Two genes encoding an endoglucanase and a cellulose-binding protein are clustered and co-regulated by a TTA codon in *Streptomyces halstedii* JM8

Ana Lila GARDA, José M. FERNÁNDEZ-ABALOS, Pilar SÁNCHEZ, Alberto RUIZ-ARRIBAS and Ramón I. SANTAMARÍA*

Instituto de Microbiología Bioquímica, Consejo Superior de Investigaciones Científicas, (CSIC)/Universidad de Salamanca, Campus Miguel de Unamuno, Avda. Campo Charro s/n, Salamanca, Spain

Streptomyces halstedii JM8 Cel2 is an endoglucanase of 28 kDa that is first produced as a protein of 42 kDa (p42) and is later processed at its C-terminus. Cel2 displays optimal activity towards CM-cellulose at pH 6 and 50 °C and shows no activity against crystalline cellulose or xylan. The N-terminus of p42 shares similarity with cellulases included in family 12 of the β -glycanases and the C-terminus shares similarity with bacterial cellulose-binding domains included in family II. This latter domain enables the precursor to bind so tightly to Avicel that it can only be eluted by boiling in 10 % (w/v) SDS. Another open

reading frame (ORF) situated 216 bp downstream from the p42 ORF encodes a protein of 40 kDa (p40) that does not have any clear hydrolytic activity against cellulosic or xylanosic compounds, but shows high affinity for Avicel (crystalline cellulose). The p40 protein is processed in old cultures to give a protein of 35 kDa that does not bind to Avicel. Translation of both ORFs is impaired in *Streptomyces coelicolor* *bldA* mutants, suggesting that a TTA codon situated at the fourth position of the first ORF is responsible for this regulation. S1 nuclease protection experiments demonstrate that both ORFs are co-transcribed.

INTRODUCTION

The actinomycetes are common inhabitants of different ecosystems and usually have the ability to degrade several polymeric compounds from which nutrients can be obtained. They can even defend their habitats from potential intruders by producing huge amounts of different compounds with antibiotic activity. Although these antibiotics have attracted the interest of basic researchers and industries, study of the production of enzymes involved in the degradation of polymeric compounds has mainly focused on basic science. Thus in recent years several groups have published studies on the biochemistry and cloning of chitinases, amylases, cellulases, xylanases, etc.

Our group has been studying the cellulolytic and xylanolytic systems of *Streptomyces halstedii* JM8 (a strain isolated from agricultural wastes) with a view to understanding their contribution to the degradation of cellulosic and hemicellulosic wastes, and as a model for understanding the regulation of gene expression. This strain produces at least three cellulases and two xylanases whose genes have been cloned [1,2], and study of these is currently under way.

In this paper we report the molecular characterization of two proteins, one of 28 kDa (Cel2) and another of 40 kDa (p40). Cel2 is a secreted protein with endo activity towards CM-cellulose and is originated by post-translational processing from a protein of 42 kDa (p42). p40 is a protein originated from an open reading frame (ORF) situated 216 nt downstream from the one giving rise to the p42 protein. This protein binds very efficiently to crystalline cellulose (Avicel) but shows no clear hydrolytic activity towards cellulosic compounds. Western blot analyses of both proteins produced in a *bldA* mutant of *Streptomyces coelicolor* suggest that the translation of a TTA codon present in the ORF originating p42 controls the expression of both genes.

EXPERIMENTAL

Bacterial strains

The *Streptomyces* strains used in this work were *S. halstedii* JM8 original strain isolated in our laboratory [1], *S. lividans* J166, *S. albus* G J1074 and *S. parvulus* J12283, and they were used as DNA recipients for Cel2 production experiments. *S. coelicolor* J1501 (parental strain) [3], and the corresponding *bldA* mutant J1700 [4], were used to study *bldA* translation dependence. All strains were obtained from The John Innes Institute.

Media and culture conditions

Streptomyces media, culture conditions, protoplasting, transformation and selection of transformants have been previously described by Hopwood et al. [5]. Growth of cells for protoplast formation was performed in YEDS medium [10 % (w/v) yeast extract/10 % (w/v) glucose/10.3 % (w/v) sucrose] supplemented with 5 mM MgCl₂ and different concentrations of glycine, depending of the strain. In this medium all the strains used grow faster than in yeast extract/malt extract medium (YEME), the preferred medium to obtain protoplasts [5]. Basal medium (minimal medium supplemented with 2 g/l yeast extract and 1 g/l protease peptone [6]) was supplemented with different carbon sources and used for Cel2 production. All liquid cultures were made up in 250 ml four-baffled flasks with 25 ml of medium and incubated at 28 °C for 48 h at 250 rev./min. When needed, thiostrepton (given by E. R. Squibb and Sons Inc., Princeton, NJ, U.S.A.) was added (10 μ g/ml) to solid or liquid medium.

Gel electrophoresis analyses and enzyme purification

The supernatants from the different liquid cultures were separated from the mycelial mass by centrifugation at 10000 g for 10 min

Abbreviations used: CBD, cellulose-binding domain; ORF, open reading frame; pNPC, *p*-nitrophenyl β -D-cellobioside.

* To whom correspondence should be addressed.

The sequences of the genes encoding Cel2 and p40 have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number U51222.

and used as crude enzyme preparations to quantify the endo- β -1,4-glucanase activity or to analyse the proteins by denaturing SDS/PAGE. In this case, proteins were precipitated with 10% (v/v) trichloroacetic acid, resuspended in loading buffer [7], neutralized with 0.2 M NaOH, boiled for 5 min and loaded on to SDS/PAGE [15% (w/v) gel]. Electrophoresis was carried out by the method of Laemmli [7] in a Mini Protean II system (Bio-Rad) and gels were stained with silver [8] or Coomassie Brilliant Blue R-250, or electrotransferred to Immobilon P (Millipore) in a mini transblot (Bio-Rad) (for further immunoblotting analysis see below). Protein sizes were estimated using Bio-Rad low-molecular-mass markers as reference.

Cel2 protein purification was done as described elsewhere [9]. Briefly, *S. albus* transformed with the plasmid pAL16 (Figure 1A) was grown in 2 litres of basal medium supplemented with 0.5% (w/v) cellobiose and 10 μ g/ml thiostrepton, at 28 °C and 250 rev./min shaking in an orbital shaker (Adolf Kühner AG Schmeiz) for 72 h. Cultures were filtered through two glass microfibre filters (Whatman) and then through 0.45 μ m Gelman filters. The supernatant was concentrated and dialysed in an Amicon system using Polyethersulphone Filtron Omega membranes with a nominal molecular mass limit of 5 kDa. Samples were concentrated in a volume of 100 ml and then dialysed against water in the same system. Purification was performed using FPLC® (Pharmacia) equipped with a Q-HR 5/5 anion-exchange column (Pharmacia). The column was equilibrated with 20 mM piperazine, pH 9.7. Protein samples were filtered through 0.22 μ m syringe filters (Gelman Sciences), adjusted to 20 mM piperazine, pH 9.7, and loaded on to the column (flow rate 1 ml/min). Cel2 was not retained and eluted almost pure in the first fraction. Residual proteins were eluted with 1 M NaCl. The cellulolytic activity of the different fractions was detected on 1.5% (w/v) agar plates containing 0.5% (w/v) low-viscosity CM-cellulose and later developed with Congo Red [10].

Protein p40 was purified from 1 litre of the same culture supernatant as Cel2. The total culture was centrifuged for 10 min at 10000 g. Avicel (Merck) was added to a final concentration of 0.15% (w/v) and the suspension was incubated with shaking for 1 h at room temperature and centrifuged for 10 min at 10000 g. The pellet was washed twice with sterile water, resuspended in a solution containing 1% (v/v) triethylamine and 1 mM EDTA and stirred for 5 min [11]. Avicel was removed by centrifugation and the supernatant was neutralized by addition of 1 M phosphate buffer, pH 7. The solution was desalted using a column of Bio-Gel P-6DG equilibrated with water.

Enzyme assays

Protein concentrations were measured by using the DC Protein Assay (Bio-Rad), using BSA as the standard. Reducing sugar concentrations were measured by the 3,5-dinitrosalicylic acid method [12] using glucose as the standard. The reaction mixture included 100 μ l of an appropriate dilution of the previously purified protein (0.5 μ g was typically used in each reaction), 100 μ l of 4% low viscosity CM-cellulose and 200 μ l of a suitable buffer (50 mM final concentration) for each assay. Substrate and enzyme controls were always established. Reaction conditions were 50 °C over 5 min. Reactions were stopped with 1 ml of 3,5-dinitrosalicylic acid, the mixture being boiled for 5 min and then diluted with 4 ml of water. Samples were centrifuged at 5000 g for 5 min and absorbance was measured at 540 nm. One unit of carboxymethylcellulase was defined as the amount of enzyme that releases 1 μ M of glucose equivalents per min under standard assay conditions (0.5 μ g of purified Cel2, 1% low viscosity CM-

cellulose and 50 mM sodium citrate buffer, pH 5.0, in 400 μ l final assay volume and incubation at 50 °C for 5 min).

Activity against several other substrates (see the Results section) was assayed following the protocols previously described [13].

Effects of pH and temperature on Cel2 activity and stability

Cellulase activity was estimated at different pH values and temperatures. To determine the optimal pH, a pH range from 3 to 10.5 was obtained with the following (50 mM) buffers: sodium citrate, pH 3–6; sodium phosphate, pH 6–8; Tris/HCl, pH 8–9; glycine/NaOH, pH 9–10.5.

The optimal temperature and thermostability were determined in the 4–80 °C range in 50 mM sodium citrate, pH 5. To determine thermal stability, Cel2 was pre-incubated for 30 min or 1 h at the corresponding temperature; the substrate was then added and the reaction was maintained for 5 min at 50 °C.

Substrate-binding capacity

Culture supernatants of *S. albus* harbouring pAL16, pAL25 or pAL26 were incubated with 0.3% (w/v) of the following substrates: Avicel (crystalline cellulose), Whatman paper No. 1, chitin, lichenan, laminarin or xylan, for 30 min at room temperature. Proteins bound to all these substrates were released with 1% triethylamine/1 mM EDTA [11] or by boiling for 5 min in the presence of 10% SDS [14]. The original supernatants, the non-retained fractions and the bound proteins were analysed by SDS/PAGE.

Cel2 and P40 antiserum generation and purification

Cel2 was purified by electroelution from SDS/PAGE loaded with culture supernatants of *S. albus* transformed with pAL26. The p40 protein was obtained after Avicel binding and triethylamine/EDTA elution from 72 h-old culture supernatants of *S. albus* transformed with pAL16. Approx. 200 μ g of purified proteins were mixed with complete Freund's adjuvant (1:1, v/v) and injected intradermally into four New Zealand White rabbits, two for each protein. Two boosters of 100 μ g of protein mixed with incomplete Freund's adjuvant (1:1, v/v) were given intramuscularly after 24 and 60 days respectively. Blood was extracted periodically every 10 days after the first booster. Proteins (4 μ g) from different *Streptomyces* culture supernatants were fractionated by SDS/PAGE and transferred to Immobilon P (Millipore). Cel2 or p40 were immunodetected by using the ProtoBlot Western blot AP system (Promega). Antibody purification was done as indicated by Beall and Mitchell [15] with Cel2 and p40 proteins electrophoresed on Immobilon (Millipore).

Recombinant DNA techniques

All DNA manipulations were performed using established methods [16], with enzymes and reagents purchased from Amersham, Promega, Bethesda Research Laboratories and Boehringer-Mannheim Biochemicals, and used following the manufacturers' guidelines.

DNA sequencing

Overlapping deletions were generated on pBluescript KS (Stratagene), pUC118 or pUC119 [17] derivatives carrying different parts of pAL16 as described by Henikoff [18]. Fragments of appropriate size were sequenced in both strands by the dideoxy chain-termination method [19] using single-strand templates and the T7 Sequencing Kit and G/A^{T7} sequencing mixes

(Pharmacia Biotech). DNA and protein sequence analyses were performed with the DNASIS, PROSIS (Pharmacia-Biotech, Hitachi) and DNA Strider [20] programs. The deduced amino acid sequences were compared with the SWISS-PROT Data Bank using FASTA [21]. Alignment of protein sequences was performed by using the CLUSTAL program [22].

S1 mapping

Total RNA was isolated as described previously [23] and treated with RNase-free DNase (RQ1 from Promega). S1 nuclease digestions and analyses of RNA-protected fragments were performed essentially as described by Favaro et al. [24]. ³²P-labelled probes were generated using appropriate DNA fragments obtained by enzyme digestion or PCR (see Figure 8). Approx. 40 µg of total RNA extracted from *S. albus* transformed with pAL16 or pIJ702 and grown for 2 days on basal medium or YEDS media were hybridized with approx. 40000 c.p.m. of the probe in 20 µl of the sodium trichloroacetic acid buffer [25]. After hybridization at 45 °C overnight, the hybrids were digested for 60 min at 37 °C with 200 units of S1 nuclease (Boehringer) per sample. Samples were then precipitated with isopropanol, washed with 80 % (v/v) ethanol, dissolved in 6 µl of formamide loading buffer and run on 6 % (w/v) polyacrylamide sequencing gels. Size markers were M13 sequencing ladders generated with M13(–20) forward primer and M13mp18 single-strand DNA by the method of Sanger et al. [19]. *In vitro* transcripts containing the targets for the probes were synthesized with T7 RNA polymerase (on convenient DNA templates cloned in pBluescript KS+ plasmid) and used as positive controls for hybridization and S1 digestion. Yeast tRNA (40 µg) was used as a negative control.

RESULTS

Isolation of endoglucanase-producing clones

The construction of an *S. halstedii* JM8 library in *S. lividans* JI1326 has been described elsewhere. Several endoglucanase-overproducing clones were isolated [1–26]. One of them, carrying a DNA fragment of 4 kb, was named pJM2 and was used for further study here. Transformation of this plasmid in *S. albus* and *S. parvulus* conferred them the ability to overproduce three proteins of 40, 35 and 28 kDa: p40, p35 and Cel2 respectively (Figure 1B, lane 2). Only the latter protein had been observed in the original *S. lividans* transformant [1].

Several subcloning assays permitted us to determine that the left part of the insert encodes Cel2 and the right part p40 and p35 (Figure 1A and Figure 1B, lanes 3 and 4). Amino acid sequencing of p40 and p35 demonstrated that both N-termini were identical, suggesting that the smaller one originated from p40 through proteolytic cleavage in the C-terminus.

Analyses of the activity of the three proteins towards CM-cellulose confirmed that only the Cel2 protein has endoglucanase activity (Figure 1C).

Protein processing

Polyclonal antibodies raised in rabbits were obtained against purified Cel2 and p40 and used to detect these proteins and their potential precursors. When immunodetection was accomplished with Cel2 antiserum on different protein fractions obtained from cells bearing the pAL25 plasmid (this has the second ORF), none of the proteins was detected. However, when the proteins were isolated from cells carrying pAL26 (this has the first ORF), a protein of 42 kDa (p42) was detected in supernatants from

Figure 1 Cloning and expression of Cel2- and p40-encoding genes

(A) Partial restriction map of plasmids obtained during this work. In all cases the vector used was pIJ702. Detailed construction of these plasmids is not indicated. Cellulase activity (carboxymethylcellulase), Avicel binding (+ or –) and the proteins observed are indicated on the right. Plasmid names are on the left. Only the restriction enzymes used for subcloning are shown (Bg, BglI; N, NruI; S, SalI; S3A, Sau3A; Sp, SphI). (B) SDS/PAGE stained by Coomassie Brilliant Blue R-250 of culture supernatants (40 µg of protein per lane) of *S. albus* transformed with several plasmids: (1) pIJ702 (vector); (2) pAL16; (3) pAL25; (4) pAL26. Numbers on the left indicate the sizes in kDa. Arrows on the right indicate the proteins originated. (C) Supernatant assay for carboxymethylcellulase activity of the same supernatants used in (B). Protein (20 µg) was loaded in each well and incubated at 37 °C for 4 h; activities were detected with Congo Red staining.

young cultures (see below). Later (72 h) this protein disappeared completely to originate a protein of 28 kDa (Cel2) and several smaller peptides (Figure 2A, lanes 48 and 72 h). These proteins were not detected when the assay was done in cell extracts or membrane plus cell-wall fractions of this clone, suggesting that the p42 protein is secreted very efficiently and is not bound to mycelium, as occurs with *Streptomyces reticuli* Cel1 cellulase [27]. Similar experiments were carried out using p40 antiserum. In this case, a protein of 40 kDa was detected in supernatants of exponential and stationary cultures of cells carrying pAL25. Another protein of 35 kDa was detected in advanced stationary-phase supernatants (72 and 96 h) (Figure 2B). This confirms the origin of the p35 protein. No protein bands were detected in cell extract or membrane plus cell-wall fractions (results not shown). When these antibodies were used against proteins produced by the strain carrying pAL26, neither p42 nor Cel2 was detected, indicating that there is no cross-reaction between p42 and/or Cel2 and anti-p40 antibodies.

Characterization of Cel2 and p40

The Cel2 and the p40 proteins were purified by FPLC® and by Avicel binding and subsequent triethylamine/EDTA release

Figure 2 Time course analyses of p42 and p40 processing

Supernatants of *S. albus* transformed with pAL26 (ORF1 that originates p42) (**A**), or with pAL25 (ORF2 that originates p40) (**B**), were harvested at the times indicated. Western blot analyses were done after SDS/PAGE and developed with Cel2 antibodies (**A**) or p40 antibodies (**B**). Protein (4 μ g) was loaded per lane. Numbers on the left indicate sizes in kDa.

respectively. The specificity of the activity of purified Cel2 was tested on different substrates such as Avicel, CM-cellulose, soluble xylan from oat spelts, *p*-nitrophenyl β -D-cellobioside (pNPC) and *p*-nitrophenyl β -D-glucoside. Activity was detected towards CM-cellulose and a weak activity was obtained on pNPC. TLC analyses of the products generated from CM-cellulose corroborated the evidence that oligosaccharides with a degree of polymerization greater than 4 were the most abundant products and, in long incubations, a small quantity of cellobiose was formed although glucose was never detected. This would indicate an endo mode of action for this enzyme. The possible synergism between Cel1 [1] and Cel2 endoglucanase was studied but no co-operative effect was detected (results not shown). Cel2 displays optimal activity at pH 6 and 55 °C and is stable from pH 3 to 10 after incubation for 1 h (Figure 3A). Regarding the thermostability of Cel2 when incubation was carried out without substrate, its activity was abolished after 1 h at 55 °C (Figure 3B). In a more precise study, incubations of the enzyme at this temperature for periods of time as short as 2 min inactivated the enzyme; however, the addition of CM-cellulose stabilized the protein and it retained most of its activity even after incubation for 1 h at 55 °C (results not shown).

Avicel-purified p40 was assayed against the following substrates: Avicel, CM-cellulose, insoluble xylan, lichenan, agarose, starch, *p*-nitrophenyl- α -D-galactopyranoside, *p*-nitrophenyl- α -L-arabinofuranoside, *p*-nitrophenyl- β -D-xylopyranoside, pNPC

Figure 3 Effect of pH (**A**) and temperature (**B**) on activity (□) and stability (●) of Cel2

Buffers used were sodium citrate, pH 3–6; sodium phosphate and Tris/maleate/NaOH, pH 6–8; Tris/HCl, pH 8–9 and glycine/NaOH, pH 9–10.5. The concentration of each buffer in the reaction was 50 mM. For pH stability, the enzymes were pre-incubated for 1 h at 32 °C in suitable buffer for each pH and reactions then was carried out in 50 mM sodium citrate, pH 5, for 5 min at 50 °C. Results shown are from three different experiments. For thermostability, the enzymes were pre-incubated for 1 h at the indicated temperature in 50 mM sodium citrate, pH 5, thereafter performing the reactions indicated above.

and methylumbelliferyl β -D-glucoside. Very weak activity was detected after very long incubations (12 h) with methylumbelliferyl β -D-glucoside and no clear activity against any of the other substrates assayed was detected.

Sequencing analysis

The 3.5 kb insert contained in pAL16 (Figure 1A) was subcloned in several *Escherichia coli* plasmids and sequenced. The nt sequence of the whole fragment (3527 bp) has two typical *Streptomyces* ORFs (Figure 4). ORF1, from nt 1 to 1131, encodes a protein of 377 amino acids. ORF2, from nt 1351 to 2442, encodes a protein of 364 amino acids. A 14 bp inverted repeat sequence, situated between nt –84 and –71, is present upstream from the translation start codon for ORF1. An identical sequence has been described in several cellulases from *Thermomonospora fusca* [28,29] and in all *Streptomyces* cellulase-encoding genes sequenced so far [1,30,31]. The unusual codon TTA coding for leucine is situated at the 4th position in ORF1, suggesting a *bldA* dependence on translation (see below). The first amino acid of the mature Cel2 is amino acid 38 of ORF1. However, the sizes of the deduced (42 kDa) and the observed protein (28 kDa) differ. This would explain the detection of a 42 kDa protein when

The —35 and —10 sequences are located upstream from the transcriptional start point, which was determined by S1 mapping (*). The putative —35, —10 regions, RBS and TTA codons are shown in bold type. The 14 bp inverted repeat sequence is underlined. Facing arrows at the end of ORF2 mark an inverted repeat sequence that could act as a putative transcriptional terminator. Vertical arrows indicate signal peptide processing points.

Figure 5 Amino acid alignments using the CLUSTAL program

(A) Sequences of several β -1,4-endoglucanases included in family 12 (or H): Cel2 (this work); EglS from *S. rochei* A2; CelB from *S. lividans*; FI-carboxymethylcellulase (FI-CMCase) from *Aspergillus aculeatus*; cellulase CelS from *Erwinia caratovora*. (B) Sequences of different CBDs: p42 (this work); EglS from *S. rochei* A2; CelB from *S. lividans*; Cex from *C. fimi*; CelA (1) from *Microbispora bispora*; EngD from *Clostridium cellulovorans*; E2 from *Thermomonospora fusca*; CelA (2) from *S. lividans*; p40 (this work). Numbering of the amino acids starts at the N-terminus of all proteins. The introduction of gaps is indicated by dashes. Lines (|) over the

anti Cel2 (p28) antibodies were used (Figure 2A, lanes 36 and 48h).

Both p40 and p35 have the same N-terminus, corresponding to amino acid 35 of the protein deduced from ORF2. A typical signal peptide precedes this amino acid. After ORF2, there is an inverted sequence between nt 2457 and 2495 that could act as a ρ -independent transcriptional terminator.

Sequence similarity analyses and cellulose-binding experiments

The proteins deduced from both ORF1 and ORF2 were compared with the EMBL data bank. The first, p42, has two clear domains. The first 222 amino acids following the signal peptide show similarity to cellulases included in family 12 [32,33]: EglS from *S. rochei* A2 (83% similarity) [30]; CelB from *S. lividans* (79%) [31]; FI-carboxymethylcellulase from *Aspergillus aculeatus* (29%) [34]; cellulase CelS from *Erwinia caratovora* (17%) [35] (Figure 5A). The C-terminus of the protein resembles the bacterial-like cellulose-binding domains (CBDs) of the following proteins: EglS from *S. rochei* A2 (66.3% similarity) [30]; CelB from *S. lividans* (63.5%) [31]; Cex from *C. fimi* (46.2%) [36]; CelA from *Microbispora bispora* (39%) [37]; EngD from *Clostridium cellulovorans* (36.5%) [38]; E2 from *Thermomonospora fusca* (35.7%) [39]; CelA from *S. lividans* (35.6%) [40]; p40 from *S. halstedii* JM8 (28.2%) (this report) (Figure 5B). Both catalytic and theoretical CBDs are joined by a typical P- and G-rich linker region of 14 amino acids. p42 binding capacity was tested using a 48 h *S. albus*/pAL26 supernatant on Avicel, xylan and chitin. The protein p42 binds efficiently on Avicel and partially to the other two substrates assayed. Release of this protein was only possible after boiling for 5 min with 10% SDS and no elution occurred with water or triethylamine/EDTA (results not shown).

Sequence comparison of the protein deduced from ORF2 did not reveal any clear similarity in the first 204 amino acids, after the signal peptide. However, the last 104 amino acids of the protein showed a clear similarity with the same bacterial CBD as the C-terminus of p42 (Figure 5B). The similarities with the CBDs shown in Figure 5(B) are as follows: p42 from *S. halstedii* JM8 (28.2%); EglS from *S. rochei* A2 (27.2%) [30]; CelB from *S. lividans* (30.1%) [31]; Cex from *C. fimi* (29.1%) [36]; CelA from *Microbispora bispora* (34%) [37]; EngD from *Clostridium cellulovorans* (22.3%) [38]; E2 from *Thermomonospora fusca* (23.8%) [39] and CelA from *S. lividans* (28.2%) [40]. Preceding this domain there is a typical T- and P-rich linker region 23 amino acids long.

The cellulose-binding capacity of p40 was tested on different substrates such as Avicel, Whatman paper No. 1, chitin, lichenan, laminarin and xylan. The highest binding capacity was detected on Avicel and Whatman paper No. 1; a small amount of p40 was bound to chitin and lichenan. No binding capacity to laminarin or xylan was detected (Figure 6). In all cases, the protein bound was eluted with triethylamine/EDTA.

bldA translation dependence of Cel2 and p40 genes

In order to determine whether the translation of Cel2 and p40 genes depends on *bldA*, the plasmids pAL16 (containing both ORFs), pAL26 (harbouring p42 ORF) and pAL25 (bearing p40 ORF) were cloned in *S. coelicolor* J1501 (wt strain) and in *S. coelicolor* J1700 (a *bldA* mutant). The expression of both genes was studied by searching for the encoded proteins in supernatants

alignment indicate amino acid differences between the three *Streptomyces* cellulases. Asterisks and dots below the alignments indicate identical and conserved amino acid residues among all these enzymes.

Figure 6 Binding capacity of p40 to different substrates analysed by SDS/PAGE

Supernatant of *S. albus* transformed with pAL16 was incubated with different substrates and the proteins bound were released with 1% triethylamine/1 mM EDTA. S, total supernatant; odd lanes, not retained; even lanes, retained. A, Avicel; Wh, Whatman No. 1; Ch, chitin; L, lichenan; La, laminarin; X, xylan. Numbers on the left indicate sizes in kDa. Proteins are indicated by the arrows.

Figure 7 Western blot analyses showing the *bldA* translation dependence for the production of Cel2 and p40

Culture supernatants of *S. coelicolor* J1501, a wt strain (wt) and *S. coelicolor* J11700, a *bldA* mutant: lanes 1, transformed with pIJ702; lanes 2, pAL16 (carrying ORF1 and ORF2); lanes 3, pAL25 (ORF2 produces p40); lanes 4, pAL26 (ORF1 produces p42). Western blot analyses were developed with anti-Cel2 antibodies (A) or anti-p40 antibodies (B). Total protein (4 µg) was loaded in each lane. Numbers on the left indicate sizes in kDa.

with anti-p28 and anti-p40 antibodies. The results show that the translation of both genes is impaired in the *bldA* mutant (Figures 7A and 7B). These data suggest that both genes are co-transcribed and that the blockade of translation originated by the TTA codon extends its effect to the p40 translation. When p40 antibodies were used on supernatants of strains carrying pAL25 (only ORF2) a single, faint antigenic band of similar intensity in each lane was detected in both strains. This would suggest the presence of a very weak promoter situated immediately upstream from the p40 coding region that would be involved in a basal

Figure 8 S1 mapping of Cel2 and p40 genes

(A) The steps followed in the generation of both probes. (B) Results with: (left) Cel2 gene probe; (right) p40 gene probe. Lanes (a) and (b) Dideoxy sequence of M13 mp18 used as size marker: (a) A + C + G nucleotides; (b) T nucleotide; (c) *S. lividans* carrying pIJ702; (d) *S. lividans* carrying pAL16; (e) probes. Numbers on the right indicate sizes of protected fragments. Asterisks indicate the position of the label.

bldA-independent expression of this gene. However, no transcription start point was detected from this hypothetical promoter (see below).

S1 mapping

The transcription start site was analysed by S1 protection experiments with the two probes obtained as described in Figure 8(A). The putative transcriptional start site was identified at nt 97 with the *celA*₂ probe (Figure 4), although several minor bands were also present around that point. However, the p40 probe was completely protected, indicating that the transcript starts upstream from the probe used and suggesting that a unique transcript was produced for both genes (Figure 8B). No transcription start point was detected inside the 216 nt that separate both ORFs, indicating that, at least under the conditions used, there is no functional promoter in this region. However, after very long exposure a faint band was detected approx. 10 nt upstream from the p40 coding region. We consider that this band could represent a major degradation product of the whole transcript, although the presence of a very weak promoter cannot be ruled out.

DISCUSSION

The cellulase Cel2 of *S. halstedii* JM8 is a 28 kDa secreted protein that is produced as a precursor of 42 kDa, which is rapidly processed at its C-terminal end to render the mature Cel2. This protein displays hydrolytic activity against CM-cellulose and weak activity against pNPC. Other compounds such as crystalline cellulose, *p*-nitrophenyl β-D-glucoside or xylan are not hydrolysed. The protein sequence and biochemical properties of Cel2 resemble the cellulases included in family 12 of the glycanases [32,33], especially EglS from *S. rochei* [30] and CelB from *S. lividans* [31]. The general structure of Cel2, EglS and CelB (and their encoding genes; see below) suggests a common origin: all three would be produced as precursor proteins of approx. 380 amino acids that are later secreted (by

means of a similar signal peptide) and proteolytically processed at their C-terminal end to generate the 28 kDa enzyme found in culture supernatants. Purified Cel2 protein was used for the biochemical characterization; optimal activity was obtained at pH 6 and 55 °C, conditions similar to the other cellulases mentioned. These results were corroborated by differential scanning calorimetry, which demonstrated that the enzyme was irreversibly inactivated after heating at 54 °C [9].

Some aspects of the genes coding for Cel2, EglS and CelB are remarkable. All three genes have the rare TTA (leucine) codon situated at the fourth position of the coding sequence, and the *S. lividans* celB gene has a second TTA codon at position 22. Curiously, other codons coding for leucine are present at the same location in the genes coding for Cel2 and EglS. From a regulatory point of view, the meaning of the TTA codons remains to be elucidated. In our experiments, we have observed that the production of Cel2 is impaired when cloned in *S. coelicolor* J1700, a strain bearing a mutated version of the *bldA* gene (which encodes the cognate tRNA responsible for translating the TTA –UUA– codons). Furthermore, this *bldA* dependence is transmitted to the expression of the downstream gene, which codes for p40 and is co-transcribed with *celA*₂; the p40 gene lacks any TTA codon. A similar polar effect on translation has been observed in the production of some spore-associated proteins ('Saps') in *S. coelicolor*, in which the expression of *sapC*, *sapD* and *sapE* genes depends on the TTA codon present in a preceding gene (J. McCormick, unpublished results). Genes for antibiotic resistance [41] and for the pathway-specific regulator *actII-ORFIV* in actinorhodin production [42] also have TTA codons, and dependence on, but no regulatory role in, the *bldA* product has been demonstrated [43,44].

The p40 protein is processed in old cultures to originate a 35 kDa protein that does not bind to Avicel and whose sequence does not present any clear similarity to any other known protein. The catalytic activity of p40 was tested against several substrates and only a very weak activity against methylumbelliferyl β -D-glucoside was detected, suggesting residual β -glucosidase activity. An interesting hypothesis to test would be whether the non-enzymic domain of p40 might act as a link for other proteins with enzymic activity to cellulose, e.g. perhaps the p42 produced from the same transcript. Another possibility is that p40 could be an enzymically inactive descendant of an ancestral able to bind and hydrolyse cellulose. It has not been possible to confirm the hypothesis that p40-like ORFs are close downstream to the EglS and CelB genes in *S. rochei* and *S. lividans* respectively since no sequence data are available in the gene bank.

Situated in similar positions, both p42 and p40 have two CBDs separated by typical T- and P-rich inter-domain regions. The CBD of p42 has tryptophan residues corresponding to W14, W34 and W50 that are present in a consensus family II bacterial CBD. However, it does not have two other conserved W residues, W38 and W68 (residues 311 and 341 in Figure 5B), that seem to be involved in binding to crystalline cellulose (W38 is present in the p40 CBD). Mutation of these conserved W residues in the CBD isolated from CenA from *Cellulomonas fimi* [45] and that of XynA from *Pseudomonas fluorescens* [46] showed that replacement with alanine causes a greater reduction in cellulose avidity than phenylalanine substitutions. However, the effect is not identical in all systems: whereas mutation of W68A in CenA causes a 30-fold reduction in the cellulose-binding capacity relative to the wild type, in XynA these residues do not appear to play any critical role in the interactions needed for binding. Thus these conserved residues in different CBDs cannot be assumed to play an equivalent role for the ligand binding of proteins from different bacterial species. This hypothesis is

supported by the fact that CenA CBD elutes with water, whereas that of XynA only elutes with 10% SDS and boiling. It is suggested that the three-dimensional structure of these domains may not be identical. A similar case would occur in the CBDs studied in this paper; thus while p40 can be released by triethylamine/EDTA, p42 needs to be boiled with 10% SDS.

We thank Dr. F. Leal for his invaluable help in the FPLC® work and discussions and suggestions, Dr. J. McCormick for communication of results before their publication, Juan C. Villoria for his help in obtaining antibodies, E. R. Squibb and Sons Inc. (Princeton, NJ, U.S.A.) for the thiostrepton and N. Skinner for correcting the English manuscript. Supported by grants from the Comisión Interministerial de Ciencia y Tecnología (BIO92-0173, and BIO95-1558-CO2-01). A.L.G. received a fellowship from the Instituto de Cooperación Iberoamericana (ICI) and the Universidad Nacional Autónoma de México (UNAM). J.M.F.-A. and A.R.-A. had fellowships from the Ministerio de Educación y Ciencia (Spain).

REFERENCES

- Fernández-Abalos, J. M., Sánchez, P., Coll, P. M., Villanueva, J. R., Pérez, P. and Santamaría, R. I. (1992) *J. Bacteriol.* **174**, 6368–6376
- Ruiz-Arribas, A., Fernández-Abalos, J. M., Garda, A., Sánchez, P. and Santamaría, R. (1995) *Appl. Environ. Microbiol.* **61**, 2414–2419
- Chater, K. F., Hopwood, D. A., Kieser, T. and Thompson, C. J. (1982) *Curr. Top. Microbiol. Immunol.* **97**, 69–95
- Piret, J. M. and Chater, K. F. (1985) *J. Bacteriol.* **163**, 965–972
- Hopwood, D. A., Bibb, J. M., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P., Ward, J. M. and Schrepf, H. (1985) *Genetic Manipulation of Streptomyces: A Laboratory Manual*, John Innes Foundation, Norwich
- Ishaque, M. and Kluepfel, D. (1980) *Can. J. Microbiol.* **26**, 183–189
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Morrissey, J. (1981) *Anal. Biochem.* **117**, 307–310
- Garda, A. L., Santamaría, R., Marcos, M. J., Zhadan, G. G., Villar, E. and Shnyrov, V. L. (1996) *Biochem. Mol. Biol. Int.* **38**, 161–170
- Teather, R. M. and Wood, P. J. (1982) *Appl. Environ. Microbiol.* **43**, 777–780
- Hansen, C. K., Diderichsen, B. and Jørgensen, P. L. (1992) *J. Bacteriol.* **174**, 3522–3531
- Bernfeld, P. (1951) in *Advances in Enzymology* (Nord, F. F., ed.), pp. 379–428, Interscience Publishers, New York
- Wood, W. A. and Kellogg, S. T. (eds.) (1988) *Methods Enzymol.* **160**
- Millward-Sadler, S., Poole, D. M., Henrissat, B., Hazlewood, G. P., Clarke, J. H. and Gilbert, H. J. (1994) *Mol. Microbiol.* **11**, 375–382
- Beall, J. A. and Mitchell, G. F. (1986) *J. Immunol. Methods* **86**, 217–223
- Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn, pp. 000–000, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Vieira, J. and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11
- Henikoff, S. (1984) *Gene* **28**, 351–359
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Marck, C. (1988) *Nucleic Acids Res.* **16**, 1829–1836
- Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444–2448
- Higgins, D. G. and Sharp, P. M. (1988) *Gene* **73**, 237–244
- Igo, M. and Losick, R. (1986) *J. Mol. Biol.* **191**, 615–624
- Favarolo, J., Treisman, R. and Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749
- Murray, M. G. (1986) *Anal. Biochem.* **158**, 165–170
- Fernández-Abalos, J. M., Sánchez, P., Coll, P. M., Villanueva, J. R., Pérez, P. and Santamaría, R. I. (1993) *J. Bacteriol.* **175**, 1211 (corrigendum)
- Schlottermeier, A., Walter, S., Schröder, J., Moormann, M. and Schrepf, H. (1992) *Mol. Microbiol.* **6**, 3611–3621
- Jung, E. D., Lao, G., Irwin, D., Barr, B. K., Benjamin, A. and Wilson, D. B. (1993) *Appl. Environ. Microbiol.* **59**, 3032–3043
- Zhang, S., Lao, G. and Wilson, D. B. (1995) *Biochemistry* **34**, 3386–3395
- Perito, B., Hanhart, E., Irdani, T., Iqbal, M., McCarthy, A. J. and Mastromei, G. (1994) *Gene* **148**, 119–124
- Wittmann, S., Shareck, F., Kluepfel, D. and Morosoli, R. (1994) *Appl. Environ. Microbiol.* **60**, 1701–1703
- Henrissat, B. (1991) *Biochem. J.* **280**, 309–316
- Henrissat, B. and Bairoch, A. (1993) *Biochem. J.* **293**, 781–788
- Ooi, T., Shinmyo, A., Okada, H., Hara, S., Ikenaka, T. and Murao, S. (1990) *Curr. Genet.* **18**, 217–222

-
- 35 Saarihtti, H. T., Henrissat, B. and Palva, E. T. (1990) *Gene* **90**, 9–14
- 36 O'Neill, G., Goh, S. H., Warren, R. A. J., Kilburn, D. G. and Miller, Jr., R. C. (1986) *Gene* **44**, 325–330
- 37 Yablonsky, M. D., Bartley, T., Elliston, K. O., Kahrs, S. K., Shalita, Z. P. and Eveleigh, D. E. (1988) *FEMS Symp.* **43**, 249–266
- 38 Hamamoto, T., Foong, F., Shoseyov, O. and Doi, R. H. (1992) *Mol. Gen. Genet.* **231**, 472–479
- 39 Lao, G., Ghangas, G. S., Jung, E. D. and Wilson, D. B. (1991) *J. Bacteriol.* **173**, 3397–3407
- 40 Théberge, M., Lacaze, P., Shareck, F., Morosoli, R. and Kluephel, D. (1992) *Appl. Environ. Microbiol.* **58**, 815–820
- 41 Leskiw, B. K., Lawlor, E. J., Fernandez-Abalos, J. M. and Chater, K. F. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 2461–2465
- 42 Fernández-Moreno, M. A., Caballero, J. L., Hopwood, D. A. and Malpartida, F. (1991) *Cell* **66**, 769–780
- 43 Gramajo, H. C., Takano, E. and Bibb, M. J. (1993) *Mol. Microbiol.* **7**, 835–837
- 44 Leskiw, B. K., Mah, R., Lawlor, E. J. and Chater, K. F. (1993) *J. Bacteriol.* **175**, 1995–2005
- 45 Din, N., Forsythe, I. J., Burtnick, L. D., Gilkes, N. R., Miller, R. C. J., Warren, R. A. J. and Kilburn, D. G. (1994) *Mol. Microbiol.* **11**, 747–755
- 46 Poole, D. M., Hazlewood, G. P., Huskisson, N. S., Virden, R. and Gilbert, H. J. (1993) *FEMS Microbiol. Lett.* **106**, 77–84
-

Received 26 July 1996/23 December 1996; accepted 10 January 1997