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

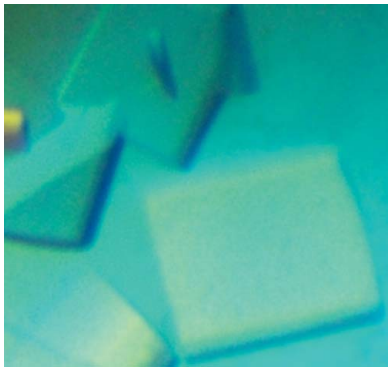
# Cloning, expression, purification, crystallization and preliminary X-ray diffraction studies of the catalytic domain of a hyperthermostable endo-1,4-β-D-mannanase from *Thermotoga petrophila* RKU-1

**Camila Ramos Santos,[a] Fabio Marcio Squina,[b] Andreia Meza Navarro,[a] Roberto Ruller,[b] Rolf Prade[c] and Mario Tyago Murakami[a]***

[a]Laboratório Nacional de Biociências (LNBio), Centro Nacional de Pesquisa em Energia e Materiais, 13083-970 Campinas-SP, Brazil, [b]Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE), Centro Nacional de Pesquisa em Energia e Materiais, 13083-970 Campinas-SP, Brazil, and [c]Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK 74078, USA

Correspondence e-mail: mtmurakami@lnls.br

Received 13 May 2010
Accepted 21 July 2010





Endo-1,4-β-D-mannanases play key roles in seed germination and fruit ripening and have recently received much attention owing to their potential applications in the food, detergent and kraft pulp industries. In order to delineate their structural determinants for specificity and stability, X-ray crystallographic investigations combined with detailed functional studies are being performed. In this work, crystals of the catalytic domain of a hyperthermostable endo-1,4-β-D-mannanase from *Thermotoga petrophila* RKU-1 were obtained from three different conditions, resulting in two crystalline forms. Crystals from conditions with phosphate or citrate salts as precipitant (CryP) belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 58.76$, $b = 87.99$, $c = 97.34$ Å, while a crystal from a condition with ethanol as precipitant (CryE) belonged to space group $I2_12_12_1$, with unit-cell parameters $a = 91.03$, $b = 89.97$, $c = 97.89$ Å. CryP and CryE diffracted to resolutions of 1.40 and 1.45 Å, respectively.

## 1. Introduction

Endo-1,4-β-D-mannanases (EC 3.2.1.78) catalyze the hydrolysis of β-1,4-D-mannopyranoside bonds in a variety of mannose-containing polysaccharides, including glucomannans and galactomannans (Dekker & Richards, 1976; Hazelwood & Gilbert, 1993). These enzymes have been studied widely owing to their roles in seed germination (Black, 1996) and fruit ripening (Pressey *et al.*, 1989); however, in the last decade they have become biotechnologically important owing to their ability to reduce complex polysaccharides into simple molecules such as manno-oligosaccharides and mannoses (Dhawan & Kaur, 2007). Their industrial applications include the biobleaching of softwood pulps (Tenkanen *et al.*, 1997; Montiel *et al.*, 2002), the extraction of oil from coconut meat (Dhawan & Kaur, 2007), reduction of the viscosity of coffee extract (Sachslehner *et al.*, 2000), improvement of the nutritional value of livestock feeds (Petty *et al.*, 2002) and use as an additive in the detergent industry (Sreekrishna *et al.*, 2000). For agro-industrial purposes, the use of enzymes with a high stability under a broad range of pH and temperature is highly desirable. In many cases, protein engineering is required to achieve these goals and the elucidation of the three-dimensional structure is an essential step in rational design (Tobin *et al.*, 2000). Archaea and bacteria isolated from harsh environments such as thermal hot springs and deep-sea hydrothermal vents are an excellent source of extremophilic enzymes (Burgess *et al.*, 2007). *Thermotoga petrophila* strain RKU-1 is a hyperthermophilic bacterium isolated from the production fluid of the Kubiki oil reservoir in Niigata, Japan which grows optimally at 353 K (Takahata *et al.*, 2001). This bacterium produces a repertoire of hyperthermostable enzymes of great industrial interest, including cellulases, arabinofuranosidases, arabinanases and mannanases, and has proved to be a suitable source of enzymes for biotechnological applications and protein engineering of glycoside hydrolases.

In order to understand the substrate specificity and thermal stability of β-mannanases, detailed knowledge of their structures at

the atomic level is crucial. Accordingly, we have cloned, overexpressed, purified and crystallized the catalytic domain of an endo-1,4-$\beta$-D-mannanase from the hyperthermophilic bacterium *T. petrophila* strain RKU-1.

## 2. Materials and methods

### 2.1. Cloning, protein expression and purification

The catalytic domain of the endo-1,4-$\beta$-D-mannanase (residues 32–394) from *T. petrophila* RKU-1 (TpMan$\Delta$CT; GenBank accession code YP_001245126) was amplified from genomic DNA of *T. petrophila* by a standard PCR method using two oligonucleotide primers (forward, 5′-**CATATG**CTTAATGGAAAAGAATTCAGATTC-3′; reverse, 5′-**AAGCTT**TCAGAACAGCTTCGCGTATTCTC-3′). The amplified TpMan$\Delta$CT gene was initially cloned into pGEM-T vector and further subcloned into the *Nde*I and *Hin*dIII restriction-enzyme sites of the pET-28a vector with a hexahistidine tag at the N-terminus. The plasmid was transformed into *Escherichia coli* BL21 (DE3) $\Delta$SlyD cells and plated in selective solid LB medium. One colony was picked up and grown in liquid LB/antibiotics for 16 h at 310 K and 200 rev min$^{-1}$. The culture was diluted at 2%(*v*/*v*) into fresh LB/antibiotics and grown under the same conditions to an $OD_{600\,nm}$ of 0.8. Overexpression was induced with 0.5 m*M* IPTG at 310 K for 4 h. The harvested cells were resuspended in lysis buffer (20 m*M* sodium phosphate pH 7.4, 50 m*M* NaCl, 5 m*M* imidazole, 1 m*M* PMSF) and then sonicated with six 30 s pulses at 500 W using a VC750 Ultrasonic Processor (Sonics Vibracell). The solution was centrifuged at 10 000*g* for 30 min and the supernatant was loaded onto a nickel-affinity column (GE Healthcare). The column was washed with five bed volumes of lysis buffer at a flow rate of 1 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 150 m*M* imidazole. The fractions were then combined and concentrated to a final volume of 1 ml for subsequent size-exclusion chromatography (SEC) on Superdex 75 (GE Healthcare) which had been pre-equilibrated with 20 m*M* sodium phosphate buffer pH 7.4 containing 200 m*M* NaCl. The sample purity was confirmed by polyacrylamide gel electrophoresis under denaturing conditions (Laemmli, 1970). The sample was dialyzed against 25 m*M* Tris–HCl buffer pH 7.5 and concentrated to 12 mg ml$^{-1}$ using Amicon centrifugal filter units (Millipore). The protein concentration was determined by absorption spectroscopy at 280 nm using a molar extinction coefficient of 110 465 $M^{-1}$ cm$^{-1}$, assuming that all of the cysteine residues form disulfide bonds.

### 2.2. Dynamic light scattering

Dynamic light-scattering (DLS) experiments were carried out using a DynaPro MS/X (Wyatt Technology Corporation) device. The wavelength of the laser light and the output power were set to 830 nm and 30 mW, respectively. Around 20 measurements were made at 15 s intervals for each run. DLS experiments were repeated twice with interval of 1 h to check the sample stability. DLS measurements were carried out at low (1.0 mg ml$^{-1}$) and high (10.0 mg ml$^{-1}$) protein concentration to verify aggregate formation. The experiments were conducted at 291 K. The hydrodynamic parameters were determined using the software *DYNAMICS* v.6.3.40.

### 2.3. Crystallization

Crystallization experiments were performed by the sitting-drop vapour-diffusion method using a Cartesian HoneyBee 963 system (Genomic Solutions) at 291 K. 544 different formulations based on commercial crystallization kits were tested, including those from Hampton Research (SaltRX, Crystal Screen and Crystal Screen 2), Emerald BioSystems (Precipitant Synergy and Wizard I and II) and Qiagen/Nextal (PACT and JCSG+). For initial screening, 0.5 µl protein solution at a concentration of 12 mg ml$^{-1}$ (in 25 m*M* Tris–HCl buffer pH 7.5) was mixed with an equal volume of screening solution and equilibrated over a reservoir containing 80 µl of the latter solution. Small crystals grew from three different conditions: (i) 0.1 *M* citrate pH 5.5, 1 *M* ammonium phosphate and 0.2 *M* sodium chloride; (ii) 0.1 *M* Tris–HCl pH 8.5 and 0.7 *M* sodium citrate; and (iii) 0.1 *M* phosphate pH 4.2, 5%(*w*/*v*) PEG 1000 and 40%(*w*/*v*) ethanol. These conditions were optimized and large crystals were achieved by using 10%(*v*/*v*) glycerol as an additive and decreasing the precipitant concentrations in small steps. The best crystals were obtained by mixing 0.5 µl protein solution with an equal volume of reservoir solution consisting of (i) 0.1 *M* citrate pH 5.5, 0.85 *M* ammonium phosphate, 0.2 *M* sodium chloride and 10%(*v*/*v*) glycerol, (ii) 0.1 *M* Tris–HCl pH 8.5, 0.6 *M* sodium citrate and 10%(*v*/*v*) glycerol or (iii) 0.1 *M* phosphate pH 4.2, 5%(*w*/*v*) PEG 1000, 36%(*w*/*v*) ethanol and 10%(*v*/*v*) glycerol. Crystals from the conditions with salt as a precipitant grew within 3 d, whereas those from the condition with organic solvent (ethanol) as a precipitant grew within 5 d.

### 2.4. X-ray diffraction data collection and analysis

Crystals were directly flash-cooled in a nitrogen-gas stream at 100 K for data collection without any additional cryoprotectant since the mother solution already contained 10%(*v*/*v*) glycerol. Diffraction data were collected on the MX2 beamline at the Brazilian Synchrotron Light Laboratory (Campinas, Brazil) with the wavelength of the radiation set to 1.458 Å. For both crystalline forms a total of 360 images were collected using an oscillation angle of 1° and an exposure time of 30 s per image. The crystal-to-detector distance was set to 50 mm, which resulted in a maximum resolution of 1.35 Å; however, depending on the crystalline system the data resolution was cut off to give adequate statistics. A MAR Mosaic 225 mm (MAR Research) charge-coupled device (CCD) detector was used to record the intensities. The data were indexed and scaled using the *DENZO* and *SCALEPACK* programs from the *HKL*-2000 package (Otwinowski & Minor, 1997). Molecular-replacement calculations were performed using the program *Phaser* (McCoy *et al.*, 2007).

## 3. Results and discussion

The endo-1,4-$\beta$-D-mannanase from *T. petrophila* consists of 667 amino-acid residues encompassing a catalytic domain belonging to glycoside hydrolase family 5 (GH5) joined to a carbohydrate-binding module (CBM family 27) by a linker unit (approximately 100 amino-acid residues). This domain composition has been well established for a number of bacterial and fungal carbohydrases (Gilkes *et al.*, 1991); however, ancillary modules can be found at both the N- and the C-termini. The majority of studies on CBMs to date have demonstrated their role in targeting the enzyme to the appropriate substrate or regions of substrate (Bolam *et al.*, 1998; Boraston *et al.*, 2002).

Crystallization of the full-length protein proved to be unsuccessful, probably owing to the inherent flexibility between domains caused by the long linker (results not shown). Thus, further studies were carried out with a truncated construct consisting of the first 362 amino-acid residues, which correspond to the catalytic core (TpMan$\Delta$CT) of endo-1,4-$\beta$-D-mannanase from *T. petrophila*.

### 3.1. Overexpression and purification

TpManΔCT was overexpressed in BL21 (DE3) ΔSlyD cells in the soluble fraction at 310 K and was purified using two chromatographic steps, nickel-affinity chromatography followed by SEC, yielding approximately 10 mg of high-purity protein per litre of expression culture. The protein eluted as a single peak from SEC; however, each fraction was separately analyzed by DLS. Only fractions with a low polydispersity index (<30%) were pooled, dialyzed and concentrated for crystallization trials. Both SEC and DLS experiments indicated that TpManΔCT behaves as a monomer in solution, with a hydrodynamic radius of 2.5 + 0.1 nm and an averaged polydispersity of 15%. The activity was tested using mannan as substrate, confirming the functional state of the recombinant TpManΔCT. On SDS–PAGE, the purified TpManΔCT migrated as a single band with a corresponding molecular weight of 44 kDa.

### 3.2. Crystallization, data collection and processing

The crystals obtained from the conditions with phosphate or citrate salts as precipitant belonged to the same crystalline system with similar unit-cell parameters. Thus, data collection was carried out from a crystal with approximate dimensions of 0.2 × 0.2 × 0.5 mm obtained from a condition consisting of 0.1 *M* citrate pH 5.5, 0.85 *M* ammonium phosphate, 0.2 *M* sodium chloride and 10%(*v*/*v*) glycerol (CryP; Fig. 1*a*). CryP diffracted to a resolution of 1.40 Å and was indexed in the orthorhombic crystal system, with unit-cell parameters $a$ = 58.76, $b$ = 87.99, $c$ = 97.34 Å. An examination of the systematic absences indicated that the crystals belonged to space group $P2_12_12_1$. In contrast, the crystals from a condition containing 0.1 *M* phosphate pH 4.2, 5%(*w*/*v*) PEG 1000, 36%(*w*/*v*) ethanol and 10%(*v*/*v*) glycerol (CryE; Fig. 1*b*) diffracted to a resolution of 1.45 Å and were indexed in the centred orthorhombic crystal system, with unit-cell parameters $a$ = 91.03, $b$ = 89.97, $c$ = 97.89 Å. In this case, the space-group ambiguity between $I222$ and $I2_12_12_1$ could not be solved by scaling and merging of intensities and only the correlation function and *R*-factor values after the translational search in molecular-replacement calculations will indicate which space group this crystalline form belongs to. Taking into consideration the molecular weight of 43 994 Da, one molecule is present in the asymmetric unit in both crystalline forms. CryP had a solvent content of 58% with a corresponding Matthews coefficient of 2.94 Å$^3$ Da$^{-1}$, whereas CryE had a solvent content of 47% with a corresponding Matthews coefficient of 2.34 Å$^3$ Da$^{-1}$ (Matthews, 1968). Data-processing statistics for both data sets are presented in Table 1.

**Table 1**
Data-collection statistics.

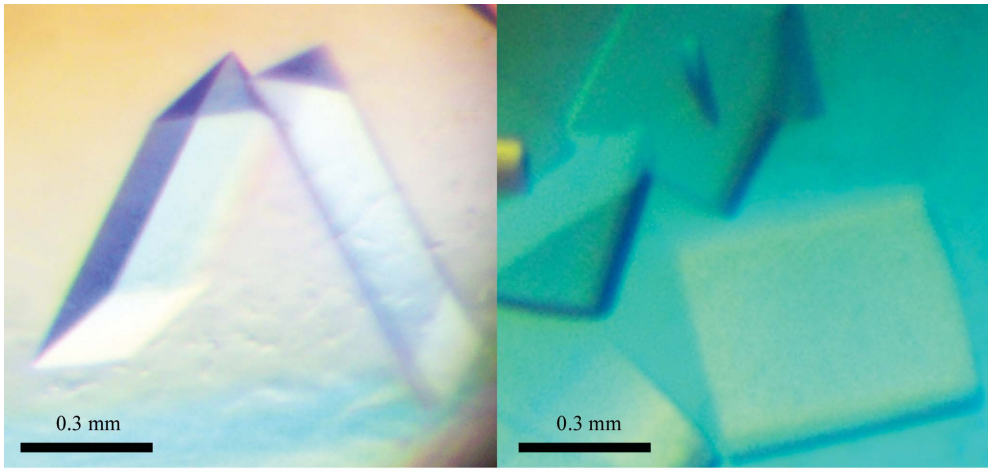
Values in parentheses are for the last resolution shell.

| | CryP | CryE |
|---|---|---|
| Data collection | | |
| Temperature (K) | 100 | 100 |
| Radiation source | Brazilian Synchrotron Light Laboratory | Brazilian Synchrotron Light Laboratory |
| Beamline | W01B-MX2 | W01B-MX2 |
| Wavelength used (Å) | 1.458 | 1.458 |
| Detector | MAR Mosaic 225 mm | MAR Mosaic 225 mm |
| Space group | $P2_12_12_1$ | $I2_12_12_1$ |
| Unit-cell parameters (Å, °) | $a$ = 58.76, $b$ = 87.99, $c$ = 97.34, $\alpha = \beta = \gamma$ = 90 | $a$ = 91.03, $b$ = 89.97, $c$ = 97.89, $\alpha = \beta = \gamma$ = 90 |
| Resolution range (Å) | 30.0–1.40 (1.45–1.40) | 35.0–1.45 (1.50–1.45) |
| $R_{\rm merge}$† (%) | 6.5 (32.1) | 5.1 (33.7) |
| $\langle I/\sigma(I)\rangle$ | 35.6 (3.9) | 40.3 (4.9) |
| Data completeness (%) | 99.5 (99.2) | 96.5 (91.7) |
| No. of measured reflections | 1282371 | 921691 |
| No. of unique reflections | 103955 | 76427 |
| Multiplicity | 12.4 (10.3) | 12.1 (9.5) |
| Data analysis | | |
| $V_{\rm M}$ (Å$^3$ Da$^{-1}$) | 2.94 | 2.34 |
| Solvent content (%) | 58 | 47 |
| Molecules per asymmetric unit | 1 | 1 |

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

### 3.3. Molecular replacement

TpManΔCT displayed low sequence identity (<37%) to other β-mannanases available in the Protein Data Bank (PDB; http://www.pdb.org) and a *PSI-BLAST* search resulted in only two hits, the catalytic domains of β-mannanases from *Trichoderma reesei* (Sabini *et al.*, 2000; PDB code 1qno) and *Lycopersicon esculentum* (Bourgault *et al.*, 2005; PDB code 1rh9), with 36 and 32% sequence identity, respectively. Thus, different strategies were used in the molecular-replacement calculations for initial model building. The best results were obtained using a polyalanine model of the β-mannanase structure from *T. reesei* (PDB code 1qno), with an *R* factor and $R_{\rm free}$




(*a*) (*b*)

**Figure 1**
Microphotographs of TpManΔCT crystals obtained from (*a*) 0.1 *M* citrate pH 5.5, 0.85 *M* ammonium phosphate, 0.2 *M* sodium chloride and 10%(*v*/*v*) glycerol (CryP) and (*b*) 0.1 *M* phosphate pH 4.2, 5%(*w*/*v*) PEG 1000, 36%(*w*/*v*) ethanol and 10%(*v*/*v*) glycerol (CryE). The approximate dimensions of CryP and CryE were 0.5 × 0.2 × 0.2 and 0.5 × 0.4 × 0.2 mm, respectively.

of 45% and 56%, respectively. A translational search solved the space-group ambiguity present in CryE, indicating that it belonged to space group $I2_12_12_1$. Visual inspection of the initial model showed that the 80-residue-long C-terminal segment was not properly fitted in the electron density and extensive manual building was necessary to trace the complete chain. In parallel with structural studies, site-directed mutagenesis and functional investigations are being carried out in order to shed light on the specificity and stability of hyperthermostable bacterial $\beta$-mannanases.

This research was supported by grants from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the National Renewable Energy Laboratory and the US Department of Energy.

## References

Black, M. (1996). *Seed Sci. Res.* **6**, 39–42.
Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. & Gilbert, H. J. (1998). *Biochem. J.* **331**, 775–781.
Boraston, A. B., Kwan, E., Chiu, P., Warren, R. A. & Kilburn, D. G. (2002). *J. Biol. Chem.* **278**, 6120–6127.
Bourgault, R., Oakley, A. J., Bewley, J. D. & Wilce, M. C. (2005). *Protein Sci.* **14**, 1233–1241.
Burgess, E. A., Wagner, I. D. & Wiegel, J. (2007). *Physiology and Biochemistry of Extremophiles*, edited by C. Gerday & N. Glansdorff, pp. 13–29. Washington: ASM Press.
Dekker, R. F. & Richards, G. N. (1976). *Adv. Carbohydr. Chem. Biochem.* **32**, 277–352.
Dhawan, S. & Kaur, J. (2007). *Crit. Rev. Biotechnol.* **27**, 197–216.
Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. Jr & Warren, R. A. (1991). *Microbiol. Rev.* **55**, 303–315.
Hazelwood, G. P. & Gilbert, H. J. (1993). *Hemicellulose and Hemicellulases*, edited by M. P. Coughlan & G. P. Hazelwood, pp. 103–126. London University Press.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
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.
Montiel, M. D., Hernandez, M., Rodriguez, J. & Arias, M. E. (2002). *Appl. Microbiol. Biotechnol.* **58**, 67–72.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Petty, L. A., Carter, S. D., Senne, B. W. & Shriver, J. A. (2002). *J. Anim. Sci.* **80**, 1012–1019.
Pressey, R. (1989). *Phytochemistry*, **28**, 3277–3280.
Sabini, E., Schubert, H., Murshudov, G., Wilson, K. S., Siika-Aho, M. & Penttilä, M. (2000). *Acta Cryst.* D**56**, 3–13.
Sachslehner, A., Foidl, G., Foidl, N., Gübitz, G. & Haltrich, D. (2000). *J. Biotechnol.* **80**, 127–134.
Sreekrishna, K., Johnstone, K., Saunders, C. & Bettiol, J. (2000). US Patent 6060299.
Takahata, Y., Nishijima, M., Hoaki, T. & Maruyama, T. (2001). *Int. J. Syst. Evol. Microbiol.* **51**, 1901–1909.
Tenkanen, M., Makkonen, M., Perttula, M., Viikari, L. & Teleman, A. (1997). *J. Biotechnol.* **57**, 191–204.
Tobin, M. B., Gustafsson, C. & Huisman, G. W. (2000). *Curr. Opin. Struct. Biol.* **10**, 421–427.