

Properties and gene structure of a bifunctional cellulolytic enzyme (Cela) from the extreme thermophile '*Anaerocellum thermophilum*' with separate glycosyl hydrolase family 9 and 48 catalytic domains

Vladimir Zverlov,¹† Sabine Mahr,² Kathrin Riedel²
and Karin Bronnenmeier²

Author for correspondence: Karin Bronnenmeier. Tel: +49 89 2892 2637. Fax: +49 89 2892 2360.
e-mail: bronnen@biol.chemie.tu-muenchen.de

¹ Institute of Molecular Genetics, Russian Academy of Science, Kurchatov Sq. 46, 123 182 Moscow, Russia

² Institute for Microbiology, Technical University Munich, Arcisstr. 21, D-80290 München, Germany

A large cellulolytic enzyme (Cela) with the ability to hydrolyse microcrystalline cellulose was isolated from the extremely thermophilic, cellulolytic bacterium '*Anaerocellum thermophilum*'. Full-length Cela and a truncated enzyme species designated Cela' were purified to homogeneity from culture supernatants. Cela has an apparent molecular mass of 230 kDa. The enzyme exhibited significant activity towards Avicel and was most active towards soluble substrates such as CM-cellulose (CMC) and β -glucan. Maximal activity was observed between pH values of 5 and 6 and temperatures of 95 °C (CM-cellulase) and 85 °C (Avicelase). Cellobiose, glucose and minor amounts of celotriose were observed as end-products of Avicel degradation. The Cela-encoding gene was isolated from genomic DNA of '*A. thermophilum*' by PCR and the nucleotide sequence was determined. The *cela* gene encodes a protein of 1711 amino acids (190 kDa) starting with the sequence found at the N-terminus of Cela purified from '*A. thermophilum*'. Sequence analysis revealed a multidomain structure consisting of two distinct catalytic domains homologous to glycosyl hydrolase families 9 and 48 and three domains homologous to family III cellulose-binding domain linked by Pro-Thr-Ser-rich regions. The enzyme is most closely related to Cela of *Caldicellulosiruptor saccharolyticus* (sequence identities of 96 and 97% were found for the N- and C-terminal catalytic domains, respectively). Endoglucanase CelZ of *Clostridium stercorarium* shows 70.4% sequence identity to the N-terminal family 9 domain and exoglucanase CelY from the same organism has 69.2% amino acid identity with the C-terminal family 48 domain. Consistent with this similarity on the primary structure level, the 90 kDa truncated derivative Cela' containing the N-terminal half of Cela exhibited endoglucanase activity and bound to microcrystalline cellulose. Due to the significantly enhanced Avicelase activity of full-length Cela, exoglucanase activity may be ascribed to the C-terminal family 48 catalytic domain.

Keywords: '*Anaerocellum thermophilum*', cellulase, endo-1,4- β -glucanase, exo-1,4- β -glucanase, thermostability

†Present address: Institute for Microbiology, Technical University Munich, Arcisstr. 21, D-80290 München, Germany.

Abbreviations: CMC, CM-cellulose; CMCase, CM-cellulase.

The GenBank accession number for the nucleotide sequence reported in this paper is Z86105.

INTRODUCTION

'*Anaerocellum thermophilum*' is an extremely thermophilic, obligately anaerobic, chemoorganotrophic bacterium which was isolated from a hot spring with residues of higher plants in the Valley of Geysers on the

Kamchatka peninsula, Russia (Svetlichnyi *et al.*, 1990). It utilizes a wide spectrum of polymeric carbohydrates and sugars. The preferred substrate is cellulose, which is quite rapidly and efficiently metabolized. Phylogenetic analysis revealed that '*A. thermophilum*' strain Z-1320 is closely related to strain Tp8T6331 isolated from New Zealand thermal springs which was tentatively named '*Caldocellum saccharolyticum*' and has now been designated *Caldicellulosiruptor saccharolyticus* (Rainey *et al.*, 1994). The phylogenetic position of both strains as a novel lineage within the *Bacillus/Clostridium* sub-phylum of the Gram-positive bacteria has recently been reported (Rainey *et al.*, 1993).

To use cellulose or hemicellulose as sources of carbon and energy, micro-organisms produce a repertoire of hydrolytic enzymes, with various specificities, which act cooperatively to convert these substrates to their constituent sugars (Warren, 1996). These enzymes commonly show a modular organization and consist of a single catalytic domain linked to one or more non-catalytic domains. However, bifunctional polysaccharidases composed of two dissimilar catalytic domains joined by linker regions have been identified via gene cloning in *Ruminococcus flavefaciens* (Zhang & Flint, 1992; Flint *et al.*, 1993), *Clostridium thermocellum* (Ahsan *et al.*, 1996) and *Ca. saccharolyticus*. In the latter organism, a cellulase-hemicellulase gene cluster composed of the genes *celA* (Te'o *et al.*, 1995), *celB* (Saul *et al.*, 1990), *manA* (Gibbs *et al.*, 1992) and *celC* (Morris *et al.*, 1995) has been characterized.

In this paper, we describe purification and characterization of the predominant cellulolytic enzyme produced by '*A. thermophilum*' and show that this large enzyme, which contains two separate catalytic regions within the same polypeptide, is encoded by a single gene.

METHODS

Bacterial strains, plasmids and growth conditions. '*A. thermophilum*' Z-1320 (Svetlichnyi *et al.*, 1990) was obtained from the Laboratory of Lithotrophic Microorganisms at the Institute for Microbiology, RAS, Moscow, Russia. Cells were grown at 70 °C and pH 6.7 in a 7 l fermenter under anaerobic conditions in CM5 medium (Weimer *et al.*, 1984) supplemented with 0.2% yeast extract and 0.3% tryptone. Cellobiose (0.4%) was added as carbon source.

The cloning vector/host strain combinations used for cloning and sequencing were pGEM-T (Promega) or pUC18/*Escherichia coli* DH5 α (Hanahan, 1983; Yanisch-Perron *et al.*, 1985). Recombinant *E. coli* cells were grown at 37 °C in Luria-Bertani broth containing 100 μ g ampicillin ml⁻¹.

Purification of CelA and CelA' from '*A. thermophilum*'. Culture supernatant (20 l) was concentrated by ultrafiltration employing a Minisette tangential flow system (Filtron) equipped with Nova filters with a nominal molecular mass limit of 30 000 Da. The retentate (660 ml) was desalted by five cycles of ultrafiltration following a twofold dilution with 20 mM Tris/HCl, pH 8.0. The chromatographic steps of the purification were performed with a Pharmacia FPLC system.

The crude exoenzyme preparation (507 mg protein) was loaded on a Pharmacia XK 26/20 column filled with 32 ml Q Sepharose Fast Flow which had been equilibrated with 20 mM Tris/HCl, pH 8.0. Fractions in the flow-through were pooled, adjusted with 1.2 M ammonium sulfate and applied to a Pharmacia Phenyl Sepharose HP HiLoad 16/10 hydrophobic interaction column equilibrated with 20 mM Tris/HCl, 1.2 M ammonium sulfate, pH 8.0. Elution was performed with a 420 ml linear gradient (1.2–0.0 M ammonium sulfate) in 20 mM Tris/HCl (pH 8.0) at a flow rate of 2 ml min⁻¹. Two peaks exhibiting Avicelase as well as CM-cellulase (CMCase) activity were eluted. The CelA-containing fractions of the peak eluted at the end of the salt gradient were pooled and concentrated in a Macrosep centrifugal microconcentrator (Filtron). The concentrate (2.5 ml) was applied to a Pharmacia Superdex 200 prep-grade HiLoad 16/60 gel filtration column equilibrated with 100 mM sodium succinate (pH 6.0) containing 1 M NaCl. The column was eluted at a flow rate of 1 ml min⁻¹ with equilibration buffer. Two repeat runs were performed.

The CelA'-containing fractions of the peak eluted at 0.3 M ammonium sulfate from the Phenyl Sepharose column described above were subjected to chromatofocusing on a Pharmacia Mono P HR 5/20 column equilibrated with 75 mM Tris/HCl, pH 9.5. The column was eluted with 50 ml Pharmacia Polybuffer 96, diluted 1:10 and adjusted to pH 7.0 at a flow rate of 1 ml min⁻¹. Fractions eluting at a pH of 8.65 were applied to a Pharmacia HR 5/5 column filled with POROS Self Pack 20 ET (PerSeptive Biosystems) and equilibrated with 50 mM sodium phosphate buffer, 1.2 M ammonium sulfate, pH 7.0. Elution was performed with a 15 ml linear gradient (1.2–0.0 M ammonium sulfate) in 50 mM sodium phosphate (pH 7.0) at a flow rate of 10 ml min⁻¹.

Enzyme assays. Enzyme assays were done at 72 °C in 0.5 or 0.75 ml reaction mixtures containing 0.1 M sodium succinate (pH 6.0) and either 0.5% (w/v) soluble polysaccharide substrates or 1% (w/v) non-soluble substrates. The enzymic liberation of reducing groups was determined with the dinitrosalicylic acid reagent (Wood & Bhat, 1988). One unit of enzyme activity was defined as the amount of enzyme needed to release 1 μ mol glucose-equivalent reducing groups min⁻¹.

CM-cellulose (CMC; low viscosity) was purchased from Sigma. Avicel (microcrystalline cellulose 0.02 mm) was from Serva. Oat spelt xylan was from Fluka. Barley β -glucan was from Megazyme. Acid-swollen Avicel was prepared by incubating Avicel in concentrated HCl with continuous agitation for 2 h (Sakamoto *et al.*, 1984).

Analytical methods. Protein concentrations were measured by the method of Sedmak & Grossberg (1977). SDS-PAGE was performed in 7.5% polyacrylamide slab gels in the presence of 0.1% SDS according to Laemmli (1970). CMCase and β -glucanase bands were visualized by a modification of the zymogram technique described by Schwarz *et al.* (1987): gels were incubated in 1% CMC or β -glucan solutions instead of incorporation of these substrates into the polyacrylamide gels. Avicelase bands were detected by the filter paper affinity blotting technique of Montgomery & Fu (1988). Cellulose-binding studies were performed as described by Hall *et al.* (1995). Protein glycosylation was analysed with the GlykoTrack kit (Oxford GlykoSystems). N-terminal amino acid sequences were determined by Edman degradation using a gas-phase amino acid Sequenator model 477A (Applied Biosystems). Cellodextrins were analysed by HPLC at 85 °C on an Animex HPX-42 A column (Bio-Rad) with water as eluent.

Table 1. Nucleotide sequences of PCR primers and biotinylated oligonucleotides

For oligonucleotides with mixed bases (wobbles) the following letters were used: w represents a/t; y, c/t; s, g/c; r, a/g; and n, a/g/c/t.

Primer	Sequence	Annealing temp. (°C)
<i>APr1</i>	5'-ggnwsnttyaaytayggngargcnyntcaraa-3'	50
<i>APr2c</i>	5'-gcrtrtrartcrcangcncac-3'	50
<i>APr3</i>	5'-actgctcatagctcatgggca-3'	58
<i>APr5(SalI)</i>	5'-ggagatgtgaaggtatgggtcgcggaccag-3'	62
<i>APr6(BamHI)</i>	5'-gtatctgttcataaccacagatcctctgcc-3'	62
<i>APr7</i>	5'-ggctggaggaacgggta-3'	55
Oligonucleotide		Hybridization temp. (°C)
<i>APr4-Bio</i>	5'-tatggagcgcttgttggtg-3'	55
<i>APr8-Bio</i>	5'-gctatcatggtattatgcatg-3'	55

Recombinant DNA techniques, sequencing and PCR.

Preparation of chromosomal and plasmid DNA, endonuclease digestion, ligation, transformation and Southern hybridization analysis were carried out using standard procedures (Sambrook *et al.*, 1989). Digoxigenin labelling was performed with the DIG DNA Labeling and Detection kit (Boehringer Mannheim). Enzymes for DNA modification were purchased from Boehringer Mannheim. The DNA sequence was determined from supercoiled double-stranded plasmid DNA by cycle sequencing of both strands (Amersham Thermo-sequenase Cycle Sequencing kit) with biotinylated primers. DNA fragments were detected with GATC 1500 Direct-Blotting Electrophoresis apparatus using streptavidin-conjugated alkaline phosphatase and nitro blue tetrazolium-5-bromo-4-chloro-3-indolyl phosphate as chromogenic substrate (TROPIX). Sequence data were analysed and compared with the DNASIS/PROSIS for Windows package (Hitachi Software Engineering). Nucleotide and protein sequence databases were screened using the BLAST software at the NCBI server (<http://www.ncbi.nlm.nih.gov>).

PCR was carried out using the synthetic oligonucleotide primers *APr1*, *APr2c*, *APr3*, *APr5(SalI)*, *APr6(BamHI)* and *APr7* (Table 1 and Fig. 4) with chromosomal DNA from '*A. thermophilum*' as a template and an Expand High Fidelity PCR system (Boehringer Mannheim) used according to the supplier's protocol. In asymmetric PCR (L. Richter & W. Ludwig, unpublished), only one primer is used which binds specifically to one DNA strand; the complementary strand is synthesized by unspecific binding, analogous to the random primer method. The template strands were separated at 94 °C (2 min) before starting 30 reaction cycles (94 °C/15 s, 50–62 °C/30 s, 72 °C/2 min plus elongation of 10 s for each cycle). For asymmetric PCR, 0.75 µl of the Expand High Fidelity PCR system enzyme mix was added to the reaction mixture after 30 reaction cycles and the cycle profile was repeated.

RESULTS

Enzyme purification

Culture supernatants of '*A. thermophilum*' grown on cellobiose exhibited CMCase, β -glucanase and Avicelase

activities. SDS-PAGE analysis of crude exoenzyme preparations revealed a prominent protein band with a molecular mass of approximately 230 kDa (Fig. 1). This protein band was shown to exhibit CMCase and β -glucanase activities by *in situ* activity staining. Affinity binding to filter paper was taken as evidence for Avicelase activity. Purification of the 230 kDa enzyme confirmed the notion that all three activities reside in the same protein (Fig. 1). As summarized in Table 2, a fivefold purification by FPLC chromatographic methods was sufficient to yield a homogeneous protein (Fig. 1). This implies that the large cellulolytic enzyme, later identified as CelA, constitutes at least 20% of the total extracellular protein produced by cellobiose-grown '*A. thermophilum*'.

During hydrophobic interaction chromatography on Phenyl Sepharose, a 90 kDa enzyme exhibiting CMCase activity was separated from the 230 kDa cellulase. The 90 kDa enzyme could be purified to electrophoretic homogeneity by chromatofocusing on a Mono P column and hydrophobic interaction chromatography on a POROS ET column (Fig. 2). Based on the molecular mass, substrate specificity (Table 3) and N-terminal amino acid sequence (Fig. 3), the 90 kDa enzyme species was thought to have arisen from limited proteolysis of the CelA cellulase. This truncated derivative consisting of the N-terminal half of CelA was designated CelA'. Like the intact protein, CelA' was able to bind to microcrystalline cellulose (data not shown).

N-terminal amino acid sequence

N-terminal sequencing of the 230 kDa cellulase purified from '*A. thermophilum*' by Edman degradation yielded the sequence GSFNYGEALQKAIMFYEFQMSGK for the first 23 amino acids. This sequence is found at positions 24 through 46 of the primary structure deduced

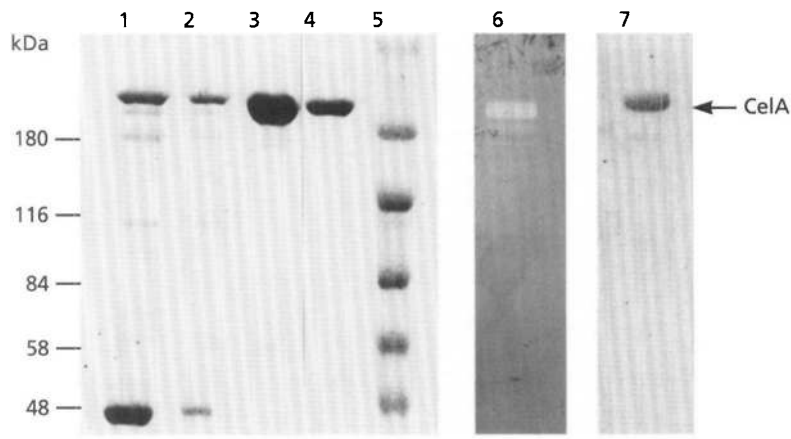


Fig. 1. SDS-PAGE of the fractions obtained during purification of the '*A. thermophilum*' cellulase CelA. The gel was stained for protein (lanes 1–5) and CMCase activity (lane 6). A filter paper affinity blot is shown in lane 7. Lanes: 1, concentrated culture supernatant; 2, flow-through of the Q Sepharose column; 3, pooled fractions from the Phenyl Sepharose column; 4, pooled fractions from the Superdex column; 5, molecular mass markers (values in kDa on the left); 6 and 7, pooled fractions from the Superdex column.

Table 2. Purification of '*A. thermophilum*' cellulase from concentrated culture supernatant

Activity was determined under standard assay conditions at 72 °C with Avicel as the substrate.

Purification step	Volume (ml)	Protein (mg)	Activity (U)	Specific activity (mU mg ⁻¹)	Purification (-fold)	Yield (%)
Culture supernatant	2000	810	45	60	1.0	100
Ultrafiltration concentrate	660	507	33	65	1.1	73
Q Sepharose	750	248	19	77	1.3	42
Phenyl Sepharose	46	45	7	156	2.6	16
Superdex	10	14	4	285	4.8	9

from the nucleotide sequence of the *celA* gene of *Ca. saccharolyticus* (Te'o *et al.*, 1995; GenBank accession no. L32742). The signal peptidase cleavage site of the *Ca. saccharolyticus* CelA protein was predicted to lie between residues 20 and 21 or 23 and 24, respectively. Thus, the mature proteins from '*A. thermophilum*' and *Ca. saccharolyticus* show identical sequences at their N-termini. Since the N-terminal catalytic domain of CelA from *Ca. saccharolyticus* belongs to glycosyl hydrolase family 9, CelA from '*A. thermophilum*' was also thought to be a member of this family. Accordingly, other family 9 cellulases, such as CelZ of *Clostridium stercorarium* (Jauris *et al.*, 1990; GenBank accession no. X55299) and CelI of *Cl. thermocellum* (Hazlewood *et al.*, 1993; GenBank accession no. L04735), are also highly homologous at their N-termini with the N-terminal sequence determined for '*A. thermophilum*' CelA (Fig. 3).

PCR cloning of the *celA* gene region

The 5' region of the *celA* gene was amplified from '*A. thermophilum*' genomic DNA by PCR with two oligonucleotide primers. The first 11 N-terminal amino acids provided sufficient information to deduce the sequence of the forward primer *APr1* (Table 1 and Fig. 4). The reverse primer *APr2c* (Table 1 and Fig. 4) was derived from the complementary nucleotide sequence deduced

from a conserved sequence region identified in the family 9 cellulases CelA of *Ca. saccharolyticus*, CelZ of *Cl. stercorarium* and CelI of *Cl. thermocellum* (Fig. 3). The 1.3 kb PCR products from two independent reactions were cloned into the *E. coli* vector pGEM-T. The DNA inserts from plasmids pATP1 and pATP2 isolated from two different recombinant clones were sequenced and had the same nucleotide sequence, thus eliminating the possibility of errors being introduced during PCR amplification.

Cloning of the missing part of the *celA* gene was achieved by two asymmetric PCR reactions using the single primers *APr3* and *APr7* (Table 1 and Fig. 4). The amplified DNA fragments were screened with biotinylated oligonucleotide probes (*APr4-Bio* and *APr8-Bio*; Table 1 and Fig. 4) by Southern hybridization. PCR products of 1.1, 1.8 and 2.3 kb were selected from the reaction with primer *APr3* and a 1.5 kb product was selected from the reaction with primer *APr7*. These fragments were cloned into vector pGEM-T and the resulting plasmids were designated pATP3-1, pATP3-2, pATP3-3 and pATP7, respectively (Fig. 4). DNA sequencing revealed that pATP3-2 and pATP3-3 had lost part of their DNA inserts, probably as a result of recombination events. The gap was filled by PCR with the two primers *APr5*(*SalI*) and *APr6*(*Bam*HI) containing recognition sites for *SalI* and *Bam*HI, which were used for cloning of the resulting PCR fragment into

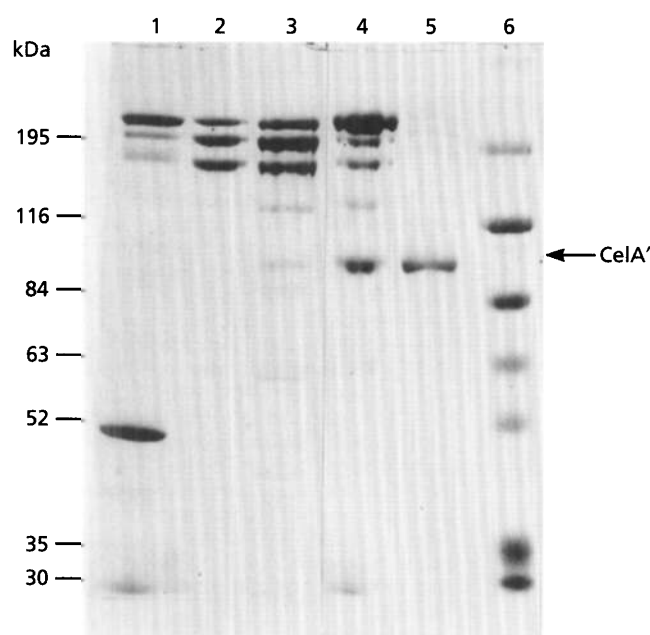


Fig. 2. SDS-PAGE of the fractions obtained during purification of the truncated cellulase CelA'. Lanes: 1, concentrated culture supernatant; 2, flow-through of the Q Sepharose column; 3, pooled fractions from the Phenyl Sepharose column; 4, pooled fractions of the Mono P column; 5, pooled fractions of the POROS 20 ET column; 6, molecular mass markers (values in kDa on the left).

Table 3. Substrate specificity of the '*A. thermophilum*' cellulase

Standard assays were performed at 72 °C using CelA concentrations of 7.4 ng ml⁻¹ or 74 pg ml⁻¹ (in the case of β -glucan). CelA' concentrations were 7.1 ng ml⁻¹ or 71 pg ml⁻¹ (in the case of β -glucan). Incubations were carried out for 15 min for CMC and β -glucan, 45 min for xylan and up to 5 h for Avicel. The results are the means of nine separate determinations in which values did not deviate by more than 15% of the mean.

Substrate	Activity (mU nmol ⁻¹)	
	CelA	CelA'
Avicel	55	18
Avicel, acid-swollen	82	76
CMC	1758	1801
β -Glucan (barley)	100510	88244
Xylan (oat spelts)	372	123

pUC18 yielding plasmid pATP5 (Fig. 4). In the cases of pATP5 and pATP7, the nucleotide sequences were again determined for two independent PCR clones.

Nucleotide sequence analysis

The origin and structural integrity of the sequenced 5.5 kb DNA fragment were checked with Southern blot

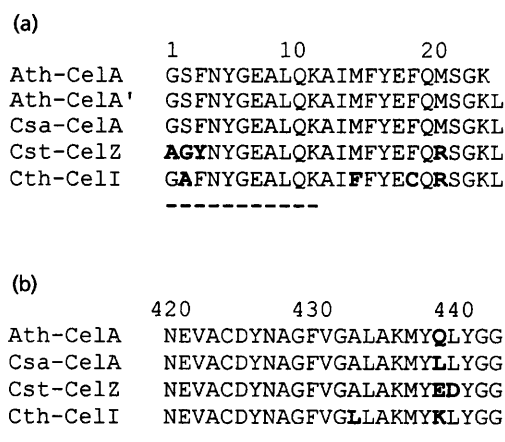


Fig. 3. N-terminal amino acid sequences (a) and highly conserved sequence regions (b) used for the design of PCR primers. CelA' represents the 90 kDa truncated derivative of the CelA enzyme. Dissimilar amino acids are indicated by bold letters. The amino acid sequence regions used to deduce primer sequences are underlined. Ath, '*A. thermophilum*'; Csa, *Ca. saccharolyticus*; Cst, *Cl. stercorarium*; Cth, *Cl. thermocellum*.

experiments. A digoxigenin-labelled *Hind*III-*Hind*III fragment from pATP1, a *Nsi*I-*Nsi*I fragment from pATP7 and an *Eco*RI-*Nsi*I fragment from pATP3-3 were used to probe chromosomal DNA of '*A. thermophilum*' digested with *Hind*III, *Nsi*I, *Nsi*I plus *Hind*III, and *Nsi*I plus *Eco*RI. The hybridization signals obtained were in full accordance with the anticipated results (data not shown).

The nucleotide sequence of a 5513 bp fragment of '*A. thermophilum*' chromosomal DNA has been deposited in the GenBank database (accession no. Z86105). Two open reading frames, ORF1 (5136 bp) and incomplete ORF2 (222 bp), separated by 156 bp, were found on the same strand. ORF1 is succeeded by a possible transcription terminator (nt 5156-5202) capable of forming a stem-loop structure with a calculated ΔG value of -30.2 kcal (-126.84 kJ). Preceding the putative ATG initiation codon (nt 292-294) of ORF2, a potential ribosome-binding site (AGGGGGT, nt 5274-5280) and promoter-like -10 and -35 sequences (TAGTAC, nt 5247-5253; TATAAT, nt 5228-5232) were identified. Mean G+C contents of 44.5 and 33.8 mol% were determined for ORF1 and ORF2, respectively. For total genomic DNA of '*A. thermophilum*' Z-1320, a G+C content of 36.7 mol% has been reported (Svetlichnyi *et al.*, 1990).

Protein structure

The protein translated from ORF1, later designated *celA*, has a predicted molecular mass of 190 kDa and starts with the amino acid sequence obtained for the N-terminus of CelA purified from '*A. thermophilum*'. The discrepancy between the calculated molecular mass and the apparent molecular mass determined for the enzyme purified from '*A. thermophilum*' by SDS-PAGE

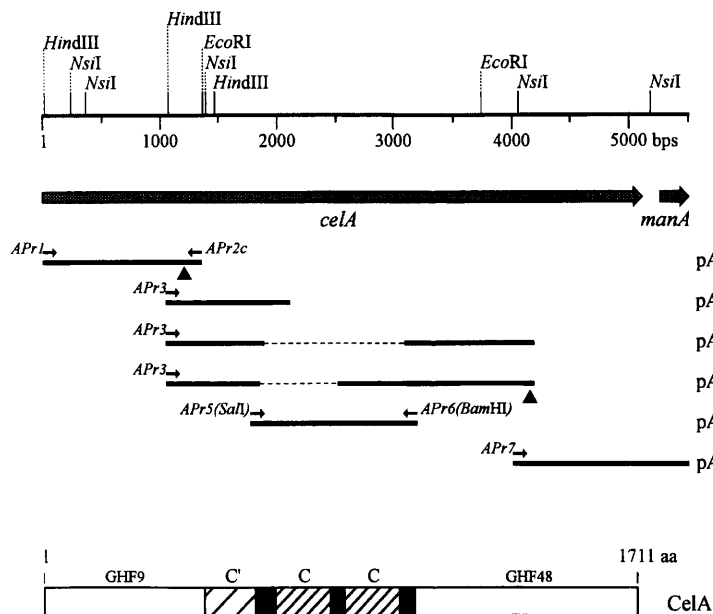


Fig. 4. Physical map of the *celA* region of the '*A. thermophilum*' genome, outline of the cloning strategy and structure of the CelA protein. The cleavage sites of restriction endonucleases used for chromosomal DNA mapping are shown. Shaded arrows indicate the positions of the open reading frames *celA* and *manA*. Segments of these cloned in the plasmids specified on the right are drawn as solid lines. Small arrows indicate the positions and directions of the oligonucleotide primers used for PCR amplifications and filled triangles indicate the positions of the biotinylated oligonucleotide probes. The domain organization of the CelA protein is shown in the lower part of the figure. Open boxes represent catalytic domains (numbers refer to glycosyl hydrolase families), hatched boxes represent family III cellulose-binding domains and black boxes Pro-Thr-Ser linker regions.

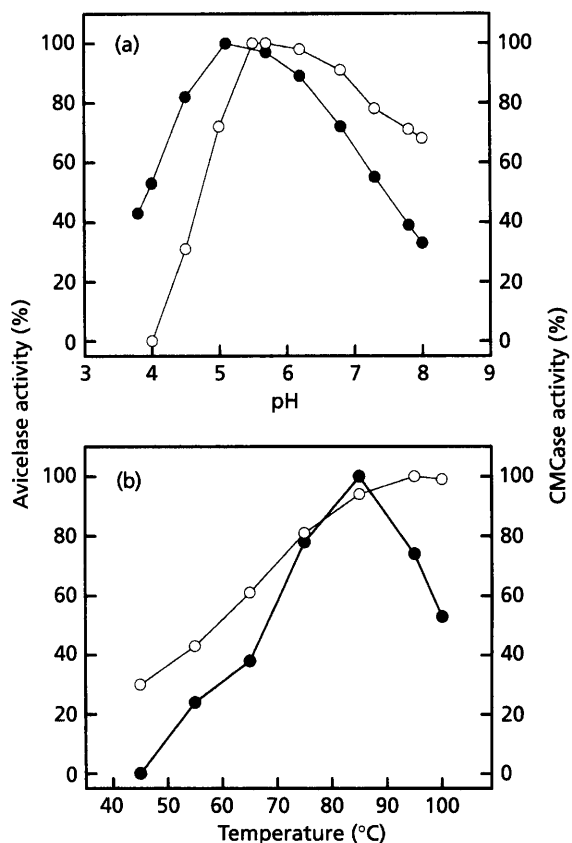


Fig. 5. Effect of pH and temperature on the activity of CelA from '*A. thermophilum*'. Purified enzyme was incubated at various pH values (a) and temperatures (b) with the substrates Avicel (●) and CMC (○). Incubations were carried out for 60 min (Avicel) or 20 min (CMC) with 35 µg (Avicel) or 1.4 µg (CMC) purified enzyme. The buffers used were sodium succinate (pH 3.8–6.0) and sodium phosphate (pH 5.0–8.0).

(230 kDa) is probably due to glycosylation of the native enzyme. CelA and CelA' blotted onto nitrocellulose both stained positively with a glycoprotein detection kit (not shown).

Further sequence analysis identified CelA from '*A. thermophilum*' as a modular glycosyl hydrolase consisting of (i) two different catalytic domains belonging to glycosyl hydrolase families 9 and 48, (ii) three family III cellulose-binding domains (Béguin & Aubert, 1994) showing homology to domains C' and C found in CelY and CelZ of *Cl. stercoararium* (Bronnenmeier *et al.*, 1997) and (iii) three Pro-Thr-Ser linker regions (Fig. 4).

The incomplete reading frame ORF2 encodes 74 amino acids of a protein with high similarity (90% identity) to the N-terminal region of the β-mannanase from *Ca. saccharolyticus* (Gibbs *et al.*, 1992). ORF2 was therefore termed *manA*.

Enzyme properties

Characterization of the multidomain cellulase CelA was carried out with the native enzyme purified from '*A. thermophilum*'. The influence of pH and temperature on enzyme activity was investigated for the substrates CMC and Avicel (Fig. 5). The pH and temperature activity profiles of the CMCase and the Avicelase both showed a shift against one another. This phenomenon may be attributed to the presence of two separate catalytic domains with presumably different properties. With both substrates, maximal activity was observed between pH 5 and 6. At lower pH values, the CMCase activity declined rapidly, whereas the Avicelase activity at pH 4.0 was still 50% of the maximum. At pH 8.0, the purified enzyme displayed 70% of the maximal CMCase

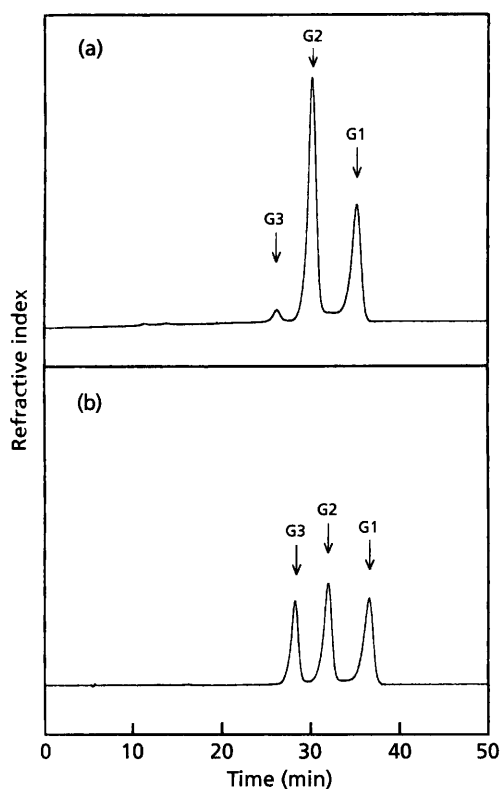


Fig. 6. HPLC analysis of degradation products released from Avicel by purified CelA (a) and CelA' (b) from '*A. thermophilum*'. A standard reaction mixture containing a 1% (w/v) suspension of Avicel was incubated for 20 h at 72 °C with purified CelA and CelA', respectively. G1, glucose; G2, cellobiose; G3, cellotriose.

activity but only 30% of the maximal Avicelase activity. A 'temperature optimum' of 85 °C was determined in 60 min assays for Avicelase activity. At this temperature, the substrate was degraded at a constant rate for approximately 4 h. Maximum CMC hydrolysis rates were measured between 95 and 100 °C. However, at these temperatures the enzyme was rapidly inactivated showing a half-life of only 40 min.

The relative rates of hydrolysis of a variety of substrates are summarized in Table 3. In accordance with the identification of CelA as an endo-exo-cellulase fusion protein by sequence analysis, the purified enzyme displayed activity towards commonly used endoglucanase substrates such as CMC and β -glucan as well as towards the microcrystalline exoglucanase substrate Avicel. Like many other endoglucanases, CelA shows much higher activity towards the mixed-linkage substrate β -glucan than towards the soluble cellulose derivative CMC. Hydrolysis of xylan could also be detected, although the catalytic efficiency was low compared with that of typical xylanases. Remarkably, removal of the C-terminal half of the enzyme as in the purified derivative CelA' leads to a significant reduction of the activity towards the microcrystalline substrate Avicel.

The products formed during hydrolysis of Avicel and cellodextrins were investigated by HPLC. The results of a kinetic analysis may be summarized as follows: (i) cellotetraose (G4), cellotriose (G3) and cellobiose (G2) were identified after short incubations (2.5–5.0 min) of Avicel and larger cellodextrins; (ii) G4 was rapidly cleaved into G3, G2 and glucose (G1); (iii) G3 was hydrolysed into G2 and G1 at a lower rate; (iv) after prolonged incubations (20 h), G2, G1 and minor amounts of G3 were detected as end-products formed by CelA action on Avicel and larger cellodextrins (Fig. 6a). The same products were released from Avicel by the truncated derivative CelA', but in a different ratio (Fig. 6b). Compared with full-length CelA, the relative amount of G3 was enhanced, whereas the formation of G2 was significantly reduced.

DISCUSSION

The multidomain cellulase CelA from '*A. thermophilum*' described in this report is the first bifunctional cellulolytic enzyme which has been isolated from its native host, an extreme thermophile. For *Ca. saccharolyticus* (formerly named '*Caldocellum saccharolyticum*'), several *cel* genes encoding bifunctional cellulolytic and/or hemicellulolytic enzymes have been described (Saul *et al.*, 1990; Gibbs *et al.*, 1992; Morris *et al.*, 1995; Te'o *et al.*, 1995). However, purification of the encoded large gene products has not been reported from either the heterologous host or the authentic organism. CelJ, the only known *Cl. thermocellum* cellulase containing two catalytic domains on one polypeptide chain, could not be isolated in an intact form because of strong proteolytic processing in an *E. coli* recombinant (Ahsan *et al.*, 1996). CelA has been identified as the major cellulolytic enzyme produced by '*A. thermophilum*'. Despite its complex architecture and concomitant large size, the enzyme does not suffer from severe proteolysis in this microorganism. Presumably, CelA is protected against proteases of its authentic host by glycosylation. The enzyme contains three Pro-Thr-Ser linker regions (Fig. 4). Such sequence elements have been identified as sites of glycosylation in bacterial cellulolytic enzymes (Langsford *et al.*, 1987; Ong *et al.*, 1994). Accordingly, dense glycoprotein staining after SDS-PAGE has provided preliminary evidence for glycosylation of CelA from '*A. thermophilum*'. Upon expression in an *E. coli* recombinant, proteolytic degradation rendered the isolation of the full-length gene product impossible. This may result from a deficiency in glycosylation (Sandercock *et al.*, 1994; Herrmann *et al.*, 1996) or from enhanced sensitivity to the proteolytic enzymes produced by the heterologous host.

Purified CelA is able to degrade the microcrystalline cellulose substrate Avicel. Remarkably, this Avicelase activity is significantly higher than that reported for fungal and bacterial cellobiohydrolases (Tomme *et al.*, 1988; Bronnenmeier *et al.*, 1991; Kruus *et al.*, 1995). The activity towards the cellulosic substrate is of the same

order of magnitude as that observed for synergistic admixtures of endoglucanase CelZ (Avicelase I) and exoglucanase CelY (Avicelase II) of *Cl. stercorarium*, provided the molecular masses of the two individual enzymes and that of the large bifunctional enzyme are taken into account (Riedel *et al.*, 1997). The combined action of Avicelase I and II produces predominantly cellobiose and minor amounts of cellotriose and glucose as end-products of Avicel hydrolysis (Riedel *et al.*, 1997). The same pattern of reaction products is found after prolonged incubations of Avicel with purified CelA enzyme (Fig. 6). Actually, CelA from the extreme thermophile '*A. thermophilum*' can be regarded as a naturally occurring fusion protein of the two enzymes effecting cellulose hydrolysis in the less thermophilic organism *Cl. stercorarium*. The different temperature and pH activity profiles determined for CelA with the substrates CMC and Avicel provided further evidence for a bifunctional organization of the large cellulolytic enzyme (Fig. 5). Finally, molecular cloning and sequence analysis of the CelA-encoding gene confirmed the above hypothesis. Homology analysis revealed a multidomain structure composed of two distinct catalytic domains from different glycosyl hydrolase families separated by binding domains and linker regions (Fig. 4). On the primary structure level, the N-terminal family 9 catalytic domain of CelA is highly similar (70.4% identity) to the catalytic region of the CelZ endoglucanase (Jauris *et al.*, 1990) and the C-terminal family 48 domain shows 69.2% sequence identity to the catalytic domain of the CelY exoglucanase (Bronnenmeier *et al.*, 1997). The enzyme most closely related to CelA of '*A. thermophilum*' is CelA of *Ca. saccharolyticus* (Te'o *et al.*, 1995). Both proteins share a similar organization and an extremely high degree of sequence identity (96% for the N-terminal domain, 94% for the binding domains and 97% for the C-terminal domain).

Identification of the N-terminal CelA region as an endoglucanase domain was confirmed by the isolation of a truncated CelA derivative (CelA'). The 90 kDa CelA' is supposed to contain the glycosyl hydrolase family 9 catalytic domain joined to domains C' and C, which are homologous to family III cellulose-binding domains. The substrate specificity and the degradation pattern of this N-terminal part of the CelA cellulase are comparable to those of other family 9 cellulases. The activity towards Avicel is of the same order of magnitude as that reported for CelZ of *Cl. stercorarium* (Bronnenmeier & Staudenbauer, 1990). However, the Avicelase activity of full-length CelA is significantly higher. Since the cellulose-binding capacity of the intact enzyme is maintained in the truncated derivative, an ability to degrade microcrystalline cellulose may be ascribed to the C-terminal catalytic domain. Accordingly, an enhanced amount of cellobiose has been detected in Avicel hydrolysates of full-length CelA (Fig. 6). This is in line with the identification of the C-terminal domain as a member of glycosyl hydrolase family 48 which contains exclusively bacterial exoglucanases. The enhanced Avicelase activity of CelA

presumably results from intramolecular synergism between the endoglucanase and exoglucanase domains.

ACKNOWLEDGEMENTS

The authors are grateful to F. Lottspeich (MPI Martinsried) for N-terminal protein sequence analysis. We also wish to thank L. Richter and W. Ludwig (TU München) for advice in using asymmetric PCR prior to publication. The work was supported by grants from the Deutsche Forschungsgemeinschaft (Br 1472/2) and from the Volkswagen-Stiftung to K. B. V. Z. thanks the Deutsche Forschungsgemeinschaft for a travel grant.

REFERENCES

- Ahsan, M. M., Kimura, T., Karita, S., Sakka, K. & Ohmiya, K. (1996). Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. *J Bacteriol* **178**, 5732–5740.
- Béguin, P. & Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiol Rev* **13**, 25–58.
- Bronnenmeier, K. & Staudenbauer, W. L. (1990). Cellulose hydrolysis by a highly thermostable endo-1,4- β -glucanase (Avicelase I) from *Clostridium stercorarium*. *Enzyme Microb Technol* **12**, 431–436.
- Bronnenmeier, K., Rücknagel, K. P. & Staudenbauer, W. L. (1991). Purification and properties of a novel type of exo-1,4- β -glucanase (Avicelase II) from the cellulolytic thermophile *Clostridium stercorarium*. *Eur J Biochem* **200**, 379–385.
- Bronnenmeier, K., Kundt, K., Riedel, K., Schwarz, W. H. & Staudenbauer, W. L. (1997). Structure of the *Clostridium stercorarium* gene *celY* encoding the exo-1,4- β -glucanase Avicelase II. *Microbiology* **143**, 891–898.
- Flint, H. J., Martin, J., McPherson, C. A., Daniel, A. S. & Zhang, J.-X. (1993). A bifunctional enzyme, with separate xylanase and $\beta(1,3-1,4)$ -glucanase domains, encoded by the *xynD* gene of *Ruminococcus flavefaciens*. *J Bacteriol* **175**, 2943–2951.
- Gibbs, M. D., Saul, D. J., Luthi, E. & Bergquist, P. L. (1992). The β -mannanase from *Caldocellum saccharolyticum* is part of a multidomain enzyme. *Appl Environ Microbiol* **58**, 3864–3867.
- Hall, J., Black, G. W., Ferreira, L. M., Millward-Sadler, S. J., Ali, B. R., Hazlewood, G. P. & Gilbert, H. J. (1995). The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochem J* **309**, 749–756.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **16**, 557–580.
- Hazlewood, G. P., Devidson, K., Laurie, J. I., Huskisson, N. S. & Gilbert, H. J. (1993). Gene sequence and properties of CellI, a family E endoglucanase from *Clostridium thermocellum*. *J Gen Microbiol* **139**, 307–316.
- Herrmann, J. L., O'Gaora, P., Gallagher, A., Thole, J. E. R. & Young, D. B. (1996). Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*. *EMBO J* **15**, 3547–3554.
- Jauris, S., Rücknagel, K. P., Schwarz, W. H., Kratzsch, P., Bronnenmeier, K. & Staudenbauer, W. L. (1990). Sequence analysis of the *Clostridium stercorarium celZ* gene encoding a thermostable cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. *Mol Gen Genet* **223**, 258–267.

- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
- Langsford, M. L., Gilkes, N. R., Singh, B., Moser, B., Miller, R. C., Warren, R. A. J. & Kilburn, D. G. (1987). Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains. *FEBS Lett* **225**, 163–167.
- Kruus, K., Wang, W. K., Ching, J. & Wu, J. H. D. (1995). Exoglucanase activities of the recombinant *Clostridium thermocellum* CelS, a major cellulosome component. *J Bacteriol* **177**, 1641–1644.
- Montgomery, L. & Fu, Y.-K. (1988). Detection of cellulose-binding proteins in electrophoresis gels by filter paper affinity blotting. *Anal Biochem* **174**, 204–208.
- Morris, D. D., Reeves, R. A., Gibbs, M. D., Saul, D. J. & Bergquist, P. L. (1995). Correction of the β -mannanase domain of the *celC* pseudogene from *Caldocellulosiruptor saccharolyticus* and activity of the gene product on kraft pulp. *Appl Environ Microbiol* **61**, 2262–2269.
- Ong, E., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1994). *Streptomyces lividans* glycosylates the linker region of a β -1,4-glycanase from *Cellulomonas fimi*. *J Bacteriol* **176**, 999–1008.
- Rainey, F. A., Ward, N. L., Morgan, H. W., Toalster, R. & Stackebrandt, E. (1993). Phylogenetic analysis of anaerobic thermophilic bacteria: aid for their reclassification. *J Bacteriol* **175**, 4772–4779.
- Rainey, F. A., Donnison, A. M., Janssen, P. H., Saul, D., Rodrigo, A., Bergquist, P. L., Daniel, R. M., Stackebrandt, E. & Morgan, H. W. (1994). Description of *Caldicellulosiruptor saccharolyticus* gen. nov., spec. nov.: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. *FEMS Microbiol Lett* **120**, 263–266.
- Riedel, K., Ritter, J. & Bronnenmeier, K. (1997). Synergistic interaction of the *Clostridium stercorarium* cellulases Avicelase I (CelZ) and Avicelase II (CelY) in the degradation of microcrystalline cellulose. *FEMS Microbiol Lett* **147**, 239–243.
- Sakamoto, R., Arai, M. & Murao, S. (1984). Enzymatic properties of hydrocellulase from *Aspergillus aculeatus*. *J Ferment Technol* **62**, 561–567.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sandercock, L. E., MacLeod, A. M., Ong, E. & Warren, R. A. J. (1994). Non-S-layer glycoproteins in eubacteria. *FEMS Microbiol Lett* **118**, 1–8.
- Saul, D. J., Williams, L. C., Grayling, R. A., Chamley, L. W., Love, D. R. & Bergquist, P. L. (1990). *celB*, a gene coding for a bifunctional cellulase from the extreme thermophile '*Caldocellum saccharolyticum*'. *Appl Environ Microbiol* **56**, 3117–3124.
- Schwarz, W. H., Bronnenmeier, K., Gräbnitz, F. & Staudenbauer, W. L. (1987). Activity staining of cellulases in polyacrylamide gels containing mixed linkage β -glucans. *Anal Biochem* **164**, 72–77.
- Sedmak, J. J. & Grossberg, S. E. (1977). A rapid, sensitive assay for protein using Coomassie brilliant blue G250. *Anal Biochem* **79**, 544–552.
- Svetlichnyi, V. A., Svetlichnaya, T. P., Chernykh, N. A. & Zavarzin, G. A. (1990). *Anaerocellum thermophilum* gen. nov. sp. nov.: an extremely thermophilic cellulolytic eubacterium isolated from hot springs in the valley of geysers. *Microbiology* (English translation of *Mikrobiologiya*) **59**, 598–603.
- Te'o, V. S. J., Saul, D. J. & Bergquist, P. L. (1995). *celA*, another gene coding for a multidomain cellulase from the extreme thermophile *Caldocellum saccharolyticum*. *Appl Microbiol Biotechnol* **43**, 291–296.
- Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. & Claeyssens, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. *Eur J Biochem* **170**, 575–581.
- Warren, R. A. J. (1996). Microbial hydrolysis of polysaccharides. *Annu Rev Microbiol* **50**, 183–212.
- Weimer, P. J., Wagner, W., Knowlton, S. & Ng, T. K. (1984). Thermophilic anaerobic bacteria which ferment hemicellulose: characterization of organisms and identification of plasmids. *Arch Microbiol* **138**, 31–36.
- Wood, T. M. & Bhat, K. M. (1988). Methods for measuring cellulase activities. *Methods Enzymol* **160**, 87–112.
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.
- Zhang, J.-X. & Flint, H. J. (1992). A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavifaciens* 17 comprises two dissimilar domains linked by an asparagine/glutamine-rich sequence. *Mol Microbiol* **6**, 1013–1023.

Received 7 August 1997; revised 3 October 1997; accepted 17 October 1997.