Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification, crystallization and preliminary X-ray crystallographic studies of a psychrophilic cellulase from *Pseudoalteromonas haloplanktis*

Sébastien Violot,[a] Richard Haser,[a]* Guillaume Sonan,[b] Daphné Georlette,[b] Georges Feller[b] and Nushin Aghajari[a]

[a]Laboratoire de BioCristallographie, Institut de Biologie et Chimie des Protéines, CNRS et Université Claude Bernard Lyon 1, UMR 5086, F-69367 Lyon CEDEX 07, France, and [b]Laboratoire de Biochimie, Institut de Chimie B6, Université de Liège, B-4000 Liège Sart-Tilman, Belgium

Correspondence e-mail: r.haser@ibcp.fr

Received 12 February 2003
Accepted 17 April 2003

The Antarctic psychrophile *Pseudoalteromonas haloplanktis* produces a cold-active cellulase. To date, a three-dimensional structure of a psychrophilic cellulase has been lacking. Crystallographic studies of this cold-adapted enzyme have therefore been initiated in order to contribute to the understanding of the molecular basis of the cold adaptation and the high catalytic efficiency of the enzyme at low and moderate temperatures. The catalytic core domain of the psychrophilic cellulase CelG from *P. haloplanktis* has been expressed, purified and crystallized and a complete diffraction data set to 1.8 Å has been collected. The space group was found to be $P2_12_12_1$, with unit-cell parameters a = 135.1, b = 78.4, c = 44.1 Å. A molecular-replacement solution, using the structure of the mesophilic counterpart Cel5A from *Erwinia chrysanthemi* as a search model, has been found.

1. Introduction

Life on earth displays a wide capacity for adaptation. Physical limits consistent with biology range from 233 to 388 K in temperature (in the stratosphere and in hydrothermal vents, respectively), up to 120 MPa in pressure (hydrostatic pressures in the deep sea), from pH 1 to 11 and up to 4 *M* in salt concentration (Jaenicke & Böhm, 1998).

Organisms growing under such conditions have been classified as thermophiles, psychrophiles, barophiles (or piezophiles), acidophiles, alkalophiles and halophiles. These organisms present adaptations to high temperatures (>328 K), cold temperatures (around 273 K), high pressures, acidic or alkaline conditions and high ionic strength, respectively.

'Cold enzymes' from psychrophilic microorganisms are generally characterized (i) by having a higher catalytic activity and catalytic efficiency than their mesophilic counterparts in the temperature range 273–303 K, (ii) by a limited thermostability owing to denaturation at moderate and high temperatures and (iii) by an activity curve displaced towards low temperatures compared with mesophilic counterparts (Feller *et al.*, 1996).

Cellulases catalyze the hydrolysis of cellulose, an unbranched homopolymer of β-1,4-linked glucose, which is the major polysaccharidic component of plant biomass.

Cellulases have been classified according to their activity on the substrate into endocellulases (EC 3.2.1.4) and exocellulases (EC 3.2.1.91), which attack the cellulose chain randomly or at the non-reducing extremity, respectively.

Pseudoalteromonas haloplanktis, a psychrophilic Gram-negative bacterium collected in Antarctic seawater, produces the endoglucanase CelG, the gene sequence (Genbank accession No. Y17552) of which shows that the mature CelG protein is made up of 494 residues and comprises a catalytic domain, a proline/serine/threonine-rich linker and a carbohydrate-binding module. Sequence alignments clearly show that the catalytic domain of CelG belongs to family 5 of the glycoside hydrolases from the 90 known families which have so far been classified (Henrissat & Bairoch, 1993; Carbohydrate-Active Enzymes server, http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html). In family 5, which contains more than 170 glycoside hydrolases, only eight residues are found to be strictly conserved (Wang *et al.*, 1993). Alignment of the primary structures within this family revealed that the cellulase Cel5A from *Erwinia chrysanthemi* (Py *et al.*, 1991) is a mesophilic homologue of the psychrophilic CelG, with which it displays 64% sequence identity.

Determination of the three-dimensional structure of this psychrophilic cellulase will allow detailed analyses and comparative studies with the three-dimensional structure of the mesophilic cellulase (Chapon *et al.*, 2001) in order to obtain insights into protein adaptation to temperature on the molecular level.

2. Materials and methods

2.1. Construction and expression of the recombinant gene

Construction of the *celG* gene lacking the coding sequence for the linker and the

cellulose-binding domain was carried out in two main steps. In order to insert the gene in the expression vector pET22b (Novagen), the nucleotide sequence including the start codon of the wild-type gene (accession No. Y17552) was modified by PCR using Vent DNA polymerase to CATATG, creating an *Nde*I restriction site. In the second step, the catalytic domain was amplified by reverse PCR using a mutating antisense primer which introduces the stop codon TGA at nucleotide 975. The silent sense primer corresponded to the polylinker sequence of the plasmid, therefore allowing the removal of the linker and CBD coding region after amplification and circularization of the product. The nucleotide sequence of the truncated gene was checked on an Amersham Biosciences ALF DNA sequencer.

Escherichia coli Epicurian BL21 (DE3) cells (Stratagene) carrying the recombinant plasmid were grown in LB–ampicillin medium at 291 K. Expression was induced at an OD_{550nm} of ∼3 by 0.1 m*M* isopropyl thio-β-D-galactoside (IPTG) and the culture was grown for an additional 7 h.

2.2. Purification of the catalytic domain of CelG endoglucanase

After centrifugation of the culture at 277 K, periplasmic proteins were extracted by osmotic shock of the cells (Ausubel *et al.*, 1989) using 100 m*M* Tris–HCl, 0.5 m*M* EDTA, 0.5 *M* saccharose, 0.1 m*M* PMSF (phenylmethylsulfonyl fluoride) pH 8.0 as the hypertonic buffer and 1 m*M* $MgCl_2$, 1 m*M* PMSF as the hypotonic solution. When present, nucleic acids were removed by precipitation after overnight stirring in the presence of 0.1%(*w*/*v*) protamine sulfate. Ammonium sulfate (75% saturation) was added to the supernatant and the resulting protein precipitate, collected by centrifugation, was dissolved in a minimal volume of 25 m*M* PIPES pH 6.5. This sample was loaded onto a phenyl-Sepharose CL-4B (Pharmacia) column (20 × 3 cm) equilibrated in 25 m*M* PIPES, 25% saturation $(NH_4)_2SO_4$ pH 6.5 (buffer *A*) and the column was then washed with 2 l buffer *A*. The elution was performed with a decreasing $(NH_4)_2SO_4$ gradient in buffer *A*. The active fractions were pooled and buffer exchange was carried out by ultrafiltration on an Amicon concentrator fitted with a PTGC membrane with sequential addition of 5 volumes of 10 m*M* HEPES pH 7.5. The sample was loaded onto a Macro-prep high Q (Biorad) column (20 × 3 cm) and elution was carried out with a linear gradient from 0 to 0.35 *M* KCl in 10 m*M* HEPES pH 7.5. In the final step, the active fractions were buffer-exchanged with 10 m*M* HEPES pH 7.5 and loaded onto an FPLC system (Pharmacia) equipped with a Mono-Q HR 5/5 column. The enzyme was eluted with a linear gradient from 0 to 0.35 *M* KCl in 10 m*M* HEPES pH 7.5. The purified catalytic domain was buffer-exchanged with 10 m*M* HEPES, 0.04% NaN_3 pH 7.5, concentrated to 18 mg ml^{-1} and stored at 203 K.

Table 1
X-ray diffraction data.

Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 135.1, b = 78.4, c = 44.1
Resolution (Å)	25–1.8
Measured reflections	102004
Unique reflections	42621
Redundancy	2.4 (2.1)
Completeness (%)	94 (94)
$I/\sigma(I)$	8.2 (3.7)
R_{sym}† (%)	5.8 (19)

† $R_{sym} = \sum_{hkl}\sum_i |I(hkl)_i - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I(hkl)$, where $I(hkl)$ are the intensities of symmetry-redundant reflections and $\langle I(hkl)\rangle$ is the average intensity over all observations.

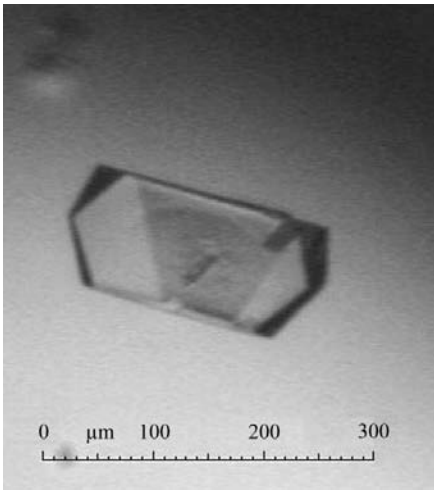

Figure 1
Crystal of the catalytic domain of *P. haloplanktis* CelG.

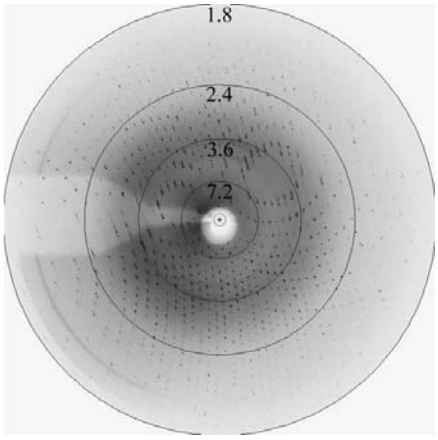

Figure 2
A 1° oscillation image collected in-house at 100 K on a native CelG crystal.

2.3. Crystallization

Initial crystallization trials were performed using the sparse-matrix sampling method. The screening was conducted using the hanging-drop vapour-diffusion technique in 24-well Linbro plates, employing Crystal Screen kits I and II (Hampton Research, Laguna, CA, USA; Jancarik & Kim, 1991). 4 µl droplets were equilibrated against 500 µl of reservoir solution at 277 and at 292 K.

Crystals suitable for X-ray diffraction studies (Fig. 1) were obtained in 1.3 *M* trisodium citrate dihydrate, 10%(*v*/*v*) glycerol and 0.1 *M* HEPES buffer pH 7.5 at 292 K. The protein-to-mother liquor ratio was 2:1 in 4 µl drops and the initial protein concentration was 9 mg ml^{-1}. Crystals grew within two weeks to dimensions of 0.4 × 0.2 × 0.2 mm.

2.4. X-ray data collection and processing

Diffraction data were collected on a 345 mm MAR Research image-plate system and the X-ray radiation used was Cu $K\alpha$ radiation from a Nonius FR 591 rotating-anode generator operated at 44 kV and 90 mA and equipped with Osmic confocal mirrors. The crystal was flash-frozen in supercooled N_2 gas produced by an Oxford Cryosystems Cryostream (600 series) and maintained at 100 K during the data collection.

The crystal-to-detector distance was 140 mm, the oscillation range per image was 1°, the total oscillation angle was 60° and the exposure time per image was 10 min (Fig. 2).

Determinations of unit-cell parameters and the integration of reflections were performed with the program *MOSFLM* (Leslie, 1991), whereas scaling was performed with the program *SCALA* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The diffraction data for the crystal of the native enzyme, which were 94% complete to 1.8 Å resolution, display good statistics (see Table 1).

3. Results

The nucleotide sequence of the truncated *celG* gene encodes a polypeptide of 325 amino acids including the signal peptide and

the catalytic domain. N-terminal amino-acid sequencing of the purified gene product shows that the signal peptide (32 amino acids) has been correctly processed in *E. coli*. Accordingly, the isolated catalytic domain contains 293 amino acids, with a predicted mass of 31 890 Da. This truncated protein retains the catalytic properties of the full-length protein, since its specific activity towards carboxymethylcellulose and *p*-nitrophenyl β-D-cellobioside is unmodified.

Crystals of this catalytic domain were obtained and diffraction data were collected. The space group of the crystals was unambiguously determined to be $P2_12_12_1$ owing to systematic extinctions along the three twofold axes. The refined unit-cell parameters are $a = 135.1$, $b = 78.4$, $c = 44.1$ Å. Assuming a molecular weight of 31 890 Da, this gives a solvent content of 67 or 33% and a volume-to-mass ratio, V_M, of 3.7 or 1.8 Å^3 Da^{-1} for one or two molecules in the asymmetric unit, respectively (Matthews, 1968). The 1.8 Å resolution data (Table 1) were used for molecular replacement with the program *AMoRe* (Navaza, 1994). The refined coordinates of the cellulase Cel5A from *E. chrysanthemi* (PDB code 1egz), with which CelG displays 64% sequence identity (Chapon *et al.*, 2001), were used as a search model. Diffraction data in the resolution range 15–2.8 Å were used throughout the search. After the rotation-function search, two peaks with correlation factors of 14.8 and 14.7%, and *R* factors of 56.3 and 56.4%, respectively, were found (the next solution gave a correlation coefficient of 10% and an *R* factor of 57.6%). A correctly positioned molecule of Cel5A corresponding to the first solution after the translation-function search (correlation coefficient of 40% and an *R* factor of 51.2%, compared with the next solution which had a correlation factor of 29.8% and an *R* factor of 55.5%) was used to locate the second molecule. The correlation coefficient subsequently increased to 53.7%. The two Cel5A molecules were then subjected to rigid-body refinement, resulting in a correlation coefficient of 56.4% and an *R* factor of 41.2% for the two molecules in the asymmetric unit.

Refinement of the core domain of CelG is in progress.

This work was partly financially supported by EU Contract No. BI04-CT97-0131; support from the CNRS (Centre National de la Recherche Scientifique) is also gratefully acknowledged.

References

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1989). Editors. *Current Protocols in Molecular Biology*. New York: John Wiley & Sons.

Chapon, V., Czjzek, M., El Hassouni, M., Py, B., Juy, M. & Barras, F. (2001). *J. Mol. Biol.* **310**, 1055–1066.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Feller, G., Narinx, E., Arpigny, J. L., Aittaleb, M., Baise, E., Genicot, S. & Gerday, C. (1996). *FEMS Microbiol. Rev.* **18**, 189–202.

Henrissat, B. & Bairoch, A. (1993). *Biochem. J.* **293**, 781–788.

Jaenicke, R. & Böhm, G. (1998). *Curr. Opin. Struct. Biol.* **8**, 738–748.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Leslie, A. G. W. (1991). *Crystallographic Computing 5. From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J.-C. Thierry, pp. 50–61. IUCR/Oxford University Press.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Py, B., Bortoli-German, I., Haiech, J., Chippaux, M. & Barras, F. (1991). *Protein Eng.* **4**, 325–333.

Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A. J., Aebersold, R. & Withers, S. G. (1993). *J. Biol. Chem.* **268**, 14096–14102.