Arun Goyal,^{a,b,*‡} Shadab
Ahmed,^{a‡} Carlos M. G. A.
Fontes^b and Shabir Najmudin^{b*}

^aDepartment of Biotechnology, Indian Institute of Technology Guwahati, Guwahati, 781 039, India, and ^bCIISA—Faculdade de Medicina Veterinária, Universidade de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisbon, Portugal

‡ These authors made equal contributions.

Correspondence e-mail: arungoyal@iitg.ernet.in, shabir@fmv.ulisboa.pt

Received 22 January 2014
Accepted 22 March 2014

Crystallization and preliminary X-ray crystallographic analysis of a novel α -L-arabinofuranosidase (CtGH43) from *Clostridium thermocellum* ATCC 27405

The truncated carbohydrate-active enzyme belonging to family 43 glycoside hydrolase from *Clostridium thermocellum* (CtGH43) is an α -L-arabinofuranosidase that in combination with endoxylanase leads to complete breakdown of L-arabinosyl-substituted xylans. The recombinant enzyme CtGH43 from *C. thermocellum* was overexpressed in *Escherichia coli* and purified by immobilized metal-ion affinity chromatography. The recombinant CtGH43 has a molecular mass of 35.86 kDa. Preliminary structural characterization was carried out on CtGH43 crystallized from different conditions, which gave either cube-shaped or brick-shaped crystals. These diffracted to a resolution of 1.65 Å for the cubic form and 1.1 Å for the monoclinic form. Molecular replacement was used to solve the CtGH43 structure.

1. Introduction

α -L-Arabinofuranosidase (α -L-Araf; EC 3.2.1.55) catalyzes the hydrolysis of L-arabinosyl linkages attached to the β -1,4-xylopyranose backbone (Saha, 2003). The latest classification based on amino-acid sequences, primary-structure similarities and hydrophobic cluster analysis classifies α -L-arabinofuranosidases (α -L-Araf) into five glycoside hydrolase (GH) families, *i.e.* GH30, GH43, GH51, GH54 and GH62 (Zhou *et al.*, 2012; Cartmell *et al.*, 2011; Sørensen *et al.*, 2006; Guais *et al.*, 2010; Hashimoto *et al.*, 2011). Although all GH43 family members belong to clan GH-F and have a fivefold β -propeller core architecture, they have different activities (see the CAZY website, <http://www.cazy.org>; Lombard *et al.*, 2013).

The truncated 903 bp gene encoding the N-terminus of CtGH43 (family 43 glycoside hydrolase from *Clostridium thermocellum*) is part of a full-length family 43 glycoside hydrolase (Ct43Araf) which also contains two tandem family 6 carbohydrate-binding modules, CtCBM6A (405 bp) and CtCBM6B (402 bp), at the C-terminus (Ahmed *et al.*, 2013). CtGH43 was expressed, biochemically characterized and shown to be predominantly an α -L-arabinofuranosidase (Ahmed *et al.*, 2013). The truncated CtGH43 shows a unique ability to catalyze both *p*-nitrophenol- α -L-arabinofuranoside (pNPaf) and *p*-nitrophenol- α -L-arabinopyranoside (pNPap), although the catalytic efficiency was greater with pNPaf (Ahmed *et al.*, 2013). Phylogenetic analysis of CtGH43 from *C. thermocellum* shows a close evolutionary relationship with GH43s linked to family 6 carbohydrate-binding modules (CBM6s) from *C. cellulolyticum*, *C. papyrosolvens*, *C. cellulyticum* and *Acetovibrio cellulolyticus* (Ahmed *et al.*, 2012). CtGH43 shows maximum activity at 50°C and at pH 5.4 with pH stability in the range 5.0–6.2 (Das *et al.*, 2012). Determining the three-dimensional structure of CtGH43 will further provide understanding of its activity towards different substrates. In the present study, recombinant CtGH43 was purified and crystallized. X-ray diffraction data were collected for preliminary structure analysis.

2. Materials and methods

2.1. Overexpression and purification of CtGH43

α -L-Arabinofuranosidase (CtGH43) from *C. thermocellum* was overexpressed in *Escherichia coli* BL21 pLysS (DE3) as described earlier (Das *et al.*, 2012). The recombinant CtGH43 containing a His₆

tag was purified from the cell-free extract by immobilized metal-ion affinity chromatography (IMAC) using a Sepharose column (HiTrap Chelating, GE Healthcare) and was eluted using 50 mM HEPES buffer containing 300 mM imidazole, 1 M NaCl, 3 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ pH 7.5. Purified CtGH43 was buffer-exchanged into 50 mM HEPES buffer pH 7.5 containing 200 mM NaCl, 5 mM CaCl_2 using PD-10 columns and subjected to gel filtration on a HiLoad 16/60 Superdex-75 column at a flow rate of 1 ml min^{-1} . Purified CtGH43 was concentrated using a 10 kDa cutoff centrifugal concentrator (Millipore, USA) and washed three times with 0.5 mM CaCl_2 . Its concentration was estimated using a molar extinction coefficient (ϵ) of $68\,823 \text{ M}^{-1} \text{ cm}^{-1}$. The recombinant CtGH43 contains an N-terminal His₆ tag (MGSSHHHHHHSSGLVPRGSHMAS), giving a total of 324 amino-acid residues with an approximate molecular mass of 35.86 kDa, as analyzed by SDS-PAGE (Das *et al.*, 2012).

2.2. Crystallization of CtGH43

Commercial screens such as Crystal Screen, Crystal Screen 2 and PEG/Ion from Hampton Research (California, USA), JCSG from Molecular Dimensions and an in-house-prepared sparse-matrix screen of 80 conditions were screened by the sitting-drop vapour-diffusion method for crystallization of CtGH43 using an Oryx8 nanodrop robotic dispensing system (Douglas Instruments). Drops consisting of 0.8 μl of either 60 or 120 mg ml^{-1} CtGH43 solution and 0.8 μl reservoir solution were prepared at 19°C. Cube-shaped or

brick-shaped crystals of CtGH43 grew after 10 d in different conditions in the initial screen as shown in Fig. 1. We also obtained cube-shaped crystals in just 2 M ammonium sulfate. These crystals were isomorphous, but diffracted less well than those from the condition given in Fig. 1(a).

The crystals were cryocooled in liquid nitrogen after soaking in cryoprotectant [30% (v/v) glycerol added to the crystallization buffer] for a few seconds except for the crystal grown in 0.1 M Tris pH 8.5, 20% PEG 2000 monomethyl ether, 0.2 M trimethyl N-oxide, for which Paratone-N was used.

2.3. Data collection and processing of CtGH43 crystals

The data from all the CtGH43 crystals obtained from different conditions were collected at the Diamond Light Source at the National Synchrotron Science Facility (Harwell, England) using a PILATUS 6M detector with the crystals cooled to -173°C using a Cryostream (Oxford Cryosystems). The data were processed by *fast DP* and *xia2* (Winter, 2010) using the programs *iMosflm* (Battye *et al.*, 2011) and *AIMLESS* (Evans, 2011) from the CCP4 suite (Winn *et al.*, 2011) or *XDS* (Kabsch, 2010). For the cube-shaped crystals (Fig. 1a), 120° of data were collected with a $\Delta\phi$ of 0.2° . These gave diffraction data between 1.65 and 2.0 Å resolution; the data sets were highly isomorphous and the crystal belonged to the cubic space group $P432$. Thus, the data from the best diffracting crystal (1.65 Å resolution) were used to solve the CtGH43 structure. The Matthews coefficient of $3.17 \text{ Å}^3 \text{ Da}^{-1}$ (Matthews, 1968) indicates the presence of one molecule in the asymmetric unit with a solvent content of 61%.

For the brick-shaped crystal (Fig. 1b), 400° of data were collected with a $\Delta\phi$ of 0.2° in two passes, first to 1.5 Å resolution and then to 1.1 Å resolution (Fig. 2). This crystal belonged to the monoclinic space group $P2_1$. The Matthews coefficient of $2.0 \text{ Å}^3 \text{ Da}^{-1}$ (Matthews, 1968) also indicates the presence of one molecule in the asymmetric unit, with a solvent content of 39%. Data-collection statistics for both

Figure 1
Crystals of CtGH43 produced by sitting-drop vapour diffusion grown with 120 mg ml^{-1} protein under the following conditions: (a) 0.1 M sodium acetate pH 4.5, 2 M ammonium sulfate and (b) 0.1 M Tris pH 8.5, 20% PEG 2000 monomethyl ether, 0.2 M trimethyl N-oxide. The largest cube-shaped crystals (Fig. 1a) are approximately $0.10 \times 0.10 \times 0.10 \text{ mm}$ in size, whereas the brick-shaped crystal (Fig. 1b) is $\sim 0.30 \times 0.15 \times 0.10 \text{ mm}$ in size.

Figure 2
Representative diffraction pattern from the CtGH43 crystal grown in 0.1 M Tris pH 8.5, 20% PEG 2000 monomethyl ether, 0.2 M trimethyl N-oxide. The outer circle corresponds to 1.1 Å resolution.

Table 1

Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Crystal	Cube-shaped	Brick-shaped
Beamline	I04, Diamond	I04, Diamond
Space group	<i>P</i> 432	<i>P</i> 2 ₁
Wavelength (Å)	0.9795	0.9795
Unit-cell parameters		
<i>a</i> (Å)	140.0	45.43
<i>b</i> (Å)	140.0	67.70
<i>c</i> (Å)	140.0	50.99
β (°)	90	113.11
Resolution limits (Å)	57.15–1.65 (1.68–1.65)	27.45–1.10 (1.16–1.10)
Average mosaicity (°)	0.30	0.38
No. of observations	1175673 (31832)	707665 (75874)
No. of unique observations	56751 (2732)	104966 (13587)
Multiplicity	20.7 (11.7)	6.7 (5.6)
Completeness (%)	100 (100)	91.5 (81.2)
$\langle I/\sigma(I) \rangle$	20.0 (1.8)	13.4 (4.9)
CC _{1/2} [†]	0.99 (0.67)	0.997 (0.944)
$R_{\text{merge}}^{\ddagger}$ (%)	9.1 (115.4)	12.6 (6.3/46.6)
$R_{\text{p.i.m.}}^{\S}$ (%)	2.1 (36.5)	8.0 (3.5/30.4)

[†] CC_{1/2} is the half data-set correlation coefficient (Diederichs & Karplus, 2013). [‡] $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} |F_{\text{obs}}|$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl*, including symmetry-related reflections, and $\langle I(hkl) \rangle$ is its average. [§] $R_{\text{p.i.m.}} = \sum_{hkl} \{1/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} |F_{\text{obs}}|$, where $\langle I(hkl) \rangle$ is the average of symmetry-related observations of a unique reflection.

CtGH43 crystalline forms are given in Table 1. *BALBES* was used to carry out molecular replacement (Long *et al.*, 2008). Five structures with a sequence identity of greater than 20% were found. The best solution was found using the GH43 arabinoxylan arabinofuranohydrolase from *Bacillus subtilis* (PDB entry 3c7e; Vandermarliere *et al.*, 2009; sequence identity 41%). This gave a monomer in space group *P*432, with a final *R* factor and *R*_{free} of 35.9 and 37.4%, respectively, and a *Q* factor of 0.767 after *REFMAC5* (Murshudov *et al.*, 2011). *Phaser* (McCoy *et al.*, 2007) was used to fit the PDB model output from *BALBES* into the monoclinic 1.1 Å resolution data, with a TFZ score of 27.5 and an LLG of 2062. As well as the difference in packing, there is a significant difference in some of the loops. Further structure refinement and analysis for both crystalline forms are ongoing.

This research was supported in part by a Cutting-edge Research Enhancement and Scientific Training (CREST) Fellowship from the Department of Biotechnology, Ministry of Science and Technology to AG. A scholarship from IIT Guwahati and UGC (MANF) to SA is

gratefully acknowledged. The authors also acknowledge financial support from Fundação para a Ciência e a Tecnologia, Portugal through projects PTDC/BIA-PRO/103980/2008 and PTDC/BIAPRO/100359/2008. We acknowledge the Diamond Light Source for provision of synchrotron-radiation facilities (beamline I04) and the European Community's Seventh Framework Programme (FP7/2007–2013) under BioStruct-X (grant agreement No. 283570, proposal No. Biostruct-X_4399) for funding. We would also like to thank Cecília Bonifácio and Pedro Bule for help with crystallization and Drs James Sandy and Joana L. A. Brás for help with data collection.

References

- Ahmed, S., Charan, R., Ghosh, A. & Goyal, A. (2012). *J. Proteins Proteomics*, **3**, 31–38.
- Ahmed, S., Luis, A. S., Bras, J. L. A., Ghosh, A., Gautam, S., Gupta, M. N., Fontes, C. M. G. A. & Goyal, A. (2013). *PLoS One*, **8**, e73575.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Cartmell, A., McKee, L. S., Peña, M. J., Larsbrink, J., Brumer, H., Kaneko, S., Ichinose, H., Lewis, R. J., Viksø-Nielsen, A., Gilbert, H. J. & Marles-Wright, J. (2011). *J. Biol. Chem.* **286**, 15483–15495.
- Das, S. P., Ravindran, R., Ahmed, S., Das, D., Goyal, D., Fontes, C. M. G. A. & Goyal, A. (2012). *Appl. Biochem. Biotechnol.* **167**, 1475–1488.
- Diederichs, K. & Karplus, P. A. (2013). *Acta Cryst.* **D69**, 1215–1222.
- Evans, P. R. (2011). *Acta Cryst.* **D67**, 282–292.
- Guais, O., Tourrasse, O., Dourdoigne, M., Parrou, J. L. & Francois, J. M. (2010). *Appl. Microbiol. Biotechnol.* **87**, 1007–1021.
- Hashimoto, K., Yoshida, M. & Hasumi, K. (2011). *Biosci. Biotechnol. Biochem.* **75**, 342–345.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2013). *Nucleic Acids Res.* **42**, D490–D495.
- Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 125–132.
- 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.
- 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.* **D67**, 355–367.
- Saha, B. C. (2003). *J. Ind. Microbiol. Biotechnol.* **30**, 279–291.
- Sørensen, H. R., Jørgensen, C. T., Hansen, C. H., Jørgensen, C. I., Pedersen, S. & Meyer, A. S. (2006). *Appl. Microbiol. Biotechnol.* **73**, 850–861.
- Vandermarliere, E., Bourgois, T. M., Winn, M. D., van Campenhout, S., Volckaert, G., Delcour, J. A., Strelkov, S. V., Rabijns, A. & Courtin, C. M. (2009). *Biochem. J.* **418**, 39–47.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Zhou, J., Bao, L., Chang, L., Zhou, Y. & Lu, H. (2012). *J. Ind. Microbiol. Biotechnol.* **39**, 143–152.