Transcription of *Clostridium cellulovorans* Cellulosomal Cellulase and Hemicellulase Genes

Sung Ok Han,1 Hideaki Yukawa,2 Masayuki Inui,2 and Roy H. Doi1*

Section of Molecular and Cellular Biology, University of California, Davis, California 95616, 1 and Molecular Microbiology and Genetics Lab, RITE Institute, Kyoto 619-0292, Japan2

Received 10 January 2003/Accepted 3 February 2003

Transcription of the cellulosomal cellulase/hemicellulase genes of *Clostridium cellulovorans* has been investigated by Northern blot, reverse transcriptase PCR (RT-PCR), primer extension, and S1 nuclease analysis. Northern hybridizations revealed that the cellulosomal *cbpA* gene cluster is transcribed as polycistronic mRNAs of 8 and 12 kb. The 8-kb mRNA coded for *cbpA* and *exgS*, and the 12-kb mRNA coded for *cbpA*, *exgS*, *engH*, and *engK*. The sizes of the mRNAs were about 3 kb for *engE*, 1.8 kb for *manA*, 2.7 kb for *xynA*, and 4 kb for *pelA*, indicating monocistronic transcription of these genes. Primer extension and S1 nuclease analysis of *C. cellulovorans* RNA showed that the transcriptional start sites of *cbpA*, *engE*, *manA*, and *hbpA* were located 233, 97, 64, and 61 bp upstream from the first nucleotide of each of the respective translation initiation codons. Alignment of the *cbpA*, *engE*, *manA*, and *hbpA* promoter regions provided evidence for highly conserved sequences that exhibited strong similarity to the ρ5 consensus promoter sequences of gram-positive bacteria.

Cellulosic biomass is a widely available substrate, and utilization of the biomass for production of energy carriers is considered an environmentally friendly process (10). *Clostridium cellulovorans* (ATCC 35296), an anaerobic, mesophilic, and spore-forming bacterium, produces cellulasmes and hemicellulasmes, which have the ability to degrade cellulosic materials either as free enzymes or as an enzyme complex called the “cellulosome” (3, 9, 23). The cellulosomal enzymes include family 5 and 9 endoglucanases, a family 48 exoglucanase, a mannanase, a xylanase, and a pectate lyase (9, 11, 16-19, 26, 34). Members of the following gene cluster have been cloned and sequenced (9, 22, 25): cellulosomal subunits coding for the scaffolding protein CbpA; the exoglucanase ExgS; the endoglucanases EngH, EngK, EngL, EngM, and EngN; and a mannanase, ManA; as well as a gene coding for the small protein HbpA, which contains a single SLH domain at its N terminus and a single cohesin at its C terminus.

Despite this rather detailed account of the components of the *C. cellulovorans* cellulosome, there is comparatively little information available on how these genes are regulated and what their pattern of transcription is. In addition, the inducers of these enzymes and the regulation of the assembly of cellulomasmes have not been studied in depth. In order to gain some understanding of the expression pattern of the genes in this cluster.

**MATERIALS AND METHODS**

**Bacterial strains and growth conditions.** *C. cellulovorans* ATCC 35296 was used as the source of genomic DNA and total RNA. The organism was grown under strictly anaerobic conditions at 37°C in 500- or 3,000-ml round-bottom flasks containing a previously described medium (21, 23), which included 0.5% (wt/vol) of several carbon sources (cellulose [Avicel], cellulbiose, locust bean gum, pectin, and xylan [oat spelt or birchwood]).

**Nucleic acid isolation.** Chromosomal DNA of *C. cellulovorans* was isolated by using the Promega genomic DNA purification kit according to the manufacturers’ instructions. Total RNA was extracted from *C. cellulovorans* broth cultures with the Qiagen RNaseasy kit (Qiagen), with an additional step of treatment with RNase-free DNase (Promega) according to the manufacturers’ instructions.

**Northern hybridization.** RNA samples (up to 20 μg) were denatured in RNA sample buffer, which consisted of 150 μl of formamide, 83 μl of 37% (wt/vol) formaldehyde, 83 μl of Promega blue/orange loading dye, and 50 μl of 10× MOPS (morpholinopropanesulfonic acid) buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA [pH 7.0], 34 μl of distilled water) at 65°C for 10 min and separated through 1% agarose gels in MOPS buffer with 17% (vol/vol) formaldehyde. DNA probes were synthesized by PCR with specific oligonucleotides derived from the *C. cellulovorans* sequence as a template (Table 1). The probes were nonradioactively labeled by random priming with digoxigenin (DIG)-High Prime (Roche). RNA was transferred overnight to a positively charged nylon membrane (Roche) by capillary transfer with 20× SSC (3 M NaCl, 300 mM trisodium citrate [pH 7.0]). Hybridization was carried out for 16 to 20 h at 50°C in DIG Eazy Hyb buffer solution (Roche). Washing of the membrane and detection of specific transcripts on the blots were carried out with the Roche DIG luminescent detection kit according to the kit’s protocol.

**RT-PCR and primer extension analysis.** Reverse transcriptase (RT) reactions were performed on total RNA with a commercially available reverse transcription system (Promega) with slight modifications to the recommended protocol. RT reactions were performed in a final volume of 20 μl, which contained 5 mM MgCl2, RT buffer (10 mM Tris-HCl [pH 9.0], 50 mM KCl, 0.1% Triton X-100), 1 mM (each) deoxynucleoside triphosphates, 1 U of recombinant RNasin RNase inhibitor, 15 U of avian myeloblastosis virus (AVM) RT, 0.25 μM oligonucleotide primer, and 10 μg of substrate RNA. Reaction mixtures were incubated at 42°C for 60 min, and reactions were terminated by heating the mixtures at 95°C for 5 min, followed by incubation on ice for 5 min. The cDNA products were then character-
amplified in 25-μl PCR mixtures with 2.5 μl of the RT reaction mixture as the template. Primer extension reactions were carried out with AMV RT according to the manufacturer’s instructions (Promega). IRDye700 or IRDye800-labeled oligonucleotides were obtained from Li-cor and were used to determine a transcription start point (Table 1). The annealing of 1 pmol of each primer was performed with approximately 10 to 100 μg of total RNA for 18 h at 55°C. Hybridization products were ethanol precipitated and analyzed on a 4% polyacrylamide sequencing gel running on a Li-cor model 4200L DNA sequencer (Li-cor) as instructed by the manufacturer. The length of the labeled fragments was prepared by PCR amplification with the template DNA and reverse oligonucleotide (complementary to the region about 100 bp downstream of the 5’ end of the proposed coding region) labeled at the 5’ end by IRDye700 or IRDye800-labeled oligonucleotides and were used to determine a transcription start point (Table 1). The annealing of 1 pmol of each primer was performed with approximately 10 to 100 μg of total C. cellulovorans RNA.

**S1 nucleic acid protection analysis.** The labeled fragments were prepared by PCR amplification with the template DNA and reverse oligonucleotide (complementary to the region about 100 bp downstream of the 5’ end of the proposed coding region) labeled at the 5’ end by IRDye700 or IRDye800-labeled oligonucleotides and were used to determine a transcription start point (Table 1). The annealing of 1 pmol of each primer was performed with approximately 10 to 100 μg of total C. cellulovorans RNA.

**Nonradioactive detection of the transcription start point.** The primer extension reaction and S1 nuclease digestion products were ethanol precipitated and analyzed on a 4% polyacrylamide sequencing gel running on a Li-cor model 4200L DNA sequencer (Li-cor) as instructed by the manufacturer. The length of each of the loading samples was determined by running sequence reactions with the same primers on the same gel. The sequencing reactions were performed with the SequiTherm cycle sequencing kit for Li-cor (Epicentre) by using a thermocycler.

**RESULTS**

Transcriptional analysis of the cellulase gene cluster. The tandem organization of the cellulase and hemicellulase genes in the *cbpA* gene cluster (Fig. 1A) allowed us to predict that the genes would be transcribed as mono- or polycistronic mRNAs. To examine this prediction and to find the lengths of the transcripts, the biosynthesis of mRNA under various conditions was measured by Northern hybridization and RT-PCR analyses. We isolated total RNA from *C. cellulovorans* cells grown in medium containing cellulose, cellobiose, locust bean gum, pectin, or xylan as the sole energy source. Although usually the same concentration of RNA was used for comparisions, the biosynthesis of mRNA under various conditions was measured by Northern hybridization and RT-PCR analyses.
tical. These experiments primarily demonstrated the sizes of the transcripts and the relative locations of the genes.

In Northern blot analyses, membranes containing RNA were hybridized with intragenic probes derived from the cellulase and hemicellulase genes. In these experiments, we used the same amount of RNA during the tests for comparative purposes. With the cbpA or exgS probes, a prominent band corresponding to an 8-kb transcript was detected (Fig. 1B, lanes 1 and 2). Also a weak additional hybridizing band 12 kb in length was detected under all conditions. The levels of both the 8- and 12-kb transcripts were enhanced when cellulose (Avicel) was included in the medium (Fig. 1B, lanes 1 and 2). In addition, the 12-kb transcript was detected when total RNA was hybridized with an engH internal probe (Fig. 1B, lane 3). The prominent shorter 8-kb band may represent mRNA processing of the full-length 12-kb cbpA-exgS-engH-engK transcript and not an independent transcript, since no transcription termination signal was evident. No additional promoters could be detected downstream of the cbpA, exgS, and engH promoter regions (see next section).

To determine whether 12-kb RNA processing occurred, RNAs prepared at different time points after the addition of rifampin were analyzed by Northern hybridization. If processing occurred following the termination of transcription by addition of rifampin, then the amount of the 12-kb transcript should decrease, and the amount of the 8-kb transcript should increase over time. However, both the 8- and 12-kb transcripts decreased together after the addition of rifampin (Fig. 2). This indicated strongly that the 8-kb transcript was not a processed product of the longer 12-kb message and that transcription termination could yield either an 8-kb product or a longer run-through product of 12 kb. The 8-kb mRNA was large enough to code for CbpA and ExgS. The 12-kb mRNA coded for CbpA, ExgS, EngH, and EngK (see the summary of these results in Fig. 6).

The RT-PCR analysis was also performed with the primers EngH-SIMP-R and EngH-DIG-R (Table 1), which were specific for identifying the putative exgS-engH transcript. The mRNA products from the RT-PCR analysis were found 112
and 1,044 nucleotides (nt) downstream of the engH start codon, respectively (Fig. 3A and B). The RT-dependent products indicated that the exgS and engH genes were transcriptionally coupled (Fig. 3B, lanes 1 and 2). The amplified products were similar to that seen in the control reaction with chromosomal DNA as a template with the same PCR primers. A negative control using the RNA that was not reverse transcribed was included to ensure that there was no contaminating DNA present in the RNA preparation (data not shown). These results, summarized in Fig. 3A suggested that the cbpA and exgS genes were in an operon and that the transcription of this operon could continue through engH to the engK gene (see the summary in Fig. 6).

Probes that consisted of the regions of hbpA and engM hybridized to 6.0- and 2.7-kb mRNA species, respectively (Fig. 3B, lanes 4 and 6). However, it was difficult by Northern blot analyses to probe for the hbpA and engM genes, since the apparent low level of transcription of these genes was sometimes below the level of detection by the Northern blot assay and/or the RNA transcripts were unstable. Therefore, the alternative method of RT-PCR analysis was performed with specific primers such as HbpA-S1MP-F2 and EngM-S1MP-R for the detection of the putative hbpA and engM transcripts.

The resultant cDNA molecules were then amplified by PCR with primers HbpA-S1MP-F2 and HbpA-S1MP-R, which produced a product of 560 bp (Fig. 3B, lane 3). With primers HbpA-DIG-F and HbpA-DIG-R, the cDNA molecules were also amplified and produced a product of 648 bp (Fig. 3B, lane 4). The oligonucleotide HbpA-S1MP-F2 binds within engK, 498 bp upstream of the hbpA start codon (Fig. 3A, lane 3). These RT-dependent PCR products were observed, indicating that the engK and hbpA genes were linked (see summary in Fig. 6). To determine whether a promoter was present upstream of the hbpA gene, primer extension analysis was carried out with RNA from cellulose and cellobiose cultures by using primer HbpA-S1MP-R, which was complementary to a region 62 nt downstream of the initiation codon of hbpA. Contrary to the RT-PCR results, a transcriptional start point was detected (as

FIG. 3. RT-PCR analysis of engH, hbpA, engM, and xynA RNA. (A) Schematic showing of PCR-amplified transcripts. The locations and extents of the exgS, engH, engK, hbpA, engL, manA, engM, and xynA genes are shown. The oligonucleotide primers used in RT-PCR analysis are numbered as follows: 1, EngH-S1MP-F2 and EngH-S1MP-R; 2, EngH-DIG-F and EngH-DIG-R; 3, HbpA-S1MP-F2 and HbpA-S1MP-R; 4, HbpA-DIG-F and HbpA-DIG-R; 5, EngM-S1MP-F2 and EngM-S1MP-R; 6, EngM-DIG-F and EngM-DIG-R; 7, XynA-S1MP-F2 and XynA-S1MP-R; and 8, XynA-DIG-F and XynA-DIG-R. (For more details, see Table 1.) The locations and sizes of the PCR-amplified transcripts are indicated by the bars. The black and white bars represent the transcription status shown in the presence and absence of PCR amplification, respectively. (B) Agarose gel electrophoresis of RT-PCR products. RT-PCR analysis was performed with total RNA isolated from C. cellulovorans in the medium containing cellulose (Avicel). In negative controls, the reactions were performed in the absence of RT or RNA template (data not shown).
given in more detail in the next section). These results indicated that the engK and hbpA genes formed an operon with a promoter in front of the engK gene as well as an internal promoter proximal to the hbpA gene (see summary in Fig. 6).

In confirmation of the Northern blot analysis (Fig. 1B, lane 6), RT-PCR detected the engM transcripts (Fig. 3B, lane 6). Upon analysis of the upstream region of EngM by RT-PCR, no cDNA products were detected (Fig. 3B, lane 5). This suggested that engM was transcribed as a monocistronic gene.

Northern hybridizations of the RNA with a manA probe showed a single transcript of 1.8 kb (Fig. 1B, lane 5). This mRNA was in good agreement with the size of the manA gene (1,275 nt) and indicated that manA was a monocistronic gene within the cbpA gene cluster.

Transcriptional analysis of cellulosomal genes unlinked to the cbpA cluster. Transcriptional analyses were carried out with cellulosomal genes that were not linked with the cbpA gene cluster. This included engE, xynA, and pelA. The results of Northern analysis with an intragenic probe derived from engE showed that engE formed a transcript of about 3 kb under all conditions (Fig. 1B, lane 7). This mRNA size was in good agreement with the size of the engE gene (3,039 nt) and indicated that engE was a monocistronic gene.

The transcript that included the xynA gene was 2.7 kb in length and was found to be expressed constitutively (Fig. 1B, lane 8). Two additional hybridizing bands 2.3 and 1.6 kb in length for xynA were also detected. These shorter bands probably represented mRNA processing of the full-length xynA transcript or transcription termination events (i.e., putative terminators were found downstream of xynA at 1,693 and 2,335 nt), since no cDNA could be detected upstream of the xynA region (Fig. 3B, lane 7); in addition, xynA mRNA could be detected with an internal probe (Fig. 3B, lane 8). The 1.6-kb mRNA corresponded well in size with the xynA gene (1,560 nt) (Fig. 1B, lane 8).

When C. cellulovorans was grown with pectin as a carbon source, a unique transcript of about 4 kb was detected with a probe internal to the pelA gene (Fig. 1B, lane 9). The longest size of the transcript (4 kb) indicated that it may be part of an operon, because the size of the pelA gene is 2,745 kb.

Determination of transcription start points and promoters for cellulase genes. In order to localize the promoters, the transcription start sites of selected cellulase genes were determined by primer extension analysis and high-resolution S1 nuclease protection analysis. Our initial studies focused on the promoter location for cbpA, the first gene in the cbpA gene cluster. An 866-bp PCR fragment (Cbpa-S1MP-F/CbpA-S1MP-R1) was generated for use as a probe with 512 bp complementary to the upstream and potential promoter region of cbpA. This probe was annealed to total RNA isolated from C. cellulovorans, cultivated with cellulose or cellobiose as the sole carbon source. One clear signal in position bp -233, relative to the cbpA start codon, was obtained by both S1 nuclease protection analysis and primer extension analysis (Fig. 4). In addition, this transcription start point was confirmed by S1 nuclease protection analysis and primer extension analysis with two additional oligonucleotides (Cbpa-S1MP-F/CbpA-S1MP-R2 and CbpA-S1MP-F/CbpA-S1MP-R3) (Fig. 5). The cbpA mRNA start point suggested a putative promoter region, TAAAAAN_{17}TATTAT, with high similarity to the σ^{\text{A}} consen-
The main objective of this study was to examine the transcription of the cellulosomal cellulase and hemicellulase genes of \textit{C. cellulovorans}. We have shown that the \textit{cbpA} and \textit{engX} genes comprise an operon that is transcribed from a single promoter located 233 bp upstream of the translational initiation codon of the \textit{cbpA} gene. The monocistronic transcript of \textit{manA} is located at -64 bp from its first codon. For the polycistronic transcript of \textit{hbpA}, the putative promoter site was found at -61 bp from its first codon by primer extension. In addition, RT-PCR analyses (Fig. 3B, lane 3) indicated that another transcriptional start point for \textit{hbpA} was present upstream of the \textit{hbpA} promoter in the \textit{engK} promoter region. This conclusion is based on the results obtained from Northern hybridization, RT-PCR, primer extension analysis, and S1 nuclelease mapping analysis. The monocistronic transcript of \textit{engE} was located at -97 bp from its first codon.

Alignment of these promoters revealed that they were similar to the consensus \textit{B. subtilis} (\textit{\sigma^70}) promoter (7, 33), and the consensus sequence contains only two mismatches to the \textit{E. coli} \textit{\sigma^70} -35 consensus sequence (15). In all of the putative promoters, the spacing between the -35 and -10 regions was 17 bp, which is the optimal spacing observed for \textit{B. subtilis} and \textit{E. coli} promoter sequences (7, 14). This spacing appears to be a stringent constraint on promoter function, because it is required for recognition by the RNA polymerase holoenzyme (2, 30). The close similarity of the promoter to the consensus \textit{\sigma^70} sequence suggests that these promoters, if not subjected to any regulatory constraints, would act as strong promoters in vivo (8, 15). In addition, comparison of these promoters with a previously determined \textit{C. cellulovorans} promoter revealed some similarity to the \textit{C. cellulovorans} \textit{engB} promoter (1), although the -10 sequence was quite different.

**TABLE 2.** Alignment of putative \textit{C. cellulovorans} promoters located at -10 and -35 bases upstream of the start point of transcription

<table>
<thead>
<tr>
<th>Gene</th>
<th>Leader</th>
<th>Sequence alignment*</th>
<th>Gap (17 bp)</th>
<th>(-10)</th>
<th>TSP$^a$</th>
<th>Coordinate$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{cbpA}</td>
<td>TTTTTA AAAA AAAAA</td>
<td>TAaaAA</td>
<td>aaaaGTTATca TsGATcT</td>
<td>20</td>
<td>AATGAaaAcTaaAAAA</td>
<td>-233</td>
</tr>
<tr>
<td>\textit{engE}</td>
<td>AAAATTTAAAAA</td>
<td>aaaa</td>
<td>TTTA AATGATATGAA</td>
<td>10</td>
<td>TTTAC</td>
<td>-97</td>
</tr>
<tr>
<td>\textit{manA}</td>
<td>TTTATTTAAA</td>
<td>aaaa</td>
<td>TTTA AATGATATGAA</td>
<td>10</td>
<td>TTTAC</td>
<td>-64</td>
</tr>
<tr>
<td>\textit{hbpA}</td>
<td>TTTTTA</td>
<td>aaaa</td>
<td>TTTA AATGATATGAA</td>
<td>10</td>
<td>TTTAC</td>
<td>-61</td>
</tr>
</tbody>
</table>

* The consensus sequence derived from this alignment is given at the bottom. It consists of nucleotides that are present in any given position in more than 50% of the sequences. Promoter sequence nucleotides that match those of the consensus sequence are capitalized and in boldface; those that do not are lowercase and in lightface.

$^a$ TSP, transcription start point (indicated by underline).

$^c$ Nucleotide numbering begins from the first codon. The coordinates refer to the following published nucleotide sequences: \textit{cbpA} (22), GenBank accession no. M73817; \textit{engE} (27), GenBank accession no. AF105331; and \textit{manA} and \textit{hbpA} (25, 28), GenBank accession no. AF132735.
Northern hybridization analyses of the \( cbpA-exgS-engH-engK \) cluster showed two polycistronic transcripts of 8 and 12 kb for \( cbpA \) and a polycistronic \( engH \) transcript of 12 kb that is the same size as that of the 12-kb \( cbpA \) mRNA. The data obtained with rifampin inhibition studies indicated that the smaller 8-kb mRNA was not a processed product of the larger 12-kb mRNA and was an independent transcript. The intergenic region between \( exgS \) and \( engH \) contains an inverted repeat (11 nt) forming a stem-loop structure and a short poly(U) tail (−20 bp from first codon of \( engH \) gene; GCGGCAAAUAUCAAGU AUAUGCUACUACUUU) that may work as a transcriptional terminator. The presence of this putative terminator structure is in agreement with the formation of the transcript (8 kb) from \( cbpA-exgS \). However, in gram-positive bacteria, such as \( B. subtilis \), certain genes are regulated by transcriptional attenuation, translational coupling, or a repressor protein (4, 5, 12, 13, 31, 32). The results suggest that in \( C. cellulovorans \), the putative transcriptional termination signal upstream of \( engH \) may be bypassed by an antitermination mechanism giving rise to a lengthier transcript (12 kb) from \( cbpA-exgS-engH-engK \).

The transcriptional data showed that whereas \( cbpA-exgS-engH-engK, \) \( manA, engE, \) and \( xynA \) were efficiently expressed, relatively low levels of \( hbpA \) and \( engM \) transcripts were present. The ratios of the different cellulase transcripts remained similar irrespective of the culture media, with most of the genes apparently expressed not only in the presence of cellulose, pectin, or xylan, but also in the presence of cellulobiose. Significantly, the major cellulase genes, such as \( cbpA-exgS \) and \( engE \), were constitutively expressed under all conditions tested. On the other hand, \( pelA \) transcripts were only transcribed when the medium contained pectin. Thus, there appears to be regulation of expression of certain hemicellulase genes.

In the general model for induction of cellulase expression, a sensor enzyme is constitutively expressed that hydrolyzes cellulose into oligosaccharides that can enter the bacterium and activate the expression of all the cellulase genes (6, 24, 29). It is possible that one or more of the constitutively expressed \( C. cellulovorans \) cellulases function as the sensor enzyme.

Since the cellulases, especially those coded for by the \( cbpA \) gene cluster, and hemicellulases are involved in the degradation of plant cell wall (9), we wished to determine whether they were coordinately regulated at the transcriptional level. The simultaneous transcription of cellulase (i.e., \( cbpA \) and \( engE \)) and hemicellulase (\( manA \) and \( xynA \)) genes suggests that the expression of the cellulase and hemicellulase is coordinated (S. O. Han and R. H. Doi, unpublished data). Further studies are clearly required to characterize the expression of the cellulase and hemicellulase genes under different conditions and at different growth phases and to determine the mechanism by which cellulose induces expression of the \( C. cellulovorans \) cellulases. Future work will also include studies of catabolite repression and induction of cellulase genes in \( C. cellulovorans \).

More work is needed to identify the factors determining promoter recognition by the RNA polymerase from \( C. cellulovorans \). We hope to gain insight into the properties of \( C. cellulovorans \) promoter regions in order to genetically engineer and to control expression of these genes. Overall, this research may lead to improved understanding of how cellulase and hemicellulase genes are regulated in \( C. cellulovorans \).

ACKNOWLEDGMENTS

This research was supported in part by the Research Institute of Innovative Technology for the Earth (RITE), Japanese Ministry of Economy, Trade and Industry (METI), and grant DE-920067 from the U.S. Department of Energy.