Hyperthermostable Endoglucanase from *Pyrococcus horikoshii*

Susumu Ando,^{1,2*} Hiroyasu Ishida,³ Yoshitsugu Kosugi,³ and Kazuhiko Ishikawa^{1*}

National Institute of Advanced Industrial Science and Technology (Kansai), Ikeda, Osaka 563-8577,¹ Rakuto Kasei Industrial Co. Inc., Otsu, Shiga 520-2277,² and National Institute of Advanced Industrial Science and Technology (Tsukuba), Tsukuba, Ibaraki 305-8566,³ Japan

Received 25 June 2001/Accepted 25 October 2001

An endoglucanase homolog from the hyperthermophilic archaeon *Pyrococcus horikoshii* was expressed in *Escherichia coli*, and its enzymatic characteristics were examined. The expressed protein was a hyperthermostable endoglucanase which hydrolyzes celluloses, including Avicel and carboxymethyl cellulose, as well as β -glucose oligomers. This enzyme is the first endoglucanase belonging to glycosidase family 5 found from *Pyrococcus* species and is also the first hyperthermostable endoglucanase to which celluloses are the best substrates. This enzyme is expected to be useful for industrial hydrolysis of cellulose at high temperatures, particularly in biopolishing of cotton products.

In the textile industry, cellulases have been used in large quantity for biopolishing of cotton products. This process is essential for removing fuzz and giving a soft touch and clean appearance to the fabrics. The enzymes that are presently used for this purpose are mesophilic cellulases from fungi, and their optimum reaction temperatures are between 50 and 55°C. If they are replaced by a hyperthermostable enzyme with an optimum temperature close to 100°C, which will make it possible to treat cotton products in steam, the processing will be much more simple, quick, and efficient than in the presently employed method. Desizing, the step to remove starch from the fabrics, is performed at temperatures at least 70°C, and higher temperatures are preferred. Because amylases active at these temperatures are available, this process is performed at temperatures higher than 70°C. However, cellulase treatment, which usually follows desizing, is performed at lower temperatures, since a cellulase that is active and stable in this temperature range has not been available. If such a hyperthermostable cellulase is introduced, it will be possible to combine desizing and biopolishing in a single step.

For the purpose of producing such a hyperthermostable cellulase, we investigated the possibility of utilizing the genetic resources of hyperthermophilic archaea. *Pyrococcus horikoshii* OT3 is one of those organisms with the optimum growth temperature above 95°C (4). Based on its complete published genome sequence (8, 9), we have expressed some of the enzymes from this organism and characterized them, most of which have proven to be highly thermostable (1, 7). It is quite likely, therefore, that other useful enzymes that are active and stable at extremely high temperatures remain to be found.

According to the classification of glycosidic hydrolases proposed at the CAZy website (<http://afmb.cnrs-mrs.fr/~pedro/CAZy/db.html>), the glycosidic hydrolases known to date have been classified into more than 80 families. Among the three

species belonging to the genus *Pyrococcus*, two endoglucanases showing regions of homology with family 12 (2) and family 16 (5) have been found in *P. furiosus*. It has been reported, however, that the hydrolytic activity of the former enzyme toward celluloses was lower than that toward β -glucose oligomers by at least 2 orders of magnitude (2) and that the latter had no detectable activity toward celluloses (5). There has been no report so far on the glucanases from *Pyrococcus abyssi* and *P. horikoshii*.

The genome sequence data of *P. horikoshii* (8, 9) suggest that the open reading frame PH1171 (1,377 nucleotides) is homologous with genes from some other organisms encoding endoglucanase belonging to family 5 (Fig. 1). There has been no report on hyperthermophilic endoglucanase belonging to this family. Therefore, we cloned this gene, expressed it in *Escherichia coli*, and characterized the expressed protein.

Preparation of recombinant protein. *P. horikoshii* cells were a gift from Y. Kawarabayasi of the National Institute of Technology and Evaluation (Tokyo, Japan). The cells were cultivated as reported previously (4), and their genomic DNA was prepared according to the method described previously (10).

The open reading frame PH1171 was amplified using the PCR with primers having *Nde*I and *Bam*HI restriction sites according to the methods reported previously (7). The sequences of the primers used were 5'-TTTGAATTCTTTCATATGGAGGGGAATACTATTCTTAAATC-3' (upper primer, containing an *Nde*I site as underlined) and 5'-TTTTCTAGATTTGGATCCTTTGGGCTACCTGGGAGCCCTTCTTAA-3' (lower primer, containing a *Bam*HI site as underlined). Although the 5' terminal sequence of PH1171 suggested the presence of a signal peptide sequence (12) in the translated product, the primers were designed to amplify the whole open reading frame. The amplified gene was digested with *Nde*I and *Bam*HI and was inserted into pET11a plasmid digested with the same restriction enzymes. The nucleotide sequence of the inserted gene was determined using an LICOR Model LIC-4200L(S) sequencer (Aloka, Mitaka, Tokyo, Japan) to verify identity with the anticipated sequence. The amplified gene was expressed using the pET11a vector system

* Corresponding author. Mailing address: Special Division for Human Life Technology, National Institute of Advanced Industrial Science and Technology (Kansai), 1-8-31 Midorigaoka, Ikeda, Osaka 563-8577, Japan. Phone: 81-727-51-9526. Fax: 81-727-51-9628. E-mail for Susumu Ando: andosusumu@hotmail.com. E-mail for Kazuhiko Ishikawa: kazu-ishikawa@aist.go.jp.

TABLE 1. Substrate specificity of EGPh^a

Substrate (concn)	Sp act (k_{cat}) (1/s)
CMC (0.25% [wt/vol])	1.286 ± 0.082
Avicel SF (0.25% [wt/vol])	0.212 ± 0.050
Lichenan (0.25% [wt/vol])	0.565 ± 0.147
Xylan (0.25% [wt/vol])	<0.001
Xyloglucan (0.25% [wt/vol])	<0.001
Curdlan (0.25% [wt/vol])	<0.001
Cellobiose (2 mM)	0.0047 ± 0.0003
Cellotriose (2 mM)	0.0076 ± 0.0019
Cellotetraose (2 mM)	0.0250 ± 0.0001
Cellopentaose (2 mM)	0.0252 ± 0.0017

^a The k_{cat} values were determined as the initial velocity (increase in reducing ends) divided by enzyme concentration and reaction time. The initial velocity was measured in the presence of excess concentration ($>K_m$) of the substrate.

in the host *E. coli* BL21(DE3) pLysS, according to the instructions provided by the manufacturer.

The expressed protein was purified by heating the crude extract at 85°C for 30 min, precipitation with 90% saturated ammonium sulfate, ion-exchange chromatography using Hi-

trap Q (Pharmacia, Uppsala, Sweden), and gel filtration chromatography using HiLoad 26/60 Superdex 200 (Pharmacia).

The detailed procedures for purification have been described previously (7).

Characterization of recombinant protein. The glucanase activity was assayed by measuring the reducing sugars according to the modified Somogyi-Nelson method (6). The substrates used were Avicel SF (microcrystalline cellulose; Asahi Kasei, Tokyo, Japan), carboxymethyl cellulose (CMC; Sigma), lichenan (from *Cetraria islandica*; Sigma), xylan (from birch wood; Sigma), curdlan (from *Alcaligenes faecalis*; Sigma), and xyloglucan (from tamarind; Megazyme International Ireland), and β -glucose oligomers (from cellobiose to cellopentaose; Seikagaku Kogyo, Tokyo, Japan).

The hydrolytic activity toward various substrates was measured at 85°C in 100 mM acetate buffer (pH 5.6). As listed in Table 1, this protein hydrolyzed celluloses such as CMC and Avicel SF. Lichenan, a β -1,3 and -1,4 glucan, was also hydrolyzed. β -Glucose oligomers (from cellobiose to cellopentaose) were hydrolyzed, although to much lesser extent. No activity

```

EGPh  1  ENT TY-QTPTGIYEVVRGDTIYMINVTSGEETPIHLFGVNWFGFETPNHVVHGLWKRNWE  59
EGAc  1  AGGGYWHT-SG-----R-E-I-L-D---ANNVPVRIAGINWFGFETCNVYVHGLWSRDYR  47
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

EGPh  60  DMLLQIKSLGFNAIRLPCTESVKPGTQPIGID-YSKNPDRLGLDSLQIMEKIIKKAQDL  118
EGAc  48  SMLDQIKSLGYNTIRLPYDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAAQOI  107
      **  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      #

EGPh  119  GIFVLLDYHRIGCTHIEPLWYTEDFSEEDFINTWIEVAKRF-GKYWNVIGADLKNEPHSV  177
EGAc  108  GLRIILDRHRPDCSGQSALWYTSVSEATWISDLQALAQRYKGNP-TVVGFDLHNEPH--  164
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      #

EGPh  178  TSPPAAAYTDGTGATWGMGNPATDWNLAERIGKAILKVAPHWLIFVEGTQ-FTNPKTDS  236
EGAc  165  -D-P-AC-----WCGDPSIDWRLAAERAGNAVLSVNPNLIFVEGVQSY-N--GD-S  210
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

EGPh  237  YKWGYNAAWGGNLMMAVDYVNL-PRNKLVSYPHYGPDVYNGPYFGPAKGFPDNLDPDIW  295
EGAc  211  ---Y--WWGGNLQAGQYPVVLNPNRLVYSAHDYATSVYPGTWFSDF-TFPNMPGIW  263
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      #  #

EGPh  296  YHHFGYVKLELGYSVV-IGFEGGKYGHGGDPRDVIWQNKLVDM-IENKF-CD-F-F-YW  349
EGAc  264  NKNWGYL-FNQNIAPVVLGEFGTTL-Q-STT-DQWLKTLVQYLRPTAQYGAASFQWTFW  319
      **  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
      #

EGPh  350  SWNPDSGDTGGILQDDWTIWD-K--Y-NNLK-RLMDSC-SKSSSSTQ---SVIRSTTP  400
EGAc  320  SWNPDSGDTGGILKDDWQTV-DTVKDGYLAPIKSSIFDPVGASAPSSQSPSPSPSPSP  378
      *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

EGPh  401  TKSNTSKKICGPAILIILAVFSLLLRRAPR  430
EGAc  379  SPSASRTPTPTPTPTASPTPTLTPTATPTPTASPTPSPTAASGARCTASYQVNSDWGNGF  438

EGAc  439  TVTVAVTNSGSVATKTWTVSWTFGGNQITNSWNAAVTQNGQSVTARNMSYNNVIQPGQN  498

EGAc  499  TTFGFQASYTGSNAAPTAVACAAS  521

```

FIG. 1. Comparison of the amino acid sequences of the endoglucanases from *P. horikoshii* (EGPh) and from *A. cellulolyticus* catalytic domain (EGAc) (43% identity). The first line shows the putative signal peptide sequence of EGPh. The sequences are aligned with dashes to indicate gaps. Asterisks (*) indicate that the amino acid is identical in both EGAc and EGPh sequences. The residues conserved in the family 5 endoglucanases are indicated by number signs (#). The active-site residues are shown in boldface. The C-terminal regions of both protein (residues 401 to 430 in EGPh and 379 to 521 in EGAc) were excluded from the alignment because of their low homology in these regions.

FIG. 2. Mode of hydrolysis of EGPh. The viscosity values measured were plotted against the reducing sugar concentrations as measured by Somogyi-Nelson reaction using glucose as the standard. A solution of CMC (3.5% in 100 mM acetate buffer, pH 5.6, 50 ml) was subjected to hydrolysis by 4 U of EGPh/ml for 30 to 120 min at 90°C. After 30, 60, and 120 min of incubation, the viscosity of the solution was measured at 40°C, using Viscometer RE110L (Toki Sangyo, Tokyo, Japan). The initial value of viscosity was 18.92 mPa.

was detected toward curdlan (β -1,3 glucan), xylan (β -1,4 glycan of xylose), and xyloglucan (mixed β -1,4 glycan of glucose and xylose). From these observations it was concluded that this protein is a glucanase (hereinafter referred to as EGPh) which specifically hydrolyzes β -1,4 glucosidic bonds. This was as expected from the amino acid sequence homology of EGPh with the family 5 endoglucanases (8), except that xyloglucan, which is usually hydrolyzed by endoglucanases, was not hydrolyzed by EGPh. In order to determine the mode of hydrolysis of EGPh, the relationship between the viscosity and the reducing power of the hydrolysis products by EGPh was determined. As shown

in Fig. 2, the decrease in viscosity of the CMC solution incubated with EGPh in relation to the increase in the reducing sugar showed a pattern characteristic of the endo-type cellulases (3). From these results, it was concluded that EGPh was an endo-type cellulase. The k_{cat} values for the substrates examined are listed in Table 1. The optimum pH was between 5.4 and 6.0. The total yield of EGPh was 0.27 mg per liter of the transformant culture. The specific activity of the purified EGPh was 8.5 U/mg, where 1 U is defined as the amount of enzyme producing reducing power equivalent to 1 mg of glucose per min from CMC in the routine assay condition (0.5% CMC in 500 μ l of 100 mM acetate buffer, pH 5.6, 85°C).

The sequence of the deduced translation product of this gene was aligned with that of the endoglucanase gene of *Acidothermus cellulolyticus*, a member of the family 5 endoglucanases, the catalytic region of which has a known crystallographic structure (11). The alignment was performed using a Genetyx-Mac program (Software Development, Tokyo, Japan). As shown in Fig. 1, all the seven residues characteristic of the family 5 endoglucanases, including the catalytic residues, are conserved in EGPh. From the result of the sodium dodecyl sulfate-polyacrylamide gel electrophoresis of the purified enzyme, the molecular size of the monomer of this enzyme was estimated to be approximately 43 kDa. Since this was significantly smaller than the value (52 kDa) calculated on the basis of the sequence of the open reading frame, it was suspected that EGPh was an incomplete product of the gene. In order to make this clear, the N-terminal amino acid analysis was performed at the Custom Service Center of Takara Shuzo (Kusatsu, Japan). The result indicated that the first 5 amino acids from the N terminus were ENTYY, suggesting that the first 28 amino acid residues of the predicted translation product of the open reading frame were missing in EGPh. The sequence of these 28 amino acid residues removed had all the characteristics unique to the signal peptide sequences, such as a high ratio of the hydrophobic amino acid residues (12), implying that

FIG. 3. (A) Effect of the temperature on the hydrolytic activity of EGPh on CMC. The hydrolytic activity was measured in 100 mM acetate buffer (pH 5.6). The assay was measured for 15 min. (B) Effect of heating on EGPh activity. EGPh solution was incubated at 97°C in 100 mM acetate buffer (pH 5.6). At the time shown, aliquots were taken out and the activity was measured in the same buffer at 85°C using CMC as the substrate.

EGPh is possibly a protein destined for secretion. In order to see whether EGPh was also produced in the parent organism, *P. horikoshii* was cultivated according to the method described previously (4) and glucanase activity was checked for both the extract of *P. horikoshii* cell mass obtained by sonication and the concentrated culture supernatant. The extract of the cell mass of this culture gave barely detectable activity of glucanase. However, no glucanase activity was detected from the concentrated culture supernatant, showing that this enzyme is not secreted in the medium. Thus, the use of *E. coli* transformant is a much better choice for production of EGPh than the use of the parent organism, *P. horikoshii*, considering the difficulty of cultivating *P. horikoshii* under strictly anaerobic conditions at 95°C. The N-terminal sequence of EGPh purified from *P. horikoshii* cells was also determined, and it proved to be identical to that of EGPh from *E. coli* transformant, reinforcing the possibility that the 28-residue segment removed from the protein is a signal sequence.

The temperature dependence of EGPh was examined by measuring the hydrolytic activity in 100 mM acetate buffer (pH 5.6). The assay was measured for 15 min. As shown in Fig. 3A, the optimum temperature for the glucanase reaction was higher than 97°C.

Thermostability of EGPh was measured by incubating aliquots of enzyme solution (0.1 mg/ml) at 97°C in 100 mM acetate buffer (pH 5.6) for up to 3 h, followed by measurement of residual activity. As shown in Fig. 3B, the residual activity after heating for 3 h at 97°C was 80% of that of EGPh that was not heated.

Because of its ability to hydrolyze celluloses at high temperatures above 90°C, as well as its thermostability, EGPh is expected to be an excellent tool for industrial hydrolysis of cellulose, particularly for biopolishing of cotton products. The yield of this enzyme from the *E. coli* transformant, however, is rather low for industrial application. In order to achieve more efficient production, a study is in progress for extracellular production of this enzyme using *Bacillus brevis* as the host cell.

We are grateful to Akio Shimomura of Rakuto-Kasei Industrial Co. Ltd. for his helpful suggestions and instructions.

REFERENCES

1. Ando, S., K. Ishikawa, H. Ishida, Y. Kawarabayasi, H. Kikuchi, and Y. Kosugi. 1999. Thermostable aminopeptidase from *Pyrococcus horikoshii*. *FEBS Lett.* **447**:25–28.
2. Bauer, M. W., L. E. Driskill, W. Callen, M. A. Snead, E. J. Mathur, and R. M. Kelly. 1999. An endoglucanase, EglA, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes β -1,4 bonds in mixed-linkage (1 $\rightarrow$ 3),(1 $\rightarrow$ 4)- β -D-glucans and cellulose. *J. Bacteriol.* **181**:284–290.
3. Gilkes, N. R., M. L. Langsford, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren. 1984. Mode of action and substrate specificities of cellulases from cloned bacterial genes. *J. Biol. Chem.* **259**:10455–10459.
4. Gonzalez, J. M., Y. Masuchi, F. T. Robb, J. W. Ammerman, D. L. Maeder, M. Yanagibayashi, J. Tamaoka, and C. Kato. 1998. *Pyrococcus horikoshii* sp. nov., a hyperthermophilic archaeon isolated from a hydrothermal vent at the Okinawa Trough. *Extremophiles* **2**:123–130.
5. Gueguen, Y., W. G. B. Voorhorst, J. van der Oost, and W. M. de Vos. 1997. Molecular and biochemical characterization of an endo- β -1-3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **272**:31258–31264.
6. Hiromi, K., Y. Takahashi, and S. Ono. 1963. Kinetics of hydrolytic reaction catalyzed by crystalline bacterial α -amylase. The influence of temperature. *Bull. Chem. Soc. Jpn.* **36**:563–569.
7. Ishikawa, K., H. Ishida, Y. Koyama, Y. Kawarabayasi, J. Kawahara, E. Matsui, and I. Matsui. 1998. Acylamino-acid releasing enzyme from the thermophilic archaeon *Pyrococcus horikoshii*. *J. Biol. Chem.* **273**:17726–17731.
8. Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, T. Yoshizawa, Y. Nakamura, F. T. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**:55–76.
9. Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S.-I. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, T. Yoshizawa, Y. Nakamura, F. T. Robb, K. Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3 (supplement). *DNA Res.* **5**:147–155.
10. Meade, H. M., S. R. Long, G. B. Ruvkun, S. E. Brown, and F. M. Ausubel. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *J. Bacteriol.* **149**:114–122.
11. Sakon, J., W. S. Adney, M. E. Himmel, S. R. Thomas, and P. A. Karplus. 1996. Crystal structure of thermostable family 5 endocellulase E1 from *Acidothermus cellulolyticus* in complex with cellotetraose. *Biochemistry* **35**:10648–10660.
12. Watson, M. E. E. 1984. Compilation of published signal sequences. *Nucleic Acids Res.* **12**:5145–5164.