Cloning and sequencing of the \textit{celA} gene encoding endoglucanase A of \textit{Butyrivibrio fibrisolvens} strain A46

GEOFFREY P. HAZLEWOOD,¹ KEITH DAVIDSON,¹ JUDITH I. LAURIE,¹ MAREK P. M. ROMANIEC² and HARRY J. GILBERT³

¹Department of Biochemistry, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, UK
²Fermentation Microbiology Laboratory, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
³Department of Agricultural Biochemistry and Nutrition, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

(Received 5 April 1990; revised 3 June 1990; accepted 4 July 1990)

Genomic DNA from \textit{Butyrivibrio fibrisolvens} strain A46 was digested with \textit{EcoRI} and ligated into λgt11. Two recombinant phages isolated from the gene bank hydrolysed carboxymethylcellulose and were shown to contain the same 2.3 kb \textit{EcoRI} restriction fragment, which was cloned into pUC12 to generate pBA46. \textit{Escherichia coli} JM83 harbouring pBA46 expressed an endoglucanase (EGA) which hydrolysed a range of other substrates including barley p-glum, Avicel, filter paper and \textit{p}-nitrophenyl \textit{p}-D-cellobioside. Nucleotide sequencing of the \textit{B. fibrisolvens} strain A46 DNA cloned in pBA46 revealed a single open reading frame (ORF) of 1286 Da, encoding a protein of 48863 Da. Confirmation that the ORF coded for EGA was obtained by comparing the N-terminal sequence of the purified endoglucanase with that deduced from the nucleotide sequence. EGA contains a typical prokaryotic signal peptide at its N-terminus and shows some homology with the \textit{Bacillus} family of cellulases. The enzyme does not contain distinct functional domains, which are prevalent in cellulases from \textit{Pseudomonas fluorescens} subsp. \textit{cellulosa} and \textit{Cellulomonas fimi}.

Introduction

\textit{Butyrivibrio fibrisolvens} plays an important role in the ruminal breakdown of plant structural polysaccharides. In ruminants such as the high-arctic Svalbard reindeer, \textit{B. fibrisolvens} is the major culturable cellulytic bacterium in both summer and winter, representing 66% and 55% of the cellulytic population respectively (Orpin et al., 1985). In view of the well-documented involvement of members of this genus in a variety of rumen processes, it is surprising that their cellulase system has received so little attention. Kopecky (1986) reported that the cellulase of \textit{B. fibrisolvens} UC142 comprised extracellular and cell-associated enzymes, including an endo-1,4-β-glucanase (EC 3.2.1.4) and three β-glucosidases (EC 3.2.1.21). More recently, an endoglucanase gene (\textit{end-I}) from \textit{B. fibrisolvens} H17c was cloned and sequenced (Berger et al., 1989); the encoded protein was homologous with an endoglucanase (EGA) from \textit{Clostridium thermocellum}.

Cellulase genes from a range of micro-organisms have now been sequenced, revealing considerable diversity in the molecular structure of these proteins (see Béguin, 1990, for a review). Some bacteria synthesize enzymes whose structure is highly conserved, while others express cellulases which show little homology at the amino acid level. Cellulases produced by \textit{Pseudomonas fluorescens} subsp. \textit{cellulosa} (Gilbert et al., 1990), \textit{Cellulomonas fimi} (Ong et al., 1989) and \textit{Bacteroides succinogenes} (McGavin & Forsberg, 1989) typically contain distinct structural regions responsible for cellulose binding and catalytic activity. In contrast, endoglucanases from \textit{Bacillus} spp., \textit{Clostridium} spp. (Béguin, 1990) and \textit{Ruminococcus albus} (Poole et al., 1990) in general do not appear to contain substrate-binding domains distinct from the active site. Similarities in molecular architecture, and the conserva-

Abbreviations: EGA, endoglucanase A; ORF, open reading frame.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M37031.

0001-6172 © 1990 SGM
tion of sequence between cellulases expressed by *B. fibrisolvens* remain to be elucidated.

A detailed knowledge of cellulases will contribute greatly to our understanding of fibre digestion in the rumen, and may also facilitate commercial exploitation of this important group of enzymes. With these objectives in mind, we describe here the cloning and sequencing of the *celA* gene from *B. fibrisolvens* strain A46, encoding endoglucanase A (EGA).

**Methods**

**Bacterial strains, plasmids and culture conditions.** *B. fibrisolvens* strain A46 was originally isolated from the rumen of the high-arctic Svalbard reindeer (Orpin et al., 1985), and was cultured statically at 39 °C in rumen-fluid-containing medium (Hobson, 1969; medium 2) with cellulose (0.25%, w/v) and carboxymethylcellulose (CMC; sodium salt, low viscosity, DS > 0.4; 0.25%, w/v) as carbon sources. *Escherichia coli* strains HB101 (Boyer & Roulland-Dussoix, 1969) and JM83 (Norlander et al., 1983) were cultured at 37 °C in Luria broth (LB) adjusted to pH 7.5 with NaOH. Agar (1.5%, w/v), and filter-sterilized ampicillin (100 µg ml⁻¹) and tetracycline (12.5 µg ml⁻¹), were added as required. Vectors used for subcloning and sequencing the endoglucanase gene were pUC12, pUC18, pUC19, M13mp18 and M13mp19 (Norlander et al., 1983). Plasmid pNM52, containing lacZ, was described by Gilbert et al. (1986).

**Isolation of DNA.** Total DNA from *B. fibrisolvens* strain A46 was extracted and purified as described by Romaniec et al. (1987). Large-scale preparations of plasmid DNA were made from *E. coli* by alkaline lysis and purified by caesium chloride density-gradient centrifugation (Gilbert et al., 1986). Isolation of DNA from *B. fibrisolvens* strain A46 was described by Romaniec et al. (1987). Large-scale preparations of plasmid DNA were made from *E. coli* by alkaline lysis and purified by caesium chloride density-gradient centrifugation (Maniatis et al., 1982). The methods of Birnboim & Doly (1979) and Holmes & Quigley (1981) were used for small-scale plasmid DNA extraction and rapid screening of plasmid-containing clones. Large-scale preparations of λ phage DNA were made from liquid lysates essentially as described by Maniatis et al. (1982). *E. coli Y1090* (Young & Davis, 1983), infected with recombinant λ phage which had been recovered in agar plugs from CMCase⁺ plaques, was cultured in NZB medium until lysis occurred. NZB medium contained (g l⁻¹): tryptone, 2; yeast extract, 5; maltose, 2; NaCl, 5; MgCl₂, 7H₂O, 2.

**Construction and screening of the genomic library.** Genomic DNA from *B. fibrisolvens* strain A46 was digested with EcoRI and fragments 2-9 kb in size were recovered by electrodialution and ligated into EcoRI-digested λgt11 DNA (Young & Davis, 1983) (Protocline λgt11 system, Promega Biotec), using conditions specified in the manufacturer's technical bulletin. In situ packaging of the λ DNA was done using protocol II of Maniatis et al. (1982). Recombinant phage (LacZ⁻) were enumerated by pelting dilutions of phage suspension, mixed with *E. coli Y1090*, on NZB soft agarose medium supplemented with isopropyl β-D-thiogalactoside (IPTG; 100 µg ml⁻¹) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal, 250 µg ml⁻¹). For screening the gene bank, λ recombinants (20000 in total) were cultured on lawns of *E. coli Y1090* to a plaque density of 10 cm⁻². Each plate was overlaid with medium containing CMC (0.5%, w/v) and agarose (1.5%, w/v) in PC buffer (50 mm-K₂HPO₄, 12 mm-citric acid, pH 6.5), and incubated at 39 °C for 16 h. Plaques producing CMCase activity were identified by staining with Congo Red (1%, w/v) and containing 1 M-sodium chloride (Teather & Wood, 1982). CMCase⁺ plaques were removed in plugs of agar and phage contained therein were allowed to diffuse into 1 ml SM buffer [10 mm Tris/HCl, pH 7.8, containing 100 mm-NaCl, 20 mm-MgCl₂ and 0.05% (w/v) gelatin].

**General recombinant DNA procedures.** Agarose gel electrophoresis, transformation of *E. coli* and the modification of DNA using restriction enzymes and T4 DNA ligase were done essentially as described by Gilbert et al. (1987). Southern blot hybridization was performed according to Romaniec et al. (1987).

**Nucleotide sequencing.** The 2.3 kb EcoRI fragment carrying the *B. fibrisolvens* strain A46 *celA* gene in pBA46 (Fig. 1) was cleaved with PstI and the two fragments generated (0.8 and 1.5 kb) were each cloned into pUC18 and pUC19 which had been digested with EcoRI and PstI. Sequential digestion of each insert from either end, with Bal31 exonuclease, was used to generate sets of nested deletions; the deleted fragments were cloned into M13mp18 and M13mp19 for nucleotide sequencing. An internal HindIII fragment of 0.14 kb (Fig. 1) was separately cloned into M13 for determining the sequence over the region of the PstI site. Nucleotide sequence was determined using the dideoxy chain-termination method of Sanger et al. (1980), with the exception that T7 DNA polymerase (Tabor & Richardson, 1987) was substituted for the Klenow fragment of DNA polymerase I. Sequences were compiled and ordered using the computer programs described by Staden (1980). The entire sequence of *celA* was determined in both strands.

**Bal31 deletions.** To delete the S' region of *celA*, pUC18 containing full-length *celA* on a 1.4 kb SacI–ClaI fragment (Fig. 1) was linearized with EcoRI, digested with Bal31, and rendered blunt-ended with T4 DNA polymerase (Hall et al., 1988). The truncated fragment was excised by digesting with SphiI, and was ligated into pUC18 which had been cleaved with SmaI and SphiI, such that the coding sequence of *celA* was in frame with the N-terminus of lacZ. The S' end of *celA* was deleted by linearizing pUC18 containing *celA* on the same SacI–ClaI fragment with SphiI, before digesting with Bal31 and filling in with T4 DNA polymerase. The deleted fragment was excised with EcoRI and ligated into pUC18 previously digested with EcoRI and SmaI.

**Assays.** Cellulase-related enzyme activities were assayed at 37 °C in PC buffer, pH 6.5, with substrate concentration in the range 2.5 to 10 mg ml⁻¹. Hydrolysis of acid-soluble cellulose (Sigma cellulase type 50 treated according to Wood, 1971), Avicel (PH105, FMC Corporation), barley β-glucan (Biocon, UK), CMC (sodium salt, Sigma), filter paper (Whatman no. 1), laminarin (Sigma), lichenan (Sigma) and the soluble fraction of oat spelt xylan (Sigma) was measured by following the release of reducing sugar colorimetrically using 3,5-dinitrosalicylic acid (Miller, 1959). Hydrolysis of the chromogenic substrates 4-methylumbelliferyl β-D-cellobioside and p-nitrophenyl β-D-cellobioside was estimated colorimetrically. One unit (U) of enzyme activity was defined as that which resulted in the liberation of 1 µmol product per min⁻¹.

Qualitative measurement of endoglucanase activity was done either by incorporating CMC into LB agar or by overlaying LB plates of *E. coli* clones with PC buffer containing 0.5% (w/v) CMC and 1% (w/v) molten agar. After incubating at 37 °C, activity was revealed by staining with Congo Red. Protein was measured by dye binding (Sedmak & Grossberg, 1977) or by the Lowry method, with bovine serum albumin as standard.

**Fractionation of *E. coli*.** Periplasmic and cytoplasmic fractions of *E. coli* clones harbouring *celA* were prepared as described previously (Hazlewood et al., 1990).

**Purification of endoglucanase A (EGA).** *E. coli* IM83 harbouring full-length *celA* cloned in pUC18 was cultured for 16 h at 37 °C in LB containing ampicillin (100 µg ml⁻¹). Cells from 3 litres of culture were collected by centrifugation and were resuspended in 50 mm Tris/HCl buffer, pH 8, to approximately 1/30 of their original volume. Cell-free extract prepared by ultrasonication was loaded onto an anion-exchange column (2.6 × 21 cm; DEAE-Trisacryl M, IBF Biotechnics) which was eluted with a 0–500 mm-NaCl linear gradient in 50 mm-Tris/HCl buffer.
buffer, pH 8. Fractions containing endoglucanase activity were pooled, dialysed against 10 mM-Tris/HCl buffer, pH 8, and applied to a second column containing DEAE-Trisacryl M, which was eluted with a 100-400 mM-NaCl linear gradient. Active fractions were dialysed as above and concentrated by vacuum evaporation.

**Sequencing of EGA.** EGA, partially purified by anion-exchange chromatography, was fractionated by SDS-PAGE (10% acrylamide, 0-1% SDS) (Laemmli, 1970) and transferred to Immobilon PVDF (Millipore) by electroblotting. Putative EGA (47 kDa), stained with Page Blue G90, was excised from the transfer membrane and applied directly to a 470 A gas-phase sequenator equipped with a 120 A on-line PTH analyser (Applied Biosystems; Hunkapillar et al., 1983).

**Results and Discussion**

**Isolation and characterization of CMCase+ clones**

The genomic library constructed by ligating EcoRI fragments of *B. fibrisolvens* strain A46 DNA into igt11 had an insertion rate of the order of 70%. Twenty-thousand igt11 recombinants were screened and two CMCase+ plaques were found. DNA was isolated from phage recovered from each of these plaques and the *B. fibrisolvens* EcoRI restriction fragments were excised and ligated into pUC12. Restriction enzyme analysis showed that both plasmids contained the same 2.3 kb EcoRI fragment, which could be cleaved with HindIII or PvuII, but was devoid of BamHI, BglII, PstI, SalI or SmaI sites (Fig. 1). In Southern blot hybridizations, a 32P-labelled probe prepared by nick-translation of pUC12 containing the cloned fragment (pBA46) hybridized with a single 2.4 kb EcoRI fragment of *B. fibrisolvens* strain A46 genomic DNA. The same probe also hybridized with genomic DNA digested with HindIII and PvuII in a manner which confirmed the physical map and indicated that the fragment had been cloned without deletion or rearrangement (data not shown); pUC12 alone was not homologous with *B. fibrisolvens* strain A46 genomic DNA.

**Subcloning and expression of celA in E. coli JM83**

The 2.3 kb EcoRI fragment carrying the *celA* gene was subcloned from pBA46 into pUC18 and pUC19 and transformed into *E. coli* JM83. In both orientations *celA* directed the synthesis of functional EGA, which suggests that the *celA* promoter is operative in *E. coli*. A 1.4 kb ScaI–ClaI restriction fragment (Fig. 1) directed the synthesis of an active endoglucanase when cloned into pUC18, but not when inserted in the opposite orientation into pUC19; this finding indicates that the ScaI–ClaI fragment contains the protein-coding region of *celA*, but not the regulatory sequences of the gene, and confirms that *celA* is transcribed in the direction indicated in Fig. 1.

**Substrate specificity of EGA**

With respect to substrate specificity, EGA from *B. fibrisolvens* strain A46 was quite similar to End1 of *B. fibrisolvens* H17c (Berger et al., 1989). Enzyme contained in cell-free extracts made from *E. coli* harbouring *celA* was most active in hydrolysing barley β-glucan (70–75% β-1,4; 25–30% β-1,3-linkages) (Table 1). Lichenan, a
Cloning of B. fibrisolvens endoglucanase gene

Fig. 2. Nucleotide sequence of the celA gene from B. fibrisolvens strain A46. The derived amino acid sequence of EGA is given in one-letter code. The signal peptide is indicated by a solid underline, and the broken underline shows the extent of the experimentally-determined N-terminal sequence of mature EGA. The putative Shine–Dalgarno sequence is boxed.

Nucleotide sequence of celA

Translation of the nucleotide sequence of the 2.3 kb EcoRI fragment cloned from B. fibrisolvens strain A46 revealed a single open reading frame (ORF) of 1296 bp coding for a polypeptide of 432 amino acid residues, with a molecular mass of 48 863 Da (Fig. 2). The putative translation initiation codon was immediately preceded by a potential ribosome-binding site (GGGAGGA), and the N-terminus of the polypeptide had the general characteristics of a prokaryotic signal peptide, being composed of a sequence containing 19 hydrophobic amino acids, preceded by a shorter sequence of positively-charged residues. Confirmation that the ORF codes for EGA was obtained experimentally by determining the molecular size and N-terminal sequence of EGA synthesized in E. coli JM83 harbouring full-length celA. Endoglucanase activity from cell-free extract, partially purified by anion-exchange chromatography and fractionated by SDS-PAGE, contained a major band corresponding to a molecular mass of 47 kDa (data not

similar mixed-linkage polysaccharide from Cetraria islandica was degraded to about the same extent as CMC. The enzyme also released reducing sugar during prolonged incubation with recalcitrant forms of cellulose (acid-swollen cellulose, Avicel, filter paper and cotton), but the activity against these substrates amounted to only 0.1% to 4% of the activity against CMC. A low level of activity was also observed against the soluble fraction from oat spelt xylan and against the P-1,3-linked polysaccharide laminarin (Table 1). Release of reducing sugar from CMC was accompanied by a rapid decline in substrate viscosity, characteristic of endo-β-1,4-glucanase activity. Cleavage of the aglycone bond in p-nitrophenyl β-D-cellobioside and 4-methylumbelliferyl β-D-cellobioside suggests that EGA also has some endoglucanase activity, but numerous other prokaryotic exoglucanases and xylanases also show activity against these substrates, so some caution is advisable in interpreting these results. No activity was detected against the aryl-glucosides p-nitrophenyl β-D-glucoside and 4-methylumbelliferyl β-D-glucoside.
having G or C in the third position occurred in codon frequencies of 18.6% and 41.3 mol%, reported for the B. jibrisolvens H17c genome (Mannarelli, 1988) and the B. jibrisolvens strain A44 celA gene (Berger et al., 1989; BCguin, 1990). The subfamily containing cellulases with 16-17% amino acid identity to GTG, which can be subdivided into subfamilies (Henrissat et al., 1990) contains approximately 50% homology with N-terminal regions of endoglucanases from Bacillus sp. (Fukumori et al., 1989), Clostridium acetobutylicum (Zappe et al., 1988) and Erwinia chrysanthemi (Guiseppi et al., 1988) (Fig. 3). No homology was found with End-1 from B. fibrisolvens H17c. A classification based on hydrophobic cluster analysis has recently been employed to assign the cellulases of prokaryotic and lower eukaryotic organisms to six distinct families, some of which can be subdivided into subfamilies (Henrissat et al., 1989; Béguin, 1990). The subfamily A2 contains endoglucanases from Bacillus spp. and from Clostridium acetobutylicum and Er. chrysanthemi. These endoglucanases show extensive homology over the entire length of their N-terminal catalytic domain, which comprises 300 to 400

### Table 2. Codon usage of B. fibrisolvens strain A46 celA gene and B. fibrisolvens H17c end-1 gene

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Codon</th>
<th>celA</th>
<th>end-1</th>
<th>Amino acid</th>
<th>Codon</th>
<th>celA</th>
<th>end-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>TTT</td>
<td>5</td>
<td>12</td>
<td>S</td>
<td>TCT</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>F</td>
<td>TTC</td>
<td>5</td>
<td>8</td>
<td>S</td>
<td>TCC</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>L</td>
<td>TTA</td>
<td>7</td>
<td>3</td>
<td>S</td>
<td>TCA</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>L</td>
<td>TTG</td>
<td>4</td>
<td>1</td>
<td>S</td>
<td>TCG</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>CTT</td>
<td>21</td>
<td>14</td>
<td>P</td>
<td>CCT</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>L</td>
<td>CTC</td>
<td>0</td>
<td>2</td>
<td>P</td>
<td>CCC</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>L</td>
<td>CTA</td>
<td>0</td>
<td>0</td>
<td>P</td>
<td>CCA</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>L</td>
<td>CTG</td>
<td>0</td>
<td>2</td>
<td>P</td>
<td>CCG</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>I</td>
<td>ATT</td>
<td>15</td>
<td>13</td>
<td>T</td>
<td>ACT</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>I</td>
<td>ATC</td>
<td>9</td>
<td>15</td>
<td>T</td>
<td>ACC</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>ATA</td>
<td>4</td>
<td>7</td>
<td>T</td>
<td>ACA</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>M</td>
<td>ATG</td>
<td>11</td>
<td>13</td>
<td>T</td>
<td>ACG</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>V</td>
<td>GTT</td>
<td>15</td>
<td>24</td>
<td>A</td>
<td>GCT</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>V</td>
<td>GTC</td>
<td>5</td>
<td>1</td>
<td>A</td>
<td>GCC</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>V</td>
<td>GTA</td>
<td>8</td>
<td>20</td>
<td>A</td>
<td>GCA</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>V</td>
<td>GTG</td>
<td>5</td>
<td>2</td>
<td>A</td>
<td>GCG</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>TAT</td>
<td>14</td>
<td>19</td>
<td>C</td>
<td>TGT</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>TAC</td>
<td>5</td>
<td>9</td>
<td>C</td>
<td>TGC</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Ter</td>
<td>TAA</td>
<td>0</td>
<td>1</td>
<td>Ter</td>
<td>TGA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ter</td>
<td>TAG</td>
<td>1</td>
<td>0</td>
<td>W</td>
<td>TGG</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>H</td>
<td>CAT</td>
<td>6</td>
<td>5</td>
<td>R</td>
<td>CTG</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td>CAC</td>
<td>2</td>
<td>3</td>
<td>R</td>
<td>CGC</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Q</td>
<td>CAA</td>
<td>4</td>
<td>0</td>
<td>R</td>
<td>CGA</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Q</td>
<td>CAG</td>
<td>7</td>
<td>13</td>
<td>R</td>
<td>CGG</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>AAT</td>
<td>16</td>
<td>26</td>
<td>S</td>
<td>AGT</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>AAC</td>
<td>11</td>
<td>17</td>
<td>S</td>
<td>AGC</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>K</td>
<td>AAA</td>
<td>12</td>
<td>6</td>
<td>R</td>
<td>AGA</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>K</td>
<td>AAG</td>
<td>24</td>
<td>19</td>
<td>R</td>
<td>AGG</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>GAT</td>
<td>34</td>
<td>32</td>
<td>G</td>
<td>GGT</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>D</td>
<td>GAC</td>
<td>6</td>
<td>12</td>
<td>G</td>
<td>GGC</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>E</td>
<td>GAA</td>
<td>27</td>
<td>12</td>
<td>G</td>
<td>GGA</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>E</td>
<td>GAG</td>
<td>16</td>
<td>10</td>
<td>G</td>
<td>GGG</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Comparison of sequences revealed that residues 151 to 283 of EGA had approximately 50% homology with N-terminal regions of endoglucanases from Bacillus sp. (Fukumori et al., 1989), Clostridium acetobutylicum (Zappe et al., 1988) and Erwinia chrysanthemi (Guiseppi et al., 1988) (Fig. 3). No homology was found with End-1 from B. fibrisolvens H17c. A classification based on hydrophobic cluster analysis has recently been employed to assign the cellulases of prokaryotic and lower eukaryotic organisms to six distinct families, some of which can be subdivided into subfamilies (Henrissat et al., 1989; Béguin, 1990). The subfamily A2 contains endoglucanases from Bacillus spp. and from Clostridium acetobutylicum and Er. chrysanthemi. These endoglucanases show extensive homology over the entire length of their N-terminal catalytic domain, which comprises 300 to 400...
Cloning of *B. fibrisolvens* endoglucanase gene

Fig. 3. Amino acid sequence alignments for endoglucanases from *B. fibrisolvens* strain A46 (EGA; this study), *Bacillus* sp. strain N-4 (CEL-A, CEL-B and CEL-C; Fukumori et al., 1989), *Bacillus subtilis* PAP115 (Bsl; Zappe et al., 1988), *B. subtilis* DLG (Bs2; Robson & Chambliss, 1987), *Erwinia chrysanthemi* (EGZ; Guiseppi et al., 1988) and *Clostridium acetobutylicum* (Cal; Zappe et al., 1988). Residues which are identical or similar in structure in all sequences compared are enclosed within a box.
amino acid residues. Our finding that EGA displays significant homology with these enzymes suggests that the celA gene of B. fibrisolvens strain A46, like the other genes in the A2 subfamily, evolved from a common ancestral gene. Alternatively, conservation of sequence in cellulases from such a variety of species may have resulted from the convergent evolution of structures having a common function.

To date, the primary structures have been determined for five endoglucanases produced by anaerobic rumen bacteria. Cellulases from Ruminococcus albus are highly conserved (Poole et al., 1990), suggesting a common ancestral gene and strong selection pressure for the retention of conserved sequences. In contrast, the endoglucanases (EGA and End1) produced by two strains of B. fibrisolvens clearly belong to different cellulase subfamilies. This lack of homology between B. fibrisolvens cellulases could reflect the apparent ease with which genetic exchange occurs between members of this genus and other bacteria (Teather, 1985), thus providing a mechanism for the acquisition of cellulase genes from diverse sources. Alternatively, it may be further evidence of the inadequacy of the classification applied to rumen bacteria. If, as seems likely, rumen butyrivibrios comprise a genetically heterogeneous group consisting of two or more genera and numerous species (Hazlewood & Teather, 1988) it would not be particularly surprising to find that their cellulase enzymes derive from more than one evolutionary source.

Structure of EGA in relation to function

There is now clear evidence that the catalytic function of some prokaryotic cellulases is vested in a clearly defined independent catalytic domain which may account for less than half of the total gene (Béguin, 1990; Gilbert et al., 1990). In such cases, a substantial non-essential portion of the gene may be deleted without affecting the catalytic function of the encoded protein. To determine whether this is the case for EGA from B. fibrisolvens strain A46, truncated derivatives of celA were generated by digesting either the 5' or the 3' end of the gene with Bal3I exonuclease. Removal of 240 bp from the 5' end or 190 bp from the 3' end of celA resulted in complete inactivation of EGA. These results indicate that virtually the whole polypeptide is required to maintain the functional integrity of EGA. The apparent absence of a cellulose-binding domain distinct from the active site is supported by cellulose-binding studies (data not shown) which revealed that recombinant EGA produced by E. coli harbouring full-length celA did not bind strongly to Avicel in the manner typical of endoglucanases having distinct cellulose-binding domains (Gilbert et al., 1990; Ong et al., 1989).

Aspects of the molecular architecture of cellulases from the rumen species B. fibrisolvens, R. albus and Bacteroides succinogenes have now been reported. To date, only the latter organism has been shown to produce an endoglucanase containing a well-defined cellulose binding domain which is distinct from the catalytic centre (McGavin & Forsberg, 1988). From an evolutionary viewpoint, the prospect of there being major differences between the cellulase systems of rumen bacteria is surprising, since all organisms evolving in the rumen have presumably been subjected to similar environmental influences. All three of these species are widely distributed in ruminants and are effective in degrading dietary cellulose. In vivo each is found colonizing plant tissue, but unlike Bac. succinogenes, R. albus and B. fibrisolvens produce cellulases which have not as yet been shown to contain functional cellulose-binding domains distinct from the active site, and form only a loose association with plant cell walls (Chesson & Forsberg, 1988).

Future work by this group will focus on the mechanisms by which different bacterial cellulase systems hydrolyse plant cell wall polysaccharides in the rumen.

We wish to thank the AFRC (Grant no. LRG 138) for supporting this work.

References


