JOURNAL OF BACTERIOLOGY, May 1999, p. 3270–3276 Vol. 181, No. 10
0021-9193/99/$04.00+0

Three Surface Layer Homology Domains at the N Terminus of the *Clostridium cellulovorans* Major Cellulosomal Subunit EngE

YUTAKA TAMARU AND ROY H. DOI*

Section of Molecular and Cellular Biology, University of California, Davis, California 95616

Received 14 December 1998/Accepted 19 March 1999

The gene *engE*, coding for endoglucanase E, one of the three major subunits of the *Clostridium cellulovorans* cellulosome, has been isolated and sequenced. *engE* is comprised of an open reading frame (ORF) of 3,090 bp and encodes a protein of 1,030 amino acids with a molecular weight of 111,796. The amino acid sequence derived from *engE* revealed a structure consisting of catalytic and noncatalytic domains. The N-terminal-half region of EngE consisted of a signal peptide of 31 amino acid residues and three repeated surface layer homology (SLH) domains, which were highly conserved and homologous to an S-layer protein from the gram-negative bacterium *Caulobacter crescentus*. The C-terminal-half region, which is necessary for the enzymatic function of EngE and for binding of EngE to the scaffolding protein CbpA, consisted of a catalytic domain homologous to that of family 5 of the glycosyl hydrolases, a domain of unknown function, and a duplicated sequence (DS or dockerin) at its C terminus. *engE* is located downstream of an ORF, ORF1, that is homologous to the *Bacillus subtilis* phosphomethylpyrimidine kinase (*pmk*) gene. The unique presence of three SLH domains and a DS suggests that EngE is capable of binding both to CbpA to form a CbpA-EngE cellulosome complex and to the surface layer of *C. cellulovorans*.

Clostridium cellulovorans is a gram-negative, mesophilic, anaerobic, spore-forming bacterium, which utilizes not only cellulose but also xylan, pectin, and several other carbon sources (35). *C. cellulovorans* produces an extracellular cellulolytic multienzyme complex, which has been called the cellulosome (17), has a total molecular weight of about one million, and is capable of hydrolyzing crystalline cellulose.

Cellulolytic bacteria inhabit natural environments in which cellulose is present primarily in plant cell walls in association with other polysaccharides (e.g., xylan, mannan, and other hemicellulose components) (27). We have recently found that the *C. cellulovorans* enzyme complex can degrade several components of plant cell walls in addition to cellulose, such as xylan, mannan, lichenan, and pectin (36).

The cellulosome of *C. cellulovorans* is comprised of three major subunits, the scaffolding protein CbpA (4, 32, 34), an exoglucanase ExgS (20), and a third major endoglucanase subunit EngE, previously designated P100 (5, 32), with a molecular mass close to 100 kDa. In addition, the cellulosome contains several enzyme subunits including endoglucanases EngB (7, 8), EngL, and EngY and a mannanase, ManA (36), each containing a duplicated sequence (DS or dockerin) consisting of 22-amino-acid repeats.

Our ultimate goal is to elucidate the relationships between the structure and function and the regulatory systems of a variety of enzymes involved in the cellulosome system of *C. cellulovorans*. In this study, we report the properties of *engE* and EngE. The derived amino acid sequence of EngE has revealed the presence of three repeated surface layer homology (SLH) domains at its N terminus and a duplicated sequence (DS) at its C terminus. These SLHs also have some homology to the hydrophilic domains (HLDs) of the scaffolding protein CbpA (36). The interesting SLH domain structure of EngE suggests that EngE not only binds to the scaffolding protein CbpA through its DS but that it may also bind to the cell surface through its SLH domains.

* Corresponding author. Mailing address: Section of Molecular and Cellular Biology, University of California, Davis, CA 95616. Phone: (530) 752-3191. Fax: (530) 752-3085. E-mail: rhdoi@ucdavis.edu.

MATERIALS AND METHODS

Bacterial strains, cloning vectors, and media. *C. cellulovorans* (ATCC 35296) (35) was used as the source of chromosomal DNA and for the preparation of cellulosomal proteins. *Escherichia coli* XL1-Blue MRF′ and SOLR strains and λZAPII pBluescript SK(−) and pBluescript II KS(+) plasmids served as the cloning hosts and vectors (Stratagene, La Jolla, Calif.), respectively. λ plaques were formed after *E. coli* cells were transfected with λ phages in NZY medium containing 5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). *E. coli* was grown at 37°C in Luria-Bertani (LB) medium supplemented with ampicillin (100 μg/ml) when required. For agar medium, LB was solidified with 1.5% (wt/vol) agar.

Recombinant DNA techniques. Chromosomal DNA from *C. cellulovorans* was isolated as described previously (33). Plasmids from *E. coli* were purified with a Qiagen kit (Qiagen Inc.). Agarose gel electrophoresis, transformation of *E. coli*, and ligation were done by the methods of Sambrook et al. (29). Chromosomal DNA was partially digested with *Eco*RI and electrophoresed on a 0.7% agarose gel. The fragments in the range of 4 to 10 kb were excised from the gel and purified with a GeneClean kit (Bio 101 Inc.). The purified fragments were ligated into the dephosphorylated *Eco*RI site of λZAPII. The ligation mixture was packaged in vitro with GigapackIII Gold (Stratagene). The *C. cellulovorans* genomic library in λZAPII was immunoscreened with an anti-EngE antiserum (diluted 1:500) previously prepared from *C. cellulovorans* (22) and a second antibody, alkaline phosphatase-conjugated goat anti-rabbit immunoglobulin G (diluted 1:3,000) (Bio-Rad). Immunodetection was performed with the Western Detection Kit (Bio-Rad) following the manufacturer's instructions. Positive immunocross-reacting plaques were isolated, and clones were obtained by in vivo excision and rescued with ExAssist helper phage (Stratagene). Positive clones were further screened for endoglucanase activity by overlaying the clones with 0.7% soft agar containing 0.3% carboxymethyl cellulose (CMC) (low viscosity; Sigma). Colonies having carboxymethyl cellulase activity were recognized by the formation of clear halos on a red background after staining with 0.1% Congo red and destaining with 1 M NaCl (1).

Nucleotide sequence analysis. The nucleotide sequences of both strands were determined by the dideoxy chain termination method (30). The original *Eco*RI fragment (pYE1) was subcloned with restriction enzymes, and a series of nested deletion mutants from each fragment was constructed by using an *Exo*III-mung bean nuclease kit (Stratagene). Double-stranded DNA templates were sequenced with the Sequenase version 2.0 (U.S. Biochemical Co.). Both strands were sequenced. Homology searches in GenBank were performed with a BLAST program.

EngE activity and protein assays. The reaction mixture consisted of 100 μl of EngE solution, 500 μl of 1% CMC (medium viscosity; Sigma), and 400 μl of 100 mM sodium acetate buffer (pH 5.0). The mixture was incubated at 50°C for 10 min, and then the reducing sugar as D-glucose was measured by the Somogyi-Nelson method (38). One unit of enzyme activity was defined as the amount of

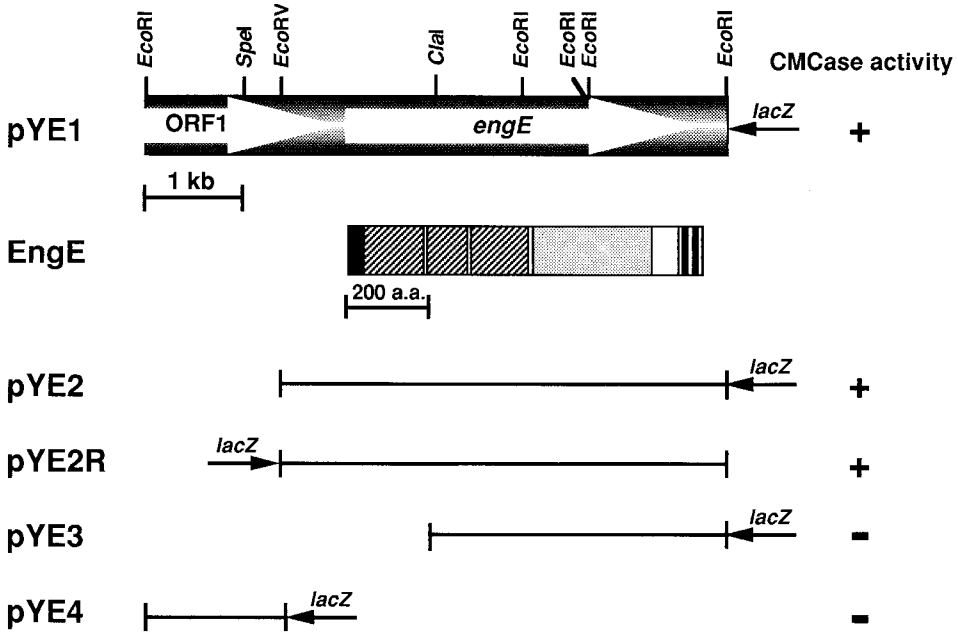

FIG. 1. Restriction enzyme map of the pYE1 insert and the domain structure of EngE. The transformants harboring the plasmids with appropriate deletions were transferred to an LB agar plate. After agar with 0.3% CMC was poured over the transformants, production of carboxymethyl cellulase (CMCase) was judged by the formation of clear halos around the colonies (+, visible halo; −, no halo). The coding sequence and the domain structure of EngE are shown at the top of the figure. Symbols: ■, signal peptide; ▨, SLH domain; ▤, catalytic domain; □, unknown domain; ▥, duplicated sequence (DS). a.a., amino acids.

enzyme that liberates 1 μmol of D-glucose per min under the above conditions. Thin-layer chromatography (TLC) analysis was performed as described previously (14). Purified EngE was incubated with cellodextrins (cellotriose [3G]; cellotetraose [4G]; cellopentaose [5G]) solutions for 16 h at 37°C. The reaction products were separated on TLC plates (Merck) with a solvent system containing 1-butanol–ethanol–water (5:5:2.5). For detection of the products, the plate was sprayed with staining reagent (5% H_2SO_4 in methanol) and baked for 10 min at 100°C.

Protein concentrations were measured by the method of Bradford (3) with a protein assay kit from Bio-Rad, using bovine serum albumin as a standard.

SDS-PAGE and Western blot analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on a 7.5% polyacrylamide gel by the method of Laemmli (16). After electrophoresis, the gel was stained with Coomassie brilliant blue R. A high-molecular-weight SDS calibration kit (Pharmacia) was used as the standard.

For Western blot (immunoblot) analysis, the PhastGel system (Pharmacia) was used. Western blot analysis was performed by using an anti-EngE antiserum (diluted 1:2,000). The procedure was done as described above. Cellulosomes from *C. cellulovorans* were prepared as described previously (22).

Purification of recombinant EngE. *E. coli* harboring pYE2R in pBluescript II KS(+) were cultured to early stationary phase at 37°C with vigorous shaking. The cells were collected by centrifugation, and the periplasmic fraction was prepared by the osmotic shock method (21). The two steps of chromatography were performed with a fast-protein liquid chromatography system (Pharmacia). The periplasmic fraction was applied to a Q Sepharose Fast Flow column (2.6 by 22 cm; Pharmacia) equilibrated with 50 mM Tris-HCl buffer (pH 7.5). After being washed with 3 bed volumes of the same buffer, the column was eluted with a linear gradient of NaCl (0 to 1.0 M) at a flow rate of 60 ml/h. The active fractions eluted around 0.6 M NaCl were collected and concentrated by saturated ammonium sulfate. After centrifugation, the precipitate obtained was dissolved in a small volume of 50 mM Tris-HCl buffer (pH 7.5). The concentrated enzyme solution was applied to a Sephacryl S-200 column (2.6 by 75 cm; Pharmacia) equilibrated with 50 mM Tris-HCl buffer (pH 7.5). The active fractions were used as purified enzyme.

Nucleotide sequence accession number. The nucleotide sequence data reported in this paper have been submitted to GenBank under accession no. AF105331.

RESULTS

Cloning and nucleotide sequence of the *engE* gene. A *C. cellulovorans* λZAPII library was immunoscreened with anti-EngE antibody and further screened for endoglucanase activity. From approximately 20,000 transformants, 5 positive clones were isolated and further characterized. All clones contained a common 6.0-kb *Eco*RI insert. A restriction enzyme map of the cloned gene is shown in Fig. 1. To determine the coding region of the enzyme, various subclones were prepared and tested for the formation of clear halos around the colonies. These results indicated that the coding region for the *engE* gene was on the 4.6-kb *Eco*RV-*Eco*RI (pYE2 or pYE2R) fragment.

Figure 2 shows the complete nucleotide sequence of the *engE* structural gene along with its flanking regions and a partial open reading frame, ORF1, upstream of the *engE* gene. The *engE* gene (nucleotides [nt] 760 to 3849) consists of 3,090 nt encoding a protein of 1,030 amino acids with a predicted molecular weight of 111,796. The putative initiation codon ATG was preceded by a spacing of 7 bp by a typical gram-positive bacterial ribosome-binding sequence, GGAGG, which was homologous to the consensus Shine-Dalgarno sequence. Upstream of the coding region, two possible promoter sequences, AGGAAA for the −35 region and TATAAT for the −10 region with a 17-bp spacing between them, were found and showed high homologies to consensus promoter sequences for σ^{70} or σ^{A} factor, TTGACA and TATAAT with a 17-bp spacing, found in *E. coli* or *Bacillus subtilis* (10). A possible transcription terminator that consisted of a 17-bp palindrome, corresponding to an mRNA hairpin loop with a ΔG of −26.3 kcal/mol (ca. −110 kJ/mol), followed by three T's was found downstream of the TAA termination codon. This structure is similar to the rho-independent terminator factor of *E. coli* (28). Upstream of the *engE* gene, an incomplete ORF1 (nt 1 to 508) encoding a peptide with partial homology to the *B. subtilis* phosphomethylpyrimidine kinase gene (*pmk*) was found.

The codon utilization of the *engE* gene shows a bias for A or T at the wobble position as shown previously for genes from *C. cellulovorans*. The percentages of codons terminating in A or T

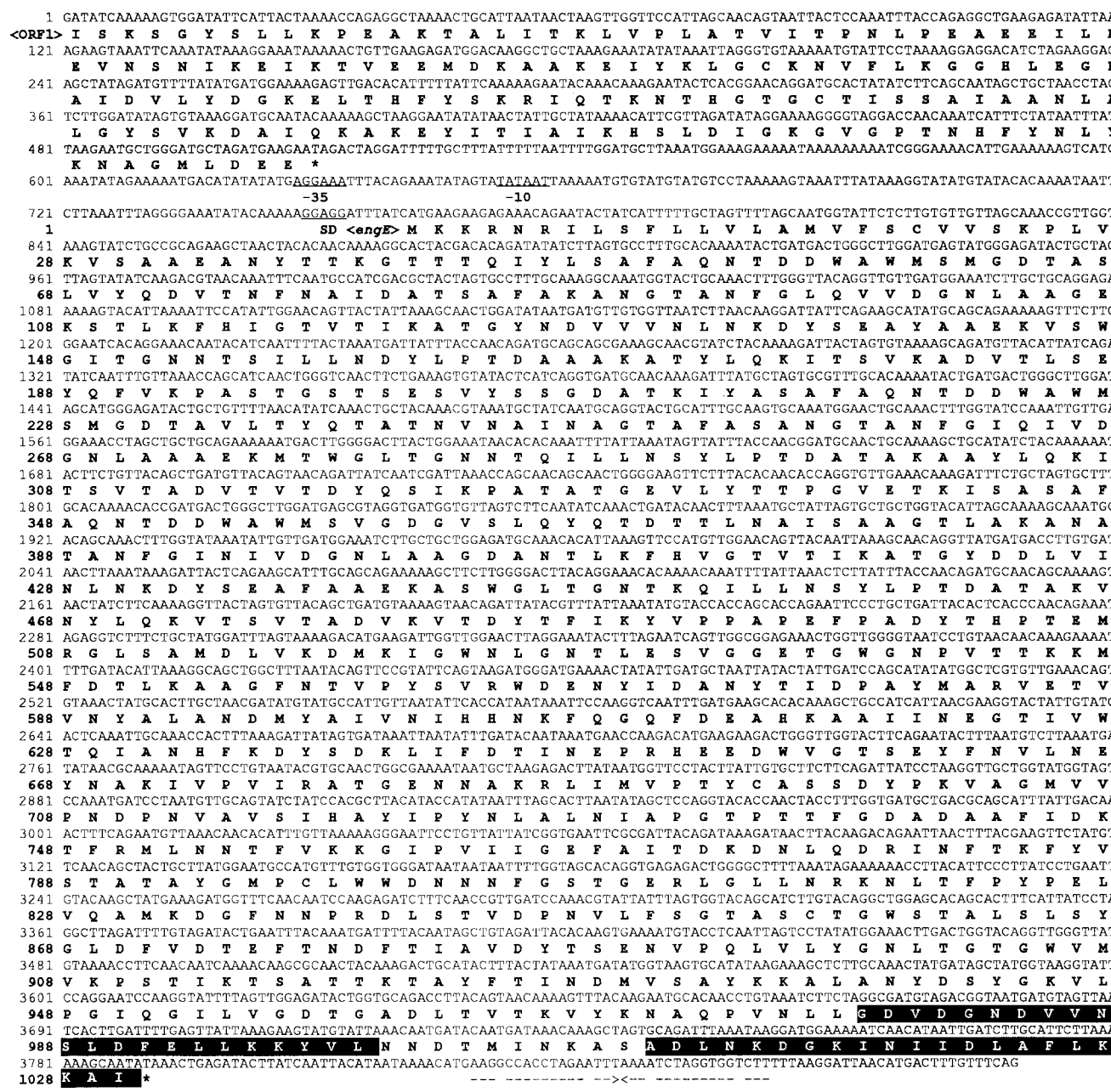

FIG. 2. Nucleotide and deduced amino acid sequences of *engE* and ORF1. The putative promoter and Shine-Dalgarno (SD) sequences are underlined. A palindrome is indicated by arrows facing each other. The stop codons are indicated by asterisks. A duplicated sequence (DS or dockerin) in EngE is shown as white letters on a black background.

are 86.0 for EngE, 79.3 for EngB (8), 80.2 for EngD (12), 81.5 for EngF (31), and 77.5 for ExgS (20), thus reflecting the low G+C content of *C. cellulovorans* DNA (35).

Domain structure of EngE. The N-terminal amino acid sequence of EngE exhibited a typical signal peptide and consensus sequence (Val-X-Ala) (37), where the predicted cleavage site is located between position 31 (Ala) and position 32 (Ala). Removal of the signal peptide yields a mature protein of 999 amino acids with a molecular weight of 108,335.

Analysis of the deduced amino acid sequence of EngE revealed a multidomain structure with a unique feature, i.e., a triplicated SLH domain at the N terminus, followed by a family 5 catalytic domain, a domain of unknown function, and a duplicated sequence (DS or dockerin) (Fig. 1).

The N-terminal half of EngE (residues 32 to 491) contained three highly conserved repeats of 163, 126, and 163 amino acids, with more than 63% identity (Fig. 3A). In addition, the three repeats of EngE showed homology with several SLH domains and most significantly with the HLDs of CbpA (34) (Fig. 3B). Furthermore, the repeated region was homologous to the S-layer protein, RsaA from *Caulobacter crescentus* (2). The region (residues 63 to 272) of the three repeats in EngE showed 16.4% identity and 70.5% similarity with the C-terminal region of RsaA (Fig. 4A), while the region (residues 304 to

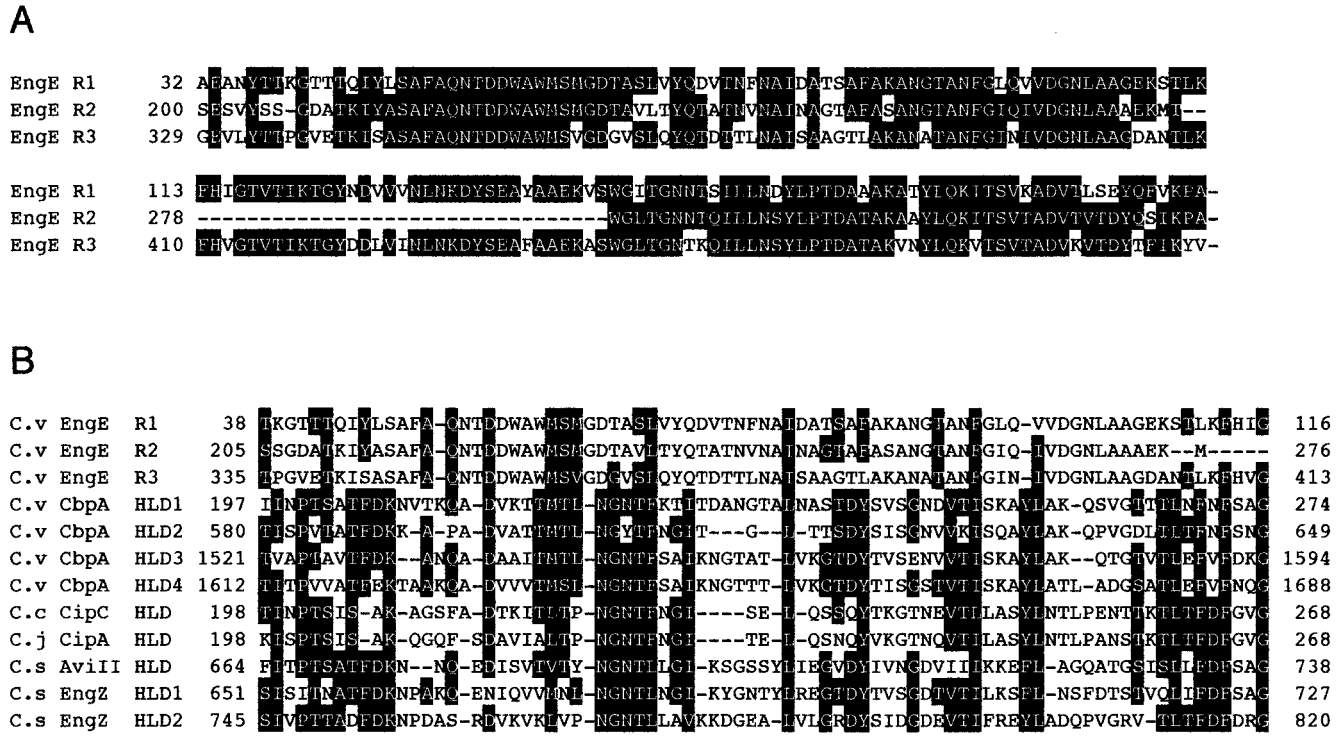

FIG. 3. (A) Alignment of the N-terminal three repeats found in EngE. (B) Alignment of the sequences shown in panel A with similar sequences of HLDs of *C. cellulovorans* (C.v) CbpA, *C. cellulolyticum* (C.c) CipC, *Clostridium josui* (C.j) CipA, and *Clostridium stercorarium* (C.s) AviII and EngZ. Amino acid residues identical or similar to those in the consensus region are shown as white letters on a black background. Similar amino acids are as follows: F, I, V, L, and M; R and K; S and T; D and E; N and Q; and F, Y, and W. Gaps left to improve alignment are indicated by dashes. The numbers refer to amino acid residues at the start and end of the respective lines; all sequences are numbered from Met-1 of the peptide.

426) had 21.1% identity and 69.5% similarity with the N-terminal region of RsaA (Fig. 4B).

The C-terminal half (residues 516 to 839) of EngE showed similarity to catalytic domains of cellulases which belong to glycosyl hydrolase family 5 (13). In particular, the region had high sequence similarity to *C. cellulovorans* EngD (44.2% identity) (12) and *C. cellulovorans* EngB (43.7% identity) (8), *Clostridium thermocellum* CelE (43.6% identity) (11), *Ruminococcus albus* EG-I (41.3% identity) (25), *Orpinomyces* CelB (41.0% identity) (19), *Clostridium longisporum* CelA (40.2% identity) (24), and *Clostridium cellulolyticum* CelA (37.0% identity) (6).

Homology search revealed that a sequence of about 130 amino acids (residues 840 to 970) downstream of the catalytic domain did not resemble any sequences in the protein databases. The function of this region therefore remains unknown.

A duplicated sequence (DS or dockerin) (residues 978 to 1,030) was also found in the C-terminal region of EngE. The DSs consisting of 22-amino-acid repeats are well conserved in cellulosomal subunits from *C. cellulovorans* and other *Clostridium* species.

Purification and characterization of rEngE. Recombinant EngE (rEngE) was purified from the periplasmic fraction of *E. coli* harboring pYE2R by using the Q Sepharose Fast Flow and Sephacryl S-200 columns. After purification and SDS-PAGE, rEngE gave a single band, and the molecular weight of the enzyme was estimated to be around 100,000 (Fig. 5A). This is in good agreement with the value (108,335), excluding the putative signal peptide, calculated from the deduced amino acid sequence. Western blot analysis indicated that rEngE immunoreacted with the anti-EngE antibody prepared from *C. cellulovorans* cellulosomes and that the size of the immunoreactive band of EngE was in good agreement with that of the cellulosomal EngE protein from *C. cellulovorans* (Fig. 5B). Some cross-reaction of the antibody was evident, perhaps due to the presence of DS in EngE, but the major band was found at the migration position of EngE. The optimum pH and temperature for activity of the purified rEngE were 5.0 and 50°C, respectively. As shown in Table 1, purified rEngE hydrolyzed CMC and lichenan but showed little or no activity with acid-swollen cellulose (ASC), Avicel, laminarin, and xylan. TLC analysis revealed that the purified EngE cleaved 3G, 4G, and 5G to form cellobiose (2G) (Table 2) but did not act on cellobiose (data not shown). rEngE seems to hydrolyze preferentially the β-1,4-cellulosidic linkages situated at position 2 and to have transglycosylation ability.

DISCUSSION

The characterization of EngE has completed the sequence analysis of the three major subunits of the *C. cellulovorans* cellulosome. Since EngE has endoglucanase activity, the three major subunits include the endoglucanase EngE, the exoglucanase ExgS (20), and the scaffolding protein CbpA (34). The combination of these three subunits was shown previously to have cellulolytic activity on ASC (4) and to comprise the core of the *C. cellulovorans* cellulosome (32).

EngE belongs to the family 5 cellulases, which comprise a large and growing family of cellulases from various microorganisms (13). It is interesting that EngB (8), EngD (12), EngF (31), and ManA (36) from *C. cellulovorans* are also members of family 5, while EngL and EngY (36) are members of family 9.

A

```
C.v EngE   63 GDTASLVYQDVTNFNAIDATSAFAKANGTANFGLQVVDGNLAAGEKSTLKFHIGTVTIKATGYNDVVVNL
              *.*.  . ..... .... * . . * .... .*. *..   *..... . *..*.*..**. ...**.
C.c RsaA  748 GTTTVTLANATGTSDVFNLTLSSSAALAAGTVALAGVETVNIAATDTNTTAHVDTLTLQATSAKSIVVTG

C.v EngE  133 NKDYSEAYAAEKVSWGITGNNTSILLNDYLPTDAAAKATYLQKITSVKADVTLSEYQFVKPASTGSTSES
              * . . . .....  ........    .   ..*..... . .* .  .. .*.. . .. . .*.....
C.c RsaA  818 NAGLNLTNTGNTAVTSFDASAVTGTAPAVTFVSANTTVGEVVTIRGGAGADSLTGSATANDTIIGGAGAD

C.v EngE  203 VYSSGDATKIYASAFAQNTDDWAWMSMGDTAVLTYQTATNVNAINAGTAFASANGTANFGIQIVDGNLAA
              .   ...*...... .... * . .. ... * . ..*.. .   .*.. ..* ....** ... *. *.
C.c RsaA  888 TLVYTGGTDTFTGGTGADIFDINAIGTSTAFVTITDAAVGDKLDLVGISTNGAIADGAFGAAVTLGAAAT
```

B

```
C.v EngE  304 LQKITSVTADVTVTDYQSIKPATATGEVLYTTPGVETKISASAFAQNTDDWAWMSVGDGVSLQY
              * * .......*...*.. ..... ...   * ...   *..* * .* ..     .....*..
C.c RsaA   17 LGKAPDAATTLTLDAYATQTQTGGLSDAAALTNTLKLVNSTTAVAIQTYQFFTGVAPSAAGLDF

C.v EngE  382 QTDTTLNAISAAGT----LAKANATANFGINIVDGNLAAGDANTLKFH-VGTVTIKATGYDDLV
               .*.* *. . ...    .*..*   **.**...*. *...* .  .  *. . . **.**...
C.c RsaA   97 LVDSTTNTNDLNDAYYSKFAQENRFINFSINLATGAGAGATAFAAAYTGVSYAQTVATAYDKII
```

FIG. 4. Alignment of N-terminal (A) and C-terminal (B) regions in the triplicated SLH domain of *C. cellulovorans* (C.v) EngE with S-layer protein of *Caulobacter crescentus* (C.c) RsaA (2). Identical and similar amino acid residues are indicated by asterisks and dots, respectively. Gaps left to improve alignment are indicated by dashes. The numbers refer to amino acid residues at the start of the respective lines; all sequences are numbered from Met-1 of the peptide.

In this study, we purified and characterized rEngE and compared some properties of EngE with those of EngB, EngD, and EngF from *C. cellulovorans*. EngE had the highest activity with CMC as the substrate, less activity with lichenan (about 60% of the activity as shown with CMC), and very little or no activity with xylan, while both EngB and EngD had much higher activity with lichenan than with CMC and xylan (7). Furthermore, the hydrolysis pattern of EngE analyzed by TLC showed that the main products from cellodextrins (3G to 5G) were only cellobiose (2G), while EngB, EngD, and EngF produced glucose (1G), 2G, and 3G, and EngF did not act on 3G (14). Thus, the properties of EngE differ from those of EngB, EngD, and EngF in specific activity, substrates degraded, and the products that are formed, although all the enzymes belong to family 5.

Electron micrographs have shown that cellulosomes are attached to the cell surface (17). In *C. thermocellum*, it has been proposed that the dockerin II present on CipA interacts with a cell surface protein Olp containing a cohesin II (18) and that

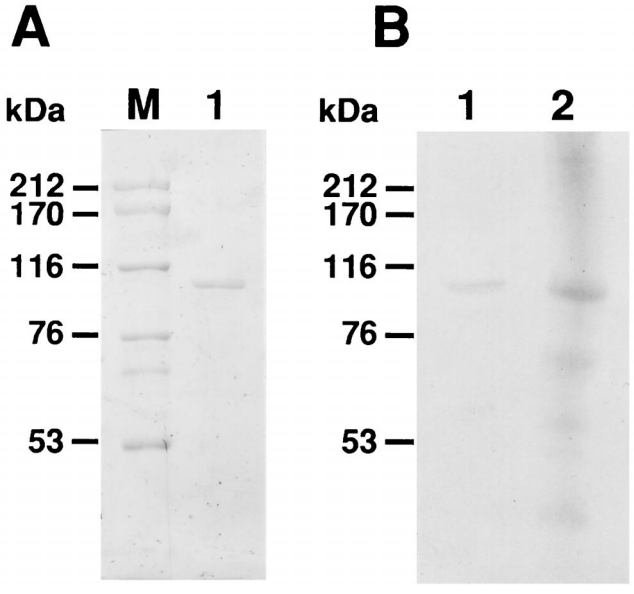

FIG. 5. Analysis of the purified EngE by SDS-PAGE (A) and Western blotting (immunoblotting) (B). Lane M, standard markers (myosin [212 kDa], α2-macroglobulin [170 kDa], β-galactosidase [116 kDa], transferrin [76 kDa], and glutamic dehydrogenase [53 kDa]); lane 1, purified EngE; lane 2, cellulosomal proteins from *C. cellulovorans*.

TABLE 1. Activity of EngE with various substrates[a]

Substrate	Sp act (U/mg of protein)
CMC	106.6
ASC	0.37
Avicel	0.14
Laminarin	0
Lichenan	64.2
Xylan (birch wood)	2.2
Xylan (oat spelt)	3.9

[a] The assay was performed with 1.0% (wt/vol) substrate and 1.34 μg of purified EngE in 1.0 ml of 50 mM sodium acetate buffer (pH 5.0), and the assay was performed at 50°C for 10 min. The reducing sugar formed was measured by the method of Somogyi-Nelson with D-glucose as the standard. One unit of polymer hydrolysis represents 1 μmol of reducing sugar liberated per min per mg of protein.

TABLE 2. TLC analysis of hydrolysis products

Enzyme	Product(s) formed with the following substrate[a]:			Reference
	3G	4G	5G	
EngB	(2G)	2G, 3G	2G, 3G	14
EngD	(1G), (2G)	(1G), 2G, 3G	(1G), 2G, 3G	14
EngF	None	(2G), (3G)	2G, 3G, 4G	14
EngE[b]	2G	2G	2G	This study

[a] Products in parentheses showed faint but significant spots on TLC plates.
[b] The reaction mixture contained 1.34 μg of purified EngE and 3G, 4G, and 5G in 50 mM sodium acetate buffer (pH 5.0). The reaction was performed at 37°C for 16 h.

this interaction binds the cellulosome to the cell surface. The *C. cellulovorans* CbpA (equivalent to CipA), on the other hand, does not contain a dockerin II but contains four conserved HLDs, which are absent from *C. thermocellum* CipA. When the sequences of the HLDs were examined with the BLAST system, it was found that partial homology existed between HLDs and surface layer proteins found on other organisms (Fig. 4). This suggested to us that the HLDs present in CbpA could play a role in binding the cellulosome to the cell surface just as the putative dockerin II-cohesin II reaction holds the cellulosome to the surface of *C. thermocellum*. This role of HLDs in CbpA has not been proven to this point but is being actively examined.

It was therefore of great interest when the analysis of the EngE domain structure revealed the presence of three repeated SLH-like domains at its N terminus. Furthermore, the triplicated SLH domain of EngE showed homology with the S-layer protein RsaA from *C. crescentus* (2). The *C. crescentus* S-layer is paracrystalline in nature, exhibiting an array of ring-like subunits (each composed of six copies of RsaA) arranged on a lattice with p6 symmetry and interlinked at a threefold rotational axis (2). Although the association between RsaA and the outer membrane is not completely understood, the protein appears to be anchored to the outer membrane via noncovalent interactions with a specific smooth lipopolysaccharide molecule (2). So far, the only known function of the *C. crescentus* S-layer is to protect cells against predation by a *Bdellovibrio*-like organism (15). RsaA, as well as the three repeated regions of EngE, contained no cysteine residues. In fact, several S-layer proteins have no cysteine residues (9). Moreover, the N-terminal region of RsaA (residues 1 to 154) was involved in cell surface anchoring, while the C-terminal region (residues 784 to 907) possessed a secretion signal (2).

Many microorganisms have been shown to have surface layer proteins (SLPs) that bind to the peptidoglycan layer (26). Even a single SLH domain can apparently bind a SLP to the

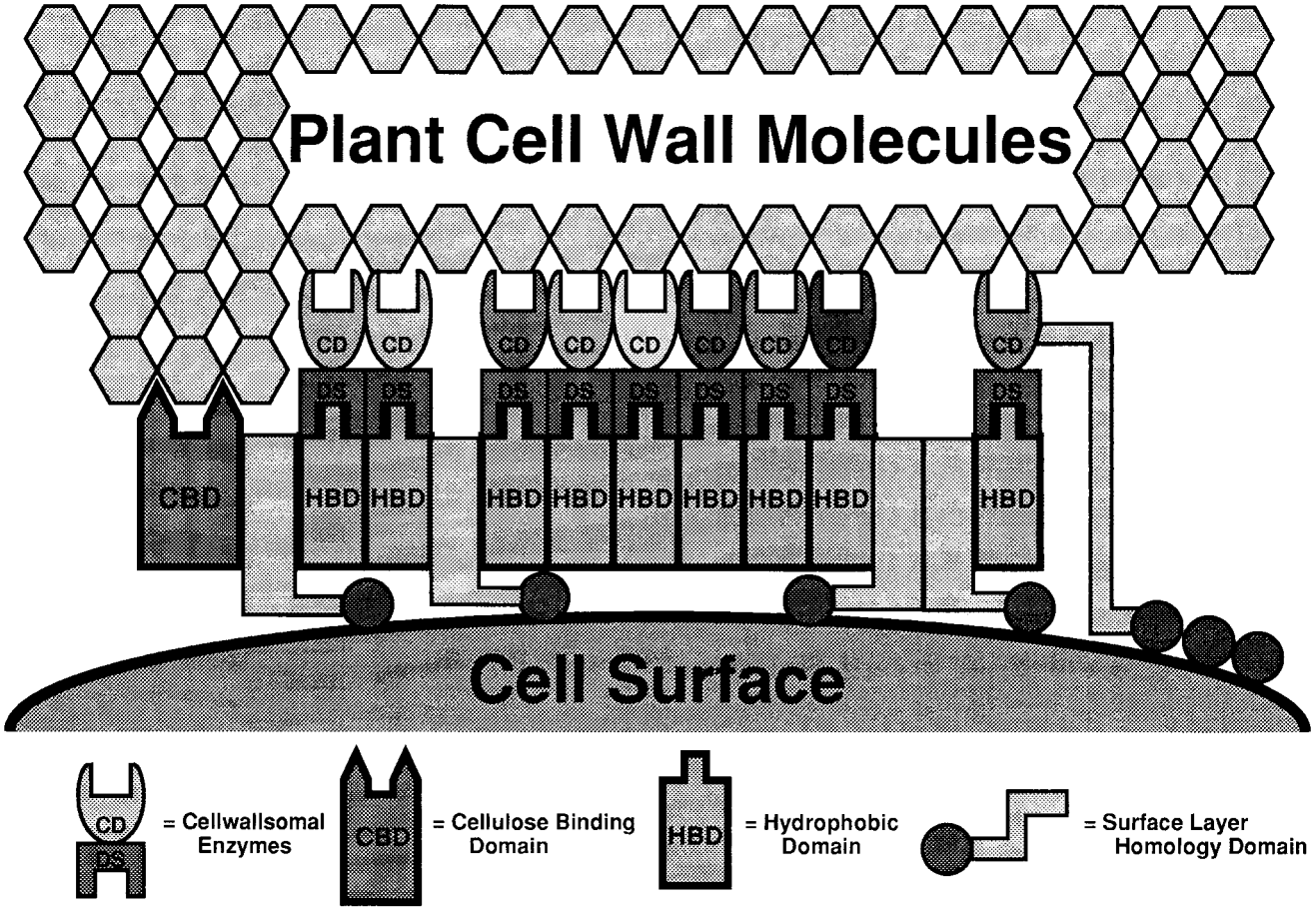

FIG. 6. Hypothetical model for attachment of EngE to the CbpA of *C. cellulovorans* and the cell surface. The drawing is not to scale. Because of its similarity to the S-layer protein and the presence of a duplicated sequence (DS or dockerin), the three SLH domains of the N terminus of EngE integrate into the lattice of the S-layer, while the C terminus of EngE is bound to CbpA through its DS. Also, CbpA integrates itself into the lattice of the S-layer through its four SLH domains. CD, catalytic domain.

peptidoglycan layer. One function proposed for these SLPs is that they serve to anchor extracellular enzymes such as pullulanase to the cell surface (23).

Since the N and C termini of the SLH domains of EngE exhibited homology with both the N- and C-terminal regions of RsaA, the SLH domains of EngE may also possess similar functions, i.e., anchoring the enzyme to the cell surface and also in some targeting function to the cell surface.

By isolating *C. cellulovorans* peptidoglycan or cell surface protein fractions and interacting them with the SLH domains from EngE and CbpA, we should be able to test the hypothesis that SLH domains are involved in binding cellulosomes and cellulases to the cell surface (Fig. 6). A positive result would further strengthen the hypothesis that the cellulosomes are attached to the cell surface by interaction of the SLH domains from both CbpA and its constituent enzymes such as EngE. This hypothesis is currently being investigated.

ACKNOWLEDGMENTS

This research was supported in part by grant DE-DDF03-92ER20069 from the U.S. Department of Energy.

Cellobiopentaose was provided by Seikagaku America, Inc.

REFERENCES

1. **Ali, B. R. S., M. P. M. Romaniec, G. P. Hazlewood, and R. B. Freedman.** 1995. Characterization of the subunits in an apparently homogeneous subpopulation of *Clostridium thermocellum* cellulosomes. Enzyme Microb. Technol. **17:**705–711.
2. **Bingle, W., J. F. Nomellini, and J. Smit.** 1997. Linker mutagenesis of the *Caulobacter crescentus* S-layer protein: toward a definition of an N-terminal anchoring region and a C-terminal secretion signal and the potential for heterologous protein secretion. J. Bacteriol. **179:**601–611.
3. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
4. **Doi, R. H., M. A. Goldstein, J.-S. Park, C.-C. Liu, Y. Matano, M. Takagi, S. Hashida, F. C. F. Foong, T. Hamamoto, I. Segel, and O. Shoseyov.** 1994. Structure and function of the subunits of the *Clostridium cellulovorans* cellulosome, p. 43–52. *In* K. Shimada, S. Hoshino, K. Ohmiya, K. Sakka, Y. Kobayashi, and S. Karita (ed.), Genetics, biochemistry and ecology of lignocellulose degradation. Uni Publishers, Tokyo, Japan.
5. **Doi, R. H., J.-S. Park, C.-C. Liu, L. M. Malburg, Y. Tamaru, A. Ichi-ishi, and A. Ibrahim.** 1998. Cellulosome and noncellulosomal cellulases of *Clostridium cellulovorans*. Extremophiles **2:**53–60.
6. **Faure, E., A. Belaich, C. Bagnara, C. Gaudin, and J. P. Belaich.** 1989. Sequence analysis of the *Clostridium cellulolyticum* endoglucanase-A-encoding gene, celCCA. Gene **84:**39–46.
7. **Foong, F., and R. H. Doi.** 1992. Characterization and comparison of *Clostridium cellulovorans* endoglucanase-xylanase EngB and EngD hyperexpressed in *Escherichia coli*. J. Bacteriol. **174:**1403–1409.
8. **Foong, F., T. Hamamoto, O. Shoseyov, and R. H. Doi.** 1991. Nucleotide sequence and characteristics of endoglucanase gene *engB* from *Clostridium cellulovorans*. J. Gen. Microbiol. **137:**1729–1736.
9. **Gilchrist, A., J. A. Fisher, and J. Smit.** 1992. Nucleotide sequence analysis of the gene encoding the *Caulobacter crescentus* paracrystalline surface layer protein. Can. J. Microbiol. **38:**193–202.
10. **Goldstein, M. A., and R. H. Doi.** 1995. Prokaryotic promoters in biotechnology. Biotechnol. Annu. Rev. **1:**105–128.
11. **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.
12. **Hamamoto, T., F. Foong, O. Shoseyov, and R. H. Doi.** 1992. Analysis of functional domains of endoglucanases from *Clostridium cellulovorans* by gene cloning, nucleotide squencing and chimeric protein construction. Mol. Gen. Genet. **231:**472–479.
13. **Henrissat, B., and A. Bairoch.** 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **293:**781–788.
14. **Ichi-ishi, A., S. Sheweita, and R. H. Doi.** 1998. Characterization of EngF from *Clostridium cellulovorans* and identification of a novel cellulose binding domain. Appl. Environ. Microbiol. **64:**1086–1090.
15. **Koval, S. F., and S. H. Hynes.** 1991. The effect of paracrystalline protein surface layers on predation by *Bdellovibrio bacteriovorus*. J. Bacteriol. **173:** 2244–2249.
16. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
17. **Lamed, R., and E. A. Bayer.** 1988. The cellulosome concept: exocellular and extracellular enzyme factor centers for efficient binding and cellulolysis, p. 101–116. *In* J.-P. Aubert, P. Béguin, and J. Millet (ed.), Biochemistry and genetics of cellulose degradation. Academic Press, Inc., San Diego, Calif.
18. **Leibovitz, E., M. Lemaire, I. Miras, S. Salamitou, P. Beguin, H. Ohayon, P. Gounon, M. Matuschek, K. Sahm, and H. Bahl.** 1997. Occurrence and function of a common domain in S-layer and other exocellular proteins. FEMS Microbiol. Rev. **20:**127–133.
19. **Li, X. L., H. Chen, and L. G. Ljungdahl.** 1997. Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. Appl. Environ. Microbiol. **63:**628–635.
20. **Liu, C.-C., and R. H. Doi.** 1998. Properties of *exgS*, a gene for a major subunit of the *Clostridium cellulovorans* cellulosome. Gene **211:**39–47.
21. **Manoil, C., and J. Beckwith.** 1986. A genomic approach to analyzing membrane protein topology. Science **233:**1403–1408.
22. **Matano, Y., J.-S. Park, M. A. Goldstein, and R. H. Doi.** 1994. Cellulose promotes extracellular assembly of *Clostridium cellulovorans* cellulosomes. J. Bacteriol. **176:**6952–6956.
23. **Matuschek, M., G. Burchhart, K. Sahm, and H. Bahl.** 1994. Pullulanase of *Thermoanaerobacterium thermosurigenes* EM1 (*Clostridium thermosurigenes*): molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell surface. J. Bacteriol. **176:**3295–3302.
24. **Mittendorf, V., and J. A. Thomson.** 1993. Cloning of an endo-(1,4)-beta-glucanase gene, *celA*, from the rumen bacterium *Clostridium* sp. ('*C. longisporum*') and characterization of its product, CelA, in *Escherichia coli*. J. Gen. Microbiol. **139:**3233–3242.
25. **Ohmiya, K., T. Kajino, A. Kato, and S. Shimizu.** 1989. Structure of a *Ruminococcus albus* endo-1,4-β-glucanase gene. J. Bacteriol. **171:**6771–6775.
26. **Olabarría, G., J. L. Carrascosa, M. A. de Pedro, and J. Berenguer.** 1996. A conserved motif in S-layer proteins is involved in peptidoglycan binding in *Thermus thermophilus*. J. Bacteriol. **178:**4765–4772.
27. **Pohlschröder, M., S. B. Leschine, and E. Canale-Parola.** 1994. Multicomplex cellulase-xylanase system of *Clostridium papyrosolvens* C7. J. Bacteriol. **176:** 70–76.
28. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13:** 319–353.
29. **Sambrook, J., J. E. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
31. **Sheweita, S., A. Ichi-ishi, J.-S. Park, C.-C. Liu, L. M. Malburg, and R. H. Doi.** 1996. Characterization of *engF*, a gene for a non-cellulosomal *Clostridium cellulovorans* endoglucanase. Gene **182:**163–167.
32. **Shoseyov, O., and R. H. Doi.** 1990. Essential 170 kDa subunit for degradation of crystalline cellulose of *Clostridium cellulovorans* cellulase. Proc. Natl. Acad. Sci. USA **87:**2192–2195.
33. **Shoseyov, O., T. Hamamono, F. Foong, and R. H. Doi.** 1990. Cloning of *Clostridium cellulovorans* endo-1,4-β-glucanase genes. Biochem. Biophys. Res. Commun. **169:**667–672.
34. **Shoseyov, O., M. Takagi, M. A. Goldstein, and R. H. Doi.** 1992. Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A (CbpA). Proc. Natl. Acad. Sci. USA **89:**3483–3487.
35. **Sleat, R., R. A. Mah, and R. Robinson.** 1984. Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp. nov. Appl. Environ. Microbiol. **48:**88–93.
36. **Tamaru, Y., and R. H. Doi.** 1999. Unpublished results.
37. **von Heijne, G.** 1985. Signal sequences. The limits of variation. J. Mol. Biol. **184:**99–105.
38. **Wood, W. A., and K. M. Bhat.** 1988. Methods for measuring cellulase activities. Methods Enzymol. **160:**87–112.