Appl Microbiol Biotechnol (2001) 57:660–666
DOI 10.1007/s00253-001-0832-4

ORIGINAL PAPER

T. Arai · H. Ohara · S. Karita · T. Kimura · K. Sakka
K. Ohmiya

Sequence of *celQ* and properties of CelQ, a component of the *Clostridium thermocellum* cellulosome

Received: 4 May 2001 / Received revision: 5 July 2001 / Accepted: 3 September 2001 / Published online: 1 November 2001

Abstract The nucleotide sequence of the *Clostridium thermocellum* F1 *celQ* gene, which codes for the endoglucanase CelQ, consists of 2,130 bp encoding 710 amino acids. The precursor form of CelQ has a molecular weight of 79,809 and is composed of a signal peptide, a family 9 cellulase domain, a family IIIc carbohydrate-binding module (CBM), and a dockerin domain. Truncated derivatives of CelQ were constructed: CelQΔdoc consisted of the catalytic domain and the CBM; CelQcat consisted of the catalytic domain only. CelQΔdoc showed strong activity toward carboxymethylcellulose (CMC) and barley β-glucan and low activity toward Avicel, acid-swollen cellulose, lichenan, and xylan. The V_{max} and K_m values were 235 μmol/min/mg and 3.3 mg/ml, respectively, for CMC. By contrast, CelQcat, which was devoid of the CBM, showed negligible activity toward CMC, i.e., about 1/1,000 of the activity of CelQΔdoc, supporting the previously proposed idea that family IIIc CBMs participate in the catalytic function of the enzyme. Immunological analysis using an antiserum raised against CelQΔdoc confirmed that CelQ is a component of the *C. thermocellum* cellulosome.

Introduction

Most cellulases consist of two or more discrete domains that function cooperatively in a polypeptide chain. The most common arrangement is a catalytic domain connected to a carbohydrate-binding module (CBM) via a linker sequence, although some cellulases contain additional functional domain(s) other than a CBM (Ohmiya et al. 1997; Tomme et al. 1995). Catalytic domains of cellulase can be classified into 12 groups in glycoside hydrolase families on the basis of amino acid sequence similarities (Henrissat and Bairoch 1996; http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html). CBMs, most of which bind preferentially to cellulose, are also classified into 26 families on the basis of amino acid sequence similarities (http://afmb.cnrs-mrs.fr/~pedro/CAZY/cbm.html). It is believed that a CBM in a cellulase molecule enhances the hydrolytic activity of a catalytic domain adjacent to the CBM by increasing the enzyme concentration on the surface of an insoluble substrate or by supplying the catalytic domain with a more easily degradable substrate, e.g., amorphous cellulose (Din et al. 1991). Among CBMs hitherto characterized, the family IIIc CBMs are unusual because they do not contribute to the primary binding to crystalline cellulose but they may participate in the catalytic function of the enzyme (Gal et al. 1997; Irwin et al. 1998).

Clostridium thermocellum secretes a high-molecular-mass cellulase complex, termed the cellulosome, composed of various catalytic components such as cellulases and hemicellulases, and at least one noncatalytic scaffolding protein variously known as SL, S1, or CipA (Béguin and Alzari 1998; Shoham et al. 1999). Catalytic components and CipA are assembled into the cellulosome by the interaction between a dockerin domain of the former and one of the cohesin domains of the latter. Thus far, many cellulase genes have been cloned, sequenced, and expressed in *Escherichia coli* in addition to several genes encoding hemicellulases and noncatalytic structural proteins, e.g., endoglucanase genes *celA*, *celB*, *celD*, *celE*, *celF*, *celG*, *celH*, and *celJ*; exoglucanase genes, *cbhA*, *celK*, and *celS* (see Shoham et al. 1999, for references). Reconstitution of a highly active cellulosome from a combination of the recombinant CipA and some recombinant catalytic components has not been reported, although an interaction and a moderate synergistic action were observed between them (Ciruela et al. 1998; Fukumura et al. 1997; Kataeva et al. 1997). These observations suggest that the cooperative action of sever-

T. Arai · H. Ohara · S. Karita · T. Kimura · K. Sakka (✉)
K. Ohmiya
Faculty of Bioresources, Mie University, 1515 Kamihamacho, Tsu 514-8507, Japan
e-mail: sakka@bio.mie-u.ac.jp
Tel.: +81-59-2319621, Fax: +81-59-2319684

Present address:
H. Ohara, Meiji Seika Kaisha, Ltd., Chiyoda, Sakado, Saitama 350–0289, Japan

al catalytic components including unidentified ones is necessary for the efficient degradation of crystalline cellulose.

We cloned eight endoglucanase genes, three xylanase genes, and a β-glucosidase gene from *C. thermocellum* F1 (Sakka et al. 1989), and we found that one of them, *celJ*, encoded the largest and major catalytic component of the cellulosome (Ahsan et al. 1996, 1997). Although cellulase genes are known to be scattered on the *C. thermocellum* chromosome DNA (Guglielmi and Béguin 1998), another cellulase gene, termed *celQ*, was found downstream of *celJ*.

Here, we report the cloning and sequencing of the full-length *celQ* gene identified downstream of *celJ*. We also deal with the characterization of CelQ, containing a family IIIc CBM, and its derivative devoid of the CBM.

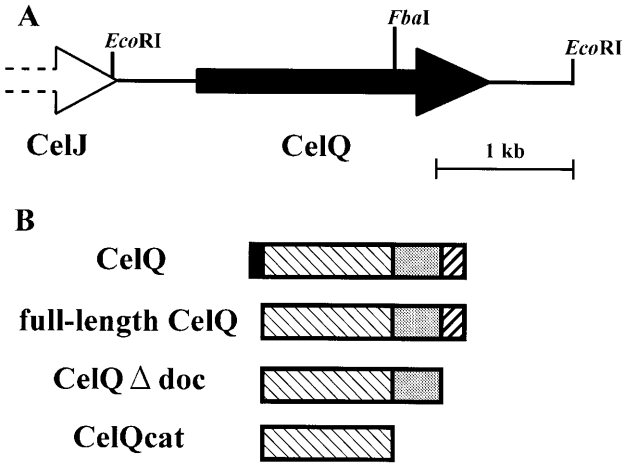

Fig. 1 Restriction enzyme sites around the *celQ* gene (**A**) and domain organization of CelQ and its derivatives expressed in *E. coli* (**B**). Patterns showing distinct domains are the same as in Fig. 3

Materials and methods

Bacterial strains and plasmids

Genomic DNA and the cellulosome fraction of *Clostridium thermocellum* strain F1 were isolated by methods described previously (Sakka et al. 1994). Plasmids pCR2.1 (Invitrogen) and pBluescript II KS(+) and KS(–) (Stratagene) were used as the cloning vectors. The *E. coli* strains XL2-Blue MRF′ and JM109 were purchased from Stratagene. *E. coli* M15 and BL21(DE3) served as the hosts for derivatives of plasmid pQE-30T (Ali et al. 2001) and pET-28a (Novagen), respectively. Both plasmids provide proteins containing a six-His tag at their N-termini.

Cloning of the full-length *celQ* gene

The unidentified region of the *celQ* gene found downstream of *celJ*, was amplified by inverse PCR (Fig. 1). *C. thermocellum* F1 genomic DNA was digested with *Eco*RI, rendered in the circular form by self-ligation, and used for PCR as the template DNA with KOD Dash DNA polymerase (Toyobo, Osaka, Japan) and two primers complementary to the nucleotide sequence downstream of *celJ*, PreCelQ1, 5′-GCTTCCCTCCAAATAGGAATCCG-3′, and PreCelQ2, 5′-ATTTTGAAGGAGGAATCATGGTG-3′. The amplified DNA fragments were ligated into pCR2.1 using the Takara Ligation Kit (Kyoto, Japan), yielding pCR-CelQ. The nucleotide sequence of the cloned DNA fragment was determined. Reliability of the nucleotide sequence of *celQ* was confirmed by sequencing four PCR products independently amplified.

DNA sequencing

Nucleotide sequence was determined on a Licor (Lincoln, Neb.) model 4000L automated DNA sequencer, with appropriate dye primers and a series of subclones. The nucleotide sequence data were analyzed with GENETYX computer software (Software Development, Tokyo). Homology searches in GenBank and DDBJ were carried out with a BLAST program.

Construction of pCelQ, pCelQΔdoc, and pCelQcat

To amplify the DNA region encoding the full-length CelQ, a combination of two synthetic oligonucleotide primers, CelQ-Few containing an artificial *Bgl*II recognition sequence, 5′-TTTAGATCTGCAGGAAGCTATAACTATGCGG-3′, and CelQ-R-full containing an artificial *Sal*I recognition sequence, 5′-TTTGTCGACCAAATGCCAAACTATTCTACCG-3′, were used for PCR with pCR-CelQ as a template. The amplified DNA fragment was digested with *Bgl*II and *Sal*I, and cloned between the *Bam*HI and *Sal*I sites of pQE-30T to generate pCelQ. This plasmid provides the full-length CelQ. The DNA region encoding the catalytic domain and the CBM of CelQ was amplified with a combination of two synthetic oligonucleotide primers, CelQ-Few and CelQ-R-doc, containing an artificial stop codon sequence and a *Sal*I recognition sequence, 5′-TTTGTCGACTATTCAAAGCCGCCTGGAACTTCTCC-3′. The amplified DNA fragment was digested with *Bgl*II and *Sal*I, and cloned between the *Bam*HI and *Sal*I sites of pQE-30T to generate pCelQΔdoc. We confirmed that the plasmid pCelQΔdoc did not contain any mutations leading to amino acid displacement. This plasmid provides a polypeptide consisting of the catalytic domain and CBM (CelQΔdoc). The stop-codon linker (Nippon Gene, Tokyo, Japan), 5′-CTAACTAATTAGCTAGCTAATTAGTTAG-3′, was inserted into the *Fba*I site between the regions encoding the catalyst domain and the CBM in pCelQΔdoc, yielding pCelQcat. This plasmid encodes the family 9 catalytic domain devoid of the CBM (CelQcat).

Purification of CelQΔdoc and CelQcat

CelQΔdoc and CelQcat were purified from the recombinant *E. coli* clones harboring pCelQΔdoc or pCelQcat. The *E. coli* clones were grown at 37 °C in 3 l of L medium (1% polypeptone, 0.5% yeast extract, 0.5% NaCl, pH 7.0) supplemented with ampicillin (50 μg/ml) or kanamycin (25 μg/ml). The production of the proteins was facilitated by the addition of isopropyl-β-D-thiogalactopyranoside. Cells were disrupted by sonication and the cell-free extracts were used for purification of the proteins with Ni-NTA resin (Amersham Pharmacia Biotech) and RESOURCE PHE (Amersham Pharmacia Biotech) column chromatography. The six-His tag was removed from the recombinant proteins by thrombin digestion after the Ni-NTA resin column chromatography.

Enzyme assays

Cellulase activity was measured by a 10-min incubation at 60 °C in 20 mM sodium phosphate buffer (pH 6.3) or Britton and Robinson's universal buffer (50 mM phosphoric acid-50 mM boric acid-50 mM acetic acid; the pH was adjusted from 2 to 12 with 1 N NaOH) in the presence of 1.5% CMC (low viscosity; Sigma). Reducing sugars released from the substrate were determined with the 3, 5-dinitrosalicylic acid reagent as described by Miller et al. (1959). One unit of activity was defined as the amount of enzyme

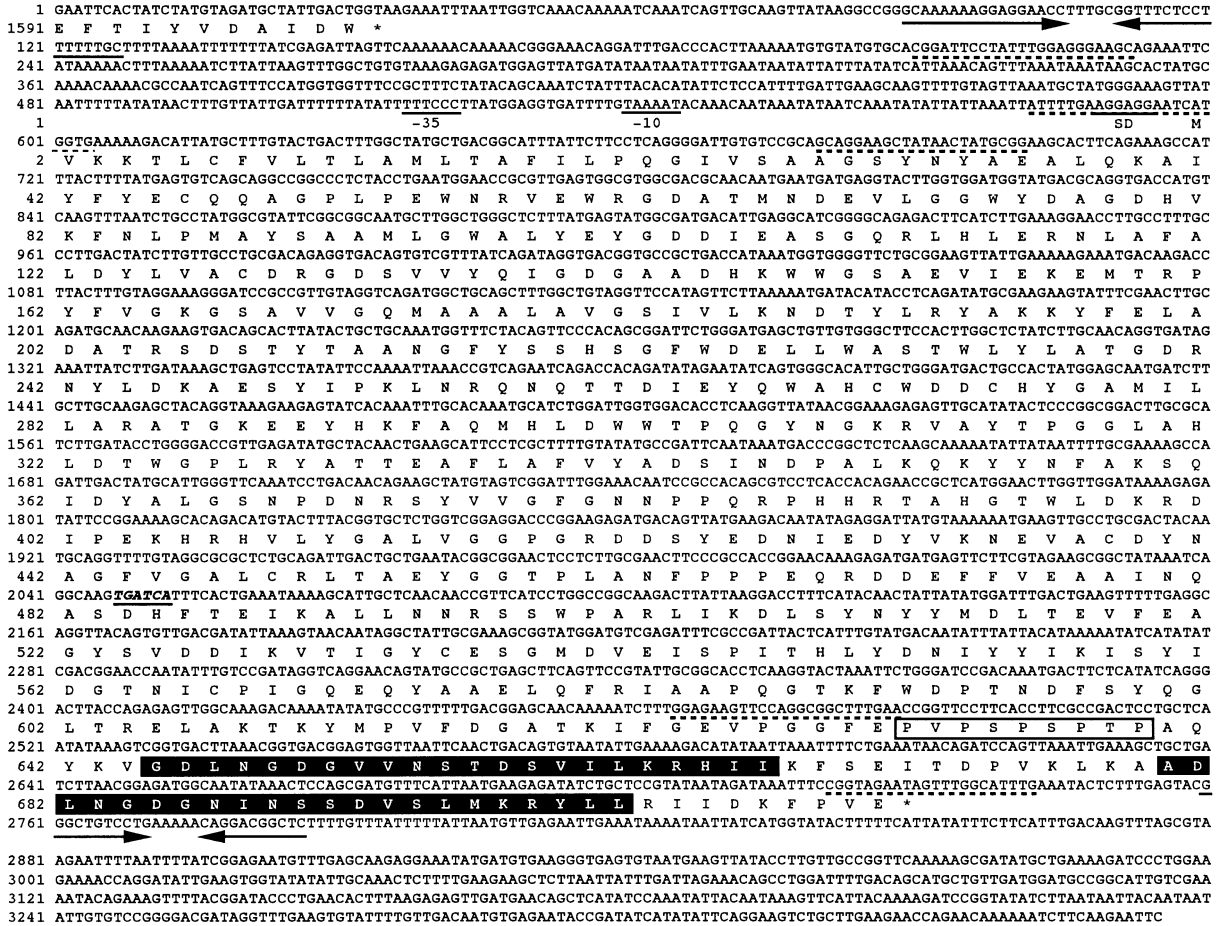

Fig. 2 Nucleotide sequence of the *celQ* gene and deduced amino acid sequence of the gene product. The –35 and –10 regions of a possible promoter sequence and the possible Shine-Dalgarno (SD) sequence are *underlined*. The *Fab*I site, in which the stop codon linker was inserted, is shown in *italics* with an *underline*. Palindromes are indicated by *arrows* facing each other. Regions corresponding to PCR primers used for amplification of the full-length *celQ* and the truncated derivative are shown with *broken lines*. Reiterated sequences of 22 amino acids of the dockerin domain are *highlighted*. A linker sequence rich in Pro between a family IIIc CBM and a dockerin domain is *boxed*. The sequence upstream of *celQ* contains the C-terminal region of *celJ*

releasing 1 µmol of glucose equivalent per min from CMC. Enzyme activities toward Avicel (Merck, Germany), acid-swollen cellulose (ASC) prepared in this laboratory (Lamed et al. 1985), ball-milled cellulose (BMC), lichenan (Sigma), laminarin (Nacalai Tesque, Japan), and oat spelt xylan (Fluka, Switzerland) were assayed as described above except that CMC was replaced with each substrate.

Analysis of hydrolysis products

Cellooligosaccharides (cellobiose to cellohexaose, each 5 µg) were incubated with 0.1 unit of the purified enzyme in 1 ml of 50 mM sodium succinate buffer (pH 5.5) at 60 °C. Thin-layer chromatography (TLC) of the hydrolysis products was performed on a DC-Fertigplatten SIL G-25 plate (Macherey-Nagel, Dorne, Germany) developed with a solvent of 1-propanol:water (85:15, v/v), and cellooligosaccharides were visualized by spraying the plate with an aniline-diphenylamine reagent (Gasparic and Churacek 1978).

Other procedures

Protein concentrations were determined by the method of Lowry et al. (1951) with bovine serum albumin as the standard. SDS-PAGE was done by the method of Laemmli (1970). Zymogram analysis was performed as described by Ali et al. (1995) using a 10% SDS-PAGE gel containing 0.1% CMC. A cellulosome fraction was prepared from the *C. thermocellum* F1 culture supernatant as described previously (Morag et al. 1992). Preparation of antiserum and immunoblotting were carried out as described previously (Hayashi et al. 1997). Antiserum was raised against the purified CelQΔdoc in a BALB/c mouse.

Nucleotide sequence accession number

The nucleotide sequence reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases with accession number AB047845.

Results

Cloning and nucleotide sequence of the full-length *celQ* gene

The full-length *celQ* gene was amplified by inverse PCR and cloned in pCR2.1. Figure 2 shows the *celQ* structural gene along with its flanking regions. There is an open reading frame composed of 2,130 nucleotides encoding a protein of 710 amino acids with a predicted molecular

weight of 79,809. The assigned ATG initiation codon at nucleotide position 599 is preceded by a putative Shine-Dalgarno (SD) sequence, GGAGG, a typical ribosome binding site in *C. thermocellum* (Wang et al. 1993). The reading frame is ended by the stop codon TGA at position 2,728. A possible promoter sequence, TTTCCC for the –35 region and TAAAAT for the –10 region with a 17-bp spacing between them, was observed. These sequences show high homologies to the consensus promoter sequences for σ^{70} factor found in *E. coli*, i.e., TTGACA and TATAAT with a 17-bp spacing (Rosenberg et al. 1979). A possible transcription terminator that consists of a 27-bp palindromic sequence, corresponding to an mRNA hairpin loop with a ΔG of –17.7 kcal/mol (Cantor et al. 1980), followed by four Ts was found downstream of the TGA termination codon. This structure is similar to the rho-factor-independent terminator of *E. coli* (Rosenberg et al. 1979).

Amino acid sequence of CelQ

The deduced N-terminal sequence of 27 amino acids contains a sequence similar to the signal peptide sequences found in prokaryotic secretory proteins (Watson 1984). Comparison of the amino acid sequence of CelQ with those registered in protein databases such as SWISS PROT and PIR clearly revealed that the mature CelQ consists of three distinct functional domains, a family 9 catalytic domain of glycoside hydrolases, a family IIIc CBM, and a dockerin domain (listed in order from the N-terminus). Figure 3 shows a schematic model of the molecular architecture of CelQ along with the related enzymes. The family 9 domain of CelQ, extending from position 32 to 451, exhibited extensive sequence homology with the catalytic domains of the other cellulases in family 9, e.g., 57% identity with CelA of *Caldicellulosiruptor* (formerly *Caldocellum*) *saccharolyticum* (Te'o et al. 1995), 57% sequence identity with CelZ of *Clostridium stercorarium* (Jauris et al. 1990), 57% identity with CelF (Navarro et al. 1991) and CelI (Hazlewood et al. 1993) of *C. thermocellum*, 53% identity with CelG of *Clostridium cellulolyticum* (Bagnara-Tardif et al. 1992), 53% identity with EngH of *Clostridium cellulovorans* (Liu and Doi 1998), and 52% identity with E4 of *Thermomonospora fusca* (Lao et al. 1991). The family IIIc CBM of CelQ, extending from position 471 to 627, exhibited 43% sequence identity with the CBM of *C. stercorarium* CelZ, 42% identity with that of *C. thermocellum* CelI, 42% identity with that of *C. saccharolyticum* CelA, 42% identity with that of *C. cellulovorans* EngH, 41% identitywith that of *T. fusca* E4, and 39% identity with that of *C. thermocellum* CelF. The third domain of CelQ, which is separated from the CBM by a short linker sequence rich in Pro, is a dockerin domain. Dockerin domains that consist of a pair of well-conserved 22-residue repeats are highly conserved in cellulases and xylanases from *C. thermocellum* and other cellulosome-forming clostridia. These proteins play a role in cellulosome assembly by docking the various catalytic subunits to a noncatalytic scaffolding protein (Pagès et al. 1997).

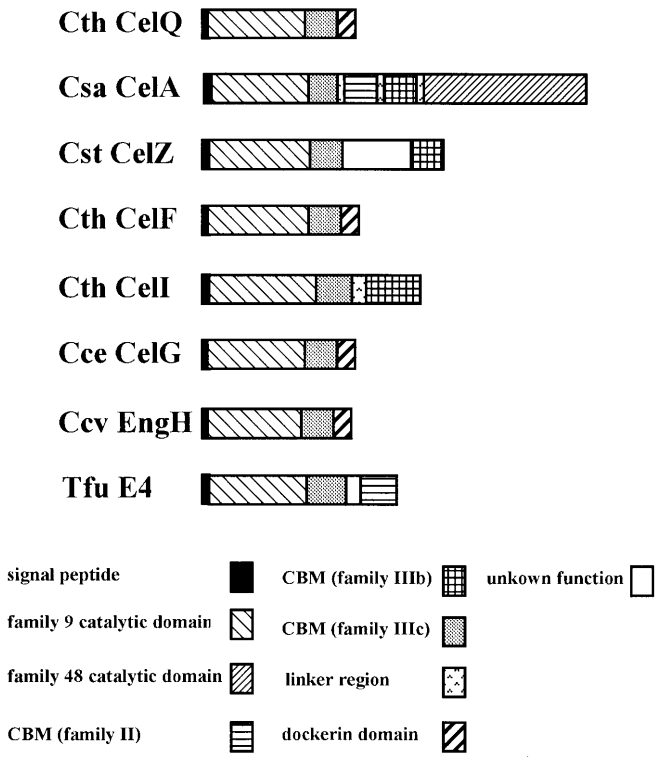

Fig. 3 Domain organization of cellulases comprising a family IIIc module adjacent to a family 9 cetaytic domain of glycoside hydrolases. Csa *C. saccharolyticus*, Cst *C. stercorarium*, Cce *C. cellulolyticum*, Ccv *C. cellulovorans*, Cth, *C. thermocellum*, Tfu, *T. fusca*

Purification and characterization of CelQΔdoc

Although we attempted to purify the full-length CelQ from *E. coli* (pCelQ), we could not obtain a sufficient amount of the purified enzyme for its characterization because of proteolysis of the enzyme at its C-terminus by a host protease(s). Therefore, we constructed pCelQΔdoc, encoding a polypeptide composed of the catalytic domain and the family IIIc CBM, and purified the truncated enzyme as described in Materials and methods. CelQΔdoc was purified six-fold from the cell-free extract of *E. coli* M15 (pCelQΔdoc) by Ni-NTA resin and RESOURCE PHE column chromatography. The final preparation gave a single band on SDS-PAGE, and the molecular weight of the enzyme was estimated to be around 67,000 (Fig. 4A), which is in good agreement with the value (68,211) estimated from the deduced amino acid sequence.

The purified enzyme had relatively high specific activity toward CMC (159 U/mg) and barley β-glucan (392 U/mg), and low activity toward several substrates, such as acid-swollen cellulose (1.4 U/mg), Avicel

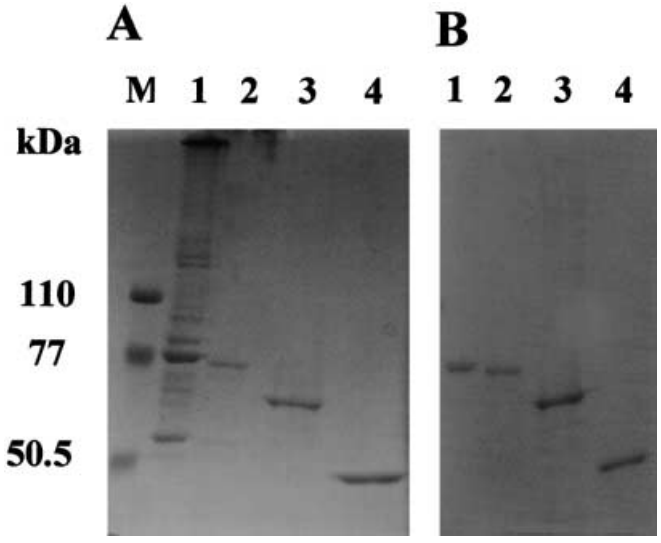

Fig. 4A, B Expression of CelQ in *C. thermocellum* F1 and *E. coli*. **A** The gel was stained with Coomassie brilliant blue. **B** CelQ proteins were analyzed by Western blot analysis using a polyclonal mouse antiserum raised against the purified CelQΔdoc. *Lane 1* Cellulosomal proteins of *C. thermocellum* F1, *lane 2* whole cell proteins of *E. coli* M15 (pCelQ), *lane 3* purified CelQΔdoc, *lane 4* purified CelQcat, *lane M* prestained SDS-PAGE standard (low range; Bio-Rad)

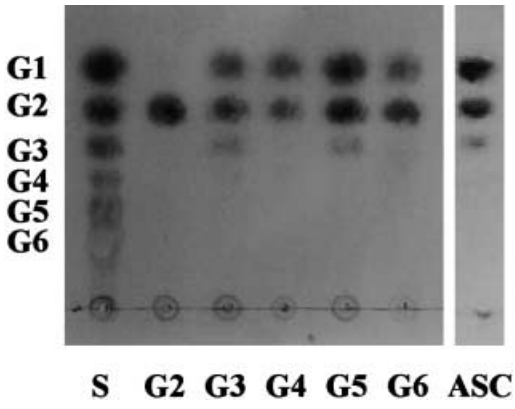

Fig. 5 TLC analysis of hydrolysis products from cellooligosaccharides. Each cellooligosaccharide (5 µg, G2–G6) was incubated with the purified enzyme (0.1 U) for 14 h, and the hydrolysates were analyzed by TLC. ASC was incubated with the enzyme for 14 h. *S* Authentic oligosaccharides, *G1* glucose, *G2* cellobiose, *G3* cellotriose, *G4* cellotetraose, *G5* cellopentaose, *G6* cellohexaose

(0.4 U/mg), lichenan (3.9 U/mg), and oat-spelt xylan (0.3 U/mg). The initial rates of the reaction were measured at 60 °C in various concentrations of CMC. From a double-reciprocal plot, the V_{max} was estimated to be 235 µmol/min/mg and K_m to be 3.3 mg/ml for CMC. The action of the enzyme on cellooligosaccharides and ASC was qualitatively analyzed by TLC (Fig. 5). When cellotetraose, cellopentaose, and cellohexaose were treated with CelQΔdoc, glucose, cellobiose, and cellotriose were produced. The enzyme showed slight activity toward cellotriose and negligible activity toward cellobiose. CelQΔdoc hydrolyzed ASC to yield mainly cellobiose and glucose, with cellotriose as a minor saccharide. The enzyme was maximally active around pH 5.5 when the enzyme activity was assayed by a 10-min incubation at 60 °C in Britton and Robinson's universal buffer solutions at various pHs. The enzyme was stable in the pH range of 5–10 when incubated at 25 °C for 24 h in the same buffer solutions without the substrate. The effects of temperature on the activity and stability of the enzyme were examined. The optimum temperature for activity was found to be 60 °C. The enzyme was stable at 55 °C for 10 min at pH 5.5 in the absence of the substrate.

Hydrolytic activity of CelQcat

CelQcat (the polypeptide of the catalytic domain of CelQ) was produced by *E. coli* M15 (pCelQcat) and purified to a homogeneous state (Fig. 4A). The purified CelQcat showed negligible activity toward CMC (0.17 U/mg). This observation suggests that the hydrolytic activity of CelQ depends on the presence of the CBM.

Detection of CelQ in the cellulosomal proteins of *C. thermocellum*

By Western blot analysis using the antiserum directed against CelQΔdoc, a immunoreactive band with an apparent molecular weight of 76,000 was detected in the cellulosomal proteins purified from *C. thermocellum* F1 (Fig. 4B). The size of the immunoreactive protein was in good agreement with that of the mature CelQ (76,918) calculated from the deduced amino acid sequence. The profiles of SDS-PAGE and Western blot analysis suggest that CelQ is one of the components of the cellulosome.

Discussion

Strong consumers of cellulose are known to produce a number of cellulases with different enzyme properties (Ohmiya et al. 1997; Tomme et al. 1995) which are thought to synergistically attack their substrate. At least 15 cellulase genes were cloned and sequenced from *C. thermocellum* and seven of them encode enzymes classified in family 9 of glycoside hydrolases (http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html), i.e., *cbhA*, *celD*, *celF*, *celI*, *celJ*, *celK*, and *celN*. CelQ is the eighth cellulase of family 9 from *C. thermocellum*.

Cellulase genes are known to be scattered on the *C. thermocellum* chromosome DNA (Guglielmi and Béguin 1998) but with some exceptions, for example, the cellulase genes *celK* and *cbhA* are arranged in tandem (Zverlov et al. 1999). The *celQ* gene was identified downstream of *celJ* and there was a relatively long spacing of 562 bp between them. A possible terminator sequence was identified between *celJ* and *celQ* (Ahsan et

al. 1996). These observations suggest that these two genes are independently transcribed from their own promoters.

CelQΔdoc efficiently hydrolyzed barley β-glucan as well as CMC. The preference for barley β-glucan is common to the enzymes consisting of a family 9 catalytic domain and a family IIIc CBM. Specific activities of *C. cellulolyticum* CelG were reported to be 1,350 and 1,170 IU/μmol for barley β-glucan and CMC, respectively (Gal et al. 1997), and those of the truncated CelZ of *C. stercorarium* were 481 and 12 U/mg for the respective substrates (Riedel et al. 1998). Although *C. thermocellum* CelI comprising a family 9 catalytic domain and a family IIIc CBM in addition to another CBM of family IIIb was active against barley β-glucan and CMC, its activities towards these substrates were much lower than those of other enzymes, i.e., 0.11 and 0.0008 U/mg for β-glucan and CMC, respectively (Hazlewood et al. 1993). *C. thermocellum* CelF (Navarro et al. 1991) has a domain organization identical to that of CelQ. It consists of a family 9 catalytic domain, a family IIIc, and a dockerin domain from the N-terminus (Fig. 3). Unfortunately, the enzyme properties of CelF have not been elucidated.

In general, the most common arrangement of cellulases is a catalytic domain connected to a CBM. It is believed that CBMs play important roles in degradation of insoluble cellulose. Two possible functions have been proposed for CBMs: to increase the enzyme concentration on the surface of an insoluble substrate, and to supply the catalytic domain with a more easily degradable substrate by decreasing the crystallinity of the substrate. Since functional domains generally can function independently, artificial removal of a CBM from a modular cellulase does not affect the enzyme activity toward soluble substrates, and the CBM isolated from the catalytic domain retains the substrate-binding ability. Family IIIc CBMs, which are found along with family 9 catalytic domains, are different from the other CBMs in that their removal reduces more or less the enzyme activity of family 9 catalytic domains. Functions of family IIIc CBMs were investigated with *C. cellulolyticum* CelG (Gal et al. 1997) and *T. fusca* E4 (Irwin et al. 1998). These studies clearly indicated the importance of the family IIIc CBMs in substrate hydrolysis. However, the results and conclusions obtained from different enzymes have not been consistent. For example, Gal et al. (1997) found that removal of the CBM from *C. cellulolyticum* CelG abolished its catalytic activity and cellulose-binding ability. A truncated CelG containing the first 91 amino acids of the CBM in addition to the catalytic domain was slightly active on CMC (0.5 IU/μ mol of protein), corresponding to about 1/2,000 of the activity of the entire protein of CelG. It is noteworthy that the entire protein of CelG was able to bind to Avicel. On the other hand, the effect of removal of the CBM from *T. fusca* E4 was less severe on hydrolysis of CMC; the absence of the CBM led to a 77% decrease in the catalytic activity toward CMC (Irwin et al. 1998). In the same study, it was shown that the ratios of soluble to insoluble reducing sugar produced after filter-paper hydrolysis depended on the presence of the family IIIc CBM, although the CBM did not bind to insoluble cellulose, suggesting that the CBM was important for E4 processivity. Removal of the CBM from CelQΔdoc reduced the enzyme activity toward CMC to about 1/1,000 of the original activity but did not abolish it. We could not detect any affinity of CelQΔdoc and the polypeptide of the family IIIc CBM for insoluble and even soluble polysaccharides such as acid-swollen cellulose, CMC, and barley β-glucan by native affinity PAGE (data not shown). CelQΔdoc produced cellobiose and glucose as the major end products from cellotetraose, cellopentaose, cellohexaose, and ASC (Fig. 5). If CelQ has strong processivity, it should produce mainly cellobiose but not glucose. However, the hydrolysis pattern observed was inconsistent with that expected from the action of cellobiohydrolases, typical processive enzyme. The function of the family IIIc CBM of CelQ remains to be studied.

Acknowledgements This work was supported in part by a grant from the Program for Promotion of Basic Research Activities for Innovative Biosciences and by a grant of Rice Genome Project PR-2211, MAFF, Japan. The experiments comply with the current laws of Japan in which the experiments were performed.

References

Ahsan MM, Kimura T, Karita S, Sakka K, Ohmiya K (1996) Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. J Bacteriol 178:5732–5740

Ahsan MM, Matsumoto M, Karita S, Kimura T, Sakka K, Ohmiya K (1997) Purification and characterization of the family J catalytic domain derived from the *Clostridium thermocellum* endoglucanase CelJ. Biosci Biotechnol Biochem 61:427–431

Ali BR, Zhou L, Graves FM, Freedman RB, Black GW, Gilbert HJ, Hazelwood GP (1995) Cellulases and hemicellulases of the anaerobic fungus *Piromyces* constitute a multiprotein cellulose-binding complex and are encoded by multigene families. FEMS Microbiol Lett 125:15–21

Ali MK, Hayashi H, Karita S, Goto M, Kimura T, Sakka K, Ohmiya K (2001) Importance of the carbohydrate-binding module of *Clostridium stercorarium* Xyn10B to xylan hydrolysis. Biosci. Biotechnol. Biochem. 65:41–47

Bagnara-Tardif C, Gaudin C, Belaich A, Hoest P, Citard T, Belaich JP (1992) Sequence analysis of a gene cluster encoding cellulases from *Clostridium cellulolyticum*. Gene 119:17–28

Béguin P, Alzari PM (1998) The cellulosome of *Clostridium thermocellum*. Biochem Soc Trans 26:178–185

Cantor CR, Schimmel PR (1980) Biophysical chemistry, part III. The behavior of biological macromolecules. W.H. Freeman, New York

Ciruela A, Gilbert HJ, Ali BR, Hazlewood GP (1998) Synergistic interaction of the cellulosome integrating protein (CipA) from *Clostridium thermocellum* with a cellulosomal endoglucanase. FEBS Lett 422:221–224

Din N, Gilkes NR, Tekant B, Miller Jr RC, Warren RAJ, Kilburn DG (1991) Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. Biotechnology 9:1096–1099

Fukumura M, Begum A, Kruus K, Wu JHD (1997) Interaction and synergism between the recombinant CelD, an endoglucanase, and the cellulosome-integrating protein (CipA) of *Clostridium thermocellum*. J Ferment Bioeng 83:146–151

Gal L, Gaudin C, Belaich A, Pages S, Tardif C, Belaich JP (1997) CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. J Bacteriol 179:6595–6601

Gasparic J, Churacek J (1978) Laboratory handbook of paper and thin-layer chromatography. Ellis Horwood, Chichester, 153–154

Guglielmi G, Béguin P (1998) Cellulase and hemicellulase genes of *Clostridium thermocellum* from five independent collections contain few overlaps and are widely scattered across the chromosome. FEMS Microbiol Lett 161:209–215

Hayashi H, Takagi K, Fukumura M, Kimura T, Karita S, Sakka K, Ohmiya K (1997) Sequence of *xyn*C and properties of XynC, a major component of the *Clostridium thermocellum* cellulosome. J Bacteriol 179:4246–4253

Hazlewood GP, Davidson K, Laurie JI, Huskisson NS, Gilbert HJ (1993) Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. J Gen Microbiol 139:307–316

Henrissat B and Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem J 316:695–696

Irwin D, Shin DH, Zhang S, Barr BK, Sakon J, Karplus PA, Wilson DB (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. J Bacteriol 180:1709–1714

Jauris S, Rucknagel K P, Schwarz W H, Kratzsch P, Bronnenmeier K, Staudenbauer W L (1990) Sequence analysis of the *Clostridium stercorarium cel*Z gene encoding a thermoactive cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. Mol Gen Genet 223:258–267

Kataeva I, Guglielmi G, Beguin P (1997) Interaction between *Clostridium thermocellum* endoglucanase CelD and polypeptides derived from the cellulosome-integrating protein CipA: stoichiometry and cellulolytic activity of the complexes. Biochem J 326:617–624

Laemmli U K (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lamed R, Kenig R, Setter E, Bayer EA (1985) Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. Enzyme Microb Technol 7:37–41

Lao G, Ghangas GS, Jung ED, Wilson DB (1991) DNA sequences of three beta-1,4-endoglucanase genes from *Thermomonospora fusca*. J Bacteriol 173:3397–3407

Liu CC, Doi RH (1998) Properties of *exgS*, a gene for a major subunit of the *Clostridium cellulovorans* cellulosome. Gene 211:39–47

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193:265–275

Miller GL (1959) Use of dinitrosalicylic acid regent for determination of reducing sugar. Anal Chem 31:426–428

Morag E, Bayer EA, Lamed R (1992) Affinity digestion for the near-total recovery of purified cellulosome from *Clostridium thermocellum*. Enzyme Microb Technol 14:289–292

Navarro A, Chebrou MC, Beguin P, Aubert JP (1991) Nucleotide sequence of the cellulase gene *cel*F of *Clostridium thermocellum*. Res Microbiol 142:927–936

Ohmiya K, Sakka K, Karita S, Kimura T (1997) Structure of cellulases and their applications. Biotechnol Genet Eng Rev 14:365–414

Pagès S, Bélaich A, Bélaich J-P, Morag E, Lamed R, Shoham Y, Bayer EA (1997) Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: Prediction of specificity determinants of the dockerin domain. Proteins 29:517–527

Riedel K, Ritter J, Bauer S, Bronnenmeier K (1998) The modular cellulase CelZ of the thermophilic bacterium *Clostridium stercorarium* contains a thermostabilizing domain. FEMS Microbiol Lett 164:261–267

Rosenberg M, Court D (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Annu Rev Genet 13:319–353

Sakka K, Furuse S, Shimada K (1989) Cloning and expression in *Escherichia coli* of thermophilic *Clostridium* sp. F1 gene related to cellulose hydrolysis. Agric Biol Chem 53:905–910

Sakka K, Karita S, Ohmiya K, Shimada K (1994) Xylanases of *Clostridium stercorarium* and *Clostridium thermocellu*m in Japan. In: Shimada K, Hoshino S, Ohmiya K, Sakka K, Kobayashi Y, Karita S (eds) Genetics, biochemistry and ecology of lignocellulose degradation. Uni, Tokyo, pp 32–38

Shoham Y, Lamed R, Bayer EA (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol 7:275–281

Te'o V S, Saul DJ, Bergquist PL (1995) *celA*, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. Appl Microbiol Biotechnol 43:291–296

Tomme P, Warren RA, Gilkes NR (1995) Cellulose hydrolysis by bacteria and fungi. Adv Microb Physiol 37:1–81

Wang WK, Wu JHD (1993) Structural features of the *Clostridium thermocellum* cellulase Ss gene. Appl Biochem Biotechnol 40:149–158

Watson MEE (1984) Compilation of published signal sequences. Nucleic Acids Res 12:5145–5164

Zverlov VV, Velikodvorskaya GA, Schwarz WH, Kellermann J, Staudenbauer WL (1999) Duplicated *Clostridium thermocellum* cellobiohydrolase gene encoding cellulosomal subunits S3 and S5. Appl Microbiol Biotechnol 51:852–859