Cloning of an endo-(1→4)- β -glucanase gene, *celA*, from the rumen bacterium *Clostridium* sp. ('*C. longisporum*') and characterization of its product, CelA, in *Escherichia coli*

VOLKER MITTENDORF and JENNIFER A. THOMSON*

Department of Microbiology, University of Cape Town, Rondebosch 7700, South Africa

(Received 16 April 1993; revised 22 June 1993; accepted 29 June 1993)

A genomic library of *Clostridium* sp. ('*C. longisporum*') ATCC 49440 in the host *Escherichia coli* was screened for endo- β -glucanases, and plasmids pCM64 and pCM4 were isolated. The nucleotide sequence of a 3620 bp fragment was found to contain a 1548 bp open reading frame (ORF), termed *celA*, which encodes an endo-(1→4)- β -glucanase, CelA, assigned to family A4. N-terminal amino acid sequence determination revealed that pCM64 encoded the full-length *celA* gene, including a signal sequence, while pCM4 carried a 5'-truncated *celA* gene expressed as an N-terminal fusion protein, CelA Δ N', without a signal sequence. CelA was secreted into the periplasm in *E. coli*. In this organism, proteolytic cleavage of CelA at or near a putative linker region resulted in the appearance of two active polypeptides of molecular masses 57 and 47 kDa. The former was the full-length enzyme, while the latter consisted of the catalytic domain from which the cellulose-binding domain (CBD) had been removed (CelA Δ CBD). The intracellularly-located CelA Δ N' was not subject to proteolytic degradation. The pH and temperature optima of CelA were pH 4.8 and 43 °C, respectively. CelA hydrolysed barley β -glucan, lichenan, carboxymethylcellulose and xylan. It showed preferential activity against the larger cello-oligosaccharides (cellohexaose and cellopentaose); cellotetraose was the smallest substrate degraded completely.

Introduction

Analysis of the structural and functional aspects of cellulose degradation by cellulases is of great interest due to the potential use of the abundant and renewable substrate for man-made purposes. A large number of cellulase genes have been cloned and their products analysed in heterologous hosts. The comparison of their DNA and amino acid sequences has provided much insight into the structural and functional aspects of these enzymes, and into their distribution in cellulolytic organisms (Béguin, 1990). Endo- and exo- β -cellulases have been assigned to six broad families according to hydrophobic cluster analysis of their catalytic domains (Gilkes *et al.*, 1991a; Henrissat *et al.*, 1989). Many

cellulases were found to have a modular structure, comprising a catalytic domain (CD) linked to a cellulose-binding domain (CBD) by a linker region (Gilkes *et al.*, 1991a). Other functional units have been identified in some cellulases that are required for the assembly of these enzymes into a large multi-protein complex, the cellulosome (Tokatlidis *et al.*, 1991). Two different strategies for cellulose degradation have been identified in bacteria and fungi. Some bacteria have cell-wall-attached cellulosomes which can be released into the medium under certain conditions, e.g. *Clostridium thermocellum* (Béguin *et al.*, 1992) and *Butyrivibrio fibrisolvens* (Lin & Thomson, 1991); other bacteria, for example *Cellulomonas*, and fungi such as *Trichoderma reesei*, make use of unassociated or transiently attached cellulases that do not form stable, high-molecular-mass complexes (Béguin *et al.*, 1992; Wood, 1992).

Here we report the first cloning and genetic analysis in *Escherichia coli* of a gene from the rumen bacterium *Clostridium* sp. ('*C. longisporum*'). The *celA* gene encodes an endo-(1→4)- β -glucanase (EC 3.2.1.4), CelA, which belongs to the endo- β -glucanase family A4 (Béguin, 1990). It is a periplasmic enzyme in *E. coli*. CelA was

* Author for correspondence. Tel. +27 21 650 3270; fax +27 21 650 4023; e-mail jat@micro.uct.ac.za.

Abbreviations: CBD, cellulose-binding domain; CD, catalytic domain; CMC, carboxymethylcellulose; NRF, non-rumen fluid.

The GenBank accession number for the sequence reported in this paper is L02868.

purified, and a *Cellulomonas fimi*-type CBD was identified at the C'-terminus of CelA which was proteolytically cleaved from the CD in *E. coli*.

Methods

Micro-organisms, plasmids and culture conditions. *Clostridium* sp. ('*C. longisporum*') ATCC 49440 was obtained from D. H. Varel (see Varel, 1989) and cultured in an anaerobic cabinet (5% H₂, 10% CO₂ and 85% N₂ gas phase) in non-rumen fluid (NRF) medium (Caldwell & Bryant, 1966) containing 2% cellobiose but without glucose or starch. Cultures were stored in Hungate tubes, at -70 °C, on NRF medium slants containing 1.2% (w/v) agar. *E. coli* LK111 (Zabeau & Stanley, 1982) and *E. coli* JM105 (Yanisch-Perron *et al.*, 1985) were used for cloning purposes. *E. coli* BL21 (F⁻ *hsdS gal*) (Studier & Moffat, 1986) was used for the enzyme localization studies. *E. coli* K514 (Wood, 1966) and *E. coli* BL21 were used for the purification of CelA and CelAΔN' respectively. DNA fragments were subcloned into pEcoR251 (Zappe *et al.*, 1986), pUC18, pUC19 (Yanisch-Perron *et al.*, 1985) or pBluescript SK (Stratagene).

Preparation of DNA from *Clostridium* sp. This organism produces copious quantities of an orange-pigmented exopolysaccharide that entraps the clostridial cells. For the extraction of DNA, the cells were grown in 1.5 l NRF medium for 20 h and the culture harvested early in the life cycle, when the cells were still motile and little of the slime had been produced. Prior to harvesting, SDS was added to a final concentration of 0.1% (w/v) and the culture was agitated moderately for 5 min to prevent cell-entrapment. The cells were then washed in 50 ml solution A [10 mM-Tris/HCl pH 8.0, 25% (w/v) sucrose], resuspended in 10 ml solution A containing 5 mg lysozyme ml⁻¹, and left for 30 min at 37 °C. EDTA and SDS were added to final concentrations of 0.1 M and 2% (w/v) respectively, and the progress of cell lysis, which took 10–30 min, monitored microscopically. The cell lysate was extracted once with 1 vol. phenol/chloroform/isoamyl alcohol (25:24:1 by vol.), and twice with 1 vol. chloroform/isoamyl alcohol (24:1, v/v) followed by H₂O-saturated diethyl ether. The solution was treated with RNAase and dialysed for 24 h against TE (10 mM-Tris/HCl pH 8.0, 1 mM-EDTA).

Construction and screening of the genomic library. The *Clostridium* sp. chromosomal DNA was partially digested with *Sau*III restriction endonuclease, and fragments between 5 and 10 kb isolated from a sucrose gradient. These were cloned into the *Bgl*II site of the positive selection vector pEcoR251, transformed into *E. coli* K514, and screened on Luria-Bertani (LB) medium (Maniatis *et al.*, 1982) containing 1% (w/v) carboxymethylcellulose (CMC). The plates were stained with Congo red (Teather & Wood, 1982) to identify CMC-degrading clones.

DNA manipulations and sequencing. CMC-positive clones pCM4 and pCM64 were isolated, mapped with restriction endonucleases and subcloned according to standard procedures (Maniatis *et al.*, 1982). The subcloning and sequencing strategy is explained in the legends to Figs 1 and 2. The complete sequence of 3620 bp, containing the *celA* gene and flanking regions, was assembled from the sequences obtained from pCMC, pCMJI and pCMJII: the former carried the 5'-truncated *celA* gene plus the downstream region; the latter two carried the 5'-part of the *celA* gene and the 5'-upstream region. The DNA was sequenced in both directions, using clones carrying overlapping deletions obtained by *Exo*III exonuclease degradation.

Enzyme localization studies. Cultures of *E. coli* BL21(pCMI) and BL21(pCML) inoculated from single colonies were grown in LB medium containing 200 µg ampicillin (Ap) ml⁻¹. IPTG was added to a final concentration of 0.5 mM during the exponential phase of growth to induce the *lac* operon, since β-galactosidase was used as a marker

enzyme. The cultures were harvested at OD₆₀₀ 1.0–1.4. Due to the instability of both plasmids, serial dilutions of the cultures were plated on CMC-containing LB medium, with and without antibiotic selection, to ascertain that all cells contained the plasmids and were expressing the *celA* gene. The cells were harvested and fractionated by the osmotic shock procedure (Nossal & Heppel, 1966). The resulting cellular fractions were assayed for CMCase, β-galactosidase and alkaline phosphatase activity.

Enzyme assays. CMCase activity was measured using the dinitro-salicylic acid method as recommended by the Commission on Biotechnology (Ghose, 1987), but the volumes were scaled down to permit the use of Eppendorf tubes. Assays were done at 43 °C in 50 mM-sodium acetate buffer pH 5.0. The purified CelA and CelAΔN' samples were diluted in a solution of 50 µg BSA ml⁻¹ to stabilize the enzymes. β-Galactosidase and alkaline phosphatase activities were assayed according to Miller (1972) and Brickman & Beckwith (1975), respectively.

Purification of CelA. *E. coli* BL21(pCML) was grown to a final OD₆₀₀ of 1.1 in 4 l LB medium containing 150 µg Ap ml⁻¹. A periplasmic extract, obtained by subjecting the cells to osmotic shock (Nossal & Heppel, 1966), was applied to a Q-Sepharose anion-exchange column (Pharmacia) and the CMCase activity eluted in a linear gradient of 50–200 mM-NaCl in 10 mM-Tris/HCl pH 7.5. CMCase-containing fractions were pooled and analysed by SDS-PAGE.

Purification of intracellular CelAΔN'. *E. coli* K514(pCMI) was grown to a density of OD₆₀₀ 1.1 in 2 l LB medium containing 150 µg Ap ml⁻¹. The cells were harvested, resuspended in 15 ml 50 mM-NaCl, disrupted in a French pressure cell, and centrifuged (39 000 g for 20 min at 4 °C) to obtain the cell-free extract. The cell-free extract was subjected to a pH-precipitation step by the dropwise addition of 0.1 M-HCl to a pH of 4.6. The suspension was centrifuged as above, and the pH of the supernatant increased to 7.0 by the addition of NaOH. The pH-precipitated extract was applied to a Q-Sepharose Fast Flow anion-exchange column (as above). CMCase-containing fractions were pooled and applied to a Bio-Gel P-200 gel filtration column (Bio-Rad). Fractions containing CMCase activity were analysed by SDS-PAGE.

N-terminal amino acid sequencing. The N-terminal amino acid sequences of CelA and CelAΔN' were determined on a gas-liquid solid-phase sequencer constructed as outlined by Hewick *et al.* (1981) and Brandt *et al.* (1984). The converted phenylthiohydantoin amino acids were identified, by an isocratic on-line HPLC system, on a 3 × 250 mm 3 µm Lichrospher C₁₈ (Bischoff) column (Lottspeich, 1985).

Cellulose-binding studies. A periplasmic extract of *E. coli* LK111(pCML) was prepared and applied to a Q-Sepharose anion-exchange column as above. CMCase-containing fractions were pooled and used for the Avicel-binding experiment, as described by Ferreira *et al.* (1990) with minor modifications. A 5 ml sample was mixed with 5 ml 5% (w/v) Avicel, suspended in 0.1 M-Tris/HCl pH 7.5 and incubated on a shaker at 0 °C for 1 h. The suspension was filtered through a small column and the eluate containing the unbound proteins was kept, while the Avicel was washed twice with 10 ml 0.1 M-Tris/HCl pH 7.5. The bound proteins were eluted with 5% (w/v) SDS at room temperature. Samples were analysed by SDS-PAGE.

Substrate specificities. The activity of purified CelA on barley β-glucan, lichenan, laminarin and oat spelt xylan (all obtained from Sigma) was determined according to the CMCase assay described above, except that the substrates were used at the following concentrations: 1% (w/v) β-glucan; 0.5% (w/v) lichenan; 1% (w/v) laminarin; 1% (w/v) xylan. In all except the xylanase assay, the release of reducing sugars was expressed as glucose equivalents using a glucose standard curve. For the xylanase assay, a xylose standard curve was used.

*Bgl*II–*Bam*HI cloning junction. CMCase activity is indicated on the right-hand side of the figure.

Divalent cation inhibition study. Purified CelA was dialysed for 24 h at 4 °C, with three buffer changes, against 2 l solution containing 50 mM-NaCl, 10 mM-Tris/HCl pH 7.5 with 0.1 mM-EDTA added to scavenge metal cations. CMCase activity was assayed as above in reactions that contained 200 μ M unchelated metal cations.

Protein assays. The Bio-Rad protein assay was used to quantify proteins. BSA fraction V (Boehringer Mannheim) was used as a standard.

Construction and screening of the genomic library

hibited the strongest CMCase phenotype and was chosen for further studies. Southern blot analysis revealed that pCM64 (Fig. 2) encoded the same cellulase gene, but the other 38 clones were different (results not shown). Southern blot analysis also confirmed that the insert DNA of pCM4 originated from the *Clostridium* sp. (results not shown).

A number of overlapping fragments of the 7-15 kb *Clostridium* sp. insert DNA of pCM4 were subcloned into pUC19, but none had CMCase activity (results not shown). Only pCMX, a deletion derivative of pCM4 (Fig. 1), was found to express CMCase activity. pCMC and pCMH contained the same *Clostridium* sp. insert DNA as pCMX, but did not express CMCase activity. Initially it was thought that the external λ promoter P_R was required for the expression of *celA* and could not be replaced by P_{lac} on pCMC or pCMH. However, later sequence determination of the first 15 N-terminal amino acid residues of the enzyme purified from pCMI showed that it was an N-terminal fusion protein, termed C ϵ LA Δ N'. Expression of the gene, termed *celA* Δ 5', was found to be initiated from an in-frame ATG codon in the vector DNA (Fig. 3a). The first seven amino acids of the mature protein were derived from the vector, pEcoR251, and the next eight corresponded to the DNA sequence of the *celA* Δ 5' gene on pCMC. This fusion protein lacked a signal sequence and the start methionine was not present in the mature protein. A similar fusion protein was not formed in pCMC or pCMH, due to the lack of a suitable ATG start codon in-frame with *celA* Δ 5'.

(a)

-114 AATTGTAGAG GTCAAAGATG ATTATGGTGA

ATGGAGAGTT ATAAATATAA GGAATGGTTT GTTAGTTGGG AAAAGAGGAG ATCAAGATT

m A A G N A I G R S A S E I V Q

ATG GCT GCT GGT AAT GCT ATC GAA AGA TCT GCA AGT GAA ATT GTG CAG

-1 1

(b)

1106 TAATTTTAT TAATAACAAT

GACTAACTTT ATGTATTAAT AGCACACTTT CACGAATTAT TATTGAATTT ATTCTTTAA

GATATTATAT TGAATTTATG TAAATAATTA CATCAAGAA ATAAGGAGA TGATTAACT

M K R S L L K T C S I I A G A T

ATG AAA AGA TCT TTA TTA AAA ACT TGT TCT ATC ATT GCT GGA GCA ACA

I I F S S L S I S R N P L E V Q

ATT ATT TTC TCA TCA CTT TCA ATT TCA AGA AAT CCT CTT GAA GTA CAA

A A S M R S A S E I V Q E M G

GCT GCT TCT ATG AGA TCT GCA AGT GAA ATT GTG CAG GAA ATG GGA 1386

This finding led us to investigate clone pCM64 (Fig. 2), which contained the same gene but had weaker CMCase activity as seen on CMC plate assays (results not shown). Restriction endonuclease analysis of pCM64, and sequence analysis of pCMJI and pCMJII, showed that pCM64 contained the full-length *celA* gene, together with 1358 bp of 5'-upstream sequence (Figs 2 and 3b). A contiguous sequence of 3620 bp of *Clostridium* sp. DNA (GenBank accession number L02868) was obtained by assembling the sequence obtained from pCMC (*celAΔ5'* and the downstream sequence), together with the sequences from pCMJI and pCMJII (*celAΔ3'* and the upstream sequence); the nucleotide sequence and the derived amino acid sequence are shown in Fig. 4. The

Fig. 2. Restriction maps of the full-length *celA* gene on various plasmids, and sequencing strategy. pCMJII and pCMJI contain the 2.65 kb *Hind*II fragment of pCM64, cloned in both orientations into the *Eco*RV site of the vector pSK. During these subclonings, 720 bp of the 3' end of *celA* were deleted as indicated by *celAΔ3'* in the diagram. Selected *Exo*III deletion derivatives were sequenced as indicated by the arrows below the maps of pCMJI and pCMJII. pCML was constructed for the purification of CelA encoded by *celA*. The 3' end of the *celA* gene was spliced onto the 5' end of the *celAΔ3'* gene by cloning the 1.4 kb *Xba*I–*Stu*I fragment of pCMC (see Fig. 1) into pCMJI-23 (the *Exo*III-shortened derivative of pCMJI shown with the solid arrow) cut with *Xba*I and *Hind*II (the latter site being in the multiple cloning site, mcs). Restriction sites of selected endonucleases are indicated as follows: B, *Bgl*II; E, *Eco*RI; Ev, *Eco*RV; H, *Hind*III; Hd, *Hind*II; K, *Kpn*I; P, *Pst*I; S, *Sac*I; St, *Stu*I; X, *Xba*I; Xh, *Xho*I. Hd/Ev and St/Hd are the *Hind*II–*Eco*RV and the *Stu*I–*Hind*II cloning junctions, respectively. mcs, multiple cloning site containing other unique restriction sites as indicated by the wedge-shaped segment.

Fig. 3. (a) Nucleotide sequence of the 5'-end of *celAΔ5'* and the experimentally determined N-terminal amino acid sequence of CelAΔN'. The methionine expressed by the start codon was not present in the mature protein and is indicated with a small letter m. The numbering is based on the beginning of the *Clostridium* sp. DNA insert in pCM4; the upstream sequence is derived from the vector pEcoR251. (b) Nucleotide sequence of the 5'-end of *celA* and the derived amino acid sequence. The sequence of the first five residues of the mature CelA protein was confirmed by N-terminal amino acid sequencing. The underlined residues constitute the signal peptide. The double-underlined amino acid sequence indicates the start of the overlap of *celA* and *celAΔ5'*. Numbering corresponds to the sequence in GenBank (accession number L02868).

celA ORF is 1548 bp long and stretches from nucleotides 1246 to 2794 in the sequence. Homology comparisons with other cellulases revealed that CelA belongs to family A4 (Béguin, 1990). The *celE* gene, encoding EGE of *C. thermocellum* (Hall *et al.*, 1988), shows 59% DNA sequence homology in a 1040 bp fragment, and 48.4% amino acid homology in a 374 amino acid fragment, with *celA* and CelA, respectively.

Enzyme localization of CelAΔN' and CelA

Fifty-eight percent of CelA was exported to the periplasm in *E. coli* BL21, while 94% of CelAΔN', lacking the signal sequence, was retained in the cytoplasm (Table 1).

1 TACTAACACATATAGAAAGCATTTCAGAATCTCTTCTTAAACCACAATCTTCAAATCTTTTCAAAAATACCATTATCATCAGGGTTTACCATGTCTACACT
101 TCCAGTAATGATTATATATGATAATAAACTACTGGTATAGGTATGCATGCTATTTAAAGCGCTAGAAAAAACTTCAGCAACTAAGACATCTTTAGTA
201 TTCTTCTTCAACCAATTTTGTAGCTCCCAATCTATAGCTTTTAATAATCTTAAAGAACTCTTATCTTTTAATATGCTTATAGGTATATTAATATTTGTGTG
301 GTTCAATGACTCAATCTTCCATCACTTAAAAATAATAAAATATTAAAGCTTAAAGGCGCAGCAATAAAATAAAGGGTATGATTTTGAACAAAATA
401 GACTGTTCAAAACTACATGCCCTTTAAATATATAGCTTTCTAGCTTAAAGCACCTTTCCAAAATCTGATACAATATCAAAGCTCATCTTTTTTCCATTAA
501 GCTTTTTTAACTGTAACCTTGGTATATTTAAAAATATCAGCTGC AAAAATAAGTCCCTCTATCTTATAATTTCTTTTAAAAAATCTTTTATAAGAAG
601 TAAAGCTTCATAAGCATATCCCTTTCTTCTTATCTTACATATCCAAAATCCCAATCTCCGGACATTCATCTTCAAATATAAAGCTCTACTTCCCA
701 ACAAACTTATCTCTTCATTGCAACAGCTAAGCATAGTATTATCACCTTTATCTTAAATTTATAAATCTCTGCAAAAACCTTCAGCTCTTTTATAG
801 ACTTAAAGAAATTTGCAAGCTGGTATTTTACTACCTCTTCAGTAAAGTTTTCATAATATAAAGCTGCCTCTTCAATTTTTGTAGAAGTTATTCTAAGCCT
901 TTTTCTTTTATTTCAATGTATATTATCATAAAAAACCTTCTTAAATATCTTAATGTATTTTAACTGATATAACCTTTTAACTCGAAATCTTTCAT
1001 GCAATACAAATCAACTATACTGTTTATTATAGAATAAATAAAAAATTATAAAAGAAAAATTTATGTTATTTTACATGCTGGTAATGCTAAAACTGGTAG
1101 TATTCTAATTTTTATTATAAACAATGACTAACTTTATGTATTAATAGCACACTTTCACGAATTTATTATTGAATTTATTTCTTTAAGATATTATATTGAAT
1201 TTATGTAAATAATTACATCAAGAAATAAGGGAGATGATTAACACTATGAAAGACTTTTATTA AAAAATCTTGTCTCATCTATTGCTGGAGCAACAATATTATT
M K R S L T L T K A C T S C I A G A T I I F
1301 TCTCATCACTTTCAATTTCAAGAAATCCTCTTGAAGTACAAGCTGCTTCTATGAGATCTGCAAGTGAAATTTGTCAGGAAATGGGAGTTGGGTGGAATCT
S S L S I S R N P L E V Q . A A S M R S A S E I V Q E M G V G W N L
1401 TGGAAATACTTTAGATGCTAAGATCTATAACCTGTCTTATAATACTTCACCAATATCTTGTGAAACAGGCTGGGGAATCCTGTTACTACAAAGGCTATG
G N T L D A K I T N L S Y N T S P I S F T C T T G W G N P V I P T Y A K
1501 ATTGATAAAATCAAAAATGCTGGATTTAAACCATAAGAAATACCAACTACTTGGGAGAACACTTAGATGGTAATAACAACTTAATGAAGAATGGGTAA
I D K I K N A G F K T I R I P T T W G E H L D G N N K L N E E W V R
1601 AGAGAGTTAAAGAAAGTTGGTGATTATGATAGCAGATGATCTTTATGTTATCTTAAATCATCATCATGAAGGAACTGGGTATTTCCAACTTACGCTAA
R V K E E V V D Y C I A D D L Y V I L N T H E G G N W V I P T Y A K
1701 AGAATCTTCAGTAACTCCAAAATCAAAAATCTTTGGACTCAAATATCTGAAGCTTCAAAGATTATGATGATCTTTAATATTGAAACTCTTAACGAA
E S S V T P K L K T L M T Q I S E A F K D Y D D H L I F E T L N E
1801 CCAAGACTTGAAGGAACTCCTTATGAATGGACAGGTTGATCAAGTGAATCTCGTATGTTGTAAATAAATAACCGCAGCTTGAAGATTTAGAATCTATAAGAA
P R L E G T P Y E W T T G G T S E S R D V V N K Y N A A C A L L E S I R K
1901 AAAGTGGTGGTAATACTTATCTAGAGCTGTTATGATGCCAACTTATGCAGCTTCTGTTTCTACTACAATGAATGATTTTAAAGTTCCAGATGATAA
T G G N N L S R A V M M P T Y A A S G S S T T M N D F K V P D D K
2001 AAATGTAATTCATCTGTTTCATGATATCTCTTACTTCTTGTCTATGGACACAAGCAGCAATTCAGTTAAACACATGGGGAGTTCTTATGCAAAATAT
N V I A S V H A Y I S P Y F F A M D T S S N S V N T W G S I T Y D K R Y
2101 TCTTTAGATGAGAATTAGATTCTTACTTAAATCTTTTAAATCTAAAGAGTTTCTGTTGTTATTGGTGAATTTGGTTCTATAAATAAAAAATAACTT
S L D A V G E L D S Y L N T F K S K G V P V V I G E F G S I N K N N T S
2201 CTTCAAGCAGAGCTTGCTGAATATTATGTTACAGCTGCTCAAAAGAGAGGTATTTCTTGTATGGTGGGATAACAACATGCTGAAACTATAAAGG
S R A E L A E Y Y V T A A Q K R G I P C V W W D N N Y A E T N K G
2301 TGAAACTTTTGGATTACTAAATAGAAGTACTTTAACTGGTACTTTAGTGATATTAAGATGCTTTAATTAGAGGATATAAAAAATGTACATCTGAAGCT
E T F G L L N R S T L N W Y F S D I K D A L I R G Y K N V H P E A
2401 ACTGAAGATGATAAACCAAGCAGATGTAACCAACCTGATTCAGGCAATACTAAACAGCAGTGGTAATCTAATCTCTGTCAGAGAACTTACTACT
T E D D K P S T D V T N P D S G N T R P D S G N T N P G T E T T T P
2501 CAACAGATAATGAAAGATTTCATAAATCTTCAAAAATTAATGATTGGGGCGGTGCTTACCAAGCTGATTTCACTTTAAAAAATAACACTTCAAGTGACAT
T D N E K I S I T S K I N D W G G A Y Y Q A D F T L K N N T S S D I
2601 TAACAACTTGGCTTTTAAAGATAAAGAAAAATGATATTGTGTTTACTTAATTTACTGGGATGTTAAGATAACTGAAGAAAAATGATATTATGTTTACACTT
N N W S F K I K R N D I V F T N Y M D V K I T E E N G Y V V T P
2701 CAAGCTTGGAAACCACTCTCTGCAAAATCTTCTATAGTATTCTTATCAAGGCACTGGAAAGTAATTAGTAATTTTGTAGTATAAATTTGACTAGA
Q A W K T T I L A N S S I V I S I Q G T G K V I S N F E Y K F D *
2801 TAACCTAAATAAAATTTAAGCCTAAACAATATAACTTCTATCTTGTTTAGGCTTAAATTATACTACTCGTCCATCATATTACGGTTATTATATGAGTACA
2901 CATTATAATCTCTCCAAAATAAGTTCAAGATTTGTAAGTATCTGAAATACTTTCCCTTTCTCTCATTCTACTGCTATTATAAGTGCAATTATTACTGA
3001 TAAAGGTGCTACTAATGAATCAACAAATGAAGCCATATTACTTTGAGCTATTAAACGTAATAATCTGCATGTGAAGCAAGTGGAATATAAATCTTGTGA
3101 AGTGCTAAAACTTAGCTCCTCTCTCTTGAAGAAAATCAAGGCATCTATAGTTTGTGTCATCTCTGGGAAACCAATCCCTATAAATCAATCTCCTT
3201 CTCCCACTAATTAATCTTTGTTCAAAATGTCTGAAATCCGAGTCTTACTATTCTTACATTTTGAAGAATATAATTCAATAAAATCCTAAGAATTCCAG
3301 AAGAGCTGTAGAATCTTAAGGCCATTATATATATTTTCTTGTCTTCAAAAATTTGAATTTAAACTTGTTCAAAGGTATAAGGGTTTATCTTTCAAGA
3401 GTTGCTCTTATATTCCCATATCTGCCTTTAATCTCCCTTTAAAGTATCTCCATCAGAAATAAGGTCCTTTGAAAGCTCTAATCTTTGAACAGTTGTTA
3501 GCTTATTTTATTTAGCTCTTGAAGAGCTTTTGTAAATTTAGGATAACCTAAAAATCCTAATTCATTTGCAAACTTACTACTGTAGATTTCAGATACAC
3601 TACAGAATCTCCAAGCTTTG

Fig. 4. Nucleotide sequence of *celA* and adjacent regions. The derived amino acid sequence is presented. The leader sequence is underlined, the putative linker region is double-underlined and the CBD is underlined with a dashed line. The arrow at position 2070 in the nucleotide sequence indicates the extent of *celA* that is retained by *celAΔΔ3'* on pCMJI and pCMJII.

This confirmed the DNA sequence of the *celA* and *celA* $\Delta 5'$ genes, as the former contains a signal sequence, while the latter does not. N-terminal amino acid sequence determination of the purified CelA protein showed that the first 26 residues encoded by *celA* constituted a typical signal sequence which is cleaved off. The sequence of the first five residues of the mature CelA protein was determined and found to be Ile-Ser-Arg-Asn-Pro (Fig. 3*b*).

Purification of CelA Δ N' and CelA, and proteolytic cleavage of CelA

To aid purification, the copy numbers of the *celA*Δ5' and *celA* genes in *E. coli* were increased by subcloning them from the low-copy-number vector pEcoR251 to the high-copy-number vectors pUC19 and pSK to obtain pCMI and pCML, respectively (Figs 1 and 2). The intracellular CelAΔN' was purified from a cell-free extract, while

Table 1. Localization of *CelA* $\Delta N'$ and *CelA* in *E. coli* BL21

Cellular fraction	β -Galactosidase*†		Alkaline phosphatase*†		CMCase*	
	pCMI‡	pCML‡	pCMI	pCML	pCMI	pCML
Culture supernatant	0.11	0.31	ND	0.55	ND	ND
First-stage osmotic shock supernatant	0.15	0.10	2.8	4.1	ND	ND
Periplasmic fraction	2.5	3.0	84.3	87.9	5.7	58
Intracellular fraction	97.2	96.6	12.9	7.5	94.3	42

ND, Not detectable.

* Enzyme activities expressed as percentage of total activity recovered.

† β -Galactosidase and alkaline phosphatase were used as intracellular and periplasmic marker enzymes, respectively.‡ pCMI and pCML encode *CelA* $\Delta N'$ and *CelA*, respectively.Table 2. Purification of (a) *CelA* $\Delta N'$ from *E. coli* K514(pCMI) and (b) *CelA* from *E. coli* BL21(pCML)

Purification step	Activity (CMCase U ml ⁻¹)*	Specific activity (U mg ⁻¹)*	Total activity (U)*	Yield (%)	Purification (-fold)
(a)					
Cell-free extract	68	2.5	1134	100	1.0
pH-precipitated extract	39	4.3	535	47	1.7
Q-Sepharose anion-exchange (pooled fractions)	4.2	47	236	21	19
Bio-Gel P-200 gel filtration	12	188	125	11	74
(b)					
Periplasmic extract	17.5	54.3	2830	100	1.0
Q-Sepharose anion-exchange					
Pool I	7.40	130	302	10.7	2.4
Pool II	8.56	147	624	22.0	2.7
Pool III	6.10	158	400	14.1	2.9

* 1 Unit (U) = 1 μ mol glucose equivalents liberated min⁻¹.

CelA was purified from the periplasmic fraction (Table 2, Fig. 5). Unlike the intracellular *CelA* $\Delta N'$, the periplasmic *CelA* was prone to proteolytic truncation when purified out of *E. coli* strains LK111 or K514. When *CelA*-containing fractions from the ion-exchange column were subjected to zymogram analysis, two CMCase bands with apparent mobilities of 57 and 47 kDa were detected (Fig. 6a). The larger band corresponded to the full-length *CelA* protein while the smaller band, as will be shown later, was the product of proteolytic cleavage at or near the putative linker region (Fig. 4), leading to the removal of the CBD to form *CelA* Δ CBD. To avoid this proteolytic degradation, *E. coli* BL21, which lacks the OmpT protease, localized to the outer membrane, and the *lon* protease (Grodberg & Dunn, 1988), was used for

the purification of full-length *CelA*, which was less contaminated with the truncated derivatives (compare lane 8, Fig. 5b, with lane 6, Fig. 6b).

Three pools of *CelA* were obtained from fractions eluted from the anion-exchange column (Fig. 5b). Pool I, which contained no *CelA* Δ CBD, corresponded to the major part of the protein peak, while pools II and III, containing small amounts of *CelA* Δ CBD, were collected before and after the main peak, respectively. Pools II and III had a slightly higher specific activity than pool I (Table 2b), because the truncated CD (*CelA* Δ CBD) had a higher activity on CMC than the full-length enzyme. This could also be seen on the zymogram of *CelA* and *CelA* Δ CBD (Fig. 6a). Higher specific activity of the truncated CD on soluble substrates has also been

Fig. 5. SDS-PAGE analysis of the purification of (a) CelA Δ N' from *E. coli* K514(pCMI) and (b) CelA from *E. coli* BL21(pCML). Lanes: 1, 6 and 7, molecular size markers; 2, cell-free extract (80 μ g); 3, pH-precipitated extract (64 μ g); 4, pooled fractions from the ion-exchange column (20 μ g); 5, pooled fractions from the gel-filtration column (3 μ g); 8, 9 and 10, pool I fractions from the ion-exchange column (5.7 μ g, 17.0 μ g and 56.8 μ g, respectively); 11, pool II (58.4 μ g); 12, pool III (38.7 μ g).

Fig. 6. Zymogram and SDS-PAGE analysis of the cellulose-binding study. These samples originated from *E. coli* LK111(pCML). (a) Zymogram of fractions containing CelA and CelA Δ CBD. Lanes 1–5 represent fractions eluting from the column corresponding to the centre of the peak of CMCase activity. (b) SDS-PAGE of the CelA and CelA Δ CBD cellulose-binding study. Lanes: 6, CelA and CelA Δ CBD from the ion-exchange column; 7, CelA Δ CBD which did not bind to the Avicel; 8, CelA eluted from the Avicel column with 5% (w/v) SDS; 9, molecular size markers.

reported for other cellulases prone to proteolytic cleavage (Durrant *et al.*, 1991; Hansen *et al.*, 1992; Fierobe *et al.*, 1991).

Cellulose-binding studies

In cellulose-binding assays the full-length CelA (57 kDa) bound to Avicel (Fig. 6b, lane 8) and had to be eluted with 5% (w/v) SDS. The proteolytic degradation product, CelA Δ CBD (47 kDa), did not bind to the Avicel (Fig. 6b, lane 7). The faint CelA Δ CBD band in lane 8 might be due to insufficient washing of the Avicel column, but it is also possible that CelA Δ CBD has another, weaker cellulose-binding region. CelA could be

eluted from the Avicel column using 1% (w/v) SDS. The least-stringent non-denaturing conditions to elute CelA from the Avicel were found to be 1% (w/v) urea at room temperature. Cold distilled H₂O did not elute CelA from the Avicel, in contrast to endoglucanase E of *C. thermocellum* (Durrant *et al.*, 1991); guanidinium chloride caused only limited desorption (Gilkes *et al.*, 1988).

pH and temperature optima and stability of CelA

pH 4.8 was optimal for CMCase activity; 50% activity was obtained at pH 4.3 and 6.8 (results not shown). The temperature optimum was 43 °C and corresponded closely to the temperature stability of CelA, which was

Table 3. Substrate specificities of CelA

Substrate	Specific activity (units mg ⁻¹)*
CMC (1→4)-β-D	188
Barley β-glucan (1→3)(1→4)-β-D	1297
Lichenan (1→3)(1→4)-β-D	310
Laminarin (1→3)-β-D	0
Oat speltis xylan (1→4)-β-D	3.5

* 1 Unit = 1 μmol glucose equivalents liberated min⁻¹.

Fig. 7. HPLC analysis of cellohexaose degradation by CelA. (a) G1–G6 mixture resolved by HPLC. (b) G6 after incubation at 43 °C for 170 min in the absence of CelA. (c, d and e) G6 + CelA after 7.5 min, 15 min and 145 min incubation at 43 °C, respectively. G2 appears as a doublet and should not be mistaken for G1 plus G2. Symbols: G1, glucose; G2, cellobiose; G3, cellotriose; G4, cellotetraose; G5, cellopentaose; G6, cellohexaose.

Table 4. Effects of divalent cations on purified CelA

The values presented are means of three assays. All metals were in the form of chloride salts at 0.2 mM. The control (100%) CMCase activity was equivalent to 9.87×10^{-3} CMCase units.

Divalent cation	CMCase activity (% of control)
Control	100
Mn ²⁺	28
Hg ²⁺	29
Ca ²⁺	85
Ni ²⁺	88
Zn ²⁺	89
Mg ²⁺	90
Cd ²⁺	90
Co ²⁺	144

determined by measuring CMCase activity after 2 h incubation of the enzyme at selected temperatures (results not shown). CelA was found to be stable at 45 °C, but higher temperatures reduced the enzyme activity rapidly. The optimal temperature for CelA activity corresponded with the temperature of the rumen environment.

Substrate specificities

CelA had the highest activity on barley β-glucan, followed by lichenan and CMC (Table 3). Avicel and ball-milled filter paper (Whatman no. 1) were not hydrolysed at all (data not shown). CelA was unable to hydrolyse (1→3)-β-D bonds, but it had low activity on xylan. The ratio of CMCase activity to xylanase activity of CelA was 50:1, while similar ratios of 10:1 and 100:1, respectively, have been reported for endoglucanases E and H from *C. thermocellum* (Béguin, 1990). Methylcellulose (0.04%, w/v) inhibited CelA activity by 55%, but 1 mM-glucose and 1 mM-cellobiose did not have an inhibitory effect (data not shown).

HPLC analysis of cellodextrin degradation

CelA had no activity on cellobiose (G2), but cellotriose (G3) was partially degraded to cellobiose and glucose (G1) after 24 h. Cellotetraose (G4) was degraded completely to G2 (85%), G3 and G1 after 16 h. Cellopentaose (G5) was rapidly converted to G3 and G2 in 2 h (results not shown). The degradation of cellohexaose (G6) proceeded very rapidly (Fig. 7). G5 and G4 were formed as intermediate breakdown products during the degradation of G6, but the final products were G3, G2 and G1. CelA clearly demonstrated a preference for larger cellodextrins.

Effects of divalent cations on CelA

Mn²⁺ and Hg²⁺ inhibited CelA, while Ca²⁺, Ni²⁺, Zn²⁺, Mg²⁺ and Cd²⁺ had no significant effect (Table 4). Co²⁺ enhanced CelA activity by 44%. The enhancing effect of Co²⁺ was also observed with the non-dialysed enzyme preparation; 10 μM-Co²⁺ was sufficient to enhance the activity by 20–30% (results not shown). The addition of 1 mM-2-mercaptoethanol or 0.5 mM-DTT did not have an effect on the CMCase activity, suggesting that Co²⁺ probably does not affect the redox potential (results not shown). The enhancement of cellulase activity by Co²⁺ has, to our knowledge, not been reported previously, but an exo-β-glucanase from *Ruminococcus flavefaciens* FD-1 was moderately inhibited by Co²⁺ (Gardner *et al.*, 1987).

Discussion

celA is the first gene from *Clostridium* sp. ATCC 49440 to be cloned and genetically analysed, since the organism has only recently been reisolated from the rumen of a bison by Varel (1989). The gene was sequenced and found to belong to the large and well-characterized endoglucanase family A4, as reviewed by Béguin (1990) and Gilkes *et al.* (1991a). The derived amino acid sequence of CelA shows high homology with EGE from *C. thermocellum* and with other endo-(1 $\rightarrow$ 4)- β -glucanases from ruminal bacteria (*Butyrivibrio fibrisolvens* H17c, *Ruminococcus albus* F-40 and SY3) that are closely clustered on the same branch in the unrooted phylogenetic tree of family A (Gilkes *et al.*, 1991a).

The DNA sequence was confirmed by the determination of the N-terminal amino acid sequence of purified CelA and CelA Δ N', and by the localization of the CelA and CelA Δ N' cellulases in *E. coli*. The molecular mass of CelA, 57 kDa as determined by SDS-PAGE, corresponded with the calculated molecular mass (57.5 kDa) of the gene product of the entire *celA* gene. This does not, however, take into account the cleavage of the signal sequence, which should increase the mobility of the mature protein on SDS-PAGE.

Endoglucanase E of *C. thermocellum*, which is 48% homologous to CelA, was found to be secreted into the periplasm of *E. coli* even when the signal sequence was removed (Hazlewood *et al.*, 1990). This was also found for the xylanase, XYLA, from *Pseudomonas fluorescens* subspecies *cellulosa*, in which serine-rich internal signal sequences were identified (Hall *et al.*, 1989). In contrast, the export of CelA into the periplasm of *E. coli* was found to be dependent on the presence of the N-terminal signal peptide. The incomplete translocation of CelA to the periplasm (58%) could be ascribed to the inefficient recognition of the *Clostridium* sp. signal peptide by the *E. coli* secretion system.

CelA was found to have a modular structure, consisting of an N-terminal CD of 408 residues (including the signal sequence) and the C-terminal CBD of 97 residues. These are joined by a putative linker region of 11 residues, containing two proline and six threonine residues. In their review of the modular structure of β -glycanases, Gilkes *et al.* (1991a) point out that the linker regions of many β -glucanases are rich in proline and hydroxyamino acid residues. The deletion of 134 residues from the C-terminal end of the CelA CD caused the loss of its hydrolytic activity, as *celA* Δ 3' on pCMJI did not produce an active endoglucanase. The cellulose-binding study suggested that the CelA CD might contain another weak cellulose-binding region, because a small but significant amount of CelA Δ CBD was still bound to the Avicel after two washes, and eluted together with the

full-length CelA in 5% SDS. This small quantity of 'bound' CelA Δ CBD could, however, also be due to inadequate washing of the Avicel column.

The duplicated 24 amino acid segments present in most *C. thermocellum* cellulases, and also in *C. cellulolyticum* EGA, (Béguin, 1990) are absent in CelA. Since these repeats are thought to be required for the binding of the cellulases to the S_L protein of the cellulosome (Tokatlidis *et al.*, 1991; Fujino *et al.*, 1992; Salamiou *et al.*, 1992), it is possible that the *Clostridium* sp. does not have such a cellulosome structure. An alternative model for bacterial cellulases, where the individual cellulases are not involved in binding to a multi-protein cellulolytic complex, has been proposed for *P. fluorescens* subsp. *cellulosa* (Béguin *et al.*, 1992; Hazlewood *et al.*, 1992).

Proteolytic cleavage between the catalytic and cellulose-binding domains has been described for a number of β -glucanases (Hall *et al.*, 1989; Hansen *et al.*, 1992; Fierobe *et al.*, 1991; Gilkes *et al.*, 1988, 1989; Jauris *et al.*, 1990). The difference in molecular masses of CelA and CelA Δ CBD (calculated from the sequence data and determined by SDS-PAGE analysis) confirms that the proteolytic cleavage of CelA takes place in or near the proposed linker region, as is the case for many of these modular-type β -glycanases (Gilkes *et al.*, 1991a). The use of *E. coli* BL21, which lacks the outer-membrane-associated OmpT protease and the lon protease (Grodberg & Dunn, 1988), greatly reduced the cleavage of CelA which was experienced with *E. coli* strains LK111 and K514 during purification. It is interesting that the intracellularly situated CelA Δ N' was not prone to proteolytic truncation in any of these *E. coli* strains.

The CBD of CelA shows amino acid sequence homology with other bacterial *C. fimi*-type CBDs (Gilkes *et al.*, 1991a,b). The CelA CBD does not, however, contain the two conserved cysteine residues which are present in all other known CBDs with the exception of that from *Butyrivibrio fibrisolvens*. These two residues have been proposed to form extended disulphide bonds in *C. fimi*-type CBDs.

The authors thank V. H. Varel for supplying *C. longisporum*, G. Lindsey for valuable advice with the purification of CelA, W. F. Brandt for the N-terminal amino acid sequence determination, R. Saunders for valuable discussions and AECI Ltd for financial support in the form of a bursary for V. M.

References

- BÉGUIN, P. (1983). Detection of cellulase activity in polyacrylamide gels using congo red-stained agar replicas. *Analytical Biochemistry* **131**, 333–336.
- BÉGUIN, P. (1990). Molecular biology of cellulose degradation. *Annual Review of Microbiology* **44**, 219–248.
- BÉGUIN, P., MILLET, J., CHAUVAX, S., SALAMITOU, S., TOKATLIDIS, K., NAVAS, J., FUJINO, T., LEMAIRE, M., RAYNAUD, O., DANIEL, M.-K. &

- AUBERT, J.-P. (1992). Bacterial cellulases. *Biochemical Society Transactions* **20**, 42–46.
- BRANDT, W. F., ALK, H., CHAUHAN, M. & VON HOLT, C. (1984). A simple modification converts the spinning cup sequencer into a vapour phase sequencer. *FEBS Letters* **174**, 228–232.
- BRICKMAN, E. & BECKWITH, J. (1975). Analysis of the regulation of *E. coli* alkaline phosphatase synthesis using deletions and pho80 transducing phages. *Journal of Molecular Biology* **96**, 307–316.
- CALDWELL, D. R. & BRYANT, M. P. (1966). Medium without rumen fluid for nonselective enumeration and isolation of rumen bacteria. *Applied Microbiology* **14**, 794–801.
- DURRANT, A. J., HALL, J., HAZLEWOOD, G. P. & GILBERT, H. J. (1991). The non-catalytic C-terminal regions of endoglucanase E from *Clostridium thermocellum* contains a cellulose-binding domain. *Biochemical Journal* **273**, 289–293.
- FERREIRA, L. M. A., DURRANT, A. J., HALL, J., HAZLEWOOD, G. P. & GILBERT, H. J. (1990). Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. *Biochemical Journal* **269**, 261–264.
- FIEROBE, H.-P., GAUDIN, C., BELAICH, A., LOUFI, M., FAURE, E., BAGNARA, C., BATY, D. & BELAICH, J.-P. (1991). Characterization of endoglucanase A from *Clostridium cellulolyticum*. *Journal of Bacteriology* **173**, 7956–7962.
- FUJINO, T., BEGUIN, P. & AUBERT, J.-P. (1992). Cloning of a *Clostridium thermocellum* DNA fragment encoding polypeptides that bind the catalytic components of the cellulosome. *FEMS Microbiology Letters* **94**, 165–170.
- GARDNER, R. M., DOERNER, K. C. & WHITE, B. A. (1987). Purification and characterization of an α -D-glucanase from *Ruminococcus flavefaciens* FD-1. *Journal of Bacteriology* **169**, 4581–4588.
- GHOSE, T. K. (1987). Measurement of cellulase activities. *Pure and Applied Chemistry* **59**, 257–268.
- GILKES, N. R., WARREN, R. A. J., MILLER, R. C. & KILBURN, D. G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *Journal of Biological Chemistry* **263**, 10401–10407.
- GILKES, N. R., KILBURN, D. G., MILLER, R. C. & WARREN, R. A. J. (1989). Structural and functional analysis of a bacterial cellulase by proteolysis. *Journal of Biological Chemistry* **264**, 17802–17808.
- GILKES, N. R., HENRISSAT, B., KILBURN, D. G., MILLER, R. C. & WARREN, R. A. J. (1991a). Domains in microbial β -1,4-glycanases: sequence conservation, function, and enzyme families. *Microbiological Reviews* **55**, 303–315.
- GILKES, N. R., CLAEYSSENS, M., AEBERSOLD, R., HENRISSAT, B., MEINKE, A., MORRISON, H. D., KILBURN, D. G., WARREN, R. A. J. & MILLER, R. C. (1991b). Structural and functional relationships in two families of β -1,4-glycanases. *European Journal of Biochemistry* **202**, 367–377.
- GRODBERG, J. & DUNN, J. J. (1988). *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology* **170**, 1245–1253.
- HALL, J., HAZLEWOOD, G. P., BARKER, P. J. & GILBERT, H. J. (1988). Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. *Gene* **69**, 29–38.
- HALL, J., HAZLEWOOD, G. P., HUSKISSON, N. S., DURRANT, A. J. & GILBERT, H. J. (1989). Conserved serine-rich sequences in xylanases and cellulases from *Pseudomonas fluorescens* subsp. *cellulosa*: internal signal sequence and unusual protein processing. *Molecular Microbiology* **3**, 1211–1219.
- HANSEN, C. K., DIDERICHSEN, B. & JØRGENSEN, P. L. (1992). *celA* from *Bacillus lautus* PL236 encodes a novel cellulose-binding endo- β -1,4-glucanase. *Journal of Bacteriology* **174**, 3522–3531.
- HAZLEWOOD, G. P., DAVIDSON, K., CLARKE, J. H., DURRANT, A. J., HALL, J. & GILBERT, H. J. (1990). Endoglucanase E, produced at high level in *E. coli* as a *lacZ'* fusion protein, is part of the *Clostridium thermocellum* cellulosome. *Enzyme and Microbial Technology* **12**, 656–662.
- HAZLEWOOD, G. P., LAURIE, J. I., FERREIRA, L. M. A. & GILBERT, H. J. (1992). *Pseudomonas fluorescens* subsp. *cellulosa*: an alternative model for bacterial cellulases. *Journal of Applied Bacteriology* **72**, 244–251.
- HENRISSAT, B., CLAEYSSENS, M., TOMME, P., LEMESLE, L. & MORNON, J.-P. (1989). Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81**, 83–95.
- HEWICK, R. M., HUNKAPILLER, M. W., HOOD, L. E. & DREYER, W. J. (1981). A gas-liquid solid phase peptide and protein sequencer. *Journal of Biological Chemistry* **256**, 7990–7997.
- 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 thermoactive cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. *Molecular and General Genetics* **223**, 258–267.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 680–685.
- LIN, L.-L. & THOMSON, J. A. (1991). An analysis of the extracellular xylanases and cellulases of *Butyrivibrio fibrisolvens* H17c. *FEMS Microbiology Letters* **84**, 197–204.
- LOTTSPREICH, F. (1985). Microscale isocratic separation of the phenylthiohydantoin amino acid derivatives. *Journal of Chromatography* **326**, 321–327.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MILLER, J. H. (1972). *Experiments in Molecular Genetics*, pp. 352–355. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- NOSSAL, N. G. & HEPPEL, L. A. (1966). The release of enzymes by osmotic shock from *E. coli* in exponential phase. *Journal of Biological Chemistry* **13**, 3055–3062.
- SALAMITOU, S., TOKATLIDIS, K., BEGUIN, P. & AUBERT, J.-P. (1992). Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. *FEBS Letters* **304**, 89–92.
- SHARMA, S. & SANDHU, D. K. (1986). An improved method for detection of isozymes of endocellulases. *Indian Journal of Experimental Biology* **24**, 732–733.
- STUDIER, F. W. & MOFFAT, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189**, 113–130.
- TEATHER, R. M. & WOOD, P. J. (1982). Use of congo red-polysaccharide interactions in enumeration and characterization of cellulolytic bacteria from the bovine rumen. *Applied and Environmental Microbiology* **43**, 777–780.
- TOKATLIDIS, K., SALAMITOU, S., BEGUIN, P., DHURJATI, P. & AUBERT, J.-P. (1991). Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. *FEBS Letters* **291**, 185–188.
- VAREL, V. H. (1989). Reisolation and characterization of *Clostridium longisporum*, a ruminal sporeforming cellulolytic anaerobe. *Archives of Microbiology* **152**, 209–214.
- WOOD, W. B. (1966). Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *Journal of Molecular Biology* **16**, 118–133.
- WOOD, T. M. (1992). Fungal cellulases. *Biochemical Society Transactions* **20**, 46–53.
- 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.
- ZABEAU, M. & STANLEY, K. K. (1982). Enhanced expression of cro- β -galactosidase fusion proteins under the control of the P_R promoter of bacteriophage λ . *EMBO Journal* **1**, 1217–1224.
- ZAPPE, H., JONES, D. T. & WOODS, D. R. (1986). Cloning and expression of a *Clostridium acetobutylicum* endoglucanase, cellobiase and amino acid biosynthesis genes in *Escherichia coli*. *Journal of General Microbiology* **132**, 1367–1372.