Cloning, purification, crystallization and preliminary X-ray studies of a carbohydrate-binding module from family 64 (StX)

Bruna Medeira Campos,^a Marcelo Vizona Liberato,^b Igor Polikarpov,^c
Ana Carolina de Mattos Zeri^a and Fabio Marcio Squina^{b*}

Received 1 December 2014

Accepted 2 February 2015

Keywords: accessory domain; carbohydrate-binding module.

^aLaboratório Nacional de Biociências (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil, ^bLaboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, São Paulo, Brazil, and ^cInstituto de Física de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil. *Correspondence e-mail: fabio.squina@bioetanol.org.br

In recent years, biofuels have attracted great interest as a source of renewable energy owing to the growing global demand for energy, the dependence on fossil fuels, limited natural resources and environmental pollution. However, the cost-effective production of biofuels from plant biomass is still a challenge. In this context, the study of carbohydrate-binding modules (CBMs), which are involved in guiding the catalytic domains of glycoside hydrolases to polysaccharides, is crucial for enzyme development. Aiming at the structural and functional characterization of novel CBMs involved in plant polysaccharide deconstruction, an analysis of the CAZy database was performed and CBM family 64 was chosen owing to its capacity to bind with high specificity to microcrystalline cellulose and to the fact that it is found in thermophilic microorganisms. In this communication, the CBM-encoding module named StX was expressed, purified and crystallized, and X-ray diffraction data were collected from native and derivatized crystals to 1.8 and 2.0 Å resolution, respectively. The crystals, which were obtained by the hanging-drop vapour-diffusion method, belonged to space group $P3_121$, with unit-cell parameters $a = b = 43.42$, $c = 100.96$ Å for the native form. The phases were found using the single-wavelength anomalous diffraction method.

1. Introduction

Despite the advances in biofuel production through the utilization of enzymes for biomass conversion, cost-effective production of biofuels is still a challenge (Bolam *et al.*, 2004; Farrell *et al.*, 2006; Santos *et al.*, 2012; Alvarez *et al.*, 2013; Kim *et al.*, 2013; Reyes-Ortiz *et al.*, 2013). The conversion of lignocellulose into sugars is a complex process and many approaches have been taken to improve it, such as the development of highly active cellulases (Graham *et al.*, 2011; Kim *et al.*, 2011; Merino & Cherry, 2007). For this reason, carbohydrate-binding modules (CBMs), which have non-hydrolytic activity, are being studied extensively (Kim *et al.*, 2009; Guillén *et al.*, 2010; Suwannarangsee *et al.*, 2012; Campos *et al.*, 2014). The function and behaviour of these auxiliary proteins, as well as their interaction with the catalytic subunits, still represents a gap in understanding carbohydrate degradation (Kim *et al.*, 2009; Guillén *et al.*, 2010; Viegas *et al.*, 2013). The CBMs are grouped into 71 families in the CAZy database based on sequence similarity (Freelove *et al.*, 2001; Cantarel *et al.*, 2009); nevertheless, many of these families have not been characterized. In order to obtain structural insight into one of these uncharacterized CBMs, we chose to study family 64, which encodes the module named StX (Angelov *et al.*, 2011).

© 2015 International Union of Crystallography

Table 1

Macromolecule cloning and expression conditions.

Source organism	<i>S. thermophila</i>
DNA source	Synthetic DNA
Cloning vector	pBlueScript
Expression vector	pET-28a
Expression host	<i>E. coli</i> ΔSlyD pRARE
Complete amino-acid sequence of the construct†	mgsshhhhhhssglvprgshSTPGGGEYTEIAL-PFSYDGAGEYYWKTDDFSTTTNWGRYVNSWNL-DLLEINGTDYANTWVPQHAIPASDGYWYIHY-KGSYPWSHVEM

† The histidine tag from the vector is represented in lower case in the amino-acid sequence. The tag was removed for the crystallization experiments.

StX has been found in the C-terminal domains of several *Spirochaeta thermophila* proteins (Angelov *et al.*, 2011), associated with glycoside hydrolase families 5, 9, 10, 12 and 48. It has also been shown that StX binds to microcrystalline cellulose with high specificity (Angelov *et al.*, 2011). The StX protein has 92 amino acids and a predicted molecular weight of 10.5 kDa. The crystallographic structure of StX will shed light on the mechanism of protein–ligand recognition and its role in biomass degradation. In this communication, we describe the expression, purification, crystallization and data processing of StX.

2. Materials and methods

2.1. Cloning, expression and purification

The gene encoding the carbohydrate-binding module StX (UniProt E0RP41) found in a xylanase from *S. thermophila* was synthesized by GenOne Biotechnologies (Rio de Janeiro, Brazil). The synthesized region based on the E0RP41 sequence contained nucleotides 1087–1353, which encode the CBM (amino acids 363–451). The synthesized gene was digested with the NdeI and NotI restriction enzymes, cloned into the expression vector pET-28a and verified by DNA sequencing

Figure 1

Crystals of StX were obtained in the presence of 7.5% PEG 1000, 17.5% PEG 8000 by the hanging-drop vapour-diffusion method.

Table 2

Crystallization conditions.

Method	Vapour diffusion
Plate type	Hanging drop
Temperature (K)	291
Protein concentration (mg ml ⁻¹)	5
Buffer composition of protein solution	20 mM sodium phosphate pH 7.2, 50 mM NaCl
Composition of reservoir solution	7.5% PEG 1000, 17.5% PEG 8000
Volume and ratio of drop	2 µl, 1:1
Volume of reservoir (µl)	200

(Table 1). The final construct encodes full-length StX fused to an N-terminal His tag with a thrombin protease cleavage site for tag removal.

The recombinant StX was expressed in *Escherichia coli* strain ΔSlyD pRARE. A single colony was used to inoculate a 10 ml Luria–Bertani (LB) starter culture supplemented with kanamycin (50 mg ml⁻¹) and chloramphenicol (25 mg ml⁻¹) and was used to inoculate 3 l LB medium, which was 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 310 K. The cells were harvested by centrifugation (15 000 rev min⁻¹), resuspended in 20 ml 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 supernatant was incubated with nickel resin for 2 h at room temperature. The beads were washed with 50 ml washing buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 10 mM imidazole, 20% glycerol) and the retained proteins were eluted with 6 ml washing buffer containing 200 mM imidazole. The 6×His tag was cleaved with thrombin (1 U µl⁻¹) at 16°C for 16 h. The protein was further purified on a HiLoad 16/60 Superdex 75 prep-grade column equilibrated with 20 mM sodium phosphate pH 7.2, 50 mM NaCl. Purified StX was stored at 4°C.

2.2. Crystallization

A highly purified StX sample was concentrated to 5 mg ml⁻¹ in 20 mM sodium phosphate pH 7.4, 50 mM NaCl. Initial crystallization experiments were performed by the sitting-drop vapour-diffusion method at 291 K using a Honeybee 963 robot (Genomic Solutions). The drop consisted of 0.5 µl StX solution plus 0.5 µl reservoir solution. As the first crystals did not diffract, manual refinement was performed using the hanging-drop vapour-diffusion method with the drop consisting of 1 µl StX solution plus 1 µl reservoir solution. The crystals grew in 24 h (Fig. 1) and were used for X-ray data collection (Table 2).

2.3. Data collection and processing

The crystals (Fig. 1) were soaked in a cryoprotection solution consisting of 15% glycerol in the crystallization solution. For derivatization, crystals were incubated for 5 min in crystallization solution containing 15% glycerol and 0.8 M sodium iodide. After incubation, the crystals were flash-cooled in a stream of gaseous nitrogen at 100 K and X-ray diffraction data

were collected on the MX2 beamline (Guimarães *et al.*, 2009) of the Brazilian Synchrotron Light Laboratory (LNLS;

(a)

(b)

Figure 2
X-ray diffraction patterns of StX crystals. (a) X-ray diffraction pattern of a native crystal of StX. (b) X-ray diffraction pattern of an iodine-derivative crystal of StX.

Table 3

Data-collection and processing statistics.

Values in parentheses are for the outer shell.

	Native	Iodine derivative
No. of crystals	1	1
X-ray source	MX2, LNLS	MX2, LNLS
Wavelength (Å)	1.459	1.8
Temperature (K)	100	100
Detector	PILATUS 2M	PILATUS 2M
Crystal-to-detector distance (mm)	151	130
Rotation range per image (°)	0.25	0.5
Total rotation range (°)	361.25	998
Exposure time per image (s)	1.2	1.2
Resolution range (Å)	30.16–1.8 (1.84–1.80)	31.53–2.0 (2.07–2.02)
Space group	$P3_121$	$P3_121$
Unit-cell parameters (Å)	$a = b = 43.42$, $c = 100.96$	$a = b = 43.65$, $c = 114.37$
Average mosaicity (°)	0.46	0.73
Total No. of reflections	136618	292111
No. of unique reflections	10529	8502
Multiplicity	13.0 (3.3)	34.4 (7.9)
$CC_{1/2}$ (%)	99.8 (90.7)	99.7 (76.1)
Mean $I/\sigma(I)$	14.0 (2.2)	15.1 (2.0)
Completeness (%)	97.9 (81.0)	96.3 (72.0)
$R_{\text{merge}}^{\dagger}$	0.094 (0.197)	0.142 (0.695)
$R_{\text{p.i.m.}}^{\ddagger}$	0.034 (0.173)	0.031 (0.367)
Anomalous completeness (%)		96.0 (70.3)
Anomalous multiplicity		18.4 (4.1)
Anomalous $CC_{1/2}$ (%)	−28.6 (−22.9)	74.5 (31.2)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, 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. $\ddagger R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, 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 .

Campinas, SP, Brazil) using a PILATUS 2M detector (Dectris). In order to collect suitable native and iodine-derivative data sets, the synchrotron-radiation wavelength was set to 1.459 and 1.8 Å, respectively. The data sets were processed with *iMosflm* (Battye *et al.*, 2011) and scaled with *AIMLESS* (Evans, 2006). The single-wavelength anomalous dispersion method was performed using *AutoSol* (Terwilliger *et al.*, 2009) from *PHENIX* (Adams *et al.*, 2010).

3. Results and discussion

According to the CAZy database, several CBMs families have no three-dimensional structures solved to date. Family 64, a cellulose-binding domain with high specificity (Angelov *et al.*, 2011), is one of them. The StX gene was synthesized by GenOne Biotechnologies, cloned into pET-28a and over-expressed in *E. coli* ΔSlyD pRARE cells. The purified protein was obtained after a two-step protocol consisting of affinity and size-exclusion chromatography and validated for purity by SDS-PAGE. Crystals suitable for X-ray analysis were obtained by the hanging-drop vapour-diffusion method in 24 h in 7.5% PEG 1000, 17.5% PEG 8000 (Fig. 1) and diffracted to 1.8 Å resolution (Fig. 2). Based on the protein molecular weight, the calculated Matthews coefficient is $2.75 \text{ Å}^3 \text{ Da}^{-1}$ (Matthews, 1968), corresponding to 55.3% solvent content with a monomer in the asymmetric unit.

The StX sequence does not have a high identity to that of any protein in the PDB (<40%), precluding the use of molecular replacement for phasing. Consequently, single-wavelength anomalous dispersion methods were applied in order to solve the crystal structure. Crystals from the same condition as used for native data collection were incubated in sodium iodide solution and a complete and redundant data set was collected at a wavelength of 1.8 Å in order to enhance the anomalous scattering. Scaling with *AIMLESS* resulted in a mid-slope anomalous normal probability value of 1.451 Å and a reasonable anomalous correlation between half-sets (Table 3). This statistics strongly pointed to detection of a pronounced anomalous signal and therefore to the presence of iodine in the structure. The first round of SAD phasing, auto-building and refinement using *AutoSol* found 15 iodine sites and built a partial model with 81 residues and a final R_{work} and R_{free} of 0.28 and 0.29, respectively. The statistics for the collection and processing of the native and derivative data sets are given in Table 3.

Phasing of the native data set by molecular replacement using the iodine-derivative structure as a model is currently in progress. In parallel with structural studies, comprehensive biochemical and functional analyses including substrate-specificity studies are being carried out.

Acknowledgements

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/CNPEM) and Robolab (LNBio/CNPEM) at the National Center for Research in Energy and Materials (CNPEM), Campinas, Brazil.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- 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.
- Angelov, A., Loderer, C., Pompei, S. & Liebl, W. (2011). *Appl. Environ. Microbiol.* **77**, 5483–5489.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Bolam, D. N., Xie, H., Pell, G., Hogg, D., Galbraith, G., Henrissat, B. & Gilbert, H. J. (2004). *J. Biol. Chem.* **279**, 22953–22963.
- Campos, B. M., Alvarez, T. M., Liberato, M. V., Polikarpov, I., Gilbert, H. J., Zeri, A. C. M. & Squina, F. M. (2014). *Acta Cryst.* **F70**, 1232–1235.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
- 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.
- Freelove, A. C., Bolam, D. N., White, P., Hazlewood, G. P. & Gilbert, H. J. (2001). *J. Biol. Chem.* **276**, 43010–43017.
- 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.
- Kim, E. S., Lee, H. J., Bang, W.-G., Choi, I.-G. & Kim, K. H. (2009). *Biotechnol. Bioeng.* **102**, 1342–1353.
- 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, T.-W., Chokhawala, H. A., Hess, M., Dana, C. M., Baer, Z., Sczyrba, A., Rubin, E. M., Blanch, H. W. & Clark, D. S. (2011). *Angew. Chem. Int. Ed.* **50**, 11215–11218.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Merino, S. T. & Cherry, J. (2007). *Adv. Biochem. Eng. Biotechnol.* **108**, 95–120.
- Reyes-Ortiz, V., Heins, R. A., Cheng, G., Kim, E. Y., Vernon, B. C., Elandt, R. B., Adams, P. D., Sale, K. L., Hadi, M. Z., Simmons, B. A., Kent, M. S. & Tullman-Ercek, D. (2013). *Biotechnol. Biofuels*, **6**, 93.
- Santos, C. R. *et al.* (2012). *Biochem. J.* **441**, 95–104.
- Suwanarangssee, S., Bunternsook, B., Arnthong, J., Paemane, A., Thamchaipenet, A., Eurwilaichitr, L., Laosiripojana, N. & Champrada, V. (2012). *Bioresour. Technol.* **119**, 252–261.
- 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. & Hung, L.-W. (2009). *Acta Cryst.* **D65**, 582–601.
- Viegas, A., Sardinha, J., Freire, F., Duarte, D. F., Carvalho, A. L., Fontest, C. M. G., Romão, M. J., Macedo, A. L. & Cabrita, E. J. (2013). *Biochem. J.* **79**, 289–300.