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

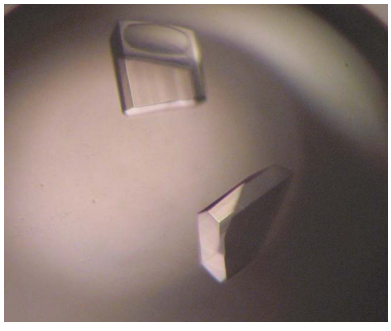
Crystallization and preliminary X-ray analysis of a hyperthermophilic endoglucanase from *Pyrococcus furiosus*

Misumi Kataoka, Han-Woo Kim and Kazuhiko Ishikawa*

Biomass Technology Research Center, National Institute of Advanced Industrial Science and Technology (AIST), 3-11-32 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-0046, Japan

Correspondence e-mail: kazu-ishikawa@aist.go.jp

Received 16 December 2011
Accepted 27 January 2012



The hyperthermophilic glycoside hydrolase family 12 endocellulase from the archaeon *Pyrococcus furiosus* (EGPf) catalyzes the hydrolytic cleavage of the β-1,4-glucosidic linkage in β-glucans in biomass. EGPf (Gene ID PF0854; EC 3.2.1.4) contains a signal sequence and proline- and hydroxyl-rich regions at the N-terminus. Truncated EGPf (EGPfΔN30) without the proline- and hydroxyl-rich regions at the N-terminus was prepared and subjected to crystallization experiments. Crystals were obtained using the hanging-drop vapour-diffusion method at 303 K. An X-ray diffraction data set was collected to 1.07 Å resolution at 100 K. The crystal belonged to space group $P2_12_12$, with unit-cell parameters a = 58.01, b = 118.67, c = 46.76 Å. The presence of one molecule of enzyme per asymmetric unit gives a crystal volume per protein mass (V_M) of 2.63 Å^3 Da^{-1} and a solvent content of 53.3%(v/v).

1. Introduction

Cellulase is one of the most important industrial enzymes in terms of biomass utilization, since the enzyme plays a key role in degradation of the β-glucan cellulose. Recent research on biofuel production from lignocellulose biomass has aimed at the development of an ideal cellulase for efficient biomass saccharification. A hyperthermophilic cellulase would be very useful in industrial applications because enzymatic reaction processes at high temperature have many merits such as a reduced risk of microbial contamination, increased solubility of substrates and improved transfer rates. Therefore, much research has focused on the development of a thermophilic cellulase with high activity.

Hyperthermophilic β-1,4-endocellulases (endo-type cellulases) from several hyperthermophilic archaea have been found in the genome database. The hyperthermophilic archaea *Pyrococcus horikoshii* and *P. furiosus* possess GH family 5 and 12 endocellulases, respectively. These enzymes showed differing substrate specificities. The first crystal structure of a hyperthermophilic endocellulase from *P. horikoshii* (EGPh; GH family 5) has been determined (Kim & Ishikawa, 2010) and its substrate-recognition mechanism has also been reported (Kim & Ishikawa, 2011). The characterization of a hyperthermophilic endocellulase from *P. furiosus* (EGPf; GH family 12) has also been reported (Bauer *et al.*, 1999). However, there is no information to date on the crystal structure of EGPf. Since Archaea are primitive microorganisms, the structures of hyperthermophilic endocellulases from Archaea contain some of the fundamental properties of this class of enzymes. We successfully prepared the first crystal of EGPf using a truncated mutant. Here, we report the preparation of an EGPf crystal suitable for X-ray analysis at atomic resolution.

2. Materials and methods

2.1. Protein preparation

We prepared recombinant EGPf using a similar method to that described previously (Kashima *et al.*, 2005). A plasmid constructed using the vector pET11a (Novagen, Madison, Wisconsin, USA) was introduced into *Escherichia coli* strain BL21 (DE3) pLysS. A gene for a truncated enzyme (EGPfΔN30) with 30 amino-acid residues (the signal sequence and the proline- and hydroxyl-rich regions) deleted

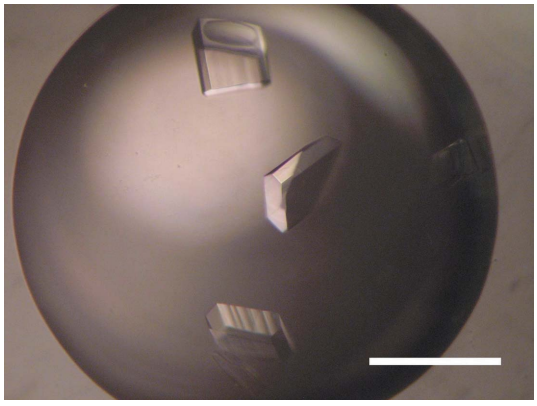


Figure 1
A photograph of the EGPfΔN30 crystals. The scale bar corresponds to 0.5 mm.

from the N-terminal region was constructed by the PCR method. The truncated gene was inserted into the expression vector pET11a. The constructed plasmids were introduced into *E. coli* strain BL21 (DE3) for recombinant protein expression. Expression and purification of the recombinant enzymes was carried out using the method reported previously (Kashima *et al.*, 2005). The purity and molecular weight of the protein sample were analyzed by SDS–PAGE. The protein concentration of EGPf was determined from the UV absorbance at 280 nm, using a molar extinction coefficient of 81 790 calculated from the protein sequence (Gill & von Hippel, 1989).

2.2. Crystallization

The purified protein was dialyzed against 50 m*M* Tris–HCl buffer pH 8.0 and concentrated to 15 mg ml^{-1} using an Amicon Centricon YM-10 (Millipore, Billerica, Massachusetts, USA). EGPfΔN30 crystals were grown at 303 K by the hanging-drop vapour-diffusion method using a reservoir solution composed of 120 m*M* CHES [2-(*N*-cyclohexylamino)ethanesulfonic acid] buffer pH 9.0, 50 m*M* lithium sulfate, 0.5 *M* potassium sodium tartrate. Typically, drops consisting of 1 µl protein solution and 1 µl reservoir solution were equilibrated against 0.4 ml reservoir solution.

2.3. Data collection and processing

Selected crystals were harvested and immersed in cryoprotectant solution consisting of 30%(*v*/*v*) glycerol in mother liquor. A soaked crystal was collected in a CryoLoop (Hampton Research, Aliso Viejo, California, USA) and immediately flash-cooled in a stream of nitrogen gas at 100 K. X-ray diffraction data were collected from a single crystal using a Quantum 315 detector (ADSC) on the BL44XU beamline at SPring-8 (Hyogo, Japan). The data were collected in two passes, at low (5.00 Å) and high (1.07 Å) resolution, to give a complete data set. The data set at low resolution was collected at a wavelength of 0.9 Å with a crystal-to-detector distance of 250 mm. The high-resolution data set was collected at a wavelength of 0.8 Å with a crystal-to-detector distance distance of 90 mm. In each pass, the crystal was rotated 360° with an oscillation angle of 1° per frame. The data collected were merged, indexed, integrated and scaled with programs from the *HKL*-2000 software package (Otwinowski & Minor, 1997).

3. Results

Using the recombinant enzyme (EGPf) with the signal sequence deleted, no crystals were obtained using the Crystal Screen (Hampton Research, Aliso Viejo, California, USA) and Wizard I and II (Emerald BioSystems, Bainbridge Island, Washington, USA) crystallization screening kits. However, recombinant truncated EGPf (EGPfΔN30) without the 30 amino-acid residues (the signal sequence and the proline- and hydroxyl-rich regions) at the N-terminal region could be crystallized. Based on the initial screening results, high-quality crystals of EGPfΔN30 were obtained using a reservoir solution consisting of 120 m*M* CHES buffer pH 9.0 containing 50 m*M* lithium sulfate and 0.5 *M* potassium sodium tartrate at 303 K. The average dimensions of the crystal were about 0.7 × 0.4 × 0.4 mm after 2 d (Fig. 1). Diffraction data were collected to a resolution limit of 1.07 Å. The data set consisted of 1 187 674 measurements and 130 828 unique reflections. The results of data collection and analysis revealed that the crystal belonged to space group $P2_12_12$, with unit-cell parameters a = 58.01, b = 118.67, c = 46.76 Å. Scaling and merging of the crystallographic data resulted in an overall $R_{\rm merge}$ of 8.5%. The data set has a completeness of 87.9%. The data-collection statistics are summarized in Table 1. The presence of one molecule of the enzyme in the asymmetric unit gives a crystal volume per protein mass ($V_{\rm M}$) of 2.63 Å^3 Da^{-1} and a solvent content of 53.3%(*v*/*v*) (Matthews, 1968). Determination of the structure of EGPfΔN30 is in progress using the molecular-replacement method with the structural data for the GH family 12 endocellulase as a model.

Table 1
Crystallographic parameters and data-collection statistics.

Data set	High resolution	Low resolution	Merged data
Wavelength (Å)	0.8	0.9	
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	a = 58.01, b = 118.67, c = 46.76	a = 58.01, b = 118.67, c = 46.76	a = 58.01, b = 118.67, c = 46.76
Matthews coefficient (Å^3 Da^{-1})	2.63	2.63	2.63
Solvent content (%)	53.3	53.3	53.3
Resolution range (Å)	2.50–1.07 (1.09–1.07)	50–2.13 (2.17–2.13)	50–1.07 (1.09–1.07)
Total No. of observed reflections	861280 (21071)	258692 (11916)	1187674 (28490)
No. of unique reflections	117385 (5017)	18591 (896)	130828 (5698)
Average $I/\sigma(I)$	25.9 (3.8)	96.6 (79.1)	18.4 (3.9)
$R_{\rm merge}$† (%)	9.6 (31.7)	4.2 (5.8)	8.5 (32.8)
Completeness‡ (%)	90.2 (77.6)	99.3 (99.7)	87.9 (75.9)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I(hkl)\rangle$ is the mean value of $I_i(hkl)$ for all i measurements. ‡ The low values for the completeness are because of the blind region (Drenth, 1999) arising from the accidental coincidence of the crystal c axis and the direction of the oscillation axis.

X-ray diffraction experiments were carried out with the approval of the Japan Synchrotron Radiation Research Institute (Hyogo, Japan).

References

Bauer, M. W., Driskill, L. E., Callen, W., Snead, M. A., Mathur, E. J. & Kelly, R. M. (1999). *J. Bacteriol.* **181**, 284–290.
Drenth, J. (1999). *Principles of Protein X-ray Crystallography*, 2nd ed., pp. 41–43. New York: Springer.
Gill, S. C. & von Hippel, P. H. (1989). *Anal. Biochem.* **182**, 319–326.
Kashima, Y., Mori, K., Fukada, H. & Ishikawa, K. (2005). *Extremophiles*, **9**, 37–43.
Kim, H.-W. & Ishikawa, K. (2010). *Proteins*, **78**, 496–500.
Kim, H.-W. & Ishikawa, K. (2011). *Biochem. J.* **437**, 223–230.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.