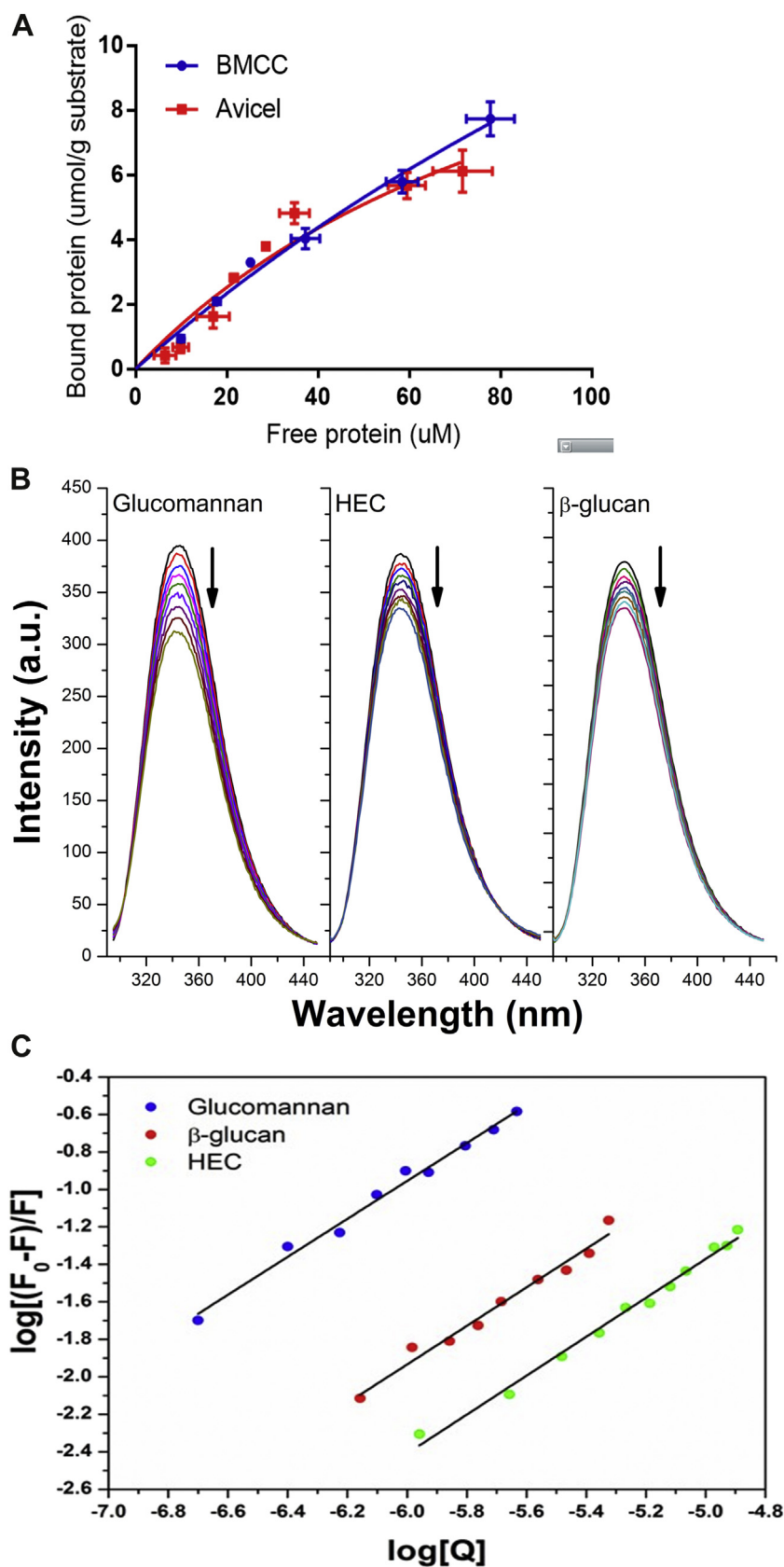
Figure 3. Quantitative assessment of binding of DdCBM8 to soluble and insoluble polysaccharides. *A*, depletion isotherms of DdCBM8 binding to insoluble forms of cellulose (Avicel and BMCC). *B*, intrinsic fluorescence spectroscopy of DdCBM08 in the absence and presence of different soluble polysaccharide concentrations (glucomannan, HEC, and β-glucan). The *arrows* indicate the reduction of fluorescence intensity caused by addition of increasing concentration of polysaccharide. *C*, double logarithmic plot of log ([F_0–F]/F]) *versus* log [Q] derived from the fluorescence quenching of DdCBM8 induced by soluble polysaccharides (Q = glucomannan, β-glucan, or HEC). CBM, carbohydrate-binding module; HEC, hydroxyethyl cellulose; BMCC, bacterial microcrystalline cellulose.
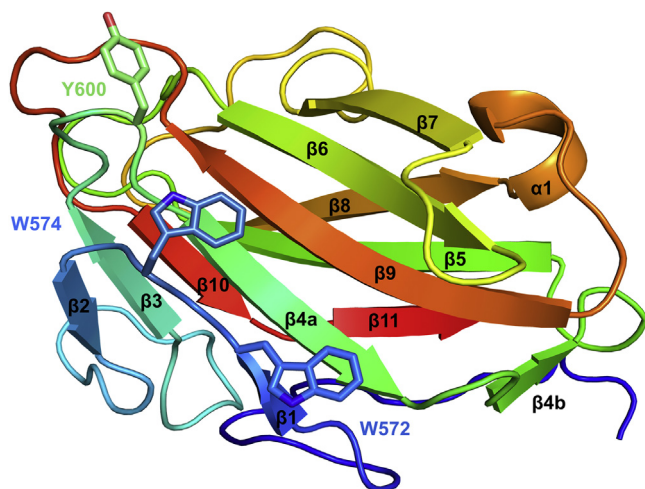
Figure 4. Cartoon showing the 3-dimensional structure of DdCBM8. The CBM has a typical β-sandwich fold, where two β-sheets, composed of 11 antiparallel β-strands, are packed against each other. The solvent-exposed aromatic residues involved in carbohydrate binding are depicted as *sticks* and labeled. CBM, carbohydrate-binding module.

CBMs were then subjected to AGE assays using known ligands for DdCBM8 (Fig. 6). As predicted from the crystallographic structure, mutations in amino acids that form the core of the binding site through CH–π interactions (W572A and Y600A) abolished DdCBM8 from binding to all polysaccharides evaluated here. The same profile would be expected for the mutation of W574 but, even after several attempts, we were not able to generate the specific mutant W574A. No protein mobility shift was observed from mutations F608A and, R634A, indicating that F608 and R634 probably does not interact with the ligand. Although R634 seems to be in a proper position for hydrogen bonding with the ligand, as indicated by the superposed structures in Figure 5C, the mobility shift effect was negligible. Finally, Q686A led to an intermediate reduction in ligand binding, confirming its role on the binding site.

Residues that comprise the ligand-binding site in DdCBM8 are conserved throughout the family

The amino acid sequence of DdCBM8 was aligned with representative members of the CBM8 family (Fig. S1). In general, the sequences showed great variability. However, the predicted residues that compose the binding site (W572, W574, Y600, and Q686) were highly conserved. The residues F608 and R634, which according to our data did not influence protein–carbohydrate interaction, were not conserved among CBM8 family members. Besides the amino acid residues involved in binding and the other conserved positions found in the alignment are buried in the protein structure and do not seem to play any role in CBM function. A second alignment (Fig. 7) performed with DdCBM8 and the two members described in CBM29 family (26, 27) showed that the aromatic residues composing the binding site were conserved as expected. However, the CBM8 and CBM29 family members display low sequence identity and, therefore, it is difficult to infer any common ancestry.

Discussion

The CBM8 family has 109 members reported to date, of which 99 belong to bacteria and the other two belong to *D. discoideum*, which is a slime mold Amoeba considered as a nonmammalian biomedical and pharmacological model owing to its cellular structure, intracellular signaling, and host–pathogen interaction similarities with mammalian cells (28, 29). The first and unique report of this family is related to the description of an endo-β-1,4-glucanase from *D. discoideum*, named CelA (23). Full-length CelA was able to bind to Avicel but not the cellulase domain itself or the cellulase with linker region (TETP repeat) (13). Thus, the C-terminal domain from CelA (described in this article as DdCBM8) was responsible for Avicel binding.

Following the previous report (23), the pull-down assay confirmed the affinity of DdCBM8 for Avicel and showed that it was able to bind to BMCC as well. Since BMCC has a significantly higher proportion of crystalline than amorphous cellulose (∼95%), as compared with Avicel (∼60%) (30), we can suppose that DdCBM8 has affinity for the crystalline portion indeed. However, the data indicate that this affinity is significantly lower than that observed for other type A CBMs (31–33). Subsequently, it was verified that DdCBM8 had a high affinity for soluble polysaccharides, especially glucomannan. The binding constant determined for glucomannan is similar ($13.3 \times 10^4 \, M^{-1}$) to that found for type B CBMs (from 0.3×10^4 to $5.8 \times 10^4 \, M^{-1}$) (26,34). Collectively, the results lead to an uncertainty in DdCBM8 classification as a type A or B.

The present study is the first structure of a CBM8 family reported to date. The crystallographic structure of DdCBM8 showed typical folding of the CBMs, but no similar structure was found in other families. Nevertheless, its binding site was inferred by structural comparison with members of the CBM29 family and further confirmed *via* mutagenesis of key amino acids. The planar binding site observed in the DdCBM8 structure would lead to its classification as a type A CBM. However, similar to CBM81, DdCBM8 is another exception to CBM-type canonical classification. It could be a type B with high affinity for soluble (single chain) polysaccharides or type A owing to the planar binding site of the protein. However, DdCBM8 lacks binding to oligosaccharides and has a low affinity for crystalline cellulose.

The unique characteristics of DdCBM8, together with its broad specificity, may contribute to a better understanding of CBM function and its future biotechnological applications. The structure of DdCBM8 has no parallel in other CBM families, which has a planar binding site that would lead to its classification as a type A CBM. However, DdCBM8 lacks binding to oligosaccharides and has a low affinity for crystalline cellulose, which are functional properties of type A CBM. On the other hand, DdCBM8 presents type B functional characteristics because of the high affinity for soluble (single chain) polysaccharides. Therefore, the structural characteristics and broad specificity of DdCBM8 will drive discussions and revision about this canonical classification and promote advances on the current understanding of CBM interaction and classification.
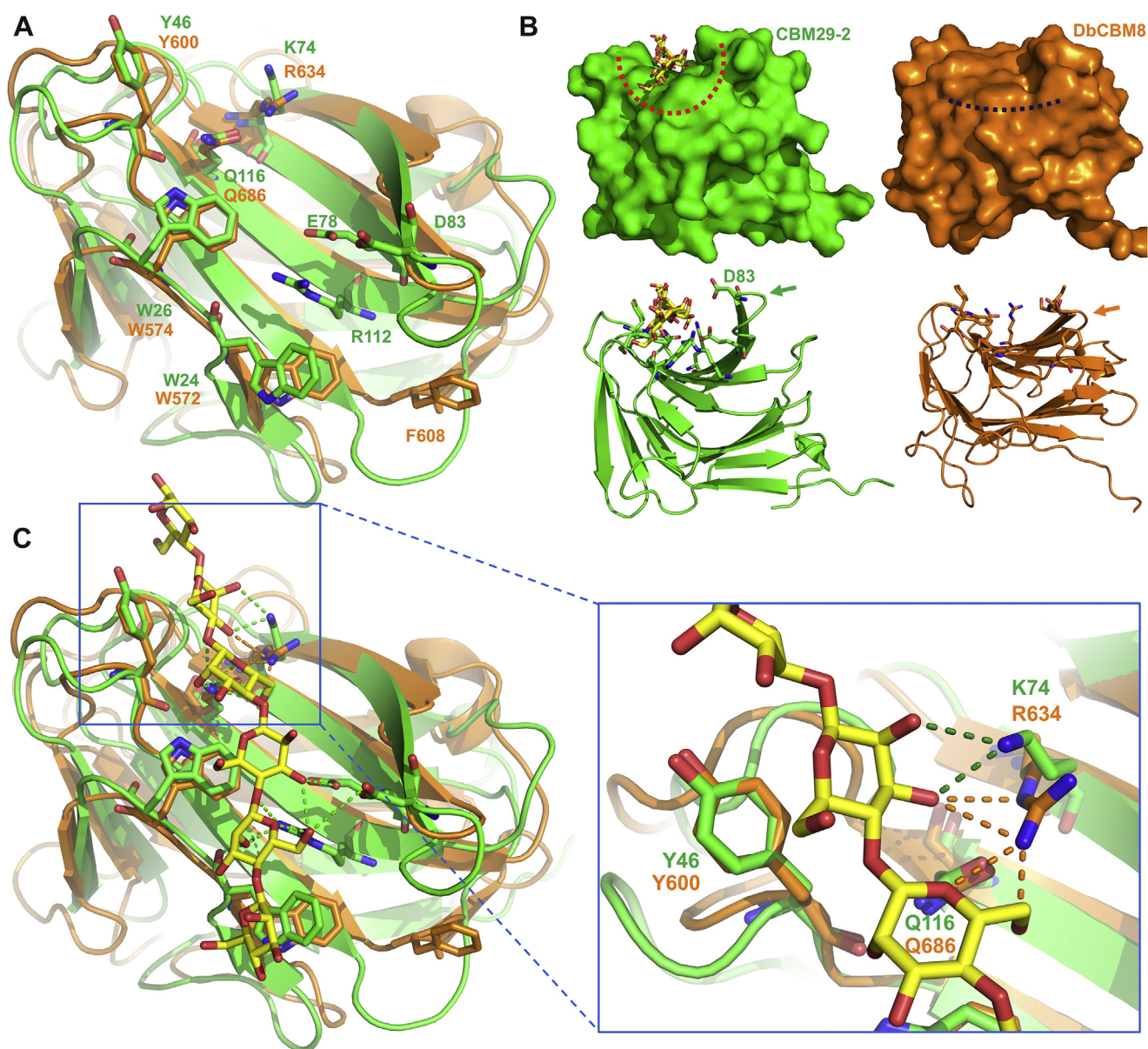
Figure 5. Comparison between DbCBM8 (*orange*) and CBM29-2 (*green*; PDB: 1GWL). *A*, superposition of both structures in *cartoon* representation, with ligand-binding site amino acids highlighted as *sticks*. Part of the binding sites (W572, W574, Y600, and Q686, in DdCBM8 numbering) has the same identity and very similar orientation, whereas other elements responsible for hydrogen bonding are absent in DdCBM8 (E78, D83, and R112, in CBM29-2 numbering). *B*, the binding site of CBM29-2 has a cleft shape (*red dashed line* in upper surface representation) to interact with ligands. However, the DdCBM8-binding site is flatter (*blue dashed line*), caused by a shorter loop and β-strand (*orange arrow*) where the amino acid D83 of CBM29-2 locates (*green arrow*). *C*, superposition of DdCBM08 (*green*) and CBM29-2 (*orange*) complexed with cellohexaose (PDB ID: 1GWL), evidencing the hydrogen bonds. CBM29-2 can form hydrogen bonds with the sugar hydroxyls *via* its K74 residue (*highlighted panel*). DdCBM08 has an arginine in the same position, R634, which theoretically could interact with ligand as well. However, the mutation of R634 into alanine did not change the mobility shift in AGE assay. AGE, affinity gel electrophoresis; CBM, carbohydrate-binding module; PDB, Protein Data Bank.

Experimental procedures

Protein expression and purification

The gene encoding the CBM DdCBM8, found in an endoglucanase from *D. discoideum* (UniProt ID: P22699), was synthesized by GenOne Biotechnologies from nucleotides 1663 to 2115 (amino acids 555–705). The commercial vector containing the gene was digested with NdeI and NotI restriction enzymes and cloned into the expression bacterial vector pET28a(+). Cloning was verified using DNA sequencing. The final construct encoded DdCBM8 fused to an N-terminal His tag with a site for thrombin protease cleavage for tag removal.

Recombinant DdCBM8 was expressed in *E. coli* strain BL21 (DE3) (Thermo Fisher Scientific). A single colony was used to inoculate 10 ml LB medium supplemented with kanamycin (50 μg/ml) as a starter culture. Then, LB medium (4 l) was cultured at 37 °C until absorbance of ≈0.6 at 600 nm, followed by induction with 0.5 mM IPTG for 16 h at 16 °C. Cells harvested by centrifugation were suspended in binding buffer (20 mM Tris−HCl, pH 8.0, 200 mM NaCl,

Figure 6. Comparison of binding ability of mutant and wildtype (*blue arrows*) DdCBM8 to soluble polysaccharides. Affinity gel electrophoresis assay experiment showed that mutants W572A, Y600A, and Q686 (*red arrows*) play an important role in ligand binding. The F608A and R634A mutations have no effect on binding. BSA was used as a control, as it does not interact with carbohydrates. BSA, bovine serum albumin; CBM, carbohydrate-binding module.

5 mM imidazole, and 20% glycerol) and incubated on ice with lysozyme (1 mg/ml) for 30 min. Cells were sonicated, and the clarified supernatants were incubated with TALON resin for 3 h at 25 °C. The beads were washed with 10 column volumes of washing buffer (20 mM Tris–HCl [pH 8.0], 200 mM NaCl, 10 mM imidazole, and 20% glycerol), and the retained proteins were eluted with wash buffer containing 250 mM imidazole. After His tag removal with thrombin, the protein was purified with a HiLoad 16/60 Superdex 75 prep grade column equilibrated with 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, and 2% glycerol. Purified DdCBM8 was stored at 4 °C.

Insoluble polysaccharide-binding assay

The pull-down assay was performed as described previously (35), with some modification. Ten micrograms of purified protein were incubated in 200 ml solution containing 35 mg/ml of Avicel or BMCC, dissolved in 25 mM ammonium acetate (pH 5.0), for 20 min at 20 °C and under 1000 rpm agitation. The mixture was centrifuged at 12,000g for 15 min, and the soluble fraction was collected, concentrated, and mixed with SDS sample buffer. The insoluble fraction was washed three times with 25 mM ammonium acetate (pH 5.0) and 1 M NaCl. After centrifugation, the pellet was resuspended in 100 ml SDS sample

Figure 7. Multiple alignment of DdCBM8 and two sequences from CBM29 family. Database source and the corresponding accession numbers are shown before the alignment. Start and end positions of CBMs and residue numbers of full-length sequences are given in parentheses. Blocks of similar amino acids (using threshold at 70%) are shown on colors background. The amino acids that compose the binding sites are indicated by *red arrows*. CBM, carbohydrate-binding module.

buffer. Soluble and insoluble fractions were analyzed using SDS-PAGE.

Ligand-binding quantification was measured using a depletion isotherm assay (31, 36). Reactions containing 1% ligand, 25 mM ammonium acetate (pH 5.0), and variable amounts of protein (4–155 μM) were incubated for 2 h at 20 °C and 1000 rpm. Samples were centrifuged at 12,000*g* at 20 °C for 5 min, and the protein concentration of the supernatant (unbound fraction) was measured using the Bradford method from Bio-Rad.

AGE

AGE was performed as described previously (37, 38). Native polyacrylamide gels were prepared consisting of 8% acrylamide utilizing 0.2% of the following soluble polysaccharides: HEC, xylan, xyloglucan, glucomannan, and β-glucan. The gels were prepared using Tris-acetate buffer (pH 8.3) plus EDTA in the absence of SDS. 2-Mercaptoethanol was excluded from the loading buffer, and proteins were not heated at 95 °C prior to loading onto the gel. In all experiments, DdCBM8 was run simultaneously in the native gel with or without incorporated soluble polysaccharides. Bovine serum albumin was used as the control. Electrophoresis was carried out at 125 V at 18 °C for 2 h. Proteins were visualized by staining with Coomassie Brilliant Blue R-250.

Intrinsic fluorescence spectroscopy

All fluorescence measurements were performed using a Cary Eclipse Fluorescence Spectrophotometer (Varian) using a 10 mm path-length quartz cuvette with DdCBM8 (2.0 μM) in 20 mM sodium phosphate (pH 7.4), 50 mM NaCl buffer, at 20 °C. The excitation wavelength was 290 nm, and the emission intensities were measured over the wavelength of 300 to 450 nm. Five scans were averaged for each experiment. The emission intensities were corrected for background fluorescence caused by buffer and carbohydrates, for dilution, and for inner filter effects, and all the experiments were done in triplicate. Quantitative binding experiments were performed by titration of the appropriate carbohydrate. The binding constant (K_a) and the number of binding sites (n) were calculated using the following relationship:

$$\log \frac{(F_0 - F)}{F} = \log K_a + n \log[Q]$$

where F_0 and F are the fluorescence intensities in the absence and presence of the ligand (Q), respectively, [Q] is the ligand concentration. In this case, K_a is determined by the linear coefficient, and n (angular coefficient) gives the number of sites. The MWs of the carbohydrates are an average value given by the manufacturers: glucomannan (MW = ∼250 kDa), β-glucan (MW = ∼179 kDa), and HEC (MW = ∼90 kDa).

Crystallization and data collection

Pure DdCBM8 was concentrated to 9.7 mg/ml and subjected to initial crystallization screening with the aid of a Honey Bee 963 robot from the ROBOLAB facility at the Brazilian Biosciences National Laboratory. The drops containing 0.5 μl of protein solution and 0.5 μl of reservoir solution were placed in sitting-drop vapor-diffusion plates and incubated at 18 °C. Commercial kits (Hampton, Qiagen, and Rigaku) were used as the initial conditions. The first hits led to the refinement of the conditions using hanging-drop vapor-diffusion plates, where the drops contained 2 μl of protein solution and 2 μl of reservoir solution and were incubated at 18 °C.

The crystals of DdCBM8 were obtained in condition containing 1.6 M disodium DL-malate, soaked in cryoprotection solution (20% ethylene glycol and crystallization solution), and flash cooled in a stream of gaseous nitrogen at 100 K. For derivatization, a DdCBM8 crystal was soaked in cryoprotection solution containing 0.6 M sodium iodide. The X-ray diffraction data were collected at the MX-2 beamline of the Brazilian Synchrotron Light Laboratory (LNLS) using a PILATUS2M detector (Dectris).

Collected data were integrated with iMosflm (39) and XDS (40) and scaled with AIMLESS (41). Initial phases were determined using the single-wavelength anomalous diffraction method using the program AUTOSOL (42) from PHENIX (43). The phases from a native dataset were identified by molecular replacement using Phaser (44) with DdCBM8/I as the search model. The models were refined using PHENIX (43) and interspersed with manual adjustment in Coot (45). Final models were deposited in the PDB with codes 7T7Z and 7T7Y, respectively, to native and derivative data.

Site-directed mutagenesis

Site-directed mutagenesis was carried out using the Quikchange kit (Agilent). The DdCBM8/pET28a vector was used as a template, and the primers carrying the mutations were as follows: W572A, 5′-CTTCCAGGATGCGAGCTGGGGC-3′ and 5′-TCGTTCTTCAGACCAGATTTG-3′; W574A, 5′-GGATTGGAGCGCGGGCGAGCAC-3′ and 5′-TGGAAGTCGTTCTTCAGAC-3′; Y600A, 5′-TCCAAAGGCCGCTGGTGCTGTG-3′ and 5′-GTGAAGGAAATAGAATTGG-3′; F608A, 5′-CCTGGGGTTGCGCCGAATGCATTG-3′ and 5′-AACACAGCACCATAGGCC-3′; R634A, 5′-GCACAGCTGCTGGCTATCACTGTGG-3′ and 5′-GCCAGAGCTGCCACCG-3′; and Q686A, 5′-GGCATCTGGATTGCGGACATTAAGGG-3′ and 5′-GTCCACTTTGCCGGAT-3′. The mutant proteins were expressed and purified as described in the previous section.

Sequence alignment of CBMs

Based on CBM8 and CBM29 families available in the dbCAN (46), the amino acid sequences of the corresponding domains were retrieved from GenBank and UniProt with a cutoff minimum E-value of 1e^{-20}. Multiple sequence alignments were performed using ClustalX (47) and the matrix Blossum62. The alignment obtained was manually refined using Bioedit (48) based on the conserved amino acid residues.

Carbohydrate sources

All soluble polysaccharides and oligosaccharides were purchased from Megazyme International, except for xylan (beechwood) and HEC, which were from Sigma. The bacterial cellulose membrane production has been described previously (22, 49).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Minic, Z. (2008) Physiological roles of plant glycoside hydrolases. *Planta* **227**, 723–740
2. Kirk, O., Borchert, T. V., and Fuglsang, C. C. (2002) Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351
3. Linares-Pasten, J., Andersson, M., and Karlsson, E. (2014) Thermostable glycoside hydrolases in biorefinery technologies. *Curr. Biotechnol.* **3**, 26–44
4. Consortium, T. Caz (2018) Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes. *Glycobiology* **28**, 3–8
5. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: Fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781
6. Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, R. A. J., and Kilburn, D. G. (1991) Non–hydrolytic disruption of cellulose fibres by the binding domain of a bacterial cellulase. *Bio/Technology* **9**, 1096–1099
7. Southall, S. M., Simpson, P. J., Gilbert, H. J., Williamson, G., and Williamson, M. P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett.* **447**, 58–60
8. Araki, R., Ali, M. K., Sakka, M., Kimura, T., Sakka, K., and Ohmiya, K. (2004) Essential role of the family-22 carbohydrate-binding modules for β-1,3-1,4-glucanase activity of *Clostridium stercorarium* Xyn10B. *FEBS Lett.* **561**, 155–158
9. Liberato, M. V., Silveira, R. L., Prates, É. T., de Araujo, E. A., Pellegrini, V. O. A., Camilo, C. M., Kadowaki, M. A., Neto, M. O., Popov, A., Skaf, M. S., and Polikarpov, I. (2016) Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Sci. Rep.* **6**, 23473
10. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495
11. Gilbert, H. J., Paul Knox, J., and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* **23**, 669–677
12. Mello, B. L., and Polikarpov, I. (2014) Family 1 carbohydrate binding-modules enhance saccharification rates. *AMB Express* **4**, 30
13. Nagl, M., Haske-Cornelius, O., Skopek, L., Bausch, F., Pellis, A., Bauer, W., Nyanhongo, G. S., and Guebitz, G. M. (2022) Mechanistic investigation of the effect of endoglucanases related to pulp refining. *Cellulose*. https://doi.org/10.1007/S10570-021-04386-5
14. Mei, X., Chang, Y., Shen, J., Zhang, Y., and Xue, C. (2020) Expression and characterization of a novel alginate-binding protein: A promising tool for investigating alginate. *Carbohydr. Polym.* **246**, 116645
15. Moller, I., Sørensen, I., Bernal, A. J., Blaukopf, C., Lee, K., Øbro, J., Pettolino, F., Roberts, A., Mikkelsen, J. D., Knox, J. P., Bacic, A., and Willats, W. G. T. (2007) High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *Plant J.* **50**, 1118–1128
16. Kim, S., Baek, S. H., Lee, K., and Hahn, J. S. (2013) Cellulosic ethanol production using a yeast consortium displaying a minicellulosome and β-glucosidase. *Microb. Cell Fact.* **12**, 14
17. Oliveira, C., Carvalho, V., Domingues, L., and Gama, F. M. (2015) Recombinant CBM-fusion technology - applications overview. *Biotechnol. Adv.* **33**, 358–369
18. Qin, Z., Lin, S., Qiu, Y., Chen, Q., Zhang, Y., Zhou, J., and Zhao, L. (2019) One-step immobilization-purification of enzymes by carbohydrate-binding module family 56 tag fusion. *Food Chem.* **299**, 125037
19. Shoseyov, O., Shani, Z., and Levy, I. (2006) Carbohydrate binding modules: Biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* **70**, 283–295
20. McGuire, B. E., Mela, J. E., Thompson, V. C., Cucksey, L. R., Stevens, C. E., McWhinnie, R. L., Winkler, D. F. H., Pelech, S., and Nano, F. E. (2022) *Escherichia coli* recombinant expression of SARS-CoV-2 protein fragments. *Microb. Cell Fact.* **21**, 21
21. Kameshwar, A. K. S., and Qin, W. (2016) Recent developments in using advanced sequencing technologies for the genomic studies of lignin and cellulose degrading microorganisms. *Int. J. Biol. Sci.* **12**, 156–171
22. Campos, B. M., Liberato, M. V., Alvarez, T. M., Zanphorlin, L. M., Ematsu, G. C., Barud, H., Polikarpov, I., Ruller, R., Gilbert, H. J., De Mattos Zeri, A. C., and Squina, F. M. (2016) A novel carbohydrate-binding module from sugar cane soil metagenome featuring unique structural and carbohydrate affinity properties *. *J. Biol. Chem.* **291**, 23734–23743
23. Ramalingam, R., Blume, J. E., and Ennis, H. L. (1992) The Dictyostelium discoideum spore germination-specific cellulase is organized into functional domains. *J. Bacteriol.* **174**, 7834–7837
24. Holm, L., and Rosenström, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
25. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268
26. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabó, L., McKie, V. A., Gilbert, H. J., and Davies, G. J. (2002) Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14077–14082

ASBMB

J. Biol. Chem. (2022) 298(5) 101891 **9**

27. Flint, J., Bolam, D. N., Nurizzo, D., Taylor, E. J., Williamson, M. P., Walters, C., Davies, G. J., and Gilbert, H. J. (2005) Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules. *J. Biol. Chem.* **280**, 23718–23726

28. Martín-González, J., Montero-Bullón, J. F., and Lacal, J. (2021) Dictyostelium discoideum as a non-mammalian biomedical model. *Microb. Biotechnol.* **14**, 111–125

29. Schaf, J., Damstra-Oddy, J., and Williams, R. S. B. (2019) Dictyostelium discoideum as a pharmacological model system to study the mechanisms of medicinal drugs and natural products. *Int. J. Dev. Biol.* **63**, 541–550

30. Hall, M., Bansal, P., Lee, J. H., Realff, M. J., and Bommarius, A. S. (2010) Cellulose crystallinity - a key predictor of the enzymatic hydrolysis rate. *FEBS J.* **277**, 1571–1582

31. Duan, C.-J., Feng, Y.-L., Cao, Q.-L., Huang, M.-Y., and Feng, J.-X. (2016) Identification of a novel family of carbohydrate-binding modules with broad ligand specificity. *Sci. Rep.* **6**, 19392

32. Georgelis, N., Yennawar, N. H., and Cosgrove, D. J. (2012) Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14830–14835

33. Pires, V. M. R., Pereira, P. M. M., Brás, J. L. A., Correia, M., Cardoso, V., Bule, P., Alves, V. D., Najmudin, S., Venditto, I., Ferreira, L. M. A., Romão, M. J., Carvalho, A. L., Fontes, C. M. G. A., and Prazeres, D. M. (2017) Stability and ligand promiscuity of type A carbohydrate-binding modules are illustrated by the structure of Spirochaeta thermophila StCBM64C. *J. Biol. Chem.* **292**, 4847–4860

34. Venditto, I., Luis, A. S., Rydahl, M., Schückel, J., Fernandes, V. O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V. M. R., Dourado, C. G., Ferreira, L. M. A., Coutinho, P. M., Henrissat, B., Knox, J. P., Baslé, A., *et al.* (2016) Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7136–7141

35. Okazaki, F., Tamaru, Y., Hashikawa, S., Li, Y.-T., and Araki, T. (2002) Novel carbohydrate-binding module of beta-1,3-xylanase from a marine bacterium, Alcaligenes sp. strain XY-234. *J. Bacteriol.* **184**, 2399–2403

36. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000) Analysis of binding of the family 2a carbohydrate-binding module from cellulomonas fimi xylanase 10A to cellulose: Specificity and identification of functionally important amino acid residues. *Protein Eng. Des. Sel.* **13**, 801–809

37. Bae, B., Ohene-Adjei, S., Kocherginskaya, S., Mackie, R. I., Spies, M. A., Cann, I. K. O., and Nair, S. K. (2008) Molecular basis for the selectivity and specificity of ligand recognition by the family 16 carbohydrate-binding modules from Thermoanaerobacterium polysaccharolyticum ManA. *J. Biol. Chem.* **283**, 12415–12425

38. Tomme, P., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (1996) Interaction of polysaccharides with the N-terminal cellulose-binding domain of *Cellulomonas fimi* CenC. 1. Binding specificity and calorimetric analysis [†]. *Biochemistry* **35**, 13885–13894

39. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 271–281

40. Kabsch, W. (2010) Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132

41. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82

42. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L.-W. (2009) Decision-making in structure solution using bayesian estimates of map quality: The *PHENIX AutoSol* wizard. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 582–601

43. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., *et al.* (2010) Phenix: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221

44. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674

45. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501

46. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451

47. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P.a., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948

48. Hall, T. A. (1999) *Bioedit: A User-Friedly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT.* Nucleic Acids Symposium Series, North Carolina State University, Raleigh, NC: 95–98

49. Pinto, E. R. P., Barud, H. S., Silva, R. R., Palmieri, M., Polito, W. L., Calil, V. L., Cremona, M., Ribeiro, S. J. L., and Messaddeq, Y. (2015) Transparent composites prepared from bacterial cellulose and castor oil based polyurethane as substrates for flexible OLEDs. *J. Mater. Chem. C.* **3**, 11581–11588

ASBMB