Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

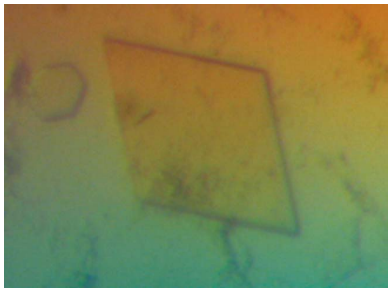
Crystallization and preliminary X-ray diffraction analysis of endoglucanase III from *Trichoderma harzianum*

Marcelo Vizoná Liberato,[a] Wesley Cardoso Generoso,[b] Wilson Malagó Jr,[b] Flávio Henrique-Silva[b] and Igor Polikarpov[a]*

[a]Instituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador Sãocarlense 400, 13566-590 São Carlos-SP, Brazil, and [b]Departamento de Genética e Evolução, Universidade Federal de São Carlos, Rodovia Washington Luís, km. 235, 13565-905 São Carlos-SP, Brazil

Correspondence e-mail: ipolikarpov@if.sc.usp.br

Received 20 September 2011
Accepted 9 January 2012



Endoglucanases are enzymes that hydrolyze cellulose and are important components of the cellulolytic complex. In contrast to other members of the complex, they cleave internal β-1,4-glycosidic bonds in the cellulose polymer, allowing cellulose to be used as an energy source. Since biomass is an important renewable source of energy, the structural and functional characterization of these enzymes is of interest. In this study, endoglucanase III from *Trichoderma harzianum* was produced in *Pichia pastoris* and purified. Crystals belonging to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 47.54, b = 55.57, c = 157.3 Å, were obtained by the sitting-drop vapour-diffusion method and an X-ray diffraction data set was collected to 2.07 Å resolution.

1. Introduction

To date, humans have mostly relied on energy provided by fossil fuels, but the increasing energy needs of the growing human population and the recognition of the problems of fossil fuels (pollution, global warming and scarcity, to mention three) have provided a driver for looking at alternative sustainable sources of energy (Singh *et al.*, 2011).

The method of producing energy through the hydrolysis of biomass has been known for a long time (Antoni *et al.*, 2007). Biomass is a renewable resource and the final product of its transformation *via* the fermentation of depolymerized sugars, cellulosic ethanol, is less toxic, more readily biodegradable and produces fewer airborne pollutants than petroleum fuel (John *et al.*, 2011). Although it is not yet possible to fulfil the world's energy demands using renewable fuels, they are increasingly becoming a feasible option.

Cellulose forms tightly interwoven crystalline fibres which are generally very resistant to degradation. However, a number of microorganisms are able to efficiently hydrolyze cellulose. For example, *Hypocrea* fungi in general and *H. jecorina* in particular are able to secrete large amounts of cellulases to hydrolyse cellulosic biomass. Another member of the same group, *H. lixii*, also known as *Trichoderma harzianum*, has been utilized in phytopathogen control (Monteiro *et al.*, 2010) and more recently has been described as a competent cellulolytic fungus which produces a balanced mixture of cellulases (de Castro, Ferreira *et al.*, 2010). It has been demonstrated that *T. harzianum* secretes large quantities of endoglucanases when cultivated on pretreated sugarcane bagasse (de Castro, Pedro *et al.*, 2010).

The cellulolytic complex produced by most fungal species is formed of at least three major components: (i) endoglucanases (EGs), which hydrolyze the internal β-1,4-glycosidic bonds in the cellulose polymer, thus introducing two new internal chain termini; (ii) cellobiohydrolases (CBHs), which act on the reduced and nonreduced termini of cellulose polymer chains, thus degrading cellulose into cellobiose; and (iii) β-glycosidases, which hydrolyze cellobiose to glucose (Medve *et al.*, 1998). A large number of endoglucanases have been identified and they have been classified into glycoside hydrolase families based on their amino-acid sequence similarity (Henrissat & Davies, 2000). Despite the differences in their three-dimensional structures, the catalytic mechanism attributed to all endoglucanases is

similar, *i.e.* they hydrolyze the β-1,4-glycosidic bond in cellulose *via* a double-displacement reaction and a glycosyl-enzyme intermediate that results in retention of the anomeric configuration in the product (Birsan *et al.*, 1998).

The present paper describes the crystallization of *T. harzianum* endoglucanase III and its preliminary X-ray diffraction studies. The enzyme belongs to the GH12 family, a group of low-molecular-weight endoglucanases that contain a single catalytic domain and lack a cellulose-binding module (Macarrón *et al.*, 1993; Nakazawa *et al.*, 2009). The members of this group have a compact β-sandwich structure that is curved to create an extensive cellulose-binding site on the concave face of the β-sheet (Sulzenbacher *et al.*, 1999). This architecture is based on a compact β-jelly-roll fold which contains one α-helix and two β-sheets consisting of a total of 15 β-strands (Sandgren *et al.*, 2001).

Several crystal structures of endoglucanase III from various organisms have previously been solved (Crennell *et al.*, 2006; Khademi *et al.*, 2002; Sulzenbacher *et al.*, 1997), including one from *T. reesei* (Sandgren *et al.*, 2001, 2004). Despite their strong structural similarity, the endoglucanases reveal considerable differences in their enzymatic activities, biochemical properties and thermal stabilities (Sandgren *et al.*, 2003). Little structural information is currently available for the *T. harzianum* glycosyl hydrolases. To at least partially fill this gap, we expressed, purified and crystallized endoglucanase III (EG3) from *T. harzianum* and collected X-ray diffraction data from the crystals.

2. Experimental methods

2.1. Protein production

The EG3 cDNA clone was obtained from a cDNA library constructed from *T. harzianum* IOC-3844 induced by microcrystalline cellulose (unpublished data). The EG3 ORF was cloned into pPICZαA (Invitrogen), excluding the putative signal peptide. The mature ORF was amplified with the forward primer 5′-CG**GAATTC**CAGACCAGCTGCGAAC-3′ and the reverse primer 5′-CGC**GTCGAC**GTTGATAGATGCGGTCC-3′ (the bases shown in bold indicate introduced restriction sites).

The amplicon and the pPICZαA plasmid were digested with *Eco*RI and *Sal*I and ligated, creating a recombinant plasmid containing the mature EG3 ORF flanked by the secretion signal peptide (α-factor) at the N-terminus and a 6×His tag at the C-terminus. The pPICZ vectors also contained an inducible promoter from the alcohol oxidase gene AOX1, which is induced by methanol.

Before transformation, the recombinant plasmid was linearized with *Pme*I endonuclease and then introduced into *Pichia pastoris* (MutS) (Invitrogen) by electroporation (1.5 kV, 25 µF, 200 Ω). Transformants were cultivated in solid yeast extract peptone dextrose (YEPD) with 1 *M* sorbitol and 100 µg ml^{-1} Zeocin. Recombinants were identified by polymerase chain reaction using the α-factor and AOX 3′ primers. For large-scale expression, a recombinant colony was inoculated in 10 ml buffered glycerol-complex medium (BMGY) and incubated at 303 K and 250 rev min^{-1} for 24 h. The 10 ml culture was added to 500 ml BMGY and incubated at 303 K and 250 rev min^{-1} for 24 h ($OD_{600\,nm}$ = 6). The cells were collected by centrifugation (1500*g* for 5 min) and the supernatant was discarded. The cells were resuspended in 100 ml buffered methanol-complex medium (BMMY) with 1% methanol and incubated at 303 K and 250 rev min^{-1} for 24 h. The culture containing the recombinant protein was centrifuged and the supernatant was vacuum-filtered through a 0.44 µm membrane.

2.2. Purification

Purification of the recombinant enzyme from the supernatant was performed by affinity chromatography using a 5 ml silica-resin column containing Ni–NTA Superflow resin (Qiagen). The resin was equilibrated with 25 ml binding buffer (10 m*M* Tris–HCl, 50 m*M* sodium phosphate, 100 m*M* sodium chloride pH 8.0). 100 ml of the supernatant was applied to 5 ml resin; after washing the resin with 15 ml binding buffer, the polyhistidine-tagged protein was eluted with 5 ml aliquots of elution buffer containing an increasing imidazole concentration (5–250 m*M*). The elution fractions were analyzed by 15% SDS–PAGE (Sambrook & Russell, 2001). The fractions containing purified protein were pooled and dialyzed against three changes of 50 m*M* Tris–HCl buffer pH 7.5 (2 l) at 277 K for 4 h. The protein concentration was determined using the BCA Protein Assay Kit (Thermo Scientific) and the samples were stored at 253 K.

2.3. Analysis of the hydrolysis of different substrates

The hydrolytic activity of EG3 was determined using the 3,5-dinitrosalicylic acid method based on quantification of the reducing-sugar content (Miller, 1959). The relative activity of the purified enzyme was checked using filter paper (Whatman, USA), carboxymethylcellulose (50–200 cps; Sigma, Germany), microcrystalline cellulose (Sigma Germany) and sawdust as substrates.

Briefly, 400 µl of 2% substrate in 100 m*M* potassium phosphate buffer pH 6.5 was mixed with 100 µl enzyme solution (0.4 µ*M* final concentration) and incubated at 315 K. After 15 min, the reaction was stopped by the addition of 1 ml dinitrosalicylic acid solution. The mixture was boiled for 5 min, diluted to 10 ml with water and the liberated reducing sugars were measured at a wavelength of 540 nm using a spectrophotometer (U-5100; Hitachi, Japan).

All assays were performed in triplicate and a reaction stopped just after addition of the enzyme was used as a blank. All experimental data and graphics were generated with the *Prism* 5 statistical program (GraphPad, USA).

2.4. Crystallization

Purified EG3 (23 781 Da; amino acids 18–236) at a concentration of 10 mg ml^{-1} was submitted to crystallization screening using the sitting-drop vapour-diffusion technique. Drops of 2 µl final volume (1:1 ratio of protein and screen solutions) were set up automatically

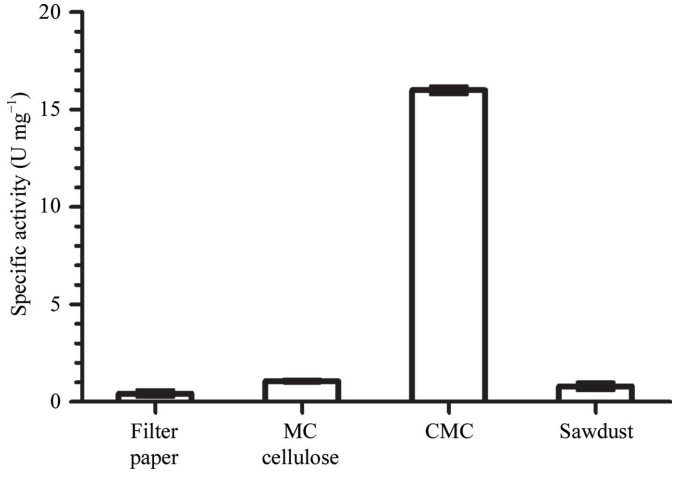


Figure 1
Substrate specificity of the purified recombinant EG3 from *T. harzianum*. Filter paper, microcrystalline cellulose (MC cellulose), carboxymethylcellulose (CMC) and sawdust were used as substrates.

using a Honeybee 931 crystallization robot (Genomic Solutions Inc.) with a variety of commercially available screens (PEGs I and II Suite, Classics Suite and MbClassics Suite from Qiagen) and maintained at a temperature of 291 K. 96-well crystallization plates from Greiner Bio-One were used.

2.5. Data collection and processing

A single EG3 crystal was transferred into a cryosolution consisting of 12% PEG 6000, 2 *M* sodium chloride, 20% ethylene glycol, mounted in a cryoloop and directly flash-cooled in a cold nitrogen stream prior to X-ray data analysis. Diffraction data were collected on beamline MX2 at the Brazilian Synchrotron Light Source (LNLS), Campinas, Brazil using a MAR charge-coupled device detector (Guimarães *et al.*, 2009). Data covering 225° were collected using a 0.5° oscillation step to a maximum resolution of beyond 2.0 Å using a MAR Mosaic225 detector. Data integration was carried out using *DENZO* and data were scaled with *SCALEPACK* in the *HKL*-2000 package (Otwinowski, 1993).

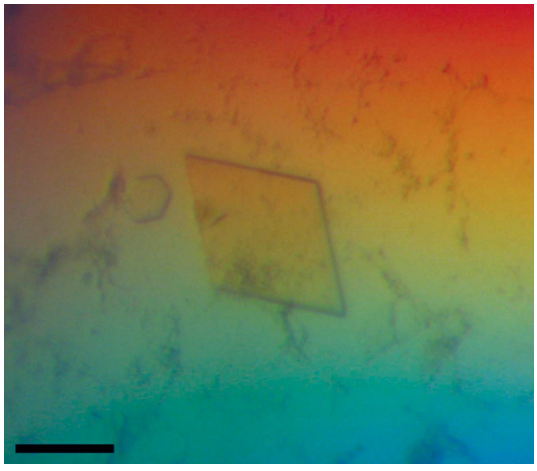


Figure 2
Crystals of EG3 grown in 25%(*w*/*v*) PEG 4000, 0.2 *M* ammonium sulfate, 0.1 *M* sodium acetate pH 4.6. The black bar corresponds to 200 µm.

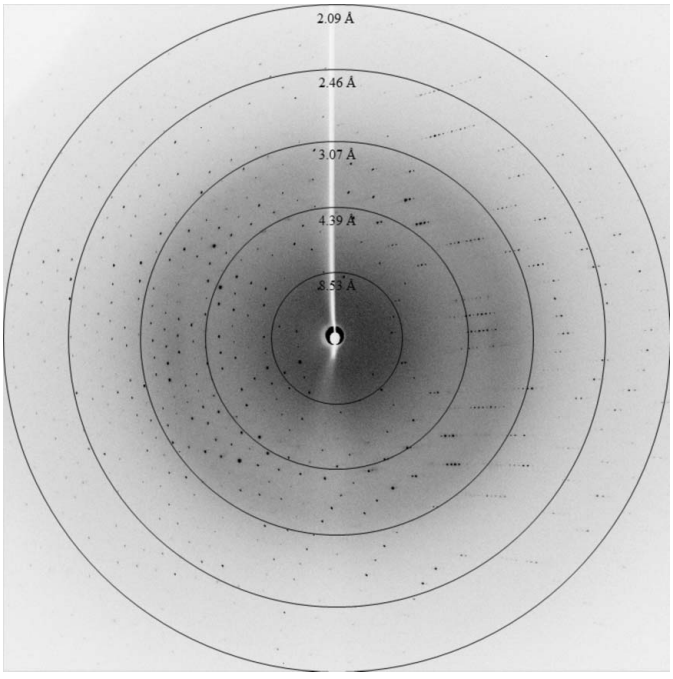


Figure 3
Typical diffraction pattern of an EG3 crystal.

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Beamline	MX2, LNLS
Wavelength (Å)	1.46
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 47.54, b = 55.57, c = 157.3
Resolution range (Å)	50–2.07 (2.14–2.07)
Unique reflections	26039
Total reflections	171836
Multiplicity	6.6 (6.3)
Completeness (%)	99.1 (97.5)
R_{merge}† (%)	12.1 (21.6)
$\langle I/\sigma(I)\rangle$	14.2 (8.1)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of reflection hkl.

3. Results and discussion

The purified EG3 was tested for enzymatic activity and proved to be active. The specific activity of the enzyme towards carboxymethyl cellulose (CMC) was 16 U mg^{-1}. This activity is comparable with that of *T. reesei* EG3 (Nakazawa *et al.*, 2009). Carboxymethyl cellulose was also preferentially consumed by EG3 compared with filter paper, microcrystalline cellulose and sawdust (Fig. 1).

Crystals of EG3 appeared after 5 d at 291 K in the original sitting-drop screen plates using the PEGs II Suite. A cluster of needles were obtained using a reservoir solution consisting of 25%(*w*/*v*) PEG 4000, 0.2 *M* ammonium sulfate, 0.1 *M* sodium acetate pH 4.6 (condition No. 65). A plate-shaped crystal (Fig. 2) was subsequently obtained in an optimization screen from a crystallization solution consisting of 12%(*w*/*v*) PEG 6000, 2.0 *M* sodium chloride (condition No. 77).

The plate-shaped crystal was tested for X-ray diffraction quality (Fig. 3) and was used to collect a complete diffraction data set to 2.07 Å resolution. Based on the data-set processing statistics, the EG3 crystal belonged to the primitive orthorhombic system ($P2_12_12_1$), with unit-cell parameters a = 47.54, b = 55.57, c = 157.3 Å. The calculated Matthews coefficient of 2.16 Å^3 Da^{-1} (Matthews, 1968) indicated that there were two molecules in the asymmetric unit; the solvent content of the crystal was 43.2%. Data-collection and processing statistics are presented in Table 1.

The structure of EG3 was solved by the molecular-replacement technique with *Phaser* v.2.1.4 (McCoy *et al.*, 2007) using the atomic coordinates of the *T. reesei* EG3 crystal structure (PDB entry 1h8v; Sandgren *et al.*, 2001) as a search model, giving a solution LLG (log-likelihood gain) of 2016.85. The amino-acid sequence identity between the EG3s from *T. reesei* and *T. harzianum* is 83%. Initial structure-refinement cycles using *REFMAC*5 (Murshudov *et al.*, 2011) resulted in a model with an R_{work} of 26.85% and an R_{free} of 30.96%. Further model building and refinement is in progress.

We are grateful to the Brazilian Synchrotron Light Source (LNLS) and to the staff members of the MX2 beamline. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) *via* research grants 2009/14333-6, 2008/56255-9 and 2009/52840-7 and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) *via* INCT do Bioetanol and grants 471834/2009-2 and 490022/2009-0.

References

Antoni, D., Zverlov, V. V. & Schwarz, W. H. (2007). *Appl. Microbiol. Biotechnol.* **77**, 23–35.

Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., Nitz, M., Rose, D. R., Tull, D., Wakarchuck, W. W., Wang, Q., Warren, R. A., White, A. & Withers, S. G. (1998). *Biochem. Soc. Trans.* **26**, 156–160.

Castro, A. M. de, Ferreira, M. C., da Cruz, J. C., Pedro, K. C., Carvalho, D. F., Leite, S. G. & Pereira, N. (2010). *Enzyme Res.* **2010**, 854526.

Castro, A. M. de, Pedro, K. C., da Cruz, J. C., Ferreira, M. C., Leite, S. G. & Pereira, N. Jr (2010). *Appl. Biochem. Biotechnol.* **162**, 2111–2122.

Crennell, S. J., Cook, D., Minns, A., Svergun, D., Andersen, R. L. & Nordberg Karlsson, E. (2006). *J. Mol. Biol.* **356**, 57–71.

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.

Henrissat, B. & Davies, G. J. (2000). *Plant Physiol.* **124**, 1515–1519.

John, R. P., Anisha, G. S., Nampoothiri, K. M. & Pandey, A. (2011). *Bioresour. Technol.* **102**, 186–193.

Khademi, S., Zhang, D., Swanson, S. M., Wartenberg, A., Witte, K. & Meyer, E. F. (2002). *Acta Cryst.* D**58**, 660–667.

Macarrón, R., Acebal, C., Castillón, M. P., Domínguez, J. M., de la Mata, I., Pettersson, G., Tomme, P. & Claeyssens, M. (1993). *Biochem. J.* **289**, 867–873.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Medve, J., Karlsson, J., Lee, D. & Tjerneld, F. (1998). *Biotechnol. Bioeng.* **59**, 621–634.

Miller, G. L. (1959). *Anal. Chem.* **31**, 426–428.

Monteiro, V. N., do Nascimento Silva, R., Steindorff, A. S., Costa, F. T., Noronha, E. F., Ricart, C. A., de Sousa, M. V., Vainstein, M. H. & Ulhoa, C. J. (2010). *Curr. Microbiol.* **61**, 298–305.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nakazawa, H., Okada, K., Onodera, T., Ogasawara, W., Okada, H. & Morikawa, Y. (2009). *Appl. Microbiol. Biotechnol.* **83**, 649–657.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 55–62. Warrington: Daresbury Laboratory.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning. A Laboratory Manual*, 3rd ed. New York: Cold Spring Harbor Laboratory Press.

Sandgren, M., Berglund, G. I., Shaw, A., Ståhlberg, J., Kenne, L., Desmet, T. & Mitchinson, C. (2004). *J. Mol. Biol.* **342**, 1505–1517.

Sandgren, M., Gualfetti, P. J., Shaw, A., Gross, L. S., Saldajeno, M., Day, A. G., Jones, T. A. & Mitchinson, C. (2003). *Protein Sci.* **12**, 848–860.

Sandgren, M., Shaw, A., Ropp, T. H., Wu, S., Bott, R., Cameron, A. D., Ståhlberg, J., Mitchinson, C. & Jones, T. A. (2001). *J. Mol. Biol.* **308**, 295–310.

Singh, A., Nigam, P. S. & Murphy, J. D. (2011). *Bioresour. Technol.* **102**, 10–16.

Sulzenbacher, G., Mackenzie, L. F., Wilson, K. S., Withers, S. G., Dupont, C. & Davies, G. J. (1999). *Biochemistry*, **38**, 4826–4833.

Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C. & Davies, G. J. (1997). *Biochemistry*, **36**, 16032–16039.