FEBS Letters 586 (2012) 1009–1013

FEBS Letters

journal homepage: www.FEBSLetters.org

Atomic resolution of the crystal structure of the hyperthermophilic family 12 endocellulase and stabilizing role of the DxDxDG calcium-binding motif in *Pyrococcus furiosus*

Han-Woo Kim, Misumi Kataoka, Kazuhiko Ishikawa *

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

ARTICLE INFO

Article history:
Received 20 December 2011
Revised 30 January 2012
Accepted 20 February 2012
Available online 6 March 2012

Edited by Miguel De la Rosa

Keywords:
Hyperthermophile
Endocellulase
Archaea
Crystal structure
DxDxDG calcium-binding motif
Pyrococcus furiosus

ABSTRACT

Hyperthermophilic glycoside hydrolase family 12 endocellulase (EGPf) from the archaeon *Pyrococcus furiosus* catalyzes the hydrolytic cleavage of β-1,4-glucosidic linkage in β-glucan cellulose. A truncated EGPf (EGPfΔN30) mutant lacking the proline and hydroxyl-residue rich region at the N terminus was constructed, and its crystal structure was resolved at an atomic resolution of 1.07 Å. Our results indicate that the structure of EGPf, which consists of a β-jelly roll, exhibits structural similarity with the endocellulase of *Thermotoga maritima*. Additionally, we further determined that the thermostability of EGPf is maintained in part by the binding of Ca^{2+} in a DxDxDG Ca^{2+}-binding motif, atypical of most archaeal proteins.

1. Introduction

Cellulase is one of the most important industrial enzymes in terms of biomass utilization owing to its key role in the degradation of β-glucan cellulose. Recent research into the production of biofuel from lignocellulose biomass has focused on developing an ideal cellulase for efficient biomass saccharification. This research suggests that a hyperthermophilic cellulase would be very useful in this context, because the enzymatic reaction process at high temperature has many advantages including reduced risk of microbial contamination, increased substrate solubility, and improved transfer rate. Therefore, many researchers have been focusing on the development of a hyperthermophilic cellulase with high activity.

Several hyperthermophilic β-1,4 endocellulases (endotype cellulase) have been identified in the genome database of hyperthermophilic archaea. The hyperthermohpilic archaea *Pyrococcus horikoshii* and *Pyrococcus furiosus* have glycoside hydrolase (GH) family 5 and 12 endocellulases, respectively. These enzymes demonstrate different substrate specificity. The first crystal structure of the hyperthermophilic endocellulase (EGPh; GH family 5 enzyme) from *P. horikoshii* has been determined [1], and its substrate recognition mechanism has also been reported by our laboratory [2]. However, the crystal structure of the hyperthermophilic endocellulase (EGPf; GH family 12 enzyme) from *P. furiosus* remains to be resolved. In this study, we reported the first successful crystallization of EGPf using the truncated mutant, and determined its crystal structure at atomic resolution. In addition, we examined the functional role of the DxDxDG Ca^{2+}-binding motif in EGPf.

2. Materials and methods

2.1. Construction and preparation of the truncated protein

A truncated enzyme gene (EGPfΔN30) with 30 amino acid residues deleted from the N-terminal region of EGPf (signal sequence and the proline and hydroxyl-residue rich regions) was constructed by the polymerase chain reaction (PCR) method. The truncated gene was inserted into the expression vector pET11a (Novagen, Madison, WI, USA) and the constructed plasmid was introduced into *Escherichia coli* strain BL21(DE3) for recombinant protein expression. Expression and purification of the recombinant enzymes was carried out using previously established methods [3]. The purity

Abbreviations: GH, glycosyl hydrolase; PCR, polymerase chain reaction; CHES, 2-(*n*-cyclohexylamino) ethanesulfonic acid; EGPf, endocellulase from *Pyrococcus furiosus*; EGPfΔN30, the truncated mutant of EGPf; DSC, differential scanning calorimetry; CMC, carboxyl-methyl cellulose

* Corresponding author.
E-mail address: kazu-ishikawa@aist.go.jp (K. Ishikawa).

0014-5793/$36.00
doi:10.1016/j.febslet.2012.02.029

and molecular weight of the protein sample were analyzed by SDS–PAGE. The protein concentration of EGPfΔN30 was determined from UV absorbance at 280 nm, based on an extinction coefficient of 81,790 calculated from the protein sequence.

2.2. Differential scanning calorimetry

Differential scanning calorimetry (DSC) measurements were carried out using a nanoDSCII instrument (TA Instruments, DE, USA) with platinum tubing cells having a volume of 0.3 mL. Proteins were dialyzed against 50 mM sodium acetate (pH 5.5). The dialysis buffer was used as a reference solution for the DSC analysis. Samples containing 1.0 mg/mL of protein were heated at 1 °C/min from 5 °C to 125 °C.

2.3. Crystallization

The purified protein (EGPfΔN30) was dialyzed against 50 mM Tris–HCl buffer (pH 8.0) and then concentrated to 15 mg/mL using an Amicon, Centricon YM-10 (Millipore, Ireland). Crystallization screening was performed using the hanging-drop vapor-diffusion method using Crystal screen (Hampton research) and Wizard 1 and 2 (Emerald Biosystems) at 298 K. Typically, drops consisting of 1 μL protein solution and 1 μL reservoir solution were equilibrated against 0.4 mL of reservoir solution. Based on the initial screening results, the best crystals of the EGPfΔN30 were obtained using a reservoir solution of 120 mM 2-(*n*-cyclohexylamino) ethanesulfonic acid (CHES) buffer (pH 9.0) containing 50 mM lithium sulfate and 0.5 M potassium sodium tartrate.

2.4. Data collection and processing

The selected crystals were harvested and immersed in the cryoprotectant solution containing 30% (v/v) glycerol in the mother liquor. The soaked crystal was collected with a cryoloop (Hampton Research, Aliso Viejo, CA, USA) and immediately flash-cooled under a stream of nitrogen gas at 100 K. X-ray diffraction data for a single crystal were collected using a Quantum 315 detector (ADSC) at the SPring-8 BL44XU beam line (Hyogo, Japan). The diffraction data were collected in two steps: low (5.00 Å) and high (1.07 Å) resolution, for the complete data set. The low resolution data set was collected at a wavelength of 0.9 Å. The distance from the crystal to the detector was 250 mm. The high resolution data set was collected at a wavelength of 0.8 Å. The distance from the crystal to the detector was 90 mm. In each process, the crystal was rotated 360° with an oscillation angle of 1° per frame. The data collected from diffraction measurements were merged, indexed, integrated, and scaled with the programs in the *HKL*-2000 software package [4]. The structure of EGPfΔN30 was solved by molecular replacement with BALBES program using PDB model 1H0B (*Rhodothermus marinus* Cel12A, 31% identity with EGPf) [5] as the search model. All model-building stages were performed with the Coot program [6]. Refinement up to 1.07 Å was performed with REFMAC5 in the CCP4 package [7]. Restraints for glycerol and CHES molecules were generated from coordinates in the HIC-UP database [8] using the PRODRG server [9]. After convergence of refinement in REFMAC5, R_{cryst} and R_{free} converged to 16.1% and 18.0%, respectively, with isotropic B-factor refinement. Subsequently, anisotropic B-factor refinement was turned on, and after extensive refinement, the R_{cryst} and R_{free} values converged to 12.7% and 15.8%, respectively. Water molecules were introduced at peaks over 3.5σ in the difference map fulfilling reasonable distance and hydrogen bonding criteria to protein residues or other water molecules. Ramachandran plot parameters were calculated by program PROCHECK [10]. The diffraction data and the crystallographic refinement statistics are summarized in Table 1. Figures were produced using PyMOL (www.pymol.org).

3. Results and discussion

3.1. Structural determination and overall structure of the truncated mutant EGPfΔN30

A hyperthermophophilic β-1,4 endoglucanase (GH family 12, endocellulase) was identified in an archaeon *P. furiosus* [11]. The first crystal structure of hyperthermophilic endocellulase (GH family 5) from *P. horikoshii* has been already determined [12]. However, no study to date has reported a structure for the family 12 hyperthermophilic endocellulase from *P. furiosus*. The protein EGPf (Gene ID: AAD54602.1, cellulase; EC3.2.1.4) contains a signal peptide and a proline and hydroxyl-residue rich regions at the N terminus. In this study, we report the crystallization and structural analysis using the truncated mutant of EGPf (EGPfΔN30) modified by deletion of 30 amino acid residues from the N-terminal region. Several crystals were obtained within 2 days at 303 K, while crystals suitable for X-ray analysis were obtained using the optimized reservoir solution as described in Section 2. The average size of the optimized crystal was approximately 0.7 mm × 0.4 mm × 0.4 mm. Diffraction data was collected to a resolution limit up to 1.07 Å, consisting of 1,187,674 measurements and 130,828 unique reflections. The crystal belonged to the orthorhombic space group $P2_12_12$ with unit cell dimensions a = 58.01, b = 118.67, and c = 46.76. One molecule of EGPfΔN30 was identified in the crystallographic asymmetric unit and its structure was determined up to 1.07 Å (Table 1). The final model, determined using the BALBES program, contained a monomer molecule comprising 270 amino acid residues organized into a β-jelly roll fold (Fig. 1A).

A structural similarity search using the DALI server identified endoglucanase as the closest structural homologue of EGPfΔN30.

Table 1
Statistics of data collection and refinement.

Data collection	
Wavelength (Å)	0.8 and 0.9
Space group	$P2_12_12$
Unit-cell parameters (Å)	$a = 58.01$, b = 118.67, $c = 46.76$
Matthews coefficient (Å^3 Da^{-1})	2.32
Solvent content (%)	47
Subunits per asymmetric unit	1
Resolution range (Å)	50–1.07 (1.09–1.07)
Number of observed reflections	1187,674
Total number of unique reflections	130,828
Redundancy	9.1 (5.0)
$\langle I/\sigma(I)\rangle$	32.5 (4.9)
R_{merge}[a]	0.062 (0.324)
Completeness (%)	92.2 (81.2)
Refinement	
No. of atoms	
Amino acid residues	2230
Glycerol	11
CHES	2
Ca^{2+}	1
Water	328
Resolution used in refinement	50–1.07
R_{work}[b]/R_{free}[c] (%)	12.7/15.8
R.M.S. bond distance (Å)	0.025
R.M.S. bond angle (deg.)	2.34
Mean overall B factor (Å^2)	15.84
Ramachandran plot	
In most favored regions (%)	97.8 0
In disallowed regions (%)	0
PDB ID	3VGI

[a] $R_{merge} = \sum_{hkl}\sum_i | I_i(hkl) - \langle I(hkl)\rangle | / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the i-th intensity measurement of reflection hkl, including symmetry-related reflections, and $\langle I(hkl)\rangle$ is their average.

[b] $R_{work} = \sum_{hkl} | F_o - F_c | / \sum_{hkl} F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection hkl, respectively.

[c] R_{free} is calculated as the R_{cryst}, using F_o that were excluded from the refinement (5% of the data). Values for the last resolution shell are given in parentheses.