**Bruna Medeira Campos,^a
Thabata Maria Alvarez,^b
Marcelo Vizona Liberato,^b Igor
Polikarpov,^c Harry J. Gilbert,^d
Ana Carolina de Mattos Zeri^a
and Fabio Marcio Squina^{b*}**

^aLaboratório Nacional de Biociências (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil, ^bLaboratório Nacional de Ciências e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil, ^cInstituto de Física de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil, and ^dInstitute for Cell and Molecular Biosciences, The Medical School, Newcastle University, Newcastle upon Tyne, England

Correspondence e-mail:
fabio.squina@bioetanol.org.br

Received 2 June 2014
Accepted 2 July 2014

Cloning, purification, crystallization and preliminary X-ray studies of a carbohydrate-binding module (CBM_E1) derived from sugarcane soil metagenome

In recent years, owing to the growing global demand for energy, dependence on fossil fuels, limited natural resources and environmental pollution, biofuels have attracted great interest as a source of renewable energy. However, the production of biofuels from plant biomass is still considered to be an expensive technology. In this context, the study of carbohydrate-binding modules (CBMs), which are involved in guiding the catalytic domains of glycoside hydrolases for polysaccharide degradation, is attracting growing attention. Aiming at the identification of new CBMs, a sugarcane soil metagenomic library was analyzed and an uncharacterized CBM (CBM_E1) was identified. In this study, CBM_E1 was expressed, purified and crystallized. X-ray diffraction data were collected to 1.95 Å resolution. The crystals, which were obtained by the sitting-drop vapour-diffusion method, belonged to space group *I*23, with unit-cell parameters $a = b = c = 88.07$ Å.

1. Introduction

One of the main challenges for commercially successful production of second-generation biofuels is the conversion of the lignocellulosic biomass (the most abundant source of renewable carbon on the planet) into glucose with high efficiency and low cost (Bolam *et al.*, 2004; Farrell *et al.*, 2006; Santos *et al.*, 2012; Kim *et al.*, 2013). A number of approaches have been undertaken for improvement of enzyme cocktails for second-generation biofuels, such as efforts for the rational design of site-directed mutagenesis, targeting enzymes for specific applications (Graham *et al.*, 2011; Kim *et al.*, 2013), engineering of multifunctional proteins with a synergistic catalytic capacity (Gonçalves *et al.*, 2012; Cota *et al.*, 2013; Damásio *et al.*, 2014) and enzymes that bind weakly to lignin (Berlin *et al.*, 2005). Recently, the search for accessory proteins that have nonhydrolytic activity during cellulose hydrolysis has attracted attention (Kim *et al.*, 2009). To access these novel proteins, metagenomics has gained attention owing to its great potential for prospecting for genes from the genomes of uncultured microorganisms, which represent approximately 99% of all microorganisms in nature (Lorenz & Eck, 2005; Amann *et al.*, 1995). With this in mind, sugarcane soil metagenomics has been used to identify novel carbohydrate-binding modules (CBMs). These macromolecules have no enzymatic activity, but are known to be auxiliary domains to cellulases and other carbohydrate-active proteins that enhance catalysis (Guillén *et al.*, 2010; Luís *et al.*, 2013), showing a range of binding capabilities to different types of polymers (Boraston *et al.*, 2004; Hashimoto, 2006).

Through the functional screening of a sugarcane soil metagenomic library, a cellulase comprising a glycoside family 5 (GH5) catalytic module (accession No. KF498957; Alvarez *et al.*, 2013) linked to an unclassified carbohydrate-binding module (CBM_E1) was isolated. In order to gain insight into its structural and functional characteristics, as well as the protein–ligand recognition properties, of this novel CBM, we decided to solve its crystal structure. In this communication, we describe the expression, purification, crystallization and preliminary X-ray analysis of CBM_E1.

Table 1

Macromolecule cloning and expression conditions.

DNA source	Sugarcane soil metagenome
Forward primer†	5'-TATATAT CA TATGAGCGCATCATGCGGTAGC-3'
Reverse primer†	5'-ATAGGATCCTTACCAGTTATCGAACTTCACATT-3'
Cloning vector	pJET
Expression vector	pET-28a
Expression host	Origami 2 (DE3) cells
GenBank accession No.	KJ917170

† Restriction sites are shown in bold and the stop codon is underlined.

2. Materials and methods

2.1. Cloning, expression and purification

The gene encoding the carbohydrate-binding module E1 (CBM_E1; KJ917170) was amplified by PCR using full-length cellulase, retrieved from a sugarcane soil metagenomic library, as a template (Alvarez *et al.*, 2013). The forward primer sequence contains an *Nde*I restriction site (in bold) and the reverse primer sequence contains a *Bam*HI restriction site (in bold) and a stop codon (underlined) (Table 1). The 282 bp product was cloned into pJET cloning vector and confirmed by DNA sequencing (data not shown). The construction CBM_E1_pJET and the expression vector pET-28a were digested with the *Nde*I and *Bam*HI restriction enzymes. The ligation mixture was transformed into *Escherichia coli* DH5 α competent cells and cloning was verified by PCR. The final construct encoded full-length CBM_E1 fused to an N-terminal His tag with a thrombin protease cleavage site for tag removal.

Recombinant CBM_E1 was expressed in *E. coli* strain Origami 2 (DE3) (Novagen). A single colony was used to inoculate a 10 ml Luria–Bertani (LB) starter culture supplemented with kanamycin (50 mg ml⁻¹) and streptomycin (25 mg ml⁻¹) and was used to inoculate 4.0 l LB medium. The bacteria were cultured at 310 K until the OD₆₀₀ reached ~0.6, followed by induction with 0.4 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) for 3 h at 303 K. The cells were harvested by centrifugation, suspended in binding buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 5 mM imidazole, 20% glycerol) and incubated on ice with lysozyme (1 mg ml⁻¹) for 30 min. The cells were sonicated and the clarified supernatants were incubated with nickel resin for 2 h at room temperature. The beads were washed with ten column volumes of wash buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 10 mM imidazole, 20% glycerol) and the retained proteins were eluted with wash buffer containing 200 mM imidazole.

Figure 1

SDS-PAGE analysis of CBM_E1 purification by size-exclusion chromatography. Lane MW, molecular-weight marker (labelled in kDa). Lane 1, soluble fraction obtained after nickel-affinity chromatography. Lane 2, sample loaded on size-exclusion column, after concentration. Lanes 3–6, elution fractions from size-exclusion chromatography containing 20 mM sodium phosphate pH 7.2, 50 mM NaCl.

Table 2

Crystallization conditions.

Method	Vapour diffusion
Plate type	Sitting-drop
Temperature (K)	291
Protein concentration (mg ml ⁻¹)	6
Buffer composition of protein solution	20 mM sodium phosphate pH 7.2, 50 mM NaCl
Composition of reservoir solution	4 M sodium formate
Volume and ratio of drop	1 μ l, 1:1
Volume of reservoir (μ l)	80

The 6 $\times$ His tag was cleaved with thrombin at 289 K for 16 h. The protein was further purified on a Superdex 75 10/300 column equilibrated with 20 mM sodium phosphate pH 7.2, 50 mM NaCl (Fig. 1). Purified CBM_E1 was stored at 277 K.

2.2. Crystallization

Highly purified CBM_E1 sample was concentrated to 6 mg ml⁻¹ in 20 mM sodium phosphate pH 7.2, 50 mM NaCl. The protein solution was incubated with cellopentaose (C5) at a molar ratio of 1 CBM_E1:2 C5. Crystallization experiments were performed using the sitting-drop vapour-diffusion method at 291 K using a HoneyBee 963 robot (Genomic Solutions). The drop consisted of 0.5 μ l of the CBM_E1–C5 complex plus 0.5 μ l of the reservoir solution. Well formed crystals were used for X-ray data collection (Table 2).

2.3. Data collection and processing

Crystals (Fig. 2) were soaked in cryoprotection solution consisting of 14% ethylene glycol and the crystallization solution. The crystal was then flash-cooled in a stream of gaseous nitrogen at 100 K. X-ray diffraction data were collected on the MX2 beamline (Guimarães *et al.*, 2009) at the Brazilian Synchrotron Light Laboratory (LNLS, Campinas-SP) using a MAR Mosaic 225 mm CCD detector (MAR Research) and a synchrotron-radiation wavelength of 1.459 Å. The data set was processed with *iMosflm* (Battye *et al.*, 2011) and scaled with *AIMLESS* (Evans, 2006).

Figure 2

A single crystal of CBM_E1 was obtained in the presence of 4 M sodium formate by sitting-drop vapour diffusion.

Table 3
Data-collection and processing statistics.
Values in parentheses are for the outer shell.

No. of crystals	1
X-ray source	MX2, LNLS
Wavelength (Å)	1.459
Temperature (K)	100
Detector	MAR Mosaic 225 CCD
Crystal-to-detector distance (mm)	80
Rotation range per image (°)	1
Total rotation range (°)	110
Exposure time per image (s)	1.2
Resolution range (Å)	22.02–1.95 (2.00–1.95)
Space group	<i>I</i> 23
Unit-cell parameters (Å)	<i>a</i> = <i>b</i> = <i>c</i> = 88.07
Average mosaicity (°)	0.52
Total No. of measured intensities	110969
Total No. of unique reflections	8435
Completeness (%)	99.9 (99.9)
Multiplicity	13.2 (13.0)
CC _{1/2} (%)	99.6 (58.5)
Mean <i>I</i> /σ(<i>I</i>)	11.3 (1.9)
<i>R</i> _{merge} [†] (%)	0.219 (1.645)
<i>R</i> _{p.i.m.} [‡] (%)	0.063 (0.473)

[†] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations. [‡] $R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the *i*th observation, $\langle I(hkl) \rangle$ is the weighted average intensity for all observations and $N(hkl)$ is the multiplicity of reflection *hkl*.

3. Results and discussion

Initially, CBM_E1 was identified after bioinformatics analysis of a metagenomic cellulase clone, which showed a region rich in tryptophan and tyrosine residues that are commonly found in CBMs. Further *BlastP* analysis revealed low homology of CBM_E1 (31% amino-acid sequence identity) to the C-terminal region of cellulase from *Pseudomonas* sp. (GenBank BAB79288.1), a region with no putative conserved domains.

Figure 3
A diffraction pattern of the recombinant CBM_E1 data set collected on beamline MX2 at the Brazilian National Synchrotron Laboratory (LNLS).

As a first step towards gaining insights into its molecular mechanism, CBM_E1 was cloned into pET-28a and overexpressed in *E. coli* Origami 2 (DE3) cells. Purified protein was obtained by a two-step protocol consisting of affinity and size-exclusion purification steps. The molecular weight of 10 kDa for CBM_E1 was confirmed by 15% SDS–PAGE (Fig. 1).

Crystals suitable for X-ray analysis were obtained by the sitting-drop vapour-diffusion method after 25–30 d in different conditions: (i) 4 M sodium formate and (ii) 0.1 M sodium acetate pH 4.5, 0.2 M lithium sulfate, 30% PEG 8000. The best crystal was grown in 4 M sodium formate (Fig. 2) and diffracted to 1.95 Å resolution (Fig. 3). Based on the protein molecular weight, the calculated Matthews coefficient is 2.85 Å³ Da^{−1} (Matthews, 1968), corresponding to 56.81% solvent content and a monomer in the asymmetric unit. Statistics for the collection and processing of the data set are given in Table 3.

The single-wavelength anomalous diffraction method is being applied in order to solve the crystal structure. In parallel with structural studies, comprehensive biochemical and functional analysis including substrate-specificity studies are being carried out. The CBM_E1 is a cellulose-binding domain that can modify cellulose or otherwise assist cellulose hydrolysis by the catalytic domain; future studies will define the role of this novel CBM in cell-wall degradation.

This work was funded by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. We gratefully acknowledge the provision of time on the MX2 beamline (LNLS/CNPq) and Robolab (LNBio/CNPq) at the National Center for Research in Energy and Materials, Campinas, Brazil.

References

Alvarez, T. M., Paiva, J. H., Ruiz, D. M., Cairo, J. P. L. F., Pereira, I. O., Paixão, D. A. A., de Almeida, R. F., Tonoli, C. C. C., Ruller, R., Santos, C. R., Squina, F. M. & Murakami, M. T. (2013). *PLoS One*, **8**, e83635.
Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995). *Microbiol. Rev.* **59**, 143–169.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
Berlin, A., Gilkes, N., Kurabi, A., Bura, R., Tu, M., Kilburn, D. & Saddler, J. (2005). *Appl. Biochem. Biotechnol.* **121–124**, 163–170.
Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B. & Gilbert, H. J. (2004). *J. Biol. Chem.* **279**, 22953–22963.
Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochem. J.* **382**, 769–781.
Cota, J., Oliveira, L. C., Damásio, A. R. L., Citadini, A. P., Hoffmam, Z. B., Alvarez, T. M., Codima, C. A., Leite, V. B. P., Pastore, G., de Oliveira-Neto, M., Murakami, M. T., Ruller, R. & Squina, F. M. (2013). *Biochim. Biophys. Acta*, **1834**, 1492–1500.
Damásio, A. R. L., Rubio, M. V., Oliveira, L. C., Segato, F., Dias, B. A., Citadini, A. P., Paixão, D. A. & Squina, F. M. (2014). *Biotechnol. Bioeng.* **111**, 1494–1505.
Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
Farrell, A. E., Plevin, R. J., Turner, B. T., Jones, A. D., O'Hare, M. & Kammen, D. M. (2006). *Science*, **311**, 506–508.
Gonçalves, T. A., Damásio, A. R. L., Segato, F., Alvarez, T. M., Bragatto, J., Brenelli, L. B., Citadini, A. P. S., Murakami, M. T., Ruller, R., Leme, A. F. P., Prade, R. A. & Squina, F. M. (2012). *Bioresour. Technol.* **119**, 293–299.
Graham, J. E., Clark, M. E., Nadler, D. C., Huffer, S., Chokhawala, H. A., Rowland, S. E., Blanch, H. W., Clark, D. S. & Robb, F. T. (2011). *Nature Commun.* **2**, 375.
Guillén, D., Sánchez, S. & Rodríguez-Sanoja, R. (2010). *Appl. Microbiol. Biotechnol.* **85**, 1241–1249.
Guimarães, B. G., Sanfelici, L., Neuenschwander, R. T., Rodrigues, F., Grizolli, W. C., Raulik, M. A., Piton, J. R., Meyer, B. C., Nascimento, A. S. & Polikarpov, I. (2009). *J. Synchrotron Rad.* **16**, 69–75.
Hashimoto, H. (2006). *Cell. Mol. Life Sci.* **63**, 2954–2967.

- Kim, I. J., Ko, H.-J., Kim, T.-W., Nam, K. H., Choi, I.-G. & Kim, K. H. (2013). *Appl. Microbiol. Biotechnol.* **97**, 5381–5388.
- Kim, E. S., Lee, H. J., Bang, W.-G., Choi, I.-G. & Kim, K. H. (2009). *Biotechnol. Bioeng.* **102**, 1342–1353.
- Lorenz, P. & Eck, J. (2005). *Nature Rev. Microbiol.* **3**, 510–516.
- Luís, A. S., Venditto, I., Temple, M. J., Rogowski, A., Baslé, A., Xue, J., Knox, J. P., Prates, J. A. M., Ferreira, L. M. A., Fontes, C. M. G. A., Najmudin, S. & Gilbert, H. J. (2013). *J. Biol. Chem.* **288**, 4799–4809.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Santos, C. R. *et al.* (2012). *Biochem. J.* **441**, 95–104.