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

Crystallization and preliminary X-ray analysis of endoglucanase from *Pyrococcus horikoshii*

Han-Woo Kim, Koshiki Mino‡ and Kazuhiko Ishikawa*

National Institute of Advanced Industrial Science and Technology (AIST), 1-18-31, Midorigaoka, Ikeda, Osaka 563-8577, Japan

‡ Present address: Nagahama Institute of Bio-Science and Technology, 1266 Tamura, Nagahama-shi, Shiga, 526-0829, Japan.

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

Received 8 September 2008
Accepted 10 November 2008

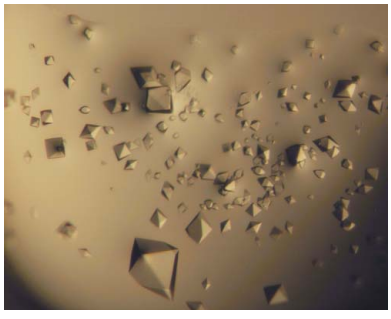
Structural information on hyperthermostable cellulases is required for bioprocess applications in the transformation of biomass. Crystals were obtained of a C-terminally five-amino-acid truncated hyperthermostable endoglucanase (family 5) from the archaeon *Pyrococcus horikoshii*. The truncated form of this enzyme showed similar enzymatic properties to the wild-type protein. The enzyme was crystallized by the sitting-drop vapour-diffusion method using ethanol as precipitant at 296 K. An X-ray diffraction data set was collected to 1.78 Å resolution at 100 K. The crystals belonged to space group $P4_12_12$ or $P4_32_12$.

1. Introduction

Cellulose is the most abundant biomass on earth and is very important as a potential renewable carbon resource. β-1,4-Glucanases (also known as endocellulases) hydrolyze the β-1,4 glycosidic bond within cellulose. It is a very potent enzyme for industrial bioprocesses, including the bio-polishing of cotton products, food processing and bioethanol production. The hyperthermostable properties of this cellulase are required for its efficient application in bioprocessing.

Recently, we reported the characterization of the hyperthermostable β-1,4-endoglucanase (EGPh; family 5) from the hyperthermophilic archaeon *Pyrococcus horikoshii*, which unlike *P. furiosus* endoglucanase can hydrolyze crystalline cellulose (Ando *et al.*, 2002; Bauer *et al.*, 1999). Amino-acid sequence analysis revealed that EGPh had 43% homology with endoglucanase from *Acidothermus cellulolyticus* (EGAc), the crystal structure of the catalytic domain of which has previously been reported (Fig. 1; Sakon *et al.*, 1996; PDB code 1ece). EGPh also possesses several conserved residues near the active site that are typical of family 5 endoglucanases (Ando *et al.*, 2002). A mutation experiment revealed that the conserved residues, with the exception of Asp385, play critical roles in the activity of EGPh, suggesting that EGPh has a similar structure and function to EGAc (Sakon *et al.*, 1996). However, the structure of the substrate-binding region and the hyperthermostability of EGPh were expected to differ from those of EGAc. In addition, no crystal structure of a family 5 endoglucanase from an archaeon has been determined to date.

Here, we report the preparation of the hyperthermostable wild-type archaeal EGPh for crystallization and preliminary crystallographic analysis.



2. Materials and methods

2.1. Protein preparation

An extensive crystallization search was conducted using recombinant wild-type EGPh (Gene ID PH1171; Kashima *et al.*, 2005) with Crystal Screen kits (Hampton Research, Laguna Niguel, California, USA) and Wizard I and II crystallization screening kits (deCode Genetics, Bainbridge Island, Washington, USA). Only needle-shaped crystals were obtained using a condition containing 0.1 *M* HEPES pH 7.5, 10%(*w*/*v*) polyethylene glycol 6000 and 5%(*v*/*v*) MPD at 296 K. However, the crystals were not suitable for X-ray diffraction analysis. Therefore, as an alternative, we aimed to produce a structurally compact protein that may yield crystals of this enzyme. To remove the

flexible part of the protein, EGPh was digested with bovine pancreatic α-chymotrypsin (Wako, Japan) for 12 h at 310 K (Kang & Ishikawa, 2007; Uegaki *et al.*, 1993). The results of proteolysis were analyzed by SDS–PAGE and the electrophoretic band was cut out for N-terminal amino-acid sequencing, which was performed using an HP G1005A protein-sequencing system (APRO, Naruto, Tokushima, Japan). Based on the results of SDS–PAGE analysis and N-terminal amino-acid sequencing, we constructed three types of truncated EGPh mutant enzyme. The resulting mutants were named EGPhΔN5, EGPhΔC5 and EGPhΔN5C5 and lacked five amino-acid residues from the N- and/or C-terminal ends of the wild type. All of the mutants maintained enzymatic activity without losing any activity or thermostability (Kang & Ishikawa, 2007). As only the EGPhΔC5 construct led to successful crystallization results, further details are limited to this protein.

EGPhΔC5 has a theoretical molecular weight of 43.97 kDa and consists of 384 amino-acid residues. The EGPhΔC5 gene was constructed by the PCR method using the wild-type gene of EGPh (PH1171; Kashima *et al.*, 2005) as a template. The resulting gene was inserted into the *Nde*I and *Bam*HI sites of the expression vector pET11a (Novagen, Madison, Wisconsin, USA). The constructed plasmid was introduced into *Escherichia coli* strain BL21(DE3)-pLysS. The transformed cells were cultured in LB broth containing 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol at 310 K and induction was carried out with 0.01 m*M* isopropyl β-D-1-thio-galactopyranoside (IPTG) at 310 K. Purification of the recombinant enzyme was carried out by the method reported previously (Kashima *et al.*, 2005). The purified protein was dialyzed against MilliQ water and concentrated to 10 mg ml^{-1} using Centricon YM-10 (Millipore, Bedford, Massachusetts, USA). The protein concentration was determined from the UV absorbance at 280 nm, using 136 270 M^{-1} cm^{-1} as the molar extinction coefficient.

2.2. Crystallization

Initial screening for crystallization conditions was performed using Wizard I and II crystallization screening kits (deCode Genetics, Bainbridge Island, Washington, USA) with the sitting-drop vapour-diffusion method at 296 K. All drops consisted of 1 μl protein solution at 10 mg ml^{-1} (in 50 m*M* Tris buffer pH 8.0) and 1 μl reservoir solution and were equilibrated against 0.1 ml reservoir solution at a temperature of 296 K.

2.3. Data collection and processing

The selected crystal was immersed in the optimum crystallization condition containing 20%(*v*/*v*) glycerol solution as a cryoprotectant and then flash-cooled in a stream of nitrogen gas cooled to 100 K. The X-ray diffraction data were collected using a Jupiter 210cs CCD detector at a wavelength of 1.0 Å at the BL38B1 experimental station of SPring-8, Japan Synchrotron Radiation Research Institute (JASRI), Hyogo, Japan. The crystal-to-detector distance was 150 mm. The crystal was rotated through 90° with an oscillation angle of 0.5° per frame and 10 s exposures. The collected diffraction data were indexed, integrated and scaled with the *HKL* programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

```
EGPh            ENTTYQTPTGIYYEVRGDTIYMINVTSGEETPIHLFGVNWFGFETPNHVVHGLWKRNWED
EGAc            -------AGGGYWHTSGREILDANN-----VPVRIAGINWFGFETCNYVVHGLWSRDYRS
                       . * *:.. *  *   *      .*::: *:******* *:******.*::..

EGPh            MLLQIKSLGFNAIRLPFCTESVKPGTQPIGIDYSKN-PDLRGLDSLQIMEKIIKKAGDLG
EGAc            MLDQIKSLGYNTIRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIG
                ** ******:*:****:. : :**** * .*:: :   **:** ***:*:**:  **::*
                             #

EGPh            IFVLLDYHRIGCTHIEPLWYTEDFSEEDFINTWIEVAKRFGKYWNVIGADLKNEPHSVTS
EGAc            LRIILDRHRPDCSGQSALWYTSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP--
                : ::** ** .*:  ..****...**  :*.    :*:*:    .*:* **:****.
                       #                                      ##

EGPh            PPAAYTDGTGATWGMGNPATDWNLAAERIGKAILKVAPHWLIFVEGTQFTNPKTDSSYKW
EGAc            ----------ACWGCGDPSIDWRLAAERAGNAVLSVNPNLLIFVEGVQSYN---------
                          * ** *:*: **.***** *:*:*.* *: ******.*  *

EGPh            GYNAWWGGNLMAVKDYPVNLPR-NKLVYSPHVYGPDVYNQPYFGPAKGFPDNLPDIWYHH
EGAc            GDSYWWGGNLQGAGQYPVVLNVPNRLVYSAHDYATSVYPQTWFSDPT-FPNNMPGIWNKN
                * . ****** .. :*** *   *:****.* *...** *.:*. .. **:*:*.** ::
                                      # #

EGPh            FGYVKLELGYSVVIGEFGGKYGHGGDPRDVIWQNKLVDWMIEN-----KFCDFFYWSWNP
EGAc            WGYLFNQNIAPVWLGEFG---TTLQSTTDQTWLKTLVQYLRPTAQYGADSFQWTFWSWNP
                :**:  :   .* :****       .. *  * :.**:::  .     .  :: :*****
                            #

EGPh            DSGDTGGILQDDWTTIWEDKYNNLKRLMDSCSKSS
EGAc            DSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV
                *********:*** *:   * . *  : .*  ..
                          #
```

Figure 1
Comparison of the catalytic domains of the endoglucanases from *P. horikoshii* (EGPh) and *A. cellulolyticus* (EGAc): asterisks (*) denote identical residues, colons (:) indicate similar residues, dots (.) indicate less similar residues and hashes (#) indicate residues that are conserved in the family 5 enzymes.

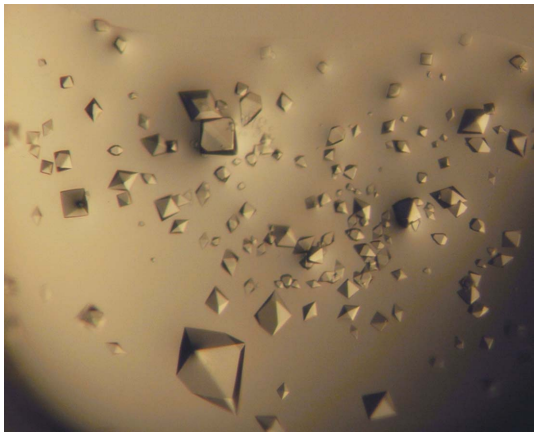


Figure 2
A photograph of the *P. horikoshii* endoglucanase crystals. The longest length of the largest crystal corresponds to 0.4 mm.

3. Results

X-ray diffraction analysis of crystals of the wild-type full-length EGPh was unsuccessful. To improve the structural stability and crystallogenesis of this protein, we attempted to produce a more compact protein, which was designed after N-terminal sequence analysis of the protein proteolyzed by α-chymotrypsin. The digested protein was shown to be about 40 kDa by SDS–PAGE (~3 kDa less than the wild-type protein) and no other protein band was detected by SDS–PAGE. N-terminal sequencing analysis revealed that five amino-acid residues from the N-terminus were cleaved off by the protease. Thus, a mutant of EGPh lacking the first five amino acids at the N-terminus (EGPhΔN5) was constructed. Analysis of the C-terminal amino-acid residues of the proteolyzed protein was ambiguous. Therefore, five and ten residues at the C-terminus (ΔC5 and ΔC10) of EGPh were removed in light of the decreased size of the digested protein. EGPhΔC10 lacking ten residues at the C-terminus showed significantly decreased activity (Kang & Ishikawa, 2007). Consequently, three truncated mutants (EGPhΔN5, EGPhΔC5 and EGPhΔN5C5) which showed no change in enzymatic properties were prepared and screened for crystallization.

Crystallization screening of EGPhΔC5 yielded tetragonal bipyramid-shaped crystals which appeared after two weeks in reservoir solution No. 20 of the Wizard II crystallization screening kit, which consisted of 0.2 *M* zinc acetate, 0.1 *M* MES buffer pH 6.0 and 15%(*v*/*v*) ethanol. Subsequent optimization of the crystallization condition yielded crystals that were suitable for X-ray analysis, with the best crystallization condition containing 0.15 *M* zinc acetate, 0.1 *M* MES buffer pH 6.0 and 15%(*v*/*v*) ethanol at 296 K. The average dimensions of the crystal were about 0.3 × 0.2 × 0.2 mm after two weeks (Fig. 2). Diffraction data were collected to a resolution limit of 1.78 Å (see Table 1). The results of data collection and analysis revealed that the crystal belonged to space group $P4_12_12$ or $P4_32_12$, with unit-cell parameters $a = b = 77.1$, $c = 161.1$ Å. The presence of one molecule of the enzyme per asymmetric unit gives a crystal volume per protein weight (V_M) of 2.71 Å^3 Da^{-1} and a solvent content of 54.68% (Matthews, 1968).

Table 1
Crystallographic parameters and data-collection statistics.

Space group	$P4_12_12$ or $P4_32_12$
Unit-cell parameters (Å)	$a = b = 77.1$, $c = 161.1$
Matthews coefficient (Å^3 Da^{-1})	2.71
Solvent content (%)	54.7
Resolution range (Å)	50–1.78 (1.84–1.78)
Total No. of observed reflections	299631
Total No. of unique reflections	47419
Redundancy	3.4 (3.3)
$\langle I/\sigma(I)\rangle$	17.1
$R_{\rm merge}$ (%)†	5.5 (28.2)
Completeness (%)	99.6 (100)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, 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.

Determination of the structure of EGPh is in progress. The structure of EGPh will be helpful in elucidating the catalytic mechanism in order to improve the activity of the enzyme.

References

Ando, S., Ishida, H., Kosugi, Y. & Ishikawa, K. (2002). *Appl. Environ. Microbiol.* **68**, 430–433.
Bauer, M. W., Driskill, L. E., Callen, W., Snead, M. A., Mathur, E. J. & Kelly, R. M. (1999). *J. Bacteriol.* **259**, 284–290.
Kang, H.-J. & Ishikawa, K. (2007). *J. Microbiol. Biotechnol.* **17**, 1249–1253.
Kashima, Y., Mori, K., Fukada, H. & Ishikawa, K. (2005). *Extremophiles*, **9**, 37–43.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sakon, J., Adney, W. S., Himmel, M. E., Thomas, S. R. & Karplus, P. A. (1996). *Biochemistry*, **35**, 10648–10660.
Uegaki, T., Shirakawa, M., Fujita, T., Taniguchi, T. & Kyogoku, Y. (1993). *Protein Eng.* **6**, 195–200.