Flávio R. Rosseto,^{a‡} Ana C. Puhl,^{a‡} Maxuel O. Andrade^b and Igor Polikarpov^{a*}

^aInstituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador Saocarlense 400, São Carlos-SP 13560-970, Brazil, and

^bInstituto de Química, Universidade de São Paulo, Avenida Prof. Lineu Prestes 748, Bloco Química Fina, Sala 10, Cidade Universitária, São Paulo-SP 05508-000, Brazil

‡ These authors contributed equally.

Correspondence e-mail: ipolikarpov@ifsc.usp.br

Received 29 October 2012

Accepted 20 December 2012

Crystallization and preliminary diffraction analysis of the catalytic domain of major extracellular endoglucanase from *Xanthomonas campestris* pv. *campestris*

Cellulases, such as endoglucanases, exoglucanases and β -glucosidases, are important enzymes used in the process of enzymatic hydrolysis of plant biomass. The bacteria *Xanthomonas campestris* pv. *campestris* expresses a large number of hydrolases and the major endoglucanase (XccEG), a member of glycoside hydrolase family 5 (GH5), is the most strongly secreted extracellularly. In this work, the native XccEG was purified from the extracellular extract and crystallization assays were performed on its catalytic domain. A complete data set was collected on an in-house X-ray source. The crystal diffracted to 2.7 Å resolution and belonged to space group C2, with unit-cell parameters $a = 174.66$, $b = 141.53$, $c = 108.00$ Å, $\beta = 110.49^\circ$. The Matthews coefficient suggests a solvent content of 70.1% and the presence of four protein subunits in the asymmetric unit.

1. Introduction

Cellulose is the most abundant renewable biological resource. Biomass, such as for example sugar cane bagasse, has a large amount of cellulose (35–50%), hemicellulose (20–35%) and lignin (10–25%) (Saha, 2004) and its enzymatic degradation holds the promise of sustainable biofuel production. The enzymatic hydrolysis of cellulose is mediated by cellulases produced by microorganisms such as bacteria and fungi. The main enzymes involved in this process are endoglucanases (endo-1,4- β -D-glucanases), exoglucanases (exo-1,4- β -D-glucanases) and β -glucosidases (1,4- β -D-glycosidases) in addition to a number of auxiliary proteins (Kleywegt *et al.*, 1997; Gao *et al.*, 2008; Serpa & Polikarpov, 2011; Horn *et al.*, 2012). Cellulases are usually organized into two domains connected by a linker region: the catalytic core domain (CCD) and cellulose-binding module (CBM). The CBM seems to enhance binding to the crystalline cellulose, while the other domain is responsible for catalytic activity (Gilkes *et al.*, 1991).

Although most industrial cellulases are of fungal origin, notably from *Trichoderma* and *Aspergillus*, scientific interest in cellulase production by cellulolytic bacteria such as *Cellulomonas* and *Clostridium* has been increasing in recent years (Lynd *et al.*, 2002). Genome analysis of *Xanthomonas campestris* pv. *campestris* (Xcc) has revealed the presence of multiple genes encoding enzymes involved in the process of cell-wall degradation, including pectinases, lyases, xylanases and nine cellulases (Van Sluys *et al.*, 2002). This shows that Xcc is capable of secreting various enzymes that degrade structural polysaccharides such as cellulose and pectin (Sutherland, 1993).

The most abundant enzymes secreted by Xcc are endoglucanases. One of them is the major endoglucanase encoded by the gene *engXCA* (Gough *et al.*, 1990). The Xcc major endoglucanase (XccEG) is a member of the GH5 (glycoside hydrolase 5) family (Cantarel *et al.*, 2009), which is one of the largest families of glycoside hydrolases in the Carbohydrate-Active Enzymes (CAZy) database.

Although the most abundant enzymes secreted by Xcc are endoglucanases, there is little information about these enzymes in the literature. In this work, we describe the isolation, purification and

plates) varying the concentration of precipitant and pH. Suitable crystals were obtained in the same condition.

2.4. Data collection and processing

A single crystal of the XccEG was harvested using a nylon loop and transferred to a cryoprotectant solution containing 20% ethylene glycol and flash-cooled to 100 K in a nitrogen stream prior to X-ray data collection. X-ray diffraction experiments were performed with an R-Axis IV⁺⁺ imaging-plate detector mounted on a Rigaku MicroMax-007 HF rotating-anode X-ray generator providing Cu K α radiation (1.54 Å) and equipped with Osmic Confocal Max-Flux optics. A data set was collected using the oscillation method. Data were processed with *XDS* (Kabsch, 2010) and reduced with the program *SCALA* from *CCP4* (Winn *et al.*, 2011). The program *POINTLESS* (Evans, 2006) was used to examine the symmetry of the diffraction pattern and scores the possible crystallographic symmetry, and the *phenix.xtriage* module of the *PHENIX* program (Adams *et al.*, 2002) was used to analyse the collected data for twinning.

3. Results and discussion

3.1. Enzyme purification and crystallization

As expected, after separation of cells in the culture medium the presence of xanthan gum was observed in the soluble fraction (Yalpani, 1987). To separate xanthan gum and purify the XccEG, precipitations with 20–40% ammonium sulfate were performed. It was observed that the majority of contaminants were removed with 20% ammonium sulfate and the major endoglucanase appeared almost pure and separated from xanthan gum with 40% of the precipitant. After this, an additional step of FPLC purification was performed using a Sephadex G-25 Superfine 5 ml column (GE Healthcare). The purity of the protein was analysed by 15% SDS-PAGE (Fig. 1, lane 1) and the yield of purification was about 1 mg per litre of culture. Protein identification was performed by mass spectrometry and the identified peptides covered a large part of the protein sequence (approximately 55%), confirming the identity of purified XccEG. The sequence of XccEG is shown in Fig. 2 and the peptides identified by mass spectrometry are shown in green. The full-length protein has a molecular weight of 50 kDa. After storage of the protein at 277 K for a week, fortuitous proteolytic cleavage of XccEG was detected. The cleavage presumably occurred in the linker

Figure 3
Crystals of XccEG grown in 0.1 M Tris–HCl pH 8.5, 2.0 M ammonium dihydrogen phosphate.

Table 1

Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.54
Space group	C2
Unit-cell parameters (Å, °)	$a = 174.66$, $b = 141.53$, $c = 108.00$, $\beta = 110.49$
Resolution range (Å)	45.33–2.70 (2.85–2.70)
Unique reflections	67461 (9782)
Multiplicity	3.6 (3.6)
Completeness (%)	100 (100)
$R_{\text{merge}}^{\dagger}$ (%)	9.6 (61.9)
$\langle I \rangle / \langle \sigma(I) \rangle$	11.6 (2.1)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $\langle I(hkl) \rangle$ is the average intensity for a set of i symmetry-related reflections and $I_i(hkl)$ is the value of the intensity for a single reflection within a set of symmetry-related reflections.

portion, separating the CBM and CCD, since it released a 38 kDa fragment corresponding to the catalytic domain (Fig. 1, lane 2).

The XccEG CCD was submitted to several crystallization trials and small crystals were obtained by a sitting-drop method in a condition consisting of 0.1 M Tris–HCl pH 8.5, 2.0 M ammonium dihydrogen phosphate. Suitable crystals were obtained in the same condition by the hanging-drop method (Fig. 3). X-ray diffraction data were collected on our in-house X-ray source from a single crystal and were processed with *XDS* (Kabsch, 2010). The crystal system and space group were defined from the processing statistics and using the program *POINTLESS* (Evans, 2006) and the data were scaled using *SCALA* from *CCP4* (Winn *et al.*, 2011). The crystal belonged to the monoclinic space group C2, with unit-cell parameters $a = 174.66$, $b = 141.53$, $c = 108.00$ Å, $\beta = 110.49^\circ$, and diffracted to 2.7 Å resolution (Fig. 4). The Matthews coefficient (Matthews, 1968) suggested the presence of four subunits in the asymmetric unit and a solvent content of 70.1%. Twinning tests were performed using the *phenix.xtriage* module of the *PHENIX* program (Adams *et al.*, 2002). The multivariate Z-score from the *L*-tests of *PHENIX* was 1.93 and is well below the expected value of 3.5, which indicates good data

Figure 4
Diffraction pattern of a single crystal of XccEG.

quality and reveals that no twinning is suspected. Data-collection and processing statistics are summarized in Table 1.

4. Conclusions

The major endoglucanase of Xcc was purified from the extracellular fraction of Xcc after cultivation in 2xTY medium for 36 h at 303 K. The catalytic domain of the enzyme was submitted to crystallization and suitable crystals were obtained in 0.1 M Tris-HCl pH 8.5, 2.0 M ammonium dihydrogen phosphate by the hanging-drop method. An X-ray diffraction data set was collected to 2.7 Å resolution from a single crystal at a home source.

We are grateful to the Brazilian Synchrotron Light Source LNLS. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) via research grant Nos. 08/56255-9, 2007/08706-9, 10/52362-5 and 2009/05349-6, by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) via INCT do Bioetanol and grant Nos. 471834/2009-2 and 301981/2011-6.

References

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1948–1954.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). *Nucleic Acids Res.* **37**, D233–D238.
- Ebina, T., Toh, H. & Kuroda, Y. (2009). *Biopolymers*, **92**, 1–8.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R. & Bateman, A. (2010). *Nucleic Acids Res.* **38**, D211–D222.
- Gao, J., Weng, H., Zhu, D., Yuan, M., Guan, F. & Xi, Y. (2008). *Bioresour. Technol.* **99**, 7623–7629.
- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. & Warren, R. A. (1991). *Microbiol. Rev.* **55**, 303–315.
- Gough, C. L., Dow, J. M., Keen, J., Henrissat, B. & Daniels, M. J. (1990). *Gene*, **89**, 53–59.
- Horn, S. J., Vaaje-Kolstad, G., Westereng, B. & Eijsink, V. G. (2012). *Biotechnol. Biofuels*, **5**, 45–??.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kleywegt, G. J., Zou, J.-Y., Divne, C., Davies, G. J., Sinning, I., Ståhlberg, J., Reinikainen, T., Srisodsuk, M., Teeri, T. T. & Jones, T. A. (1997). *J. Mol. Biol.* **272**, 383–397.
- Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. (2002). *Microbiol. Mol. Biol. Rev.* **66**, 506–577.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. (2011). *Nature Methods*, **8**, 785–786.
- Saha, B. C. (2004). *Lignocellulose Biodegradation and Applications in Biotechnology*, edited by B. C. Saha & K. Hayashi, pp. 2–34. Washington: American Chemical Society.
- Serpa, V. I. & Polikarpov, I. (2011). *Routes to Cellulosic Ethanol*, edited by M. S. S. Buckeridge & G. H. H. Goldman, pp. 97–113. New York: Springer.
- Sutherland, I. W. (1993). *Xanthomonas*, edited by J. G. Swings & E. L. Civerolo, pp. 363–388. London: Chapman & Hall.
- Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E., Menck, C. F., Da Silva, A. C., Ferro, J. A., Oliveira, M. C., Setubal, J. C., Kitajima, J. P. & Simpson, A. J. (2002). *Annu. Rev. Phytopathol.* **40**, 169–189.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Yalpani, M. (1987). *Progress in Biotechnology*, Vol. 3, *Industrial Polysaccharides*, pp. 1–5. Amsterdam: Elsevier.