Cloning and Nucleotide Sequence of celA₁, an Endo-β-1,4-Glucanase-Encoding Gene from Streptomyces halstedii JM8

José M. Fernández-Abalos, Pilar Sánchez, Pedro M. Coll, Julio R. Villanueva, PILAR PÉREZ, and Ramón I. Santamaría*

Instituto de Microbiología Bioquímica, Facultad de Biología, Consejo Superior de Investigaciones Científicas/Universidad de Salamanca, 37008-Salamanca, Spain

Received 21 May 1992/Accepted 5 August 1992

The celA₁ gene encoding an endo-β-1,4-glucanase from a mesophilic actinomycete, strain JM8, identified as Streptomyces halstedii, was cloned and expressed in S. lividans J166. From the nucleotide sequence of a 1.7-kb DNA fragment we identified an open reading frame of 963 nucleotides encoding a protein of 321 amino acids, starting at TGTA (instead of ATG). The Cel1 mature enzyme is a protein of 294 amino acids (after signal peptide cleavage) and can be included in the β-glycanase family B (N. R. Gilkes, B. Henrisat, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren, Microbiol. Rev. 55:303–315, 1991). The Cel1 enzyme lacks a cellulose-binding domain as predicted by computer analysis of the sequence and confirmed by Avicel binding experiments. The promoter region of celA₁ was identified by S1 mapping; the –35 region closely resembles those of housekeeping Streptomyces promoters. Three imperfectly repeated sequences of 15, 15, and 14 nucleotides were found upstream from celA₁ [ATTTGGGACCGCTTCC-(N85)-ATTGGGACCGCTTCC-(N2)-TGGGAGCGTCCCA]; The 14-nucleotide sequence has a perfect palindrome identical to that found in several cellulase-encoding genes from Thermomonospora fusca, an alkalophilic Streptomyces strain, and Streptomyces lividans. This sequence has been implicated in the mechanism of induction exerted by cellulose. Using an internal celA₁ probe, we detected similar genes in several other Streptomyces species, most of them cellulase producers.

Many soil bacteria, in particular actinomycetes, participate in the recycling of carbon locked in polymeric materials such as plant matter; their activity is crucial for soil fertility. Cellulose, hemicellulose, and lignin constitute a huge reservoir of polymerized sugars and aromatic compounds that can provide an abundant source of carbon and energy for microorganisms with degradative enzymatic systems. Basic knowledge of these microorganisms and their application in food and industry have attracted the interest of scientists worldwide (8). More than 60 genes encoding endo-β-1,4-glucanases (CMCases) and xylanases have been cloned and studied (reviewed in references 2, 11, and 36).

Fungal and bacterial glycanases were initially grouped into six families (14) and later into nine families (11) on the basis of the primary and secondary structural similarities of their catalytic domains. These families include enzymes from taxonomically unrelated organisms, and even a single organism can produce several enzymes that belong to different families.

In this paper we report the cloning of celA₁, a gene encoding a CMCase (EC 3.2.1.4) from an actinomycete isolated from straw. We identified this strain as Streptomyces halstedii. This actinomycete secretes several proteins with enzymatic activities against carboxymethylcellulose (CMC). Our results show that one of the proteins (Cel1) encoded by celA₁ is a new member of the B family of β-glycanases (11, 14).

MATERIALS AND METHODS

Bacterial strains and plasmids. The following Streptomyces strains were used: S. halstedii JM8 and Streptomyces sp. strain JM5 (strains isolated as cellulase producers by J. M. Fernández-Abalos [9a]); S. lividans J166 (used as DNA recipient in the cloning experiments); S. albus GR2; S. badius ATCC 39117; S. celluloflavus ATCC 29806; and S. coelicolor A3(2); S. flavogriseus ATCC 33331; S. griseus ATCC 10137; S. halstedii NRRL 2381; S. olivaceus SC 3107; S. viridosporus ATCC 39115; and Streptomyces sp. strain ATCC 11238. Escherichia coli DH5α (13) was used for subcloning and plasmid isolation, whereas E. coli JM101 (32) and MV1190 (Bio-Rad) were used to isolate single-stranded DNA. All of the plasmids used are listed in Table 1.

Media and culture conditions. Streptomyces media, culture conditions, protoplasting, transformation, and selection of transformants were as described by Hopwood et al. (17). Basal medium (20) containing 0.5 ml of Tween 80 per liter and supplemented with different carbon sources was used for cellulase production. All liquid cultures were carried out in 250-ml indented flasks with 25 ml of medium and incubated at 30°C for 72 h at 250 rpm. E. coli strains were grown in Luria broth or Luria agar and 2×YT, depending on the experiment (39), and transformed by the method of Hanahan (13).

When needed, antibiotics were added at the following final concentrations: thiostrepton (a gift from E. R. Squibb and Sons Inc., Princeton, N.J.) (50 µg/ml), ampicillin (100 µg/ml), and kanamycin (70 µg/ml).

Isolation of cellulolytic bacteria. A sample of straw was treated with sterile water, and suitable dilutions were spread on basal agar medium containing 0.5% CMC (low viscosity; Sigma). The plates were incubated at 30°C for 3 days and then stained with Congo red (44). The Streptomyces-like colonies producing clear halos were selected and studied in more detail. Strain identification was carried out as specified in Bergey's Manual of Systematic Bacteriology (47).

Enzyme preparation and gel electrophoresis. The culture supernatants of Streptomyces strains grown on basal me-
dium supplemented with 0.5% Avicel (Merck) were separated from the mycelial mass by centrifugation at 10,000 × g for 10 min. They were then used as crude enzyme preparations to quantify the CMCase activity and employed to analyze the proteins by polyacrylamide gel electrophoresis (PAGE) and analytical isoelectric focusing (IEF). For denaturing sodium dodecyl sulfate (SDS)-PAGE, proteins were precipitated with 10% trichloroacetic acid, resuspended in loading buffer (23), neutralized, boiled for 4 min, and loaded in 15% polyacrylamide gels. Electrophoresis was carried out by the method of Laemmli (23) in a Mini Protein II system (Bio-Rad), and the gels were stained with Coomassie brilliant blue R-250. Protein sizes were estimated by using Bio-Rad low-molecular-weight markers. Endogalactanase activity was detected by the method of Schwarz et al. (41) on renatured SDS-PAGE gels containing 0.1% high-viscosity CMC (Sigma) (SDS-CMC-PAGE). For native PAGE, precipitation of proteins with trichloroacetic acid was omitted, and the SDS was omitted from the loading and running buffers. The zymogram technique of Béguin (1) was used to locate CMCase bands in these gels.

IEF was carried out on a Multiphor II system with Ampholine polyacrylamide gel plates in the pH 3.5 to 9.5 range and the Broad pI Calibration Kit (Pharmacia LKB) for pI markers. Endogalactanase active bands were detected by the replica method of MacKenzie and Williams (29) on a 1% agarose gel (in 100 mM sodium phosphate [pH 7.0]) containing 0.5% CMC and supported on a plastic film (Gelbond; FMC), incubated at 37°C for 3 h, and then stained with Congo red.

Enzyme assays. CMCase activity was measured as reducing sugar production by the Somogyi-Nelson colorimetric assay with glucose as a standard (43). One CMCase unit was defined as the amount of enzyme releasing 1 μM of glucose equivalents per min under standard assay conditions (50 μl of enzyme solution, 1% low-viscosity CMC, and 100 mM sodium phosphate buffer [pH 7.0] in 200 μl [final assay volume]; incubation at 50°C for 5 min). Protein concentrations were estimated by the method of Lowry et al. (28) as modified by Peterson (35) with bovine serum albumin as a standard.

Amino-terminal amino acid sequencing. Proteins from culture supernatants of S. lividans carrying plasmid pJM1 (Table 1) were separated on SDS-PAGE gels and electro-

phoretically blotted onto Immobilon-P membranes (Millipore) by the protocol of Matsudaira (31). To minimize the possibility of modification of reactive amino acid residues during electrophoresis, 0.1 mM sodium thioglycolate was added to the running buffer in the upper chamber of the electrophoresis apparatus (18). The protein band corresponding to Cel1 enzyme was cut out, and the NH₂-terminal sequence was sequenced directly in an Applied Biosystems 470A protein sequenator.

Recombinant DNA techniques. Total DNA and plasmids were isolated from Streptomyces spp. as described previously (17). E. coli plasmids were isolated as described by Birnboim and Doly (5). Restriction endonucleases, modifying enzymes, and T4 DNA ligase were purchased from Bethesda Research Laboratories, Boehringer-Mannheim Biochemicals, or Amersham and used as specified by the manufacturers' guidelines and standard protocols (17, 39).

Construction of a genomic library of S. halstedii JM8 in plasmid pIJ702 and selection of endogalactanase-positive clones. Total DNA isolated from S. halstedii JM8 was partially digested with Sau3AI and size fractionated in a 0.8% agarose gel. Fragments between 2 and 5 kb were purified by elution (39) and ligated to plasmid pIJ702 that had been digested with BglII and dephosphorylated. DNA ligation mixtures were used to transform S. lividans J166 protoplasts, and transformants were selected on plates of R2YE medium overlaid, after incubation for 16 h at 30°C, with 50 μg of thioestreptone per ml. Sporulated transformants were harvested by adding 5 ml of sterile water to each plate and scraping the surface with a sterile bent spatula. The cotton-filtered spores were stored in 20% glycerol at −20°C.

To select for endogalactanase-producing clones, spore dilutions were plated on R2YE or basal medium with 0.5% CMC and incubated at 30°C for 3 days. The plates were then stained with Congo red, and colonies surrounded by clear halos were selected and used in further studies.

DNA sequencing. The nucleotide sequences of both DNA strands of the celA₁ coding sequence and its flanking regions were obtained by the dideoxy-chain termination method (40) after subcloning in pBluescript KS⁺ and SK⁺ plasmids (Strategene) with M13-KO7 as a helper phage and a Sequenase sequencing kit (U.S. Biochemical Corp.) or a T7 Sequencing Kit (Pharmacia LKB). DNA and protein sequences were analyzed by using the DNASIS, PROSIS

<table>
<thead>
<tr>
<th>Designation</th>
<th>Relevant phenotype</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Streptomyces plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pIJ702</td>
<td>Thioestrepton resistance and melanin production</td>
<td>This report</td>
</tr>
<tr>
<td>pIJ699</td>
<td>Thioestrepton, viomycin, and kanamycin resistance</td>
<td>This report</td>
</tr>
<tr>
<td>pJM1</td>
<td>pIJ702 carrying a 2.5-kb Sau3AI DNA fragment of S. halstedii JM8 containing celA₁ (original clone)</td>
<td>This report</td>
</tr>
<tr>
<td>pJM2</td>
<td>pIJ702 carrying a 5-kb Sau3AI DNA fragment of S. halstedii JM8 containing celA₁ (original clone)</td>
<td>This report</td>
</tr>
<tr>
<td>pJM6</td>
<td>pIJ702 carrying a 2.8-kb Sau3AI DNA fragment of S. halstedii JM8 containing celA₁ (original clone)</td>
<td>This report</td>
</tr>
<tr>
<td>pJM11</td>
<td>pIJ702 carrying a 1.7-kb DNA fragment from S. halstedii JM8 subcloned from pJM1 by using SstI and partial BamHI digestions</td>
<td>This report</td>
</tr>
<tr>
<td><strong>E. coli plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pBluescript KS⁺ and SK⁺</td>
<td>Ampicillin resistance, lacZ complementation, and single-stranded DNA replication</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pBluescript KS⁺ derivative containing the SmaI-SmaI fragment from pJM11 carrying the celA₁ gene (one SmaI site is from the vector pIJ702)</td>
<td>This report</td>
<td></td>
</tr>
</tbody>
</table>
(Pharmacia-LKB, Hitachi), FRAME (4), and BRUJENE (45a) programs. The BRUJENE program is based on the CODON PREFERENCE program (12). The amino acid sequence of Cel1 was compared with the SWISS-PROT data bank by using the FASTA program (34). Alignments of protein sequences were done with CLUSTAL programs (15).

RNA isolation. Streptomyces cells were grown for 60 h in basal liquid medium supplemented with 0.5% Avicel as a carbon source. The mycelia were harvested by centrifugation (10,000 x g, 10 min) and washed twice with chilled 0.3 M sucrose. The RNA was prepared as described by Igo and Losick (19), except that the concentrations of all reagents were scaled for 25-ml cultures. The final RNA pellet was dissolved in diethylylpyrocarbonate-treated water, quantified by spectrophotometry, and stored at -70°C until used.

High-resolution S1 mapping. S1 nuclease protection-hybridization experiments were carried out by the method of Deng et al. (9) with a hybridization temperature of 63°C. A DNA probe was prepared as follows. Plasmid pJM31 was digested with XhoI, which cuts the plasmid at two sites, one in the polylinker and the other in the celA1 region (Table 1; see Fig. 4A). The DNA was dephosphorylated; the 617-bp band containing part of the Bluescript polylinker and the region upstream from the celA1 gene was purified from an agarose gel by using QIAEX (DIAGEN GmbH) and then labeled at its 5' termini with [γ-32P]ATP and polynucleotide kinase. The labeled DNA was digested with EcoRI. This digestion generates two kinds of fragments: a small fragment containing only DNA from the Bluescript polylinker region (33-bp XhoI-EcoRI fragment labeled in the 5' end of the XhoI site; see Fig. 4A) and a larger fragment, the probe, containing 14 bp from the Bluescript polylinker extending from the EcoRI site to the SmaI site and 572 bp of S. halstedii DNA from the SmaI site to the XhoI site and labeled at the 5' end of the XhoI site (see Fig. 4A). Samples of 10 μg of total RNA were used in each S1 experiment. The hybridization mixtures were incubated for 4 h and then treated with 100 U of S1 nuclease at 37°C for 30 min; the resulting products were analyzed in denaturing 6% polyacrylamide gels. A second probe was obtained in a similar way, but the first BamHI site inside celA1 (9 bp upstream from the XhoI site in celA1; see Fig. 2) was end labeled.

Genomic Southern analysis. Southern hybridization was performed with DNA from the Streptomyces species mentioned above. Total Streptomyces DNAs were digested with BamHI or SmaI and electrophoresed on 0.8% agarose gels. The DNA fragments were blotted to Nytran membranes (Schleicher and Schuell) and hybridized, as specified by the membrane manufacturer, to a nick translation-labeled [α-32P]dCTP probe containing the 784-bp internal BamHI fragment of celA1.

Nucleotide sequence accession number. The sequence of celA1 reported in this paper has been submitted to the EMBL data base under accession number Z12157.

RESULTS

Isolation of Streptomyces JMB and characterization of its cellulolytic enzymes. Several Streptomyces colonies with cellulolytic activity were isolated after screening for actinomycetes on basal agar medium containing CMC. One of them, JMB, showed the largest halo of hydrolysis when grown on this medium and developed by Congo red staining (Fig. 1A). The taxonomic characteristics listed in Bergey's Manual of Systematic Bacteriology (47) were used to identify JMB as S. halstedii. This strain was deposited in the Spanish Collection of Microorganisms (Colección Española de Cultivos Tipo) under the number CECT 3310.

Proteins in the supernatant of JMB grown for 3 days in liquid basal medium containing 0.5% Avicel were examined by native PAGE, SDS-PAGE, and IEF. The approximate pIs of four main protein bands with activity against CMC in IEF gels and zymograms were 4.0, 5.5, 6.4, and 8.7 (Fig. 1C). The molecular masses of these proteins were calculated by comparing the mobilities on SDS-CMC-PAGE Coomassie blue-stained gels with those on the same kinds of gels revealed by activity with Congo red. Only two bands of CMCase activity corresponding to sizes of 31 and 28 kDa, were detected. These proteins were named Cel1 and Cel2, respectively. We were unable to find the optimal renaturation conditions for the other two proteins.

Cloning of DNA fragments encoding CMCases. After construction of a genomic library of S. halstedii in pIJ702 and transformation of S. lividans protoplasts (see Materials and Methods), about 8,000 transformants were analyzed; six endoglucanase-overproducing (EG+) clones were selected for further studies. Plasmid DNAs from EG+ clones were analyzed by restriction enzyme digestion, and three species of plasmids were identified. These plasmids were named pJM1, pJM2 (derived from one of four clones having the vector with related inserts), and pJM6. Another library was made with plasmid pIJ699 (22), but the number of transformants was low and the library was not used later.

Protein analysis. SDS-PAGE analysis comparing culture supernatants from S. lividans EG+ clones with those of S. halstedii JMB grown in basal medium with 0.5% Avicel showed that only cells carrying plasmid pJM1 overproduced...
a protein of approximately 31 kDa. On SDS-PAGE, none of the other *S. lividans* EG+ clones showed overproduction of any protein, as judged by Coomassie brilliant blue staining (Fig. 1B), but SDS-CMC-PAGE-renatured gels developed by activity gave rise to unique bands of activity on pJM1 and pJM2 transformants (data not shown). The sizes of these proteins, 31 and 28 kDa, respectively, correspond to Cel1 and Cel2 from JM8; however, after SDS-PAGE, we were unable to detect any band of activity in transformants bearing pJM6. IEF gels and zymograms showed that the transformants carrying pJM1 produce two major active proteins with pIs of 5.5 and 6.4. The DNA fragment from pJM2 encodes a protein with a pI of 8.7, and the transformants carrying pJM6 produce a protein with a pI of 4.0. We named the latter protein Cel3. These proteins, Cel1 (31 kDa, pI of 5.5 and 6.4), Cel2 (28 kDa, pI 8.7), and Cel3 (pI 4, unknown size), represent the four main CMCase bands produced by the strain JM8 (Fig. 1C). From this point onward, our work was focused on the cellulase (Cel1) encoded by the DNA insert contained in pJM1.

Plasmid pJM1 has a DNA insert of 2.5 kb from strain JM8 that does not hybridize with the inserts of the other selected *S. lividans* EG+ clones (data not shown). A physical map of celA1 and its flanking regions is shown in Fig. 2. A small fragment of plasmid pJM1 was removed without losing CMCase activity. The resulting plasmid, pJM11, contained only a 1.7-kb insert (Fig. 2) and was used for sequencing analysis.

**Sequence analysis.** The 1.7-kb insert from plasmid pJM11 was cloned in pBluescript KS++ and SK+ plasmids, and both strands were sequenced (Fig. 2). The nucleotide sequence of the gene celA1 and its flanking regions is shown in Fig. 3. The sequence from nucleotides 446 to 1409 corresponds to an open reading frame encoding a 321-amino-acid protein that corresponds to the Cel1 precursor (see below). The FRAME and BRUIJNE programs were used to confirm the correct open reading frame based on *Streptomyces* codon preference (12). There are two possible translation start triplets in the "right" ORF, a TTG preceded by a potential ribosome-binding site (GAAA--GAGG) (3) and, eight triplets downstream, an ATG without any Shine-Dalgarno sequence. Starting at the TTG initiation codon, Cel1 would have a putative signal peptide of 27 amino acids at its NH2 terminus, leaving a protein of 294 amino acids with a molecular mass of 31 kDa. The von Heijne rule (46) was used to predict that signal protease cleavage would occur after the Ala-His-Ala sequence (Fig. 3). The mature Cel1 protein produced by *S. lividans* cells carrying pJM1 or pJM11 was refractory to several attempts at N-terminal sequence analysis; only a partial sequence (AXPXXM), which is coincident with the predicted N terminus, was obtained. After the translation stop codon (TGA), there is a sequence resembling a rho-independent transcriptional terminator (37) between nucleotides 1415 and 1471. This sequence has the potential to form a stem-loop secondary structure with a $\Delta G^\circ$ of $-59.3$ kcal (ca. 248 kJ/mol, as predicted by the DNASIS program.

**Mapping the 5' terminus of celA1 mRNA.** The transcriptional start point was determined by S1 nuclease mapping. Two end-labeled hybridization probes were used (see Materials and Methods). With both DNA probes the 5' end of the mRNA was identified as nucleotides 392 and/or 393. Figure 4B shows the protected fragment obtained with the XhoI probe. The -35 (5'-TTGACA-3') and -10 (5'-CGATGG-3') regions upstream from the mRNA 5' end (Fig. 3) show similarity to those of *Streptomyces* housekeeping promoters (16, 30).

**Relative regulatory sequences.** Near the 5' end of the mRNA (positions 401 to 414) there is a palindromic sequence of 14 nucleotides (5'-TGGGGACCGCTCCA-3') (Fig. 3) that is identical to a sequence found in other CMCase genes and described as being involved in the induction exerted by cellobiose (see Discussion). Table 2 shows a comparison of these sequences. The celA1 gene has two identical, perfectly repeated 15-bp sequences (5'-ATTGGGACCGCTTCC (N85)-ATTGGGACCGCTTCC-3') located at positions 285 to 299 and 384 to 398 (Fig. 3). The second of these sequences overlaps the promoter -10 region. Both sequences have 11 bp that are identical to 11 bp of the palindromic sequence mentioned above and could be involved in the regulation of the expression of celA1.

**Similarity of Cell protein to other cellulases.** The deduced amino acid sequence of the protein Cel1 from *S. halstedii* JM8 was compared with other protein sequences in the SWISS-PROT database (release 21) by using the FASTA program. A significant level of similarity was observed with the catalytic domain of several endoglucanases: 40% similarity to CasA from *Streptomyces* sp. strain KSM9 (10, 33), 48.7% similarity to E2 from *Thermomonospora fusca* (24), and 48.6% similarity to CenA from *Cellulomonas fimt* (48). The alignment shown in Fig. 5 is similar to that offered by Rouvinen et al. (38) and Gilkes et al. (10), although cellobio- hydrase CBHIII from *Trichoderma reesei* (6) is not included in the figure. The sequence of CelA from *Microbi- spora bispora* (48% similarity to Cel1) was added to the alignment from reference 49. Six major regions of similarity can be noted in all of the endoglucanases of family B (boxes A to F in Fig. 5).

**Multiple celA1-like genes in Streptomyces genomes.** Southern blots of genomic DNA isolated from 13 *Streptomyces* species were hybridized to a nick-translated DNA probe corresponding to the 784-bp BamHI internal celA1 fragment, which contains all the regions of similarity (A through F) of the cellulases included in family B. Four groups were made based on the production of endoglucanases on plates of basal medium supplemented with 0.5% CMC and on the hybridization results obtained.

(i) The first group included the strains that gave strong bands of hybridization. These strains were *S. halstedii* JM8 (original strain), *S. halstedii* NRRL 2381, *S. flavogriseus* ATCC 33331, and *Streptomyces* sp. strain ATCC 11238 (Fig. 6, lanes 1, 2, 3, 5, and 6). All of these strains are CMCase
producers, and the strong hybridization bands suggested that these microorganisms produce cellulases that belong to family B.

(ii) The second group included *Streptomyces* species that showed a weak hybridization band. These strains were *S. viridosporus* ATCC 39115 (a CMCase producer) and *S. badius* ATCC 39117 (a non-CMCase producer) (Fig. 6, lanes 9 and 10, respectively).

(iii) A third group included the strains that did not hybridize to celA4, but were known CMCase producers: *S. olivaceus* SC 3107, *Streptomyces* sp. strain JM5, *S. celluloflavus* ATCC 29806, *S. lividans* J166, *S. coelicolor* A3(2), and *S. griseus* ATCC 10137 (Fig. 6, lanes 4, 7, 8, 11, 12, and 14).

(iv) One strain, *S. albus* GR2, did not hybridize to celA4 and was not a CMCase producer (Fig. 6, lane 13).

**FIG. 3.** Nucleotide and deduced amino acid sequences of the *S. halstedii* JM8 celA4 gene and its flanking regions. The -35 and -10 sequences are located upstream from the transcriptional start point, which was determined by S1 mapping (asterisks). Three sequences, two of them similar and another identical to an activator binding site present in other cellulases (see Discussion), are underlined at the 5' end of celA4. The facing arrows at the end of celA4 mark an inverted repeat sequence. The mature N terminus is indicated by a vertical arrow; this sequence is preceded by a signal peptide of 27 amino acids. The DNA sequence from the vector pIJ702 is shown in lowercase letters after the BamHII-BglII junction.

**DISCUSSION**

In this paper we report the isolation of *Streptomyces* sp. strain JM8, identified here as *S. halstedii*, and the cloning and sequencing of one of its genes, celA4, which encodes a CMCase, CelI. This protein has a size of 31 kDa and migrates as two active bands of pI 5.5 and 6.4 in IEF gels. The presence of several active bands against CMC, after IEF, on supernatants from *S. lividans* transformants carrying pJM1 or pJM11 could indicate the existence of proteolytic activity (specific or unspecific), posttranslational modification, or formation of aggregates of Cell with other proteins. We checked for possible glycosylation on secreted proteins from *S. halstedii* JM8 and *S. lividans* carrying pJM1 or pJM11 by using the DIG glycan detection kit (Boehringer-Mannheim Biochemicals), but we could not detect glycosyl-
atation in Cel1; this keeps open the possibility of other kinds of modifications that could be responsible for several active bands after IEF of Cel1.

The initiation triplet for translation of the cel\textsubscript{A\textsubscript{1}} gene is likely to be a TTG (in position 446). Preliminary results of in vitro mutations of this TTG suggest that it is the initiation codon (data not shown). At least two other \textit{Streptomyces} genes have been described to have TTG as the starting codon: the \textit{IS116} open reading frame from \textit{Streptomyces clavuligerus}, which encodes a putative integrase-replicase protein (25), and \textit{cel\textsubscript{A\textsubscript{1}}}, a recently described endoglucanase-encoding gene from \textit{S. lividans} (45).

Protein sequence comparison showed that the \textit{S. halstedii} JM8 Cel1 enzyme belongs to $\beta$-glycanase family B (11, 14); it has six major regions of similarity with the catalytic domains of all five endoglucanases in this family (Fig. 5). Region B includes the I/L-L/V-E-P domain, which is similar to the N-E-P domain present in other glycanases and involved in catalysis (10). In region C, the His is present in all of the proteins included in this family except Cel1, which has Asn at this position. This change is conservative, and thus point mutations could elucidate the effect of His or Asn at this position. Cel1 is similar to CBHII, a cellobiohydrolase from \textit{T. reesei} that also belongs to this family. The similarity with this protein is mostly restricted to the six regions, although the region E is not as well conserved in CBHII as it is in all the endoglucanases of this family (only amino acids D, R, and N are conserved). Mutations in those regions would be useful for demonstrating their function in endo- or exocatalytic activity or in their substrate specificity. Conservation of aspartic acid at critical positions suggests that the catalytic action of Cel1 is that of an inverting enzyme. This is the common mechanism for all of the enzymes included in family B (10, 38).

The existence of disulfide bonds in the catalytic domain of glucanases included in family B have been determined in CBHII and assigned to the CenA catalytic domain (10, 38). These cysteines form two disulfide bonds that help stabilize large loops forming the active site tunnel of the CBHII or small loops forming the groove-shaped active site of endoglucanases (10, 38). Cel1 and all the members of this family have four cysteine residues at similar positions (Fig. 5). This led us to assume that two disulfide bonds between those cysteines (at positions 112 through 156 and 265 through 298) are formed in Cel1. The amino acid sequence of Cel1 does not have any obvious sequence features, such as

![FIG. 4. High-resolution S1 mapping to determine the transcriptional start site for cel\textsubscript{A\textsubscript{1}} in \textit{S. halstedii} JM8 and \textit{S. lividans} J166 transformed with plasmid pJM1. (A) Generation of the DNA probe from the plasmid pJM1. (B) S1 protection analysis. A radioactive DNA probe that had been uniquely 5' end labeled at the XhoI terminus within the cel\textsubscript{A\textsubscript{1}} coding sequence (prepared as described in Materials and Methods) was denatured and hybridized with several RNA samples from the following (lanes): 1, \textit{S. lividans} carrying pIJ702 (negative control); 2, \textit{S. lividans} carrying pJM1; 3, \textit{S. halstedii} JM8. The DNA sequence ladder belongs to the cel\textsubscript{A\textsubscript{1}} gene obtained by the Sanger et al. sequencing method (40) and a 20-mer synthetic oligonucleotide with a 5' end identical to that of the probe and labeled at this end with $[^{32}\text{P}]\text{ATP}$. The sequence of the ladder is shown at the right of panel B. The transcription start point is indicated by asterisks, and the 5' end of the mRNA is shown.](http://jb.asm.org/issue/vol174/issue3/)

<table>
<thead>
<tr>
<th>Strain</th>
<th>Gene</th>
<th>Position\textsuperscript{a}</th>
<th>Sequence</th>
<th>mRNA 5' start\textsuperscript{a}</th>
<th>Translation initiation</th>
<th>Family\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{S. halstedii} JM8</td>
<td>cel\textsubscript{A\textsubscript{1}}\textsuperscript{c}</td>
<td>-161 to -147</td>
<td>ATGGGACGCGTCCC</td>
<td>-54 to -53</td>
<td>TTG</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-62 to -48</td>
<td>ATGGGACGCGTCCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-45 to -32</td>
<td>TGGGACGCGTCCC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\textit{Streptomyces} sp. strain KSM-9</td>
<td>cas\textsubscript{A}</td>
<td>-73 to -60</td>
<td>TGGGACGCGTCCC</td>
<td>-74</td>
<td>GTG</td>
<td>B</td>
</tr>
<tr>
<td>\textit{S. lividans} J166</td>
<td>cel\textsubscript{A}</td>
<td>-49 to -36</td>
<td>TGGGACGCGTCCC</td>
<td>?</td>
<td>TTG</td>
<td>A</td>
</tr>
<tr>
<td>\textit{T. fusca}</td>
<td>E2</td>
<td>-76 to -63</td>
<td>TGGGACGCGTCCC</td>
<td>?</td>
<td>ATG</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>-49 to -36</td>
<td>TGGGACGCGTCCC</td>
<td>?</td>
<td>ATG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>E5</td>
<td>-49 to -36</td>
<td>TGGGACGCGTCCC</td>
<td>156, -152, -146</td>
<td>ATG</td>
<td>A</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Counting from the first nucleotide of the initiation codon as +1.
\textsuperscript{b} Determined from the catalytic domain (11, 14).
\textsuperscript{c} cel\textsubscript{A\textsubscript{1}} has three imperfectly repeated sequences in the indicated positions.
Pro-Thr boxes and putative cellulose-binding domains, that are present in most of the proteins included in family B (11). In fact, experiments with Avicel have confirmed that this protein does not bind to crystalline cellulose (data not shown).

The 5' region of celA1 contains three copies of an imperfectly repeated nucleotide sequence that resembles operator sequences. Two of the sequences are 15 bp long and are identical; the other (14 bp) has 11 bp of identity with them. These sequences could be implicated in the regulation of the expression of celA1, perhaps acting as an operator to which a regulatory dimeric or trimeric protein binds. The 14-nucleotide sequence TGGGAGGGTCCCA is also present in a similar location in other endoglucanase genes, i.e., E2, E4 and E5 endoglucanase-encoding genes from T. fusca (24, 26, 27), casA3 from Streptomyces sp. strain KSM9 (33), and celA from S. lividans (45) (Table 2). The role of this sequence in T. fusca has been studied. In the E5 gene, this sequence has been described as part of a 21-bp binding sequence for an activator protein involved in the induction of transcription of this gene by cellobiose (27). Gel retardation assays have shown that a protein from cell extracts binds to that se-
FIG. 6. Southern hybridization of a celA probe with restriction enzyme digests (BamHI for sample 1, SmaI for samples 2 through 14) of genomic DNA from several CMCase-producing (lanes 1 through 9, 11, 12, and 14) and nonproducing (lanes 10 and 13) Streptomyces species. CMCase production was detected by colony assay on solid basal medium with CMCase. Lanes: 1, S. halstedii JM8; 2, S. halstedii JM3; 3, S. halstedii NRRL 2381; 4, S. olivaceus SC 3107; 5, S. flavogriseus ATCC 33331; 6, Streptomyces sp. strain ATCC 11238; 7, Streptomyces sp. strain JMS; 8, S. celluloloflavus ATCC 29806; 9, S. viridosporus ATCC 39115; 10, S. badius ATCC 39117; 11, S. lividans 666; 12, S. coelicolor A3(2); 13, S. albus GR2; 14, S. griseus ATCC 10137. M, positions of molecular mass markers (λ phage, HindIII digested in kilobase pairs). The blot was washed for 1 h at 70°C in 18 mM NaCl–1 mM Na2HPO4–7.0O–0.1 mM EDTA–1% SDS.

sequence in E2 and E4 genes (24). A similar sequence (12 identical nucleotides) was found in a similar position in cenC from C. fim (7) and in a β-1,3-glucanase-encoding gene from Oerskovia xanthineolytica (42) (data not shown), suggesting a similar regulatory mechanism for celA.

The strong hybridization of the celA probe to several Streptomyces DNAs indicates that family B cellulases are produced by many members of the genus Streptomyces. The SmaI-generated fragments homologous to celA have different sizes in all species analyzed, including S. halstedii NRRL 2381. This could be explained by the high genomic instability of Streptomyces spp., manifested as high mutation rates and differences in the restriction patterns. The cloning and identification of these celA homolog genes could give more information on the CMCases included in family B. The celA probe did not hybridize with DNAs from other CMCase-producing Streptomyces spp., suggesting that the cellulases produced by these strains belong to a family different from that of Cel1. The weak hybridization obtained with a nonproducer, such as S. badius, could indicate that a homologous gene is present, although it would be poorly or not expressed in the culture conditions used.

ACKNOWLEDGMENTS

We thank M. D. García and F. Uriburu for identification of strain JM8 as S. halstedii. We thank K. Chater for suggesting to us that the TGT is the more likely start codon, the FRAME analysis, and invaluable suggestions. We thank J. Vara for making the BRUJENE program available to us. We thank B. Henrissat, F. Leal, C. Hansen, P. Carlsson, R. Esteban, and T. Fujimura for helpful discussions and suggestions. We thank N. Skinner for correcting the English manuscript. S. lividans and S. coelicolor are from D. Hopwood’s laboratory. We thank M. Romanic for the gift of S. albus GR2.

This research was supported by grant BIO088-2070 from the Comisión Asesora de Investigación Científica y Técnica, J.M.F.-A. and P.M.C. had fellowships from the Ministerio de Educación y Ciencia (Spain) and from the Universidad de Salamanca.

REFERENCES

46. Vara, J. Unpublished data.