J. Biochem. 122, 683-685 (1997)

Crystallization and Preliminary X-Ray Analysis of a Truncated Family A Alkaline Endoglucanase Isolated from *Bacillus* sp. KSM-635[1]

Tsuyoshi Shirai,* Takashi Yamane,*,[2] Teruo Hidaka,* Kazuo Kuyama,* Atsuo Suzuki,* Tamaichi Ashida,*,[3] Katsuya Ozaki,† and Susumu Ito†

**Department of Biotechnology and Biomaterial Chemistry, Graduate School of Engineering, Nagoya University, Chikusa-ku, Nagoya 464-01; and* †*Tochigi Research Laboratories of Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi 321-34*

Received for publication, June 25, 1997

The catalytic domain of an alkaline endo-1,4-β-glucanase (family A) isolated from *Bacillus* sp. KSM-635 (M_r=40.2 kDa) was crystallized using the hanging drop vapor diffusion method. Two different crystal forms were obtained. Form 2 crystals (trigonal space group R3 with cell dimensions of $a=b=111.9$ and $c=207.1$ Å in a hexagonal lattice) were found to be more stable than form 1 ones upon X-ray irradiation. A full data set for form 2 crystals has been collected up to 3.3 Å resolution.

Key words: alkaline cellulase, crystallization, detergent enzyme, endo-1,4-β-glucanase, X-ray crystallography.

Cellulose is the most abundant carbohydrate on earth and thus a major ingredient of the biomass. The products of enzyme digestion of cellulose are currently generating much interest as potential substitutes for fossil oil resources. Cellulose is a β-1,4-linked polyglucose, and is naturally hydrolyzed by a number of cellulases produced by a variety of cellulolytic microorganisms. One of these types of enzyme is the endoglucanases (endo-1,4-β-glucanase, EC 3.2.1.4), which have been classified into 11 families, named A through K, on the basis of amino acid sequence similarity (*1, 2*).

Cellulases comprise a catalytic domain and usually one or more additional domains concerned with substrate binding or multienzyme complex formation (*2*). X-Ray structural studies of the catalytic domains of endoglucanases have clarified the catalytic mechanisms of the following endoglucanases with respect to their three-dimensional structures: endoglucanase CelCCA (classified into family A) from *Clostridium cellulolyticum* at 1.6 Å resolution (*3*); cellobiohydrolase II (CBHII, family B) from *Trichoderma reesei* at 2.0 Å resolution (*4*); endoglucanase E2 (family B) from *Thermomonospora fusca* at 1.8 Å resolution (*5*); cellobiohydrolase I (CBHI, family C) from *Trichoderma reesei* at 1.81 Å resolution (*6*); endoglucanase CelD (family E) from *Clostridium thermocellum* at 2.3 Å resolution (*7*); and endoglucanase V (family K) from *Humicola insolens* at 1.5 Å resolution (*8*).

In a previous study, some of the authors of the present paper isolated and purified an alkaline endoglucanase (alkaline EG) from *Bacillus* sp. KSM-635 (*9*). Sequence analysis of the alkaline EG gene revealed that the purified alkaline EG consists of 941 amino acid residues with a molecular mass of 105 kDa (*10*). The optimum pH and temperature for the catalytic activity of alkaline EG are pH 9.5 and 45°C, respectively (*9*). Due to these alkaliphilic and thermostable characteristics, the enzyme is commercially used as a component of laundry detergents. The structural analysis of alkaline EG is essential for a better understanding of the catalytic mechanism of this enzyme, and for providing a basis for future protein engineering aimed at improving its activity and stability.

The alkaline EG gene was truncated and cloned in order to extract the catalytic domain. The 584 amino-terminal residues of the alkaline EG are sufficient for expression of the full enzymatic activity, and the sequence of the C-terminal half of this fragment is homologous to those of the catalytic domains of all the enzymes in family A (*10*). The truncated gene, that encodes amino acids Ala228 to Leu584 of alkaline EG, was recently expressed in *Bacillus subtilis*. The 40.2 kDa product of the truncated alkaline EG (truncated EG) gene is composed of 357 native amino acids plus 7 amino acids from the cloning-vector (*11*). The enzymatic activity of the truncated EG is comparable to that of the intact alkaline EG, which prompted crystal structure analysis of the protein. Here we report the crystallization and preliminary crystallographic analysis of the truncated enzyme.

Crystals of truncated EG were prepared using the hanging drop vapor diffusion method. A 0.1 M ammonium acetate buffer (pH 5.5 or 6.0) containing 1.5% (w/v) NaCl, 15% (v/v) methanol, and 25% (w/v) PEG4000 was used as the reservoir solution. A 10 μl drop of a mixture of equal volumes of the reservoir solution and a 1.4% (w/v) protein solution in 10 mM phosphate buffer (pH 7.0) containing 100 mM NaCl (concentrated fractions of the Sephacryl-S200 gel filtrate) was equilibrated against 1 ml of the

[1] This work was supported by a Grant-in-Aid for Developmental Scientific Research (B) (No. 07554059) from the Ministry of Education, Science, Sports and Culture of Japan.

[2] To whom correspondence should be addressed. Phone: +81-52-789-3339, Fax: +81-52-789-3218, E-mail: yamane@hix.nagoya-u.ac.jp

[3] Present address: 1-299 Oobari, Meito-ku, Nagoya 465.

reservoir solution at room temperature. Crystals with maximum dimensions of $0.2\times0.2\times0.1$ mm^3 grew in 2-3 weeks.

Two morphologically different forms of crystals were obtained. Form 1 crystals had a rhombic prism shape with finer edges and planes than form 2 crystals, which were relatively dull in appearance and tended to grow into rather irregular shapes in the final stage of crystal growth (Fig. 1). Form 1 and 2 crystals were dominant in the drops at pH 6.0 and 5.5, respectively. Preliminary experiments revealed that form 1 crystals were highly liable to X-ray radiation damage and thus lost their diffractability within a few minutes irradiation. The highest resolution of reflections from form 1 crystals was less than 4 Å. Unexpectedly, form 2 crystals withstood X-ray exposure and diffracted better than form 1 crystals, and thus were the focus of further analysis.

The form 2 crystals belong to trigonal space group R3, and have cell dimensions of $a=b=111.9$ and $c=207.1$ Å when the crystal lattice is converted into hexagonal axes. Assuming the number of molecules in an asymmetric unit

TABLE I. **Diffraction data processing statistics.**

Resolution range (Å)	20.0-3.3
No. of total reflections	89,912
No. of independent reflections	11,790
Average $I/\sigma(I)$	4.2
Completeness $[I>\sigma(I)]$ (%)	81.1
of final shell (3.42-3.30 Å)	55.3
R_{merg}[a]	0.116
of final shell (3.42-3.30 Å)	0.217

[a] $R_{merg}=\Sigma|I_i-\langle I\rangle|/\Sigma I_i$, where I_i is the scaled intensity of the ith observation, and $\langle I\rangle$ is the average of the intensities that are equivalent to I_i.

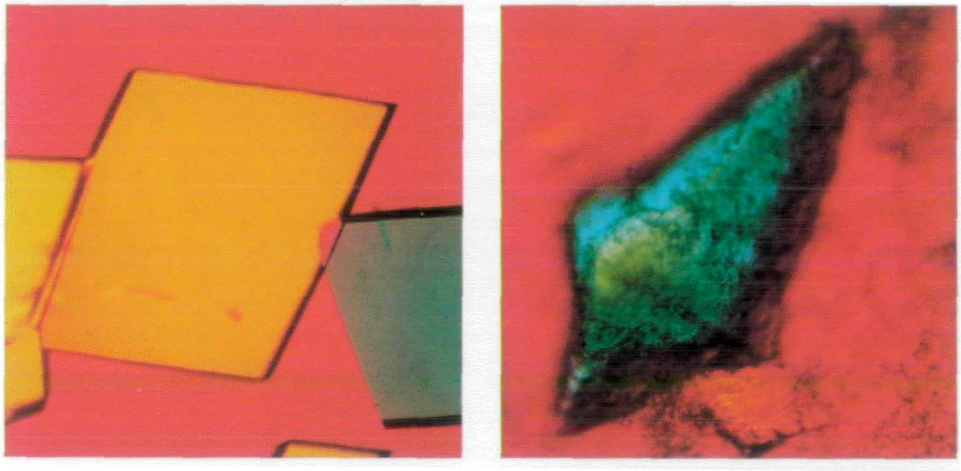

Fig. 1. **Form 1 (left) and form 2 (right) crystals of truncated EG.** The approximate size of the form 1 crystal is $0.2\times0.2\times0.05$ mm^3 and that of the form 2 one is $0.15\times0.1\times0.1$ mm^3.

```
                                                    +
Tru EG  (228-272)  ---------------AVKSPSEAGALQLVELNGQLTLAGEDGTPVQLRGMSTHGLQWFGE
                                                   | |                       | |
CelCCA  (  1- 59)  MYDASLIPNLQIPQKNIPNNDGMNFVKGLRLGWNLGNTFDAFNGTNITNELDYETSWSG-
                                          H0      S1
                                    *
Tru EG  (273-328)  IVNENAFVALSNDWGSNMIRLAMYI----GENGYATNPEVKDLVYEGIELAFEHDMYVIV
                   |          | | |          |          | |          ||||
CelCCA  ( 60-119)  IKTTKQMIDAIKQKGFNTVRIPVSWHPHVSGSDYKISDVWMNRVQEVVNYCIDNKMYVIL
                         H1          S2                        H2          S3
                    *                                                  **
Tru EG  (329-377)  DWHVH--------APGDPRADVYSGAYDFFEEIADHYKDHPKNHYIIWELANEPSP---N
                     |          |          ||          | |  |||
CelCCA  (120-177)  NTHHDVDKVKGYFPSSQYMASSKKYITSVWAQIAARFANY--DEHLIFEGMNEPRLVGHA
                                        H3                S4
                    ++
Tru EG  (378-433)  NNGGPGLTND--EKGWEAVKEYAEPIVEMLREKG--DNMILVGNPNWSQRPDLSADNPID
                   |   | |||          |   |  |          |      ||
CelCCA  (178-237)  NEWWPELTNSDVVDSINCINQLNQDFVNTVRATGGKNASRYLMCPGYVASPDGATNDYFR
                                   H4                S5        H5
                                 * *  +
Tru EG  (434-483)  AE--------NIMYSVHFYTGSHGASHIGYPEGTPSS--ERSNVMANVRYALDNGVAVFA
                           |  |||  |    |          ||       |     ||
CelCCA  (238-297)  MPNDISGNNNKIIVSVHAYCPWNFAGLAMADGGTNAWNINDSKDQSEVTWFMDNIYNKYT
                                S6                              H6
                             *                                 +
Tru EG  (484-538)  TEWGTSQANGDGGP-----YFDEADVWLNFLNKHNISWANWSLTNKNEISGAFTPFELGR
                      |     |  |                              |    |  ||     |
CelCCA  (298-355)  S-RGIPVIIGECGAVDKNNLKTRVEYMSYYVAQAKARGILCILWDNNNFSGTGELFGFF-
                           S7              H7          S8

Tru EG  (539-584)  TDATDLDPGANQVWAPEELSLSGEYVRARIKGIEYTPIDRTKFTKL
                          |    |  ||       |    |
CelCCA  (356-380)  ------DRRSCQFKFPEIIDGMVKYAFGLIN---------------
                                 H8
```

Fig. 2. **Amino acid sequence comparison between the truncated EG (TruEG in the figure) and the catalytic domain of CelCCA.** Conserved residues are indicated by vertical bars between the sequences. Residues that comprise the catalytic center of CelCCA are indicated by asterisks (*) above the sequences. Residues denoted by pluses (+) are aromatic amino acids in CelCCA and may comprise substrate binding sites, but are replaced by smaller amino acids in the truncated EG. Secondary structure elements are underlined, and labeled H0-H8 for helices and S1-S8 for strands from the N- to the C-terminal.

to be 2 or 3, the V_m value becomes 3.10 or 2.07 Å^3/Da, respectively. Since both these values are within the range generally observed for protein crystals (*12*), the number is not specified between 2 and 3.

Two oscillation data sets were collected for a form 2 crystal ($0.15 \times 0.10 \times 0.08$ mm^3) around different rotation axes using a DIP100 imaging plate detector (MAC Science). Cu K_α radiation from a Rigaku RU-300 X-ray generator operating at 45 kV × 80 mA (focus size, 0.3 × 3 mm) with a focusing double-mirror monochromater was used. For both crystal settings, the oscillation angle for one image was 3° and the total oscillation angle was 60°, with 1° overlapping of images. The X-ray exposure time was 10 min for each image at an oscillation speed of 15°/min. The crystal to detector distance was 150 mm. The intensities of the collected reflections were merged into one data set using the programs, DENZO and SCALEPACK (*13*). Although form 2 crystals diffracted up to 3.0 Å, the reflections up to 3.3 Å resolution were digitized and processed. Statistical analysis of the data is shown in Table I. The rather high R_{merg} value may be attributed to the small crystal size.

Recently, the crystal structure of the catalytic domain of endoglucanase CelCCA, which belongs to the same family as alkaline EG (family A), was determined at 1.6 Å resolution (*3*). However, several attempts to determine the form 2 crystal structure using the molecular replacement method have failed. The CelCCA crystal structure (BNL code, 1EDG) was used as a poly-Ala/Gly search model in which inserted/deleted loops have been omitted, and as a similar model in which conserved side chains have been inherited using the alignment shown in Fig. 2 as a primer. The failure of the molecular replacement was primarily due to the low amino acid sequence identity (17.8%) between the truncated EG and the catalytic domain of CelCCA.

In addition, comparison of the two amino acid sequences indicated a large difference between the alkaline EG and CelCCA in the structure of the active center. Although the catalytic sites are conserved in these two enzymes (except for Glu307 of CelCCA, which is replaced by Asp494 in alkaline EG), five (Phe42, Trp180, Trp181, Trp259, and Trp340, according to the numbering system for 1EDG) of the eight aromatic side chains that lie in the substrate binding cleft of CelCCA are replaced by smaller side chains in alkaline EG (Gly254, Gly380, Gly381, Ser447, and Thr522, respectively) (Fig. 2). Thus, the possibility exists that alkaline EG differs in the substrate binding mode from CelCCA, and thus belongs to a distinct structure class in family A. Elucidation of how these substitutions alter the active center structure and how they are related to the shift in the optimum pH of the alkaline EG (9.5) from that of CelCCA (6.0) is of particular interest. Future research will focus on a search for heavy-atom derivatives of form 2 crystals in order to determine the crystal structure using the isomorphous replacement method.

REFERENCES

1. Henrissat, B. and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781-788
2. Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C., and Warren, R.A.J. (1991) Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiol. Rev.* **55**, 303-315
3. Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H.-P., Belaich, J.-P., Davies, G.J., and Haser, R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* **3**, 939-949
4. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K.C., and Jones, T.A. (1990) Three dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* **249**, 380-386
5. Spezio, M., Wilson, D.B., and Karplus, P.A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase. *Biochemistry* **32**, 9906-9916
6. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J.K.C., Teeri, T.T., and Jones, T.A. (1994) The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524-528
7. Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeyssens, M., Béguin, P., and Aubert, J.-P. (1992) Three-dimensional structure of a thermostable bacterial cellulase. *Nature* **357**, 89-91
8. Davies, G.J., Dodson, G., Moore, M.H., Tolley, S.P., Dauter, Z., Wilson, K.S., Rasmussen, G., and Schülein, M. (1996) Structure determination and refinement of the *Humicola insolens* endoglucanase V at 1.5 Å resolution. *Acta Cryst.* **D52**, 7-17
9. Yoshimatsu, T., Ozaki, K., Shikata, S., Ohta, Y., Koike, K., Kawai, S., and Ito, S. (1990) Purification and characterization of alkaline endo-1,4-β-glucanases from alkalophilic *Bacillus* sp. KSM-635. *J. Gen. Microbiol.* **136**, 1973-1979
10. Ozaki, K., Shikata, S., Kawai, S., Ito, S., and Okamoto, K. (1990) Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from *Bacillus* sp. KSM-635. *J. Gen. Microbiol.* **136**, 1327-1334
11. Ozaki, K., Hayashi, Y., Sumitomo, N., Kawai, S., and Ito, S. (1995) Construction, purification, and properties of a truncated alkaline endoglucanase from *Bacillus* sp. KSM-635. *Biosci. Biotech. Biochem.* **59**, 1613-1618
12. Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497
13. Otwinowski, Z. (1993) Oscillation data reduction program in *Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaacs, N., and Bailey, S., eds.) pp. 56-62, SERC Daresbury Laboratory, England