APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 1989, p. 2480–2487
0099-2240/89/102480-08$02.00/0

Vol. 55, No. 10

Purification and Characterization of Endoglucanase C of *Cellulomonas fimi*, Cloning of the Gene, and Analysis of In Vivo Transcripts of the Gene

B. MOSER, N. R. GILKES, D. G. KILBURN, R. A. J. WARREN, AND R. C. MILLER, JR.*

Department of Microbiology, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada

Received 2 February 1989/Accepted 31 May 1989

Two nonglycosylated endoglucanases which bind to Sephadex were purified from culture supernatants of *Cellulomonas fimi* grown on microcrystalline cellulose. Their M_rs were 120,000 and 130,000. The N-terminal amino acid sequences of the enzymes were identical, suggesting that the enzymes were related. A DNA fragment encoding this N-terminal sequence was cloned in *Escherichia coli*. The nucleotide sequence corresponding to the N-terminal amino acid sequence was preceded by a sequence encoding a typical leader peptide. Transcripts hybridizing to the cloned fragment were detected in total RNA isolated from *C. fimi* cells grown on carboxymethyl cellulose but not from cells grown on glycerol or glucose. Transcription started at a cluster of sites 53 to 59 nucleotides upstream of a GUG translation initiation codon and terminated at either of two closely spaced C residues immediately downstream of a region of potential secondary structure. The size of the transcript was approximately 3.5 kilobases, sufficient to encode a polypeptide of 130 kilodaltons. The 130-kilodalton polypeptide is designated endoglucanase C (CenC), and the gene encoding it is designated *cenC*.

Like many cellulolytic microorganisms, the bacterium *Cellulomonas fimi* produces a complex mixture of cellulases when grown on cellulosic substrates (2, 3, 16). This complexity is partly genetic in origin, with a number of the enzymes being determined by different genes (7, 8, 34). Some of the complexity appears to be physiological, however, with enzymes in the culture supernatant being converted by deglycosylation or proteolysis to derivatives which retain enzymatic activity (9, 15, 16).

We are attempting to identify and characterize all of the components of the cellulase complex of *C. fimi*. In light of the physiological variation in the complex (16), our initial approach was to clone the genes determining these components, using an immunological screening procedure that required expression of the genes in *Escherichia coli* (7, 33). This led to the characterization of two enzymes which bind tightly to microcrystalline cellulose: an exoglucanase (Cex) encoded by gene *cex*, and an endoglucanase (CenA) encoded by gene *cenA* (25, 34). The gene (*cenB*) for a second endoglucanase (CenB) was also cloned and partially characterized (7, 8, 26). Unlike Cex and CenA, CenB has not yet been identified in culture supernatants of *C. fimi*.

An unidentified *Cellulomonas* sp. produces cellulases which bind to Sephadex (2, 3). Such enzymes were not produced by any of the *C. fimi* genes cloned previously. If the genes encoding these Sephadex-binding enzymes were expressed poorly in *E. coli*, they might not have been detected by the antibody-screening procedure used to detect the clones, since the screen is dependent on the level of expressed gene product. This paper describes the purification from *C. fimi* culture supernatants of two cellulases which bind to Sephadex. Analysis of the proteins suggested that they were closely related. A gene encoding the N-terminal sequence common to both polypeptides was cloned and partially characterized.

* Corresponding author.

MATERIALS AND METHODS

Bacterial strains and plasmids. *C. fimi* strain ATCC 484 was used. *E. coli* JM101 was used as the host for plasmids pUC13 (36), pTZ18R, and pTZ19R and their derivatives. Lambda replacement vector L47.1 (17) and its derivatives were propagated on *E. coli* NM358 or NM359(P2).

Enzymes and reagents. DNA polymerase I (Klenow fragment), T4 DNA polymerase, and T4 polynucleotide kinase were from Pharmacia LKB Biotechnology, Inc. Calf intestinal alkaline phosphatase and 7-deaza-dGTP were from Boehringer Mannheim Ltd. RNase T_1, yeast tRNA, and nuclease S1 were from Bethesda Research Laboratories, Inc. Exonuclease Bal 31, T4 DNA ligase, and DNA sequencing primers were from New England BioLabs, Inc. Lambda L47.1 DNA and the in vitro lambda packaging kit were from Amersham Canada, Ltd. The USB Genescribe-Z system (pTZ18R and pTZ19R) was from U.S. Biochemical. DNase RQ1 was from Promega Biotec. Radionucleotides were from Dupont, NEN Research Products, or from ICN Biomedicals Inc. Avicel PH101 was from FMC International, Ireland. Carboxymethyl cellulose (CM-cellulose), low viscosity, was from Sigma Chemical Co. The degree of polymerization was 400; the degree of substitution was 0.7. Concanavalin A-horseradish peroxidase was from Seikagaku America Inc.

Enzyme assays. The reduction in the viscosity and the release of reducing sugars from CM-cellulose by the cellulases were measured as described previously (8). The retention time (t_o) of an enzyme solution containing no CM-cellulose was 8.4 s at 37°C. The real incubation times (t) were calculated as the time at the start of viscometric measurements (t_x) plus half of the measured retention times (t_r). The specific fluidities (ϕ_{sp}) were calculated by the equation: $\phi_{sp} = 1/[(t_r/t_o) - 1]$ (29). The changes in reducing sugars corresponding to the $\Delta\phi_{sp}$ values were determined with dinitrosalicylic acid (8). One unit of CM-cellulase activity releases 1 μmol of glucose equivalents in 1 min at 37°C. Protein concentrations were determined with the Bio-Rad assay kit (4). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (PAGE) was performed as described previously (14).

The staining of polyacrylamide gels with concanavalin A-horseradish peroxidase for glycoprotein detection was described previously (37).

Purification of cellulases. *C. fimi* was grown in 12 liters of basal salts medium (8) containing 0.2% (wt/vol) Avicel for 5 days at 30°C. Cells and residual Avicel were removed by centrifugation. Phenylmethylsulfonyl fluoride, sodium azide, and ammonium sulfate were added to the supernatant to give final concentrations of 50 μM, 0.2% (wt/vol), and 0.2 M, respectively. Then the supernatant was agitated gently with 200 ml of preswollen Sephadex G-25 for 24 to 48 h at 4°C. The Sephadex was packed into a column to give a bed of 40 by 2.5 cm. The column was washed with several column volumes of starting buffer (40 mM potassium phosphate, pH 7.5, 0.2 M ammonium sulfate, 0.02% [wt/vol] sodium azide). Proteins then were eluted from the column with a concave gradient. The mixing chamber was a closed 250-ml Erlenmeyer flask containing 250 ml of starting buffer; the reservoir contained 2.25 liters of 0.02% sodium azide in water. Two peaks of CM-cellulase activity were eluted from the column. The fractions in each peak were pooled separately and concentrated by ultrafiltration through Amicon PM-10 membranes. Each sample was diluted fivefold with 20 mM sodium piperazine, pH 5.8–0.2 M NaCl and applied to a Pharmacia Mono Q fast protein liquid chromatography column. Bound material was eluted with a linear gradient of NaCl, 0.2 to 0.6 M. The activity from the first Sephadex band (CenC) was eluted routinely with 0.42 M NaCl; that from the second band (CenC′) was eluted with 0.44 M NaCl.

The active fractions in each preparation were pooled, and either they were desalted by chromatography on Bio-Gel P-6DG and then lyophilized or azide was added to a final concentration of 0.02% (wt/vol) and the fractions were stored at 4°C.

Isolation of an internal peptide from CenC′. A 2-nmol sample of CenC′ was digested with 5 μg of trypsin for 15 h at 37°C. The digest was fractionated by reverse-phase high-performance liquid chromatography, using an aqueous trifluoroacetic acid-acetonitrile solvent system. A discrete band was subjected to amino acid sequencing (see Table 2), which confirmed that it contained a single peptide, designated T-115.

Determination of amino acid sequences. The N-terminal amino acid sequences of CenC, CenC′, and T-115 were determined by automated Edman degradation with an Applied Biosystems 470A gas phase sequenator.

Isolation of DNA and RNA from *C. fimi*. Total DNA was isolated from *C. fimi* essentially as described previously (19), using cells grown in low-salt LB medium (19). RNA was isolated from *C. fimi* as described previously (11), using cells grown to late exponential phase in basal salts medium (8) with 1.0% (wt/vol) CM-cellulose, 0.2% (wt/vol) glycerol, or 0.2% (wt/vol) glucose as carbon source.

Cloning of gene *cenC*. Lambda L47.1 DNA was ligated and digested with *Bam*HI, and the ligated lambda arms were separated from the stuffer fragments by agarose gel electrophoresis. *Sau*3A fragments of genomic *C. fimi* DNA ranging in size from 10 to 20 kilobase pairs were isolated from an agarose gel. Prior to ligation to the lambda arms, the *Sau*3A fragments were dephosphorylated with calf intestinal alkaline phosphatase. The ligation mixture was packaged in vitro according to the instructions of the supplier and plated on *E. coli* NM539(P2) to select for recombinant phage particles (13, 17). The yield of L47.1 particles was 7×10^4 PFU/μg of ligated DNA. Approximately 3.0×10^4 particles were plated with *E. coli* NM538 on two agar plates (15 cm in diameter) and incubated to form plaques of <1.0 mm in diameter. The plaques were lifted to Biodyne P/N membranes (Pall Ultrafine Filtration Corp.) and prepared for screening with ^{32}P-labeled oligonucleotide probes (35). Four sets of synthetic oligonucleotides (sets JO.3A, B, C, and D with 32 oligonucleotides per set) comprising the fully redundant repertoire of oligonucleotides corresponding to the N termini of C3.1 and C3.2 (see Table 2) were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase; 6×10^6 dpm per set of probes (0.6×10^6 to 3.3×10^6 dpm/pmol of DNA) were used for screening the membranes (35). The nine positive plaques obtained were purified by replating and then rescreened with ^{32}P-end-labeled probes JO.3A and BM.C (see Table 2). L47.1-169 was chosen for subcloning into pTZ18R and DNA sequencing.

DNA sequence determination. Deletions of the *Sst*I/*Bam*HI fragment in pTZ18R-8/5 (Fig. 3) were created with exonuclease Bal 31 or by using the exonucleolytic activity of T4 DNA polymerase (5). The DNA probes used in S1 nuclease protection analyses were cloned into the appropriate pTZ vector prior to sequencing. Single-stranded template DNA for sequencing was isolated by the Genescribe-2 protocol. DNA was sequenced by the dideoxy chain termination method (27) with 7-deaza-dGTP instead of dGTP in all nucleotide mixes, and the nucleotide mixes were adjusted for the high G+C content of *C. fimi* DNA (22).

RNA dot blot analysis. A 5-μg sample of each *C. fimi* RNA preparation was spotted onto a Biodyne P/N membrane. A 5-μg sample of each preparation was hydrolyzed with 0.2 M NaOH at 37°C for 30 min and then spotted on the membrane. A synthetic oligonucleotide probe (5′-GTCGTCGAACGT TCCCTCCCCGAT), corresponding to the coding sequence for amino acids 4 to 11 at the N termini of the mature cellulases, was labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase, and 4×10^6 dpm of labeled probe (>10^7 dpm/pmol of DNA) was hybridized with the membrane (35).

Mapping the ends of *cenC* transcripts. The DNA probe (C5′) for mapping the 5′ ends of *cenC* transcripts was prepared as follows. pTZ18R-8/5 was digested with *Bam*HI and then calf intestinal alkaline phosphatase and labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. The labeled probe was digested with *Eco*RI, and the desired fragment was purified by nondenaturing PAGE, electroelution (19), and ethanol precipitation. To prepare the DNA probe (C3′) for mapping the 3′ ends of *cenC* transcripts, the *Sma*I-*Hin*dIII fragment of pTZ18R-8/5 was subcloned into pTZ18R to give pTZ18R-8/5-4 (see Fig. 3). pTZ18R-8/5-4 was digested with *Mlu*I, and the 3′ ends were labeled with [α-^{32}P]dCTP in the presence of excess dGTP, using DNA polymerase I (Klenow fragment). After digestion with *Hin*dIII, the desired fragment was purified as described above. A 20-μg sample of each *C. fimi* RNA preparation and a 20-μg sample of yeast tRNA were each hybridized to 2×10^5 dpm of C5′(>2×10^5 dpm/pmol) or to 10^6 dpm of C3′ (>10^6 dpm/pmol). The protocol for the S1 nuclease protection analysis was described previously (11).

RESULTS

Purification of cellulases. The procedure for the purification of the *C. fimi* cellulases which bound to Sephadex is described in Materials and Methods. Two peaks of CM-cellulase activity were eluted from the Sephadex column (Table 1). The total activity recovered from the column represented about 2% of the CM-cellulase activity in the culture supernatant. The activity in each peak was purified

TABLE 1. Purification and properties of cellulases CenC and CenC′ of *C. fimi*

Material	Total protein (mg)	Total activity (U)	Sp act (U/mg)	M_r[a]
Culture supernatant	147.00	1,899.0	12.9	
Sephadex G-25				
Pool I	0.34	8.4	24.3	
Pool II	0.82	27.1	33.0	
Mono Q				
CenC (pool I)	0.03	7.1	236.0	130,000
CenC′ (pool II)	0.06	22.0	368.0	120,000

[a] M_r was determined by sodium dodecyl sulfate-PAGE.

further by fast protein liquid chromatography on a Mono Q anion-exchange column. A single peak of activity was recovered from each Sephadex peak. The recovery from the Mono Q column was >80% in both cases. The two activities were quite homogeneous, with apparent M_rs of 130,000 and 120,000 (Fig. 1).

Properties of the cellulases. The enzymes did not react with the periodic acid-Schiff's reagent or with concanavalin A-horseradish peroxidase conjugate (Table 1). Therefore, unlike CenA and Cex (8, 9), neither enzyme appeared to be glycosylated. They were 20-fold less active than Cex on *p*-nitrophenylcellobioside (data not shown). Their actions on CM-cellulose were identical and resembled that of CenA rather than Cex (Fig. 2). Thus, they were endoglucanases, although, from the slopes of the curves (Fig. 2), their actions on CM-cellulose appeared less random than that of CenA.

The N-terminal amino acid sequences of the enzymes

TABLE 2. Amino-terminal sequences of *C. fimi* endoglucanases CenC and CenC′ and of CenC′ internal peptide T-115

Cycle	Amino acid (yield, pmol)[a]		
	CenC (2,500 pmol)[b]	CenC′ (2,500 pmol)	T-115 (400 pmol)
1	A (2,190)	A (1,860)	L (275)
2	S (120)	S (220)	L/E (168/54)
3	P (800)	P (1,310)	**E**[c] (73)
4	I (830)	I (710)	**P** (112)
5	G (660)	G (980)	**Y** (97)
6	**E** (360)	**E** (1,150)	**D** (71)
7	**G** (300)	**G** (950)	**P** (49)
8	**T** (260)	**T** (180)	**Q** (38)
9	**F** (650)	**F** (640)	L (137)
10	**D** (190)	**D** (430)	A (118)
11	**D** (150)	**D** (410)	Q (50)
12	G/D (440/270)		T (29)
13	P (340)	P (510)	L (72)
14	E (130)		L (105)
15	E/G (130/320)		E (41)
16	W (200)		A (77)
17	V (330)		A (177)
18	A (460)		
19	Y (320)		
20	G (260)		

[a] Amino acids in bold letters indicate the sequence chosen for the synthesis of four sets of amino-terminal specific oligonucleotides (JO.3A, B, C, and D).
[b] Sample amount.
[c] Sequence corresponding to the T-115 specific set of synthetic oligonucleotides (BM.C).

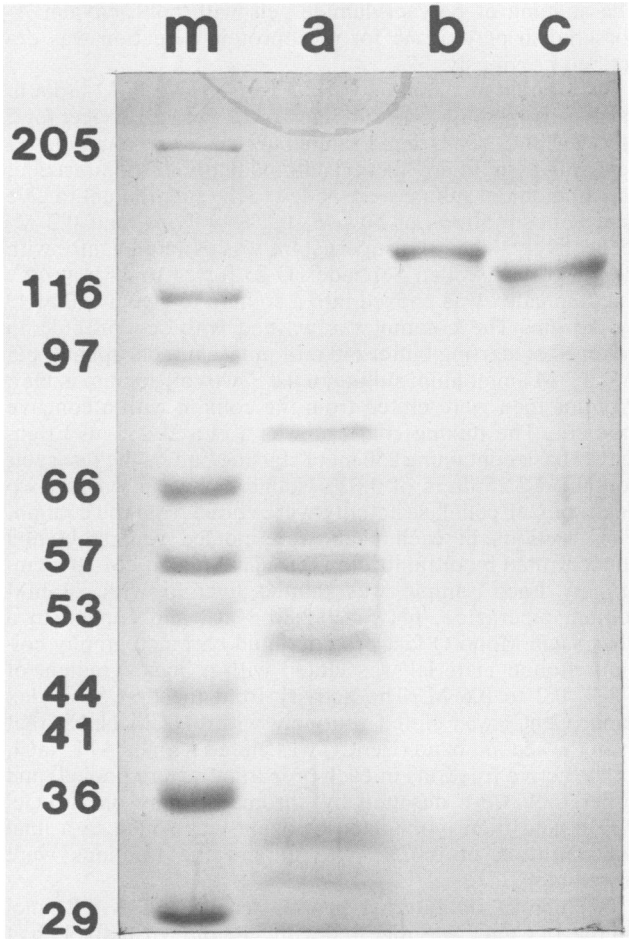

FIG. 1. Purity of CenC and CenC′ analyzed by sodium dodecyl sulfate-PAGE. Lane a, Sample of 20-fold-concentrated, clarified culture supernatant from *C. fimi* grown on 0.2% (wt/vol) Avicel. Lanes b and c, Samples of CenC and CenC′, respectively, after purification by Mono Q anion-exchange chromatography. Lane m, Calibration standards: rabbit muscle myosin (205 kDa), *E. coli* β-galactosidase (116 kDa), rabbit muscle phosphorylase *b* (97.4 kDa), bovine serum albumin (66 kDa), catalase (57.5 kDa), bovine liver glutamate dehydrogenase (53 kDa), alcohol dehydrogenase (44 kDa), ovalbumin (41 kDa), rabbit muscle glyceraldehyde-3-phosphate dehydrogenase (36 kDa), and bovine erythrocyte carbonic anhydrase (29 kDa).

were identical (Table 2), suggesting that they were related. *C. fimi* secretes a protease which cleaves Cex and CenA into discrete fragments (9, 15). The 120-kilodalton (kDa) enzyme could be a cleavage product of the 130-kDa enzyme, or both enzymes could be products of a common precursor protein. They were designated as endoglucanases C and C′ (CenC and CenC′) of *C. fimi*; the gene encoding CenC was designated *cenC*. Since it was also possible, however, that the two enzymes were encoded by different genes, it was decided to clone the gene(s) encoding the N-terminal sequence.

Cloning of *cenC*. The sequence of an internal peptide from CenC′ rather than CenC was chosen to dictate the nucleotide sequence of the second probe used in the cloning of gene *cenC*. This obviated the possibility of choosing a peptide present in CenC but not in CenC′. Three plaques out of a total of 3×10^4 hybridized to JO.3A and BM.C as described in Materials and Methods. One of them, L47.1-169, was

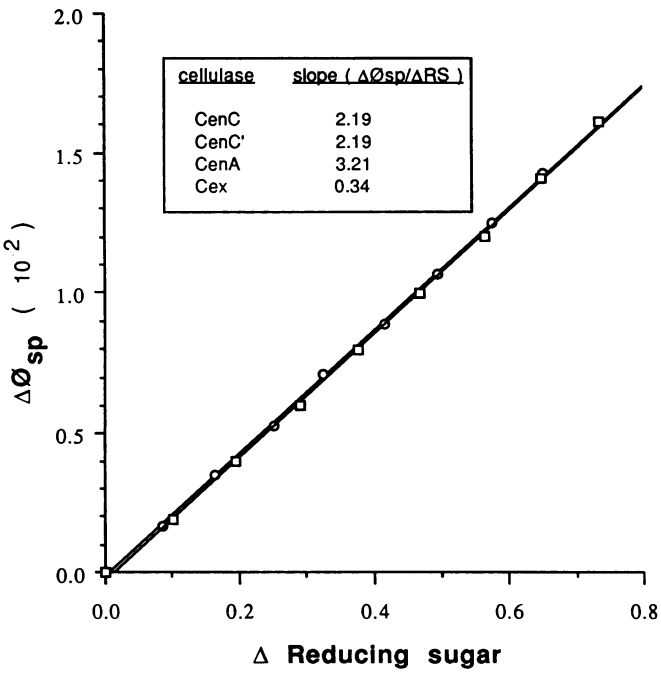

FIG. 2. Action of CenC and CenC′ on CM-cellulose. See Materials and Methods for details. The final concentrations of purified CenC and CenC′ were 26 ng/ml, and the solutions were supplemented with 0.2 mg of bovine serum albumin per ml. Symbols: ○, CenC; □, CenC′. The values for CenA and Cex are from reference 8.

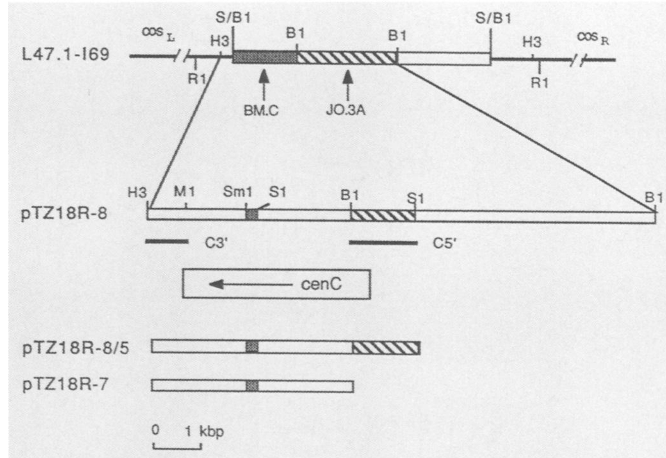

FIG. 3. Diagram of *C. fimi* DNA in lambda L47.1-169 and pTZ18R subclones. Hatched bars represent regions which hybridized to oligonucleotide probes in JO.3A, and stippled bars represent those which hybridized to oligonucleotides in BM.C. (see Table 2). 5′ to 3′ orientation and location of *cenC* are represented by an arrow and an open box. Black bars indicate the DNA probes used for mapping of the 5′ (C5′) and 3′ (C3′) ends of *C. fimi* RNA by S1 nuclease protection analysis. Abbreviations for restriction endonucleases are: B1, *Bam*HI; H3, *Hin*dIII; R1, *Eco*RI; SI, *Sst*I; Sm1, *Sma*I; S/B1, *Sau*3A-*Bam*HI fusion. Cos_L and Cos_R stand for the left and right arms of lambda L47.1, respectively. kbp, Kilobase pairs.

chosen for further study. A restriction map of L47.1-169 showed that it contained a 10.5-kilobase-pair fragment of *C. fimi* DNA (Fig. 3). To determine the orientation and approximate location of *cenC* in L47.1-169, the *Bam*HI-*Hin*dIII and *Bam*HI-*Bam*HI fragments were subcloned into pUC13 and analyzed by hybridization to JO.3A and BM.C (Fig. 3). The 7.7-kilobase-pair *Bam*HI-*Hin*dIII fragment of L47.1-169 was then subcloned into pTZ18R to give pTZ18R-8 (Fig. 3). The 5′ end of the sequence encoding mature CenC was identified by Southern transfer analysis, using the amino-terminal-specific probe (JO.3A). It was located on an 850-base-pair *Sst*I-*Bam*HI fragment of DNA. Similarly, the probe corresponding to tryptic peptide T-115 (BM.C) allowed identification of an internal sequence of *cenC* (Fig. 3). This allowed the further subcloning of the *Sst*I-*Hin*dIII fragment to give pTZ18R-8/5 (Fig. 3).

The sequence corresponding to the amino terminus of CenC was identified by matching the predicted amino acid sequence with those determined for the proteins purified from *C. fimi* (Fig. 4). A putative GTG start codon was found 96 nucleotides upstream of the amino-terminal codon of the mature cellulases. The DNA of *C. fimi* is 67 mol% G+C, so GTG translation initiation codons may be expected in this bacterium. The intervening sequence encoded a putative

```
                                                                5'-CCGGTGCGCACGCTTCCGC
GGGCGCCGCGGCCGTGCCTACCCTGCCGCCGCAGCACCGGCGTTCTCCACCGCAGGGTGGGAGCGCTCCTACGACAGGGGAGACAGA
                    550                                                    600

        -30                                 -20
 M   V   S   R   R   S   S   Q   A   R   G   A   L   T   A   V   V   A   T   L   A   L
GTG GTT TCT CGC AGG TCA TCA CAG GCG CGC GGC GCG CTC ACG GCC GTC GTC GCG ACG CTC GCC CTC
                                            650

-10                                  -1  +1                                 +10
 A   L   A   G   S   G   T   A   L   A   A   S   P   I   G   E   G   T   F   D   D   G
GCG CTC GCC GGG AGC GGC ACC GCG CTC GCC GCG TCG CCG ATC GGG GAG GGA ACG TTC GAC GAC GGG
                    700                                                             750

                             +20                                +30
 P   E   G   W   V   A   Y   G   T   D   G   P   L   D   T   S   T   G   A   L   W
CCC GAG GGG TGG GTC GCG TAC GGC ACC GAC GGC CCC CTC GAC ACG AGC ACG GGC GCG CTG TGG etc.
                                                                    800
```

FIG. 4. DNA sequence upstream of the 5′ end of *cenC*. Numbers under the sequence refer to the distance in nucleotides from the *Sst*I site proximal to the 5′ end of *cenC* (Fig. 3). The underlined sequence in close proximity to the *cenC* start codon indicates the putative ribosome-binding site. Amino acids encoded by the 5′ region of *cenC* are written above the sequence and numbered with respect to the N-terminal amino acid of mature CenC and CenC′ (+1). The amino acid sequence of the putative CenC leader peptide is underlined.

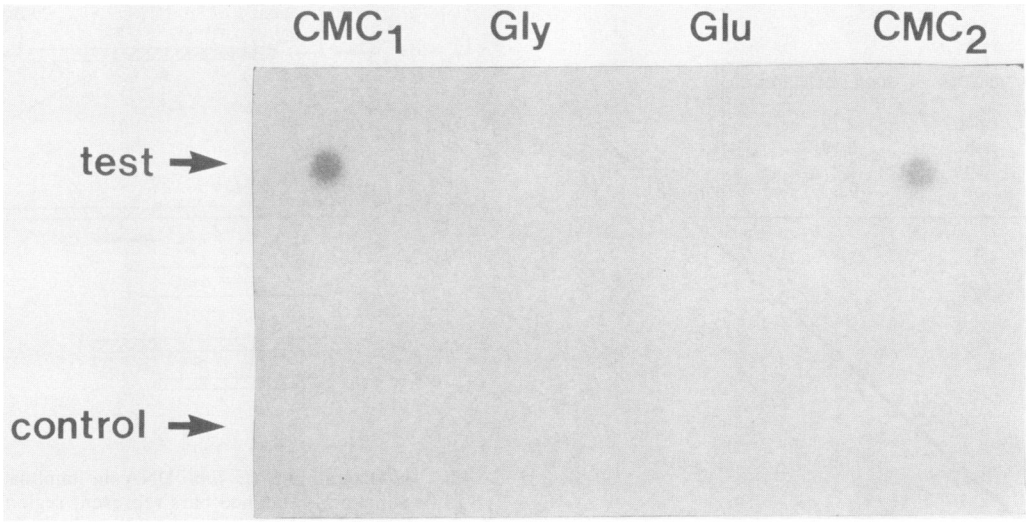

FIG. 5. Dot blot analysis of *C. fimi* RNA. Test row: 5-μg samples of total RNA extracted from *C. fimi* grown on CM-cellulose (CMC_1 and CMC_2 were two separate RNA preparations), glycerol (Gly), or glucose (Glu) were spotted onto a Biodyne membrane. Control row: 5-μg samples of the RNA preparations were hydrolyzed with 0.2 M NaOH before spotting on the membrane. The membrane was probed with a ^{32}P-labeled synthetic oligonucleotide corresponding to an amino acid sequence near the N termini of CenC and CenC′.

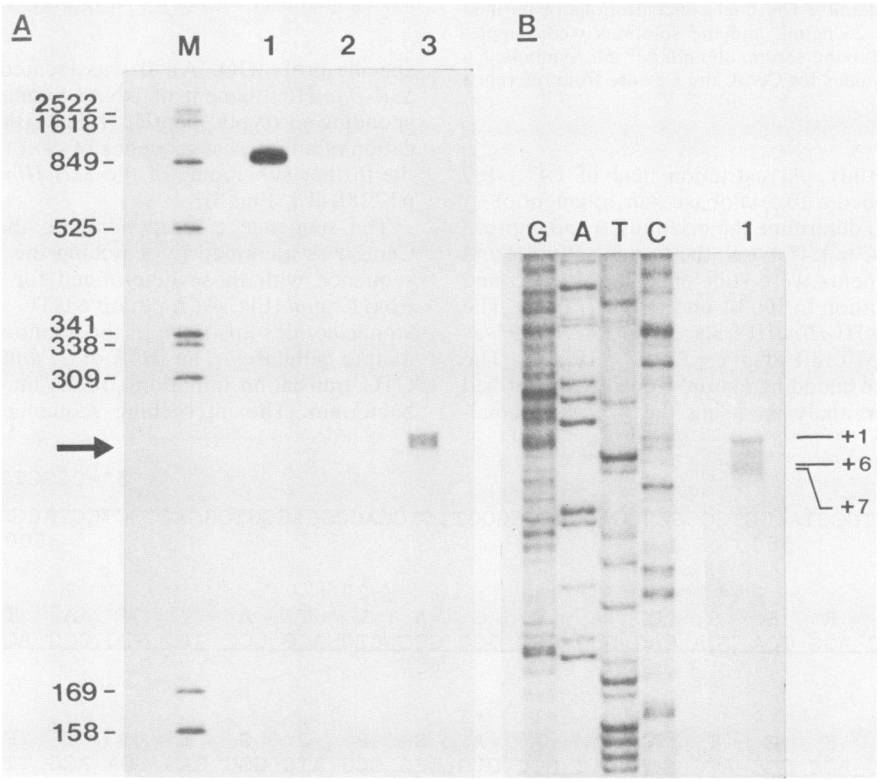

FIG. 6. Mapping of the 5′ ends of *cenC*-specific mRNA from *C. fimi*. After hybridization with RNA and S1 nuclease treatment, the protected fragments derived from the 5′-labeled DNA probe were analyzed by electrophoresis through polyacrylamide gels containing 7 M urea. (A) Lane 1, Fully protected, 5′-labeled DNA probe; lane 2, DNA fragments protected by control RNA (yeast tRNA); lane 3, DNA fragments protected by *C. fimi* RNA (see arrow); lane M, 5′-labeled, single-stranded M13mp11 *Hae*III fragments (numbers indicate the sizes of the restriction fragments). (B) High-resolution mapping of DNA fragments protected by *C. fimi* RNA. The sequencing ladder was prepared by subcloning of the *Sst*I-*Bam*HI fragment of pTZ18R-8/5 which was used as probe for the S1 nuclease protection analysis (Fig. 3) into pTZ19R prior to sequencing. The dideoxy sequencing products were digested with *Bam*HI to ensure that all fragments had the same 5′ end as the S1 nuclease protection probe before separation by denaturing PAGE (lanes G, A, T, and C). Lane 1, Same sample as in lane 3 of part A. Numbers on the right indicate the differences in lengths (in nucleotides) of the three major protected species compared with the largest fragment (+1).

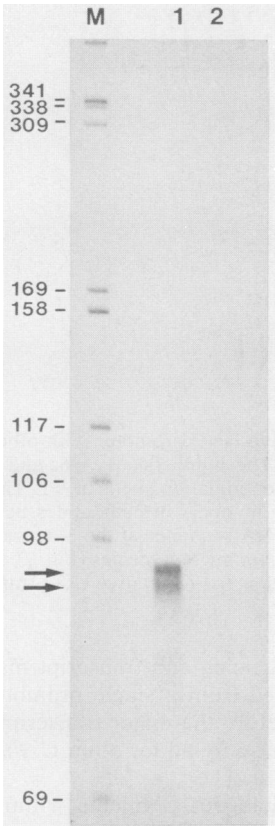

FIG. 7. Mapping of the 3′ ends of *cenC*-specific mRNA from *C. fimi*. The 3′-labeled DNA fragments protected from S1 nuclease degradation by total *C. fimi* RNA were analyzed by denaturing PAGE (lane 1). The arrows refer to the 3′ ends of the two major protected species. Lane M, Single-stranded M13mp11 *Hae*III marker fragments as in Fig. 6.

signal peptide and an Ala-Ala leader peptidase cleavage site. The sequence AGGGGA six bases upstream of the GTG start codon resembled a procaryotic ribosome-binding site.

Expression of *cenC* in *E. coli*. Although *cenC* was in the correct orientation in L47.1-169 to be transcribed from p_L, CM-cellulase activity could not be detected in the culture supernatant of *E. coli* NM358 following infection with bacteriophage. Sequence analysis of *C. fimi* DNA preceding the GTG start codon of *cenC* revealed several inverted repeats which could act as transcription termination signals in L47.1-169 (data not shown). The specific activities of crude extracts of *E. coli* JM101 carrying pTZ18R-8 or pTZ18R-8/5 (Fig. 3) were 0.008 U/mg of protein for both constructs. Again, the low-level production of cellulase might have been due to interference of *C. fimi* DNA upstream of the 5′ end of *cenC* with *lacZ-cenC* fusion transcripts. It also indicated that the *cenC* promoter was poorly recognized by *E. coli*, if at all.

Expression of *cenC* in *C. fimi*. Transcripts of *cenC* could not be detected by dot blot analysis of RNA extracted from cultures of *C. fimi* grown on glycerol or glucose as sole carbon source. However, they were readily detected in cultures grown on CM-cellulose (Fig. 5).

The transcription start sites for *cenC* were located by high-resolution nuclease S1 mapping, using the probe C5′ (Fig. 3). Three major start sites (+1, +6, and +7) were detected 59, 54, and 53 nucleotides upstream of the putative GTG translation initiation codon of *cenC* (Fig. 6). The initiating bases at the three sites were C, G, and C, respectively.

TABLE 3. Transcription of cellulase genes during growth of *C. fimi* on different carbon sources

Gene	Transcripts present[a] during growth on given carbon source		
	Glycerol	Glucose	CM-cellulose
cex			+
cenA	+		+
cenB	+	+	+
cenC			+

[a] Qualitative analysis only.

The M_r of CenC indicated the approximate location of the 3′ terminus of the coding region of *cenC*. Therefore, the 600-base-pair *Mlu*I-*Hin*dIII fragment of pTZ18R-8/5-4, termed C3′ (Fig. 3), was used as a hybridization probe to detect the 3′ termini of the *cenC* transcripts. Two major termination sites were detected, 93 and 95 nucleotides downstream of the *Mlu*I site (Fig. 7). Both transcripts terminated with G. Sequence analysis of the probe C3′ revealed inverted repeats just upstream of the termination sites (Fig. 8C). The transcript analyses suggested that the subclone pTZ18R-8/5 contained the entire *cenC* coding sequence. The mRNA was estimated to be 3.5 kilobases long, sufficient for a protein of 130 kDa.

DISCUSSION

Endoglucanases CenC and CenC′ appear to be closely related; they may be the mature *cenC*-encoded polypeptide and a product derived from it by C-terminal truncation, respectively. It is significant that all of the *C. fimi* cellulases characterized previously retain enzymatic activity when truncated: Cex (9), CenA (9), and CenB (26).

CenC and CenC′ of *C. fimi* may be analogous to endoglucanases EG2A and EG2B from *Bacteroides succinogenes* (20). EG2A and EG2B are not resolved during the initial phases of purification, but they can be separated by repeated ion-exchange chromatography. The M_rs are 118,000 for EG2B and 94,000 for EG2A. The pIs are 9.40 for EG2B and 9.18 for EG2A. The two enzymes give very similar one-dimensional cyanogen bromide peptide maps, but EG2B does give a few more peptides than EG2A. The two polypeptides appear to be closely related, with EG2A being a proteolytic degradation product of EG2B. It is significant that EG2B but not EG2A binds to cellulose. This parallels the loss of affinity for cellulose following proteolysis of Cex and CenA from *C. fimi* (9, 15).

CenC is not glycosylated, unlike Cex and CenA. It appears to be similar to endoglucanase C1, which was purified from culture supernatants of *Cellulomonas* sp. strain IIbc by binding to and elution from Sephadex G-25 (3). The M_r of C1 was 118,000, and it was not glycosylated. However, the specific activity of CenC is some 2×10^4-fold greater than that reported for C1. Other bacteria also produce a mixture of glycosylated and nonglycosylated cellulases (1). Why an organism glycosylates only certain of its cellulases is unclear at present.

CenC acts like an endoglucanase on CM-cellulose. The specific activity of CenC′ is about 50% greater than that of CenC, but both enzymes hydrolyze CM-cellulose in a similar manner. Again, this is consistent with CenC′ being a proteo-

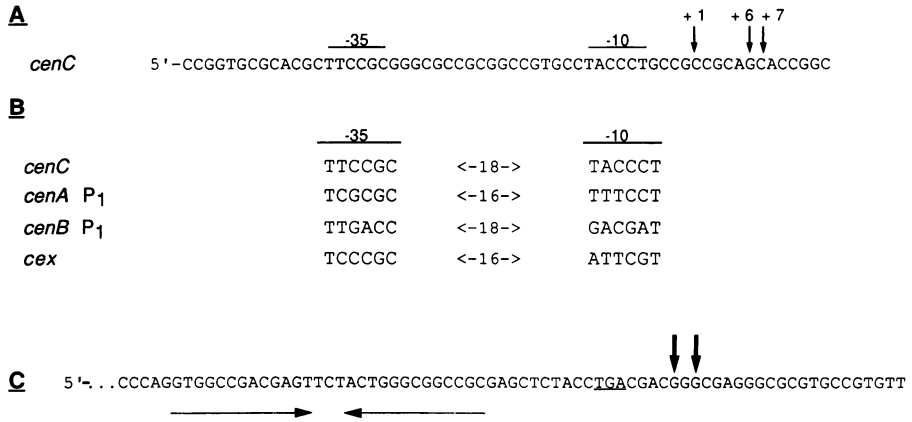

FIG. 8. Initiation and termination sites of *cenC* transcripts and comparison of the putative *cenC* promoter with other *C. fimi* promoter sequences. (A) DNA sequence corresponding to the 5'-terminal region of *cenC* mRNA. The numbers +1, +6, and +7 refer to the 3' nucleotides of the DNA probe sequences protected by *C. fimi* mRNA. The major transcription initiation site is at +1. The putative −10 and −35 sequences of the *cenC* promoter are overlined. (B) −10 and −35 sequences, with the numbers of nucleotides between them, of the promoters for *C. fimi* cellulase genes characterized to date (see references 10 and 11). (C) DNA sequence at the 3'-terminal region of *cenC*. Arrows point to the nucleotides (two G's) which correspond to the 3' ends of the mapped *C. fimi* mRNA species (Fig. 7). Arrows underlining the sequence indicate the locations and orientations of inverted repeats. The triplet corresponding to the putative UGA translation termination codon of *cenC* is underlined.

lytic product of CenC. Other cellulases retain activity when truncated by proteolysis or by manipulation of their genes (6, 9, 12, 18, 26, 30, 31, 38), and proteolysis appears to contribute to the complexity of many microbial cellulase systems (15, 28).

The cellulase genes of *C. fimi* characterized to date fall into three groups according to their responses to the carbon source supplied for growth, with *cenC* behaving like *cex* (Table 3). The *cenB* gene may be expressed constitutively to allow CenB to form inducers from cellulosics for the induction of the other cellulase genes. CenC is not made in large amounts in *C. fimi* compared with the substrate-bound cellulases CenA and Cex, but it is difficult to estimate exactly what proportions the latter represent because their activities can be determined only after elution from the residual cellulose in a culture. Nonetheless, CenC probably represents somewhat less than 2% of the total extracellular CM-cellulase activity of a *C. fimi* culture because purified CenC plus CenC' accounted for only 2% of the CM-cellulase activity in the culture supernatant (Table 1). However, it could have a major role in the overall activity of the *C. fimi* cellulase system. Its specific activity on CM-cellulose is higher than those of CenA and CenB.

The DNA sequencing and transcript mapping show that *cenC* is transcribed from a single promoter. The −10 and −35 sequences before the major transcription start site are quite similar to those found for other *C. fimi* cellulase genes (Fig. 8A and B) (10, 11).

There are homologous segments within the amino acid sequences of cellulases from several gram-positive bacteria which are more or less closely related to *C. fimi*, suggesting that they share common origins, and may have arisen by sequence shuffling (23–25, 32, 34). Although sequence data are not yet available for the entire CenC, the sequence of its leader peptide shows significant homology with the leader peptides of these cellulases, especially with that of endoglucanase Cas from an alkalophilic *Streptomyces* strain (Fig. 9). The N-terminal segment of the Cas leader peptide is homologous to that of the N-terminal segment of the Cex leader peptide, whereas its C-terminal segment is more homologous to the CenC leader peptide (Fig. 9).

The level of expression of *cenC* on pTZ18R was too low to give enough CenC for characterization. We have manipulated *cenC* to give levels of expression sufficient for characterization of the gene products from *E. coli*. Preliminary

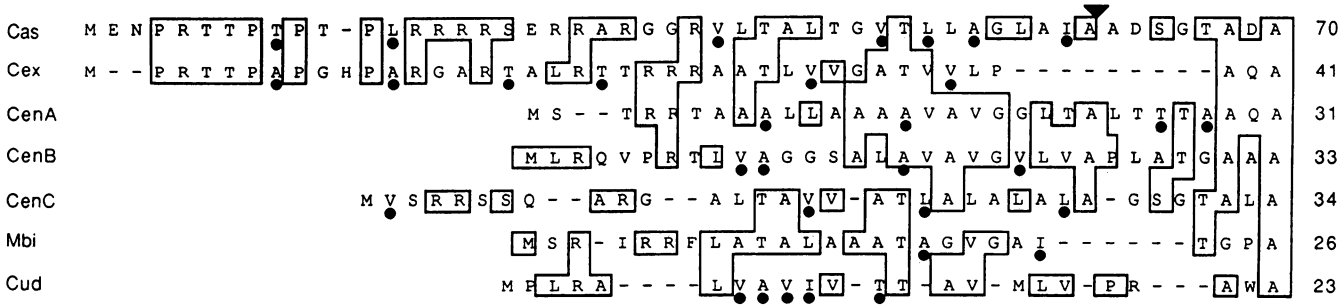

FIG. 9. Comparison of the signal peptides of various cellulases, deduced from the following nucleotide sequences: *cas* (23); *cex* (25); *cenA* (34); *cenB* (26); *cenC* (this paper); *mbi* (an endoglucanase from *Microbispora bispora* [M. D. Yablonsky, personal communication]); *cud* (24). Conserved residues are boxed. Amino acids marked with ● are conservative replacements. ▼, A sequence of 19 amino acids resembling the Pro-Thr box of Cex and CenA (23, 25, 34) is omitted at this point; the sequence is PAATGASPSPAPPASPAPS. The number of amino acids in each leader peptide is given at the right.

results show that *E. coli* also produces polypeptides of M_rs 130,000 and 120,000.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

We thank M. D. Yablonsky and D. Eveleigh for communicating results prior to publication, D. J. McKay and S. L. Kielland for amino acid sequence determinations, and T. Atkinson for oligonucleotide syntheses.

LITERATURE CITED

1. **Aubert, J.-P., P. Beguin, and J. Millet (ed.).** 1988. Biochemistry and genetics of cellulose degradation. FEMS symposium 43. Academic Press, Inc., New York.
2. **Beguin, P., and H. Eisen.** 1978. Purification and partial characterization of three extracellular cellulases from *Cellulomonas* sp. Eur. J. Biochem. **87**:525–531.
3. **Beguin, P., H. Eisen, and A. Roupas.** 1977. Free and cellulose-bound cellulases in a *Cellulomonas* species. J. Gen. Microbiol. **101**:191–196.
4. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
5. **Dale, R. M.K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13**:31–40.
6. **Fukumori, F., T. Kudo, and K. Horikoshi.** 1987. Truncation analysis of an alkaline cellulase from an alkalophilic *Bacillus* species. FEMS Microbiol. Lett. **40**:311–314.
7. **Gilkes, N. R., D. G. Kilburn, M. L. Langsford, R. C. Miller, Jr., W. W. Wakarchuk, R. A. J. Warren, D. J. Whittle, and W. K. R. Wong.** 1984. Isolation and characterization of *Escherichia coli* clones expressing cellulase genes from *Cellulomonas fimi*. J. Gen. Microbiol. **132**:1377–1384.
8. **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.
9. **Gilkes, N. R., R. A. J. Warren, R. C. Miller, Jr., and D. G. Kilburn.** 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. **263**:10401–10407.
10. **Greenberg, N. M., R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1987. Regulation and initiation of *cenB* transcripts of *Cellulomonas fimi*. J. Bacteriol. **169**:4674–4677.
11. **Greenberg, N. M., R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1987. Regulation, initiation, and termination of the *cenA* and *cex* transcripts of *Cellulomonas fimi*. J. Bacteriol. **169**:646–653.
12. **Hall, J., G. P. Hazlewood, P. J. Barker, and H. J. Gilbert.** 1988. Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. Gene **69**:29–38.
13. **Kotoujansky, A., A. Diolez, M. Boccara, Y. Bertheau, T. Andro, and A. Coleno.** 1985. Molecular cloning of *Erwinia chrysanthemi* pectinase and cellulase structural genes. EMBO J. **4**:781–785.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
15. **Langsford, M. L., N. R. Gilkes, B. Singh, B. Moser, R. C. Miller, Jr., R. A. J. Warren, and D. G. Kilburn.** 1987. Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. FEBS Lett. **225**:163–167.
16. **Langsford, M. L., N. R. Gilkes, W. W. Wakarchuk, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1984. The cellulase system of *Cellulomonas fimi*. J. Gen. Microbiol. **130**: 1367–1376.
17. **Loenen, W. A. M., and W. J. Brammar.** 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene **10**:249–259.
18. **Mackay, R. M., A. Lo, G. Willick, M. Zuker, S. Baird, M. Dove, F. Moranelli, and V. Seligy.** 1986. Structure of a *Bacillus subtilis* endo-β-1, 4-glucanase gene. Nucleic Acids Res. **14**:9159–9170.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **McGavin, M., and C. W. Forsberg.** 1988. Purification and characterization of endoglucanases 1 and 2 from *Bacteroides succinogenes* S85. J. Bacteriol. **170**:2914–2922.
21. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
22. **Mizusawa, S., S. Nishimura, and F. Seela.** 1986. Improvement of the dideoxy chain termination method of DNA sequencing by the use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. **14**:1319–1324.
23. **Nakai, R., S. Horinouchi, and T. Beppu.** 1988. Cloning and nucleotide sequence of a cellulase gene, *casA*, from an alkalophilic *Streptomyces* strain. Gene **65**:229–238.
24. **Nakamura, K., N. Misawa, and K. Kitamura.** 1986. Sequence of a cellulase gene of *Cellulomonas uda* CB4. J. Biotechnol. **4**:247–254.
25. **O'Neill, G. P., S. H. Goh, R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1986. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene **44**:325–330.
26. **Owolabi, J. B., P. Beguin, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1988. Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. Appl. Environ. Microbiol. **54**:518–523.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Stahlberg, J., G. Johansson, and G. Pettersson.** 1988. A binding-site deficient, catalytically active, core protein of endoglucanase III from the culture filtrate of *Trichoderma reesei*. Eur. J. Biochem. **173**:179–183.
29. **Thomas, R.** 1956. Fungal cellulases. Aust. J. Biol. Sci. **9**: 159–183.
30. **Tomme, P., H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens.** 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. **170**:575–581.
31. **Van Tilbeurgh, H., P. Tomme, M. Claeyssens, R. Bhikhabhai, and G. Pettersson.** 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. Separation of functional domains. FEBS Lett. **204**:223–227.
32. **Warren, R. A. J., C. F. Beck, N. R. Gilkes, D. G. Kilburn, M. L. Langsford, R. C. Miller, Jr., G. P. O'Neill, M. Scheufens, and W. K. R. Wong.** 1987. Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. Proteins **1**:335–341.
33. **Whittle, D. J., D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1982. Molecular cloning of a *Cellulomonas fimi* cellulase gene in *Escherichia coli*. Gene **17**:139–145.
34. **Wong, R. W. K., B. Gerhard, Z.-M. Guo, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterization and structure of an endoglucanase gene *cenA* of *Cellulomonas fimi*. Gene **44**:315–324.
35. **Woods, D.** 1984. Oligonucleotide screening of cDNA libraries, p. 1–3. *In* Focus, vol. 6. Bethesda Research Laboratories, Inc., Gaithersburg, Md.
36. **Yanish-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp and pUC vectors. Gene **33**:103–119.
37. **Zacharius, R. M., T. E. Zell, J. H. Morrison, and J. J. Woodlock.** 1969. Glycoprotein staining following electrophoresis on acrylamide gels. Anal. Biochem. **30**:148–152.
38. **Zappe, H., W. A. Jones, D. T. Jones, and D. R. Woods.** 1988. Structure of an endo-β-1,4-glucanase gene from *Clostridium acetobutylicum* P262 showing homology with endoglucanase genes from *Bacillus* spp. Appl. Environ. Microbiol. **54**:1289–1292.