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

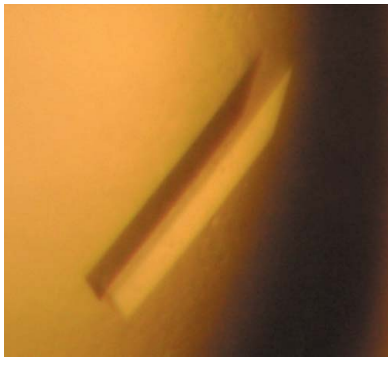
Expression, purification, crystallization and preliminary crystallographic analysis of an endo-1,5-α-L-arabinanase from hyperthermophilic *Thermotoga petrophila*

Fabio Marcio Squina,[a] Rolf Alexander Prade,[b] Hongliang Wang[b] and Mario Tyago Murakami[c]*

[a]Bioethanol Science and Technology Center, Brazilian Association for Synchrotron Light Technology, 13083-970 Campinas-SP, Brazil, [b]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA, and [c]Center for Structural Molecular Biology, Brazilian Association for Synchrotron Light Technology, 13083-970 Campinas-SP, Brazil

Correspondence e-mail: mtmurakami@lnls.br

Received 8 July 2009
Accepted 27 July 2009



The endo-1,5-α-L-arabinanases belonging to glycoside hydrolase family 43 are of great industrial interest for use in food technology, organic synthesis and biofuel production owing to their ability to catalyze the hydrolysis of α-1,5-arabinofuranosidic bonds in arabinose-containing polysaccharides. In this work, *Thermotoga petrophila* endo-1,5-α-L-arabinanase, a GH43-family member, has been cloned, overexpressed, purified and crystallized. Single crystals were obtained from a solution containing 0.1 *M* MES buffer pH 6.5, 0.8 *M* ammonium sulfate, 0.1 *M* EDTA, 0.1 *M* L-proline and 5%(*v*/*v*) dioxane. X-ray diffraction data were collected to a resolution of 2.86 Å using synchrotron radiation and the diffraction pattern was indexed in the tetragonal space group *P*422, with unit-cell parameters $a = b = 83.71$, $c = 408.25$ Å.

1. Introduction

The reduction of the cost of bioethanol production from lignocellulosic materials is the main challenge in both academic and industrial research. The critical step in the biomass-to-bioethanol process is the rapid and efficient conversion of cellulosic and hemicellulosic materials to monomeric sugars (Balat *et al.*, 2008). The enzymatic arsenal from thermophilic microorganisms has been instrumental in the discovery and development of new degradation pathways for this purpose (Hata *et al.*, 1992; Saha, 2000, 2003). As the polymers from the recalcitrant biomass are more accessible to chemical and enzymatic degradation at high temperatures, the use of thermostable enzymes is extremely desirable to attain a higher degree of efficiency on the industrial scale.

In the conversion of lignocellulosic biomass to biofuels, endo-1,5-α-L-arabinanases (EC 3.2.1.99) and α-L-arabinofuranosidases (EC 3.2.1.55) have attracted much attention owing to their ability to hydrolyze α-1,5-arabinofuranosidic bonds in hemicelluloses such as arabinoxylan and arabinan as well as in other arabinose-containing polysaccharides (Beldman *et al.*, 1997; Saha, 2000; Van Laere *et al.*, 2000). To date, only a few endo-1,5-α-L-arabinanases and α-L-arabinofuranosidases have been reported (Shallom & Shoham, 2003; Wong *et al.*, 2008) and the molecular basis of their thermostability and catalytic activity is not fully understood.

In this work, we report the cloning, overexpression, purification and crystallization of a putative endo-1,5-α-L-arabinanase, a member of glycoside hydrolase family 43 (GH43), from the hyperthermophilic bacterium *Thermotoga petrophila* strain RKU-1T, which grows optimally at 353 K (Takahata *et al.*, 2001). The structure determination of a new endo-1,5-α-L-arabinanase belonging to the GH43 family will provide essential information for the elucidation of the structural basis of its thermostability and catalytic mechanism, including its substrate specificity.

2. Materials and methods

2.1. Protein expression and purification

The full-length GH43 gene (Tpet_0637) was amplified from genomic DNA of *T. petrophila* by a standard PCR method using two

oligonucleotide primers (forward, 5′-ATGCCCATGGAGATTTCT-TTTTCTGATGATTACGC-3′; reverse, 5′-ATGCTCTAGAAATTC-TTCCACTCTTATTCCCCA-3′). The amplified GH43 gene was then cloned into the *Nco*I and *Xba*I restriction-enzyme sites of the pBAD/Myc-His B vector (Invitrogen), generating an expression plasmid in which the GH43 gene was expressed under the control of the *ara*BAD promoter with a *myc* epitope and a 6×His tag at the C-terminus. The resultant plasmid was used to transform the host strain *Escherichia coli* Top10 (Invitrogen). The transformants were grown in LB medium containing 50 µg ml^{-1} ampicillin at 310 K to an optical density of 0.6 at 600 nm. After induction with 0.02% arabinose, growth was continued for a further 16 h at 303 K and the cells were harvested by centrifugation. The harvested cells were washed and then resuspended in lysis buffer (50 m*M* Tris–HCl pH 7.5 and 100 m*M* NaCl) after an initial freeze–thaw step. The suspension was sonicated after treatment with 0.5 mg ml^{-1} lysozyme. The solution was centrifuged at 10 000*g* for 30 min and the supernatant was loaded onto an Ni^{2+}-chelating affinity column (GE Healthcare). The column was washed with five bed volumes of lysis buffer at a flow rate of 2 ml min^{-1}. The bound fractions were eluted using a linear gradient of 0.0–0.5 *M* imidazole in lysis buffer. The target protein was eluted with 300 m*M* imidazole. The fractions were then combined and concentrated to a final volume of 1 ml for subsequent molecular-exclusion chromatography on Superdex 75 (GE Healthcare), which had been pre-equilibrated with 25 m*M* Tris–HCl buffer pH 7.5. The purified GH43 was further analyzed by SDS–PAGE, which resulted in a single band. The sample was concentrated to 30 mg ml^{-1} by centrifugation using Centripep filters (Millipore). The purified GH43 is composed of 488 amino acids (molecular weight 55 410.4 Da) including the additional amino-acid sequence (FLEQKLISEEDLN-SAVDHHHHHH) at the C-terminus.

2.2. Dynamic light scattering

Dynamic light-scattering experiments were carried out using a DynaPro 810 (Protein Solutions) apparatus equipped with a Peltier module for temperature control. The wavelength of the laser light and the output power were set at 830 nm and 30 mW, respectively. About 50 measurements were made at intervals of 20 s for each run. The DLS experiments were repeated several times with intervals of 30 min to check stability. A protein solution of 1.0 mg ml^{-1} was prepared in 0.1 *M* HEPES pH 7.5. Standard curves of bovine serum albumin were used for calibration and the experiments were conducted at 291 K. Hydrodynamic parameters were determined using the software *DYNAMICS* v.6.3.40.

2.3. Crystallization

Crystallization was performed by the sitting-drop vapour-diffusion method using a Cartesian HoneyBee 963 system (Genomic Solutions). 544 conditions from commercially available crystallization kits were tested, including Crystal Screens I and II (Hampton Research) and JBScreen Classic and Basic kits (Jena Bioscience). Typically, 0.5 µl drops of protein solution (25 m*M* Tris–HCl buffer pH 7.5) at a concentration of 30 mg ml^{-1} were mixed with an equal volume of the screening solution and equilibrated over a reservoir containing 0.3 ml of the latter solution. Once initial crystallization conditions had been determined, they were optimized and large single crystals were obtained when a 1 µl protein droplet was mixed with an equal volume of reservoir solution consisting of 0.1 *M* MES buffer pH 6.5, 0.8 *M* ammonium sulfate, 0.1 *M* EDTA, 0.1 *M* L-proline and 5%(*v*/*v*) dioxane. The crystals grew to maximum dimensions of approximately 0.1 × 0.1 × 0.6 mm in 6 d.

2.4. X-ray diffraction data collection and analysis

The crystal was transferred to a cryoprotectant solution containing 20%(*v*/*v*) glycerol and flash-cooled at 100 K. Diffraction data were collected on the MX2 beamline at the Brazilian Synchrotron Light Laboratory (Campinas, Brazil). The wavelength of the radiation source was set to 1.48 Å and a total of 360 images with an oscillation range of 180° were collected using a MAR Mosaic 225 mm CCD detector (MAR Research). The data were indexed and scaled using the *DENZO* and *SCALEPACK* programs from the *HKL*-2000 package (Otwinowski & Minor, 1997). Molecular-replacement calculations were carried out using various programs, including *Phaser* (McCoy *et al.*, 2007), *AMoRe* (Navaza, 1994) and *MOLREP* (Vagin & Teplyakov, 1997). The atomic coordinates of several endo-1,5-α-L-arabinanases found in the Protein Data Bank were used as

```
T. petrophila   MRFL-FLMITLTALTGYILADEQPTFRWAVVHDPSIIKVGNMYYVFGT------HLQVAKSKDLMHWEQINTSAHDKNPIIPNINEELKETLSWA------RTRNDIWAPQVIQLSDGRY
G. stearotherm. VHFHPFGNVNFYEMD-WSLKG------DLWAHDPVIAKEGSRWYVFHTG----SGIQIKTSEDGVHWENMGWV-------F-------PSLPDWYKQYVPEKDEDHLWAPDI-CFYNGIY
B. subtilis     ------------EAAFWGAS-------NELLHDPTMIKEGSSWYALGTGLTEERGLRVLKSSDAKNWTVQKSI-------F-------TTPLSWWSNYVPNYG-QNQWAPDI-QYYNGKY
C. cellulosa    --------------------------QVDVHDPVMTREGDTWYLFSTG----PGITIYSSKDRVNWRYSDRA-------F-------ATEPTWAKRVSPSFD-GHLWAPDI-YQHKGLF

T. petrophila   YMYYCASTFGSPRSAIGIAVSDDIEGP-----YKHYAVIVKSGQVYSVDGPSEDGTPYDSRKHPNALDPGVFYDKEGNLWMVYGSWFGGIYILKLDPNTGLPLPG-QGYGKRLVG-----
G. stearotherm. YLYYSVSTFGKNTSVIGLATNQTLDPRDPDYEWKDMGPVIHST---------------AS-DNYNAIDPNVVFDQEGQPWLSFGSFWSGIQLIQLDTETMKPAAQAELLTIASRG-----
B. subtilis     WLYYSVSSFGSNTSAIGLASSTSISSG----GWKDEGLVIPST---------------SS-NNYNAIDPELTFDKDGNPWLAFGSFWSGIKLTKLDKSTMKPTG--SLYSIAARP-----
C. cellulosa    YLYYSVSAFGKNTSAIGVTVNKTLNPASPDYRWEDKGIVIESV---------------PQRDLWNAIDPAIIADDHGQVWMSFGSFWGGLKLFKLNDDLTRPAEPQEWHSIAKLERSVLM

T. petrophila   ----GNHSSMEGPYILYSPDTDYYYLFLSFGGLD--YRGGYNIRVARSKNPNGPYYDPEGKSMENCMGSKTVISNYGAKLVG--NFILSE---SNTIDFKAF------------------
G. stearotherm. ----EEPNAIEAPFIVCR--NGYYYLFVSFDFCCRGIESTYKIAVGRSKDITGPYVDKNGVSMMQGGGTILDEGND--RWIGPGHCAVYFSGVSAILVNHAYDALKNGEPTLQIRPLYWD
B. subtilis     ----NNGGALEAPTLTYQ--NGYYYLMVSFDKCCDGVNSTYKIAYGRSKSITGPYLDKSGKSMLEGGGTILDSGND--QWKGPGGQDIVN---GNILVRHAYDANDNGIPKLLINDLNWS
C. cellulosa    DDSQAGSAQIEAPFILRK--GDYYYLFASWGLCCRKGDSTYHLVVGRSKQVTGPYLDKTGRDMNQGGGSLLIKGNK--RWVGLGHNSAYTWDGKDYLVLHAYEAADNYLQKLKILNLHWD

T. petrophila   -------------GYVSPGHNSAYYDPETGKYFIFFHTRFPGRGETYQLRVHQLFLNEDGWFVMAPFPYGGETVSKLPNEEIVGEYQFINHGKEITDKIKQPVRIKLNSDGSITGAVEGR
G. stearotherm. DEGWPYL-------------SV----------------------------------------------------------------------------------------------
B. subtilis     -SGWP--------------SY----------------------------------------------------------------------------------------------
C. cellulosa    GEGWPQVDEKELDSYISQRLK----------------------------------------------------------------------------------------------

T. petrophila   WERKEHYITLKIIEGNTTVIYKGVLLKQWHYSEKKWVTVFTALSNQGVSVWGIRVEE
G. stearotherm. -------------------------------------------------------
B. subtilis     -------------------------------------------------------
C. cellulosa    -------------------------------------------------------
```

Figure 1
Multiple sequence alignment of endo-1,5-α-L-arabinanases from *T. petrophila*, *G. stearothermophilus* (PDB code 3cu9), *B. subtilis* (PDB code 1uv4) and *C. cellulosa* (PDB code 1gyd). The conserved amino-acid residues are highlighted in grey.

search models [PDB codes 1gyd (Nurizzo *et al.*, 2002), 1uv4 (Proctor *et al.*, 2005) and 3cu9 (Alhassid *et al.*, 2009)].

3. Results and discussion

Amino-acid sequence analysis of the GH43 gene from *T. petrophila* showed significant similarity to the endo-1,5-α-L-arabinanases from *Geobacillus stearothermophilus* (PDB code 3cu9; Alhassid *et al.*, 2009), *Bacillus subtilis* (PDB code 1uv4; Proctor *et al.*, 2005) and *Cellvibrio cellulosa* (PDB code 1gyd; Nurizzo *et al.*, 2002) (Fig. 1). The purified recombinant GH43 from *T. petrophila* was active towards sugar beet arabinan and debranched arabinan (Megazyme) using the DNS method (Miller, 1959). These analyses support the protein encoded by the GH43 gene from *T. petrophila* being an endo-1,5-α-L-arabinanase.

Structural homogeneity in solution was assessed by dynamic light-scattering experiments, which demonstrated a single peak (monomodal distribution) with approximate molecular weight 52 kDa corresponding to a monomer. The DLS results were reproducible after several experimental runs and the sample did not indicate aggregation or decay.

A single crystal with dimensions of 0.1 × 0.1 × 0.6 mm (Fig. 2) was used for the X-ray diffraction experiments and data were collected to 2.86 Å resolution (Fig. 3) under cryogenic conditions (100 K). The diffraction data were indexed in space group *P*422, with unit-cell parameters $a = b = 83.71$, $c = 408.25$ Å. An examination of the systematic absences indicated that the crystals belonged to either space group $P4_12_12$ or to its enantiomorph $P4_32_12$. Processing of the 890 735 measured reflections to 2.86 Å resolution led to 34 412 unique reflections with an $R_{\rm merge}$ of 11.6% (35.4% in the last shell; 2.96–2.86 Å resolution) and a completeness of 98.1% (91.2% in the last shell). Taking into consideration the molecular weight of 51 048 Da, three molecules are present in the asymmetric unit, with a corresponding Matthews coefficient of 2.35 Å^3 Da^{-1} and a solvent content of 48% (Matthews, 1968). Data-processing statistics are presented in Table 1. The atomic coordinates of several endo-1,5-α-L-arabinanases found in the Protein Data Bank (PDB codes 3cu9, 1wl7 and 1uv4), which display a sequence identity of around 32%, were used as search models with various molecular-replacement programs including *Phaser*, *AMoRe* and *MOLREP*. Different strategies were employed for molecular-replacement calculations such as the use of a polyalanine model and using only the catalytic domain. However, no clear solution was obtained and the phase problem will be solved using either multiple isomorphous replacement (MIR) or multi-wavelength anomalous diffraction (MAD). The full-length endo-1,5-α-L-arabinanase contains an extended C-terminal region of around 130 residues which is not found in other similar proteins of known three-dimensional structure and this feature, combined with the low identity with current templates in the PDB, may result in an unclear solution appearing during molecular-replacement calculations. In parallel with structural studies, comprehensive biochemical and functional analyses including substrate-specificity and hydrolysis-product patterns are being carried out.

Table 1
Data-collection statistics.

Values in parentheses are for the last resolution shell.

Data collection	
Temperature (K)	100
Radiation source	Brazilian Synchrotron Light Laboratory
Beamline	WO1B-MX2
Wavelength used (Å)	1.458
Detector	MAR Mosaic 225 mm
Space group	$P4_12_12$ or $P4_32_12$
Unit-cell parameters (Å, °)	$a = b = 83.71$, $c = 408.25$, $\alpha = \beta = \gamma = 90$
Resolution range (Å)	35.0–2.86 (2.96–2.86)
$R_{\rm merge}$† (%)	11.6 (35.4)
$\langle I/\sigma(I)\rangle$	16.2 (2.51)
Data completeness (%)	98.1 (91.2)
No. of unique reflections	34412
Data analysis	
$V_{\rm M}$ (Å^3 Da^{-1})	2.35
Solvent content (%)	48
Molecules per ASU	3

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

The crystallographic studies were supported by grants from CNPq (471192/2007-4) to MTM and FAPESP (08/58037-9) to FMS. The

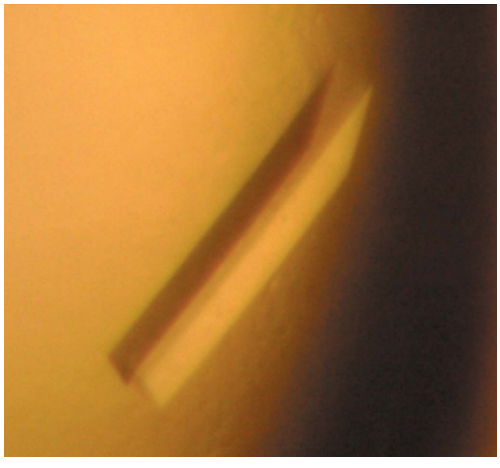


Figure 2
Microphotograph of a crystal of the endo-1,5-α-L-arabinanase from *T. petrophila*. The crystal dimensions were approximately 0.1 × 0.1 × 0.6 mm.

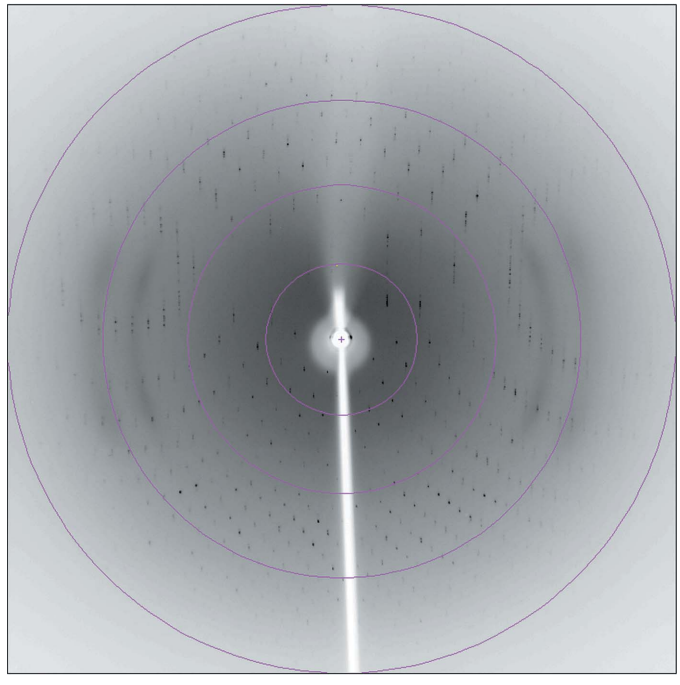


Figure 3
X-ray diffraction pattern of an endo-1,5-α-L-arabinanase crystal. The circles indicate resolution ranges of 11.8, 5.8, 3.8 and 2.8 Å.

cloning, expression, purification and biochemical studies were supported by grants awarded to RAP from the US Department of Energy and National Renewable Energy Laboratory: 06103-OKL and ZDJ-7-77608-01, respectively.

References

Alhassid, A., Ben-David, A., Tabachnikov, O., Libster, D., Naveh, E., Zolotnitsky, G., Shoham, Y. & Shoham, G. (2009). *Biochem. J.* **422**, 73–82.

Balat, M., Balat, H. & Öz, C. (2008). *Prog. Energy Combust. Sci.* **34**, 551–573.

Beldman, G., Schols, H. A., Piston, S. M., Searle-van Leeuwen, M. J. F. & Vorngen, A. G. J. (1997). *Adv. Macromol. Carbohydr. Res.* **1**, 1–64.

Hata, K., Tanaka, M., Tsumuraya, Y. & Hashimoto, Y. (1992). *Plant Physiol.* **100**, 388–396.

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.

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

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J. & Davies, G. J. (2002). *Nature Struct. Biol.* **9**, 665–668.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Proctor, M. R., Taylor, E. J., Nurizzo, D., Turkenburg, J. P., Lloyd, R. M., Vardakou, M., Davies, G. J. & Gilbert, H. J. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 2697–2702.

Saha, B. C. (2000). *Biotechnol. Adv.* **18**, 403–423.

Saha, B. C. (2003). *J. Ind. Microbiol. Biotechnol.* **30**, 279–291.

Shallom, D. & Shoham, Y. (2003). *Curr. Opin. Microbiol.* **6**, 219–228.

Takahata, Y., Nishijima, M., Hoaki, T. & Maruyama, T. (2001). *Int. J. Syst. Evol. Microbiol.* **51**, 1901–1909.

Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.

Van Laere, K. M., Hartemink, R., Bosveld, M., Schols, H. A. & Voragen, A. G. (2000). *J. Agric. Food Chem.* **48**, 1644–1652.

Wong, D., Chan, V. & Batt, S. (2008). *Appl. Environ. Microbiol.* **79**, 941–949.