Characterization of the gene *celD* and its encoded product 1,4- β -D-glucan glucohydrolase D from *Pseudomonas fluorescens* subsp. *cellulosa*

Jane E. RIXON,* Luis M. A. FERREIRA,* Alastair J. DURRANT,* Judith I. LAURIE,†
Geoffrey P. HAZLEWOOD† and Harry J. GILBERT*†

*Department of Biological and Nutritional Sciences, The Agriculture Building,
University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, and †Department of Biochemistry,
AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, U.K.

A genomic library of *Pseudomonas fluorescens* subsp. *cellulosa* DNA constructed in pUC18 and expressed in *Escherichia coli* was screened for recombinants expressing 4-methylumbelliferyl β -D-glucoside hydrolysing activity (MUGase). A single MUGase-positive clone was isolated. The MUGase hydrolysed cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexaose to glucose, by sequentially cleaving glucose residues from the non-reducing end of the cello-oligosaccharides. The K_m values for cellobiose and cellohexaose hydrolysis were 1.2 mM and 28 μ M respectively. The enzyme exhibited no activity against soluble or insoluble cellulose, xylan and xylobiose. Thus the MUGase is classified as a 1,4- β -D-glucan glucohydrolase (EC 3.2.1.74) and is designated 1,4- β -D-glucan glucohydrolase D (CELD). When expressed by *E. coli*, CELD was located in the cell-envelope fraction; a significant proportion of the native enzyme was also associated with the cell envelope when synthesized by its endogenous host. The nucleotide sequence of the gene, *celD*, which encodes CELD, revealed an open reading frame of 2607 bp, encoding a protein of M_r 92 000. The deduced primary structure of CELD was confirmed by the M_r of CELD (85 000) expressed by *E. coli* and *P. fluorescens* subsp. *cellulosa*, and by the experimentally determined *N*-terminus of the enzyme purified from *E. coli*, which showed identity with residues 52–67 of the *celD* translated sequence. The structure of the *N*-terminal region of full-length CELD was similar to the signal peptides of *P. fluorescens* subsp. *cellulosa* plant-cell-wall hydrolases. Deletion of the *N*-terminal 47 residues of CELD solubilized MUGase activity in *E. coli*. CELD exhibited sequence similarity with β -glucosidase B of *Clostridium thermocellum*, particularly in the vicinity of the active-site aspartate residue, but did not display structural similarity with the mature forms of cellulases and xylanases expressed by *P. fluorescens* subsp. *cellulosa*.

INTRODUCTION

Pseudomonas fluorescens subsp. *cellulosa* is a Gram-negative bacterium capable of hydrolysing several components of the plant cell wall, including cellulose and xylan. In common with most other cellulolytic micro-organisms, the bacterium expresses isoenzymic forms of cellulases and xylanases that are encoded by discrete genes (Gilbert *et al.*, 1987, 1988). These enzymes include two endoglucanases, two xylanases, one arabinofuranosidase and a cellobiohydrolase. There is no evidence to suggest that *P. fluorescens* subsp. *cellulosa* synthesizes a cellobiohydrolase. The nucleotide sequences of the six plant-cell-wall-hydrolase genes have been determined and the primary structures of the encoded enzymes deduced. The molecular architecture of each of the six pseudomonad proteins is similar and features a non-catalytic cellulose-binding domain (CBD), linked to a catalytic domain (CD) by sequences rich in hydroxy amino acids (Hall & Gilbert, 1988; Hall *et al.*, 1989; Gilbert *et al.*, 1990; Kellett *et al.*, 1990; Ferreira *et al.*, 1991). The CD and CBD can function independently of one another, and the linker region can be deleted without apparently affecting enzyme function (Ferreira *et al.*, 1990). The precise role of the CBD has not been firmly established. However, we have proposed that this domain provides a general mechanism by which cellulases and xylanases attach to the plant

cell wall and are thus able to hydrolyse appropriate polysaccharides as they become exposed during the degradative process (Ferreira *et al.*, 1991).

In addition to endoglucanases, cellobiohydrolases and cellobiohydrolases, most cellulolytic organisms also express β -glucosidases and 1,4- β -D-glucan glucohydrolases, which hydrolyse cellobiose and other cello-oligosaccharides to glucose. Some micro-organisms synthesize only single forms of these enzymes (Kohchi & Toh-e, 1985), whereas other bacteria and fungi express multiple β -glucosidases and 1,4- β -D-glucan glucohydrolases (Gräbnitz *et al.*, 1989). Although extracellular forms of β -glucosidases have been reported (Enari, 1983), the enzyme is normally located in the cytoplasm (Wakarchuk *et al.*, 1988). The catalytic mechanism for β -glucosidase has been investigated, and it has been suggested that the enzyme cleaves β -1,4-linkages via an acid–base catalytic mechanism, analogous to the hydrolysis of bacterial cell walls by lysozyme. Support for this view is provided by the presence of conserved aspartate and glutamate residues at the active site of β -glucosidase A_3 from *Aspergillus wentii* (Legler *et al.*, 1979). The primary structures of several fungal and bacterial β -glucosidases, which have been deduced from the nucleotide sequences of the corresponding genes (Béguin, 1990) indicate the existence of two families of enzymes. These, presumably, have evolved from distinct ancestral

Abbreviations used: MUGase, 4-methylumbelliferyl β -D-glucoside (MUG)-hydrolysing activity; CELD, 1,4- β -D-glucan glucohydrolase D; *celD*, its corresponding gene; CBD, cellulose-binding domain; CD, catalytic domain; LB, Luria broth; IPTG, isopropyl thio- β -D-galactopyranoside; CMC, carboxymethylcellulose.

† To whom correspondence and reprint requests should be sent.

The nucleotide sequence data reported will appear in the EMBL/GenBank/DBJ Nucleotide Sequences Databases under the accession number X65527.

proteins. There is no evidence that β -glucosidases consist of multiple domains, and they exhibit no sequence similarity to β -endoglucanases, xylanases or cellobiohydrolases.

Yamane *et al.* (1970) showed that *P. fluorescens* subsp. *cellulosa* expresses an intracellular enzyme that cleaves cellobiose. To investigate the structure-function relationship of the pseudomonad cellobiase and to compare it with other plant-cell-wall hydrolases expressed by the bacterium, we have cloned and sequenced a *P. fluorescens* subsp. *cellulosa* 1,4- β -D-glucan glucosylhydrolase gene and characterized the encoded protein. The enzyme did not contain a distinct non-catalytic CBD, although the N-terminal region consisted of a membrane-anchor sequence that was linked to the catalytic domain by a short stretch of hydroxy amino acids. The enzyme displays some sequence similarity to β -glucosidase B from *Clostridium thermocellum*, particularly in the vicinity of the active-site aspartate residue.

MATERIALS AND METHODS

Bacterial strains vectors and culture conditions

The *E. coli* strains employed in the present study were JM83 and JM101 (Norrrander *et al.*, 1983). The vectors used were pUC18, pUC19, M13mp18, M13mp19 (Norrrander *et al.*, 1983), pMTL22p, pMTL23p (Chambers *et al.*, 1988) and pNM52 (Gilbert *et al.*, 1986). *E. coli* strains were cultured in Luria broth (LB) or on LB agar (2%, w/v). Ampicillin (100 μ g/ml) and 5-bromo-4-chloro-3-indolyl β -D-galactoside (2 μ g/ml) were added to medium to select for transformants and recombinants respectively. Isopropyl thio- β -D-galactopyranoside (IPTG; 23 μ g/ml) was employed to induce *lacZp*. *P. fluorescens* subsp. *cellulosa* was cultured aerobically at 37°C in M9 salts (Maniatis *et al.*, 1982) containing glucose (0.2%), cellobiose (0.2%) or carboxymethylcellulose (CMC) (2%) as the sole carbon source.

General recombinant DNA procedures

Agarose-gel electrophoresis, Southern hybridization, transformation of *E. coli* and the preparation and modification of genomic and plasmid DNA using restriction enzymes and T4 DNA ligase have been described by Gilbert *et al.* (1987, 1988). *BAL31* deletions were performed as follows: pMC4 was cleaved with *SstI*, digested with *BAL31* as described by Hall *et al.* (1988) and restricted with *HindIII*. The resultant DNA fragments, containing truncated forms of *celD*, were purified by agarose-gel electrophoresis and cloned into *SmaI/HindIII*-digested pUC19. The precise extent of *BAL31* deletions was determined by the direct sequencing of recombinant plasmids. Those plasmids containing truncated *celD* in-phase with *lacZ'* were retained for further studies.

PCR was used to delete 141 bp from the 5' end of *celD*. The primers 5' GGGAATTCTCCAGGCCACCGAAAAGGTG-GAG 3' and 5' AGACTGGCTGACAAATGGGTA 3' were employed to amplify the sequence between map positions 141 and 760 of the MUGase gene (Fig. 7 below) by the PCR, using the Gene Amp kit (Perkin-Elmer/Cetus). The reaction was performed according to the manufacturers' instructions, except that primer/target sequence hybridization was carried out at 40°C. The amplified sequence was digested with *EcoRI* (*EcoRI* site created at 5' end of first primer) and *PstI* (*PstI* site within amplified sequence) and the resultant 0.6 kb *PstI-EcoRI* restriction fragment was purified by agarose-gel electrophoresis and cloned into *PstI/EcoRI*-digested pMTL23p. The nucleotide sequence of the pseudomonad insert in the resultant plasmids was determined.

One recombinant which contained no mutation (one out of five sequenced), was designated pMC12 and retained for further

studies. The plasmid pMC4 was subsequently digested with *XbaI*, the 5' overhang filled in with T4 DNA polymerase and restricted with *PstI*. The resultant 3.1 kb *PstI*/blunt-ended DNA fragment containing the 3' region of *celD* was purified and cloned into pMC12. The resultant plasmid, pMC9, contains a truncated derivative of *celD* which is in-phase with *lacZ'*, but lacks the 141 5' nucleotides.

Nucleotide sequencing

The 1.6 kb *KpnI* and 3.0 kb *PstI* restriction fragments derived from pMC4, which contained *celD*, were circularized by self-ligation. The DNA was then randomly cleaved by sonication, blunt-ended and cloned into dephosphorylated *SmaI*-restricted M13mp18 as described by Minton *et al.* (1986). The 270 bp sequence between the *KpnI* and *PstI* restriction sites was determined after cloning into M13mp18 and M13mp19. To sequence over the *KpnI* and *PstI* sites appropriate DNA fragments overlapping these restriction sites were cloned into M13-based vectors.

Recombinant clones were sequenced essentially by the dideoxy chain-termination method of Sanger *et al.* (1980), except that T7 DNA polymerase (Tabor & Richardson, 1987) was used instead of the Klenow fragment of DNA polymerase. Sequences were compiled and ordered using the computer programs published by Staden (1980). The *celD* gene was sequenced in both strands. For direct plasmid sequencing, DNA prepared by the method of Birnboim & Doly (1979) was denatured in 1 M-NaOH for 5 min, the alkali was removed and the DNA annealed with either universal or reverse primer; sequencing reactions were performed as above.

Screening of plasmid library of *P. fluorescens* subsp. *cellulosa* DNA

A gene library of *P. fluorescens* subsp. *cellulosa* DNA was constructed in pUC18 as described by Ferreira *et al.* (1991). Each recombinant plasmid contained approx. 10 kb of *P. fluorescens* subsp. *cellulosa* DNA. The gene library was screened for clones expressing β -glucosidase activity as follows. Recombinants were grown overnight at 37°C with shaking in 1.5 ml micro-centrifuge tubes containing 1 ml of LB supplemented with ampicillin. Cells were subjected to osmotic shock and then incubated with 4-methylumbelliferyl β -D-glucoside (MUG; 1 mM) in 50 mM-sodium phosphate buffer, pH 7.0, for 30 min. Recombinants expressing MUGase activity fluoresced brightly when exposed to long-wavelength u.v. light.

Cell fractionation

Cultures (100 ml) of *E. coli* and *P. fluorescens* subsp. *cellulosa* were grown to mid-exponential phase; the cells were pelleted by centrifugation at 10000 g for 5 min at 4°C, and resuspended in 5 ml of 10 mM-Tris/HCl buffer, pH 7.0. The bacteria were disrupted by sonication and the envelope fraction was pelleted by centrifugation at 100000 g for 1 h at 4°C. The cytoplasmic fraction was recovered and the cell envelope was resuspended in 5 ml of 10 mM-Tris/HCl buffer, pH 7.0. To verify that the soluble and envelope components of the bacteria had been adequately separated, the various fractions were assayed for marker enzymes located in the cytoplasm (malate dehydrogenase) and in the cell membrane (NADH oxidoreductase) as described by Minton *et al.* (1983).

Purification of MUGase and determination of N-terminal sequence

E. coli JM83 harbouring pMC4 was cultured for 16 h at 37°C in LB broth (3 litres) containing ampicillin (100 μ g/ml). Cells harvested by centrifugation and resuspended in ice-cold 50 mM-

Tris/HCl, pH 8.0 (100 ml) were sonicated and centrifuged (8500 g; 4°C; 20 min) to produce soluble and cell-envelope fractions. The cell-envelope fraction was resuspended in 10 ml of trypsin/EDTA (Gibco; Puck's saline A containing 0.5 mg of trypsin/ml and 0.2 mg of EDTA/ml) and incubated at 37°C for 30 min. The supernatant, containing MUGase activity, was dialysed against 10 mM-Tris/HCl, pH 8.0, and applied to an anion-exchange column (DEAE-Trisacryl M; 25 cm $\times$ 2.6 cm) equilibrated with the same buffer. The column was eluted at a flow rate of 30 ml/h with a linear gradient containing NaCl (0–400 mM) in 10 mM-Tris/HCl, pH 8.0. MUGase activity, which emerged between 170 and 200 mM-NaCl, was pooled, added to an equal volume of 1 M-phosphate buffer, pH 6.6, and applied to a phenyl-Sepharose column (11 cm $\times$ 2.6 cm) that had been equilibrated with 0.5 M-phosphate, pH 6.6. The column was eluted at 15 ml/h with a descending gradient from 0.5 M-phosphate, pH 6.6, to distilled water. MUGase activity emerged between 145 and 65 mM-phosphate. A sample of active fraction containing a single protein species was concentrated 10-fold by vacuum evaporation and applied directly to a 470 A gas-phase sequencer equipped with a 120 A on-line phenylthiohydantoin-derivative analyser (Applied Biosystems) for determination of *N*-terminal sequence (Hunkapillar *et al.*, 1983).

Immunological detection of the MUGase

Antiserum was raised against purified MUGase in rabbits as described by Hazlewood *et al.* (1990a). Proteins from various fractions of *E. coli* and *P. fluorescens* subsp. *cellulosa* were subjected to SDS/PAGE (Laemmli, 1970) and were electrophoretically transferred to Hybond C (Amersham) as described by Towbin *et al.* (1979). The immunological detection of the MUGase on Western blots, using anti-MUGase antiserum as the primary antibody, was as described by Hazlewood *et al.* (1990a).

Assays

MUGase was assayed at 37°C in reaction mixtures containing 1 mM-MUG in 50 mM-sodium phosphate buffer, pH 7.0. The appearance of the product, 4-methylumbelliferone, was detected at 350 nm. The absorption coefficient of the product in phosphate buffer, pH 7.0, was determined experimentally as 4350 litre mol⁻¹ cm⁻¹. One unit of enzyme activity was defined as the amount of MUGase required to produce 1 μ mol of product/min at 37°C. Assays for xylanase- and cellulase-related activities were as described by Hazlewood *et al.* (1990b). The products generated by the action of the MUGase on cello-oligosaccharides (cellobiose to cellohexaose) were identified by h.p.l.c. Reaction mixtures (250 μ l final volume) containing 6.13×10^{-4} – 6.13×10^{-3} units of purified MUGase and 0.1 mg of the appropriate cello-oligosaccharide/ml in 50 mM-potassium phosphate/12 mM-citric acid buffer, pH 6.5, were incubated for up to 1 h at 37°C. Any insoluble material was removed by centrifugation at 13000 g for 5 min, and aliquots (20 μ l) of supernatant were applied to a Dionex CarboPac PA1 (4 mm $\times$ 250 mm) column. The detection system was a Dionex pulsed amperometric detector using a 10 K detector range with the following applied potentials: E_1 , +0.05V; E_2 , +0.6V; E_3 –0.6V. The detection cell used a gold working electrode. Sugars were eluted from the column with a 5–25 mM-sodium acetate gradient in 150 mM-NaOH. Cello-oligosaccharidase activity was assayed at 37°C in reaction mixtures containing 10 μ M–10 mM-cellobiose or cellohexaose, in 50 mM-Tris/HCl buffer, pH 7.5. The appearance of the product, glucose, was detected using the Sigma Glucose (HK) kit. Liberated glucose was converted into glucose 6-phosphate using an excess of hexokinase and ATP; the phosphorylated sugar was then oxidized by glucose-6-phosphate dehydrogenase, and the resultant reduction of NAD⁺ to NADH was monitored at 340 nm.

The concentration of the reagents used to detect glucose was described in the manufacturer's instructions. Protein was determined by the Lowry method, with BSA as standard.

RESULTS

Cloning and expression of *celD*

A gene library of *P. fluorescens* subsp. *cellulosa* DNA was constructed in pUC18. Approx. 2000 recombinants were screened, as described in the Materials and methods section, for the capacity to express enzymes which cleave MUG (MUGase). Only one recombinant was found to hydrolyse the aryl β -glucoside. A restriction map of the plasmid, pMC4, harboured by this MUGase positive clone is shown in Fig. 1. To locate the position of the MUGase gene (*celD*), various restriction fragments were either deleted from pMC4 or subcloned into appropriate pUC or pMTL vectors. An evaluation of the capacity of resultant plasmids to direct the synthesis of functional MUGase (Fig. 1) indicates that *celD* is located between the *StuI* and *HindIII* restriction sites at map positions 2.05 and 5.8 respectively.

The synthesis of MUGase by cells carrying the plasmid pMC8 was found to be 3.5 times greater than by those carrying pMC4. To assess whether this increase in expression is the result of transcription from *lacZ'*, pNM52, a multicopy plasmid (30–50 copies/cell) which is compatible with *colE1* replicons and contains *lacI^q*, was transformed into cells containing pMC4 and pMC8. The *lacI^q* gene caused a 20-fold decrease in MUGase expression in cells containing pMC8, but did not significantly affect MUGase synthesis when enzyme was encoded by pMC4. MUGase production was regulated by the *lac* inducer IPTG, in cells containing both pMC8 and pNM52 (Table 1; Fig. 2). These data indicate that *celD* is transcribed from left to right in pMC4.

Location and purification of MUGase from *E. coli*

To investigate the cellular location of the MUGase encoded by *celD*, cell-envelope and cytoplasmic fractions were prepared from *E. coli* cells containing pMC4 and were assayed for MUGase activity and marker enzymes normally associated with either the cytoplasm (malate dehydrogenase) or the cell membrane (NADH oxidoreductase). Most of the MUGase and NADH oxidoreductase activity was associated with the cell-envelope fraction (Table 2). SDS/PAGE of cell-envelope proteins derived from *E. coli* harbouring pMC4 revealed a prominent polypeptide of M_r 85000 which was not apparent in the corresponding fraction from *E. coli* containing pUC18 (Fig. 2). Similarly, the cell envelope from *E. coli* harbouring pMC8 and pNM52 contained a prominent protein of M_r 85000 when grown in the presence of IPTG, which was greatly decreased in the absence of the *lac* inducer (Fig. 2). This supports the view that the 85000- M_r protein is the recombinant MUGase.

Recombinant MUGase, initially associated with the cell envelope from *E. coli* harbouring pMC4 was solubilized by treatment with trypsin and was purified by anion-exchange and hydrophobic-interaction chromatography. SDS/PAGE of the purified enzyme revealed a single polypeptide of M_r 80000 (Fig. 3); the experimentally determined *N*-terminal sequence of the protein was:

K V A W P V L N G A I K K D L A

Substrate specificity

Purified MUGase hydrolysed several aryl β -glucosides, but did not cleave xylan, carboxymethylcellulose or insoluble cellulose (Table 3). These results indicate that the enzyme has a very limited substrate range and differs from xylanases (Gilbert *et al.*,

Fig. 1. Restriction map of recombinant plasmids containing *celD*

The positions of the cleavage sites *Eco*RI (RI), *Eco*RV (RV), *Hind*III (H), *Kpn*I (K), *Nru*I (N), *Pst*I (Ps), *Pvu*II (Pv), *Sal*I (Sa), *Sph*I (Sp) and *Stu*I (St) in the pseudomonad insert of the recombinant plasmids are shown. The solid arrow shows the extent and orientation of *celD*. The deletion of the two *Eco*RV restriction fragments from pMC4, to generate pMC6 is indicated by ----. Plasmids in which the 5' region of *celD* was deleted by either *BAL*31 or by using PCR are designated pMC9-pMC12. The position of the 5' nucleotide of each truncated *celD* gene is shown in the nucleotide sequence depicted in Fig. 7. The phenotypes of *E. coli* strains harbouring the recombinant plasmids are shown.

Table 1. CELD activity in *E. coli* strains containing recombinant plasmids

IPTG was added to culture medium at the time of inoculation. MUGase activity was determined using cell-free extracts of the recombinant strains.

Plasmid(s)	Addition of IPTG to culture medium	Specific activity of CELD (units/mg of total protein)
pMC4	—	1.46×10^{-3}
pMC8/pNM52	—	1.57×10^{-4}
pMC8/pNM52	+	4.87×10^{-3}
pMC4/pNM52	—	1.65×10^{-3}
pMC4/pNM52	+	2.65×10^{-3}

1988; Kellett *et al.*, 1990) and cellulases (Gilbert *et al.*, 1987; Gilbert *et al.*, 1990) previously isolated from *P. fluorescens* subsp. *cellulosa*, which also hydrolyse MUC. To investigate whether the MUGase cleaves cello-oligosaccharides, the enzyme was incubated with appropriate substrates, and the products generated were analysed by h.p.l.c. The results (Fig. 4) showed that cellobiose, cellotriose, cellotetraose, cellopentaose and cellohexaose were all hydrolysed to glucose. In the initial stages of hydrolysis of cellohexaose, cellopentaose and, to a lesser extent, cellotetraose, were the primary products (Fig. 4). These data indicate that the MUGase successively removes glucose units from, presumably, the non-reducing end of the oligosaccharides. The K_m values for cellobiose and cellohexaose hydrolysis were 1.2 mM and 28 μ M respectively. Since the relative affinity for substrate increases (K_m decreases) as the degree of polymerization increases from 2 to 6 (i.e. cellobiose to cellohexaose), and since the enzyme acts to remove glucose residues from the non-reducing ends of such substrates, CELD may be classified as a 1,4- β -D-glucan glucohydrolase (EC 3.2.1.74) rather than as a β -glucosidase (EC 3.2.1.21).

The substrate specificity of the MUGase indicates that it is a

Fig. 2. SDS/PAGE of cell-envelope proteins from *E. coli* strains harbouring recombinant plasmids containing *celD*

Proteins present in cell-envelope fractions from recombinant *E. coli* strains were subjected to SDS/PAGE using a 10% (w/v) polyacrylamide gel. The lanes contained proteins derived from *E. coli* JM83 containing pUC18 (2); pMC4 (3); pMC4/pNM52 with the *E. coli* strain cultured with (4) and without (5) IPTG; pMC8/pNM52 with the *E. coli* strain cultured with (6) and without (7) IPTG. The M_r value for the prominent polypeptide, absent from *E. coli* containing pUC18, is shown. SDS/PAGE high- (lanes 1 and 8) and low- (lane 9) M_r standards were supplied by Sigma. Proteins were stained with Coomassie Blue.

1,4- β -D-glucan glucohydrolase which we have designated, 1,4- β -D-glucan glucohydrolase (CELD).

Gene copy number of *celD*

Previous studies have shown that the 5' regions of *P. fluorescens* subsp. *cellulosa* xylanase genes, *xynB* and *xynC*, are reiterated elsewhere in the organism's genome (Kellett *et al.*, 1990). In contrast, other *cel* and *xyn* genes are present as single copies in

Table 2. Localization of CELD in *E. coli* harbouring pMC4 or pMC9 and *P. fluorescens* subsp. *cellulosa* cultured with CMC

Fraction	Bacterium	Plasmid	CELD (%)	Malate dehydrogenase (%)	NADH oxidoreductase (%)
Envelope	<i>E. coli</i>	pMC4	91.0	1.2	91.2
Soluble	<i>E. coli</i>	pMC4	9.0	98.8	8.7
Envelope	<i>E. coli</i>	pMC9	15.3	0.5	81.2
Soluble	<i>E. coli</i>	pMC9	84.7	99.5	18.7
Envelope	<i>P. fluorescens</i> (CMC)	—	39.9	3.2	70.0
Soluble	<i>P. fluorescens</i> (CMC)	—	60.1	96.8	30.0

Fig. 3. Purification of CELD

Lane 1, SDS/PAGE, using a 10% (w/v) polyacrylamide gel, of proteins present in cell-free extract from *E. coli* containing pMC4; lane 2, proteins released by trypsin/EDTA treatment of the cell-envelope fraction; lane 3, polypeptide present in the MUGase fraction eluted from a phenyl-Sepharose column. The M_r of purified CELD is shown. Proteins were stained with Coomassie Blue.

Table 3. Substrate specificity of CELD expressed by *E. coli* harbouring pMC4

Substrate	Specific activity* (units/mg of protein)
Acid-swollen cellulose	0
Avicel	0
Barley β -glucan	0
Carboxymethylcellulose †	0
Cotton	0
Filter paper	0
Laminarin	0
Lichenan	0
MUC ‡	3.35×10^{-3}
MUG	3.98×10^{-3}
<i>p</i> -Nitrophenyl β -D-cellobioside	5.99×10^{-4}
<i>p</i> -Nitrophenyl β -D-glucoside	6.92×10^{-4}
Xylan	0

* Specific activity is the value for sonicated cells.

† Medium viscosity.

‡ Abbreviation: MUC, 4-methylumbelliferyl β -D-cellobioside.

Southern hybridization using pMC4 as the probe. The data (Fig. 5) were consistent with the insert in pMC4 being a contiguous sequence of *P. fluorescens* subsp. *cellulosa* DNA that is not reiterated elsewhere in the organism's genome. Thus *celD* is a single-copy gene.

Expression of *celD* in *P. fluorescens* subsp. *cellulosa*

To verify that *celD* is not a pseudogene, but is expressed in its endogenous host, proteins synthesized by *P. fluorescens* subsp. *cellulosa* grown with various carbon sources, were subjected to Western analysis using antiserum raised against purified CELD as a probe. The results presented in Fig. 6 showed that a protein immunoreactive with anti-CELD antiserum was present in both cell-envelope and cytoplasmic fractions from *P. fluorescens* subsp. *cellulosa* and had an M_r of 85000. Total MUGase activity and putative CELD were not subject to catabolite repression, but were constitutively expressed (Table 4 and Fig. 6).

Nucleotide sequence of *celD*

The 4.9 kb DNA region between the *Kpn*I (map position 0.9) and *Hind*III (map position 5.8) restriction sites of pMC4, which contained *celD*, was sequenced in both strands. Translation of the nucleotide sequence (Fig. 7) revealed a single open reading frame (ORF) of 2607 bp encoding a polypeptide of M_r 92000. The *N*-terminal 16 residues of purified recombinant CELD exhibited near identity with residues 52–67 of the translated sequence. The proposed translation initiation codon of *celD* is supported by the following observations: (i) upstream (7 bp) of the putative translation initiation codon is the sequence AGGT which is similar to the ribosome-binding sequence of Gram-negative bacteria. No other possible translational start codons (ATG, GTG, TTG) in this region are preceded by sequences which resemble ribosome-binding sequences; (ii) the M_r of CELD expressed by *P. fluorescens* subsp. *cellulosa* and by *E. coli* (85000) is similar to the predicted M_r of CELD.

Immediately downstream from the translational stop codon of *celD* is a G+C-rich region of dyad symmetry [$\Delta G = 179.5$ J (19 kcal)] that is followed by a succession of thymine residues and is preceded by a run of adenine nucleotides. This region is, therefore, characteristic of rho-independent transcription terminators and could, potentially, function as such in either orientation. Codon utilization of *celD* was similar to other *cel* and *xyn* genes from *P. fluorescens* subsp. *cellulosa* (Kellett *et al.*, 1990; Gilbert *et al.*, 1990) with a lack of marked bias towards G+C nucleotides in the wobble position (A+T of 36% in the wobble position) compared with genes from other pseudomonads (Minton *et al.*, 1983). This could reflect the recent transfer of genes encoding plant-cell-wall hydrolases to the organism.

The *N*-terminal region of the deduced primary structure of full-length CELD displays many of the features characteristic of a signal peptide. The hydrophilic *N*-terminus contains three basic residues and is followed by 15 small hydrophobic amino acids

the *Pseudomonas* chromosome (Gilbert *et al.*, 1987, 1988). To investigate whether *celD* was present as a single copy, genomic DNA from *P. fluorescens* subsp. *cellulosa* was subjected to

Fig. 4. H.p.l.c. analysis of sugars released from cello-oligosaccharides by purified CELD

The retention times of the sugar standards glucose (G1), cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose (G5) and cellohexaose (G6) were 2.25 min, 3.34 min, 4.75 min, 6.4 min, 8.1 min and 9.75 min respectively (6A). Sugar released after incubating CELD (6.13×10^{-3} units) for 1 h with G2 (1A), G3 (2A), G4 (3A), G5 (4A) and G6 (5A) are shown. The products generated after 6 min and 60 min in CELD (6.13×10^{-4} units)/G6 reaction mixtures are depicted in 1B and 2B respectively.

which would be capable of forming an α -helix. Signal peptides of similar structure are found at the *N*-terminus of other extracellular cellulases and xylanases synthesized by *P. fluorescens* subsp. *cellulosa*. However, CELD is not secreted into the culture medium (results not shown), despite containing an apparent signal peptide. The primary structure of CELD immediately downstream of the putative signal peptide consists of a small hydroxyamino-acid-rich region in which 70% of the 16 residues are either serine or threonine. Such sequences, frequently

Fig. 5. Hybridization of pMC4 with *P. fluorescens* subsp. *cellulosa* genomic DNA

The DNA probe was hybridized with *P. fluorescens* subsp. *cellulosa* DNA which had been restricted with *Bam*HI (1), *Eco*RI (2), *Hind*III (3) and *Pst*I (4). The filters were autoradiographed on X-ray film at -70°C for 6 h.

Fig. 6. Production of CELD by *P. fluorescens* subsp. *cellulosa* cultured with different carbon sources

Cell-envelope and cytoplasmic proteins derived from *P. fluorescens* subsp. *cellulosa* were fractionated by SDS/PAGE, using a 10% (w/v) polyacrylamide gel, and were analysed by Western blotting using anti-CELD antiserum as the primary antibody. Lanes 1, 3 and 5 contained cytoplasmic, and lanes 2, 4 and 6 envelope-associated proteins from *P. fluorescens* subsp. *cellulosa* grown in media containing glucose (1 and 2), cellobiose (3 and 4) and carboxymethyl cellulose (5 and 6) as sole carbon sources. The M_r of the major immunoreactive polypeptide is shown.

Table 4. MUGase activity expressed by *P. fluorescens* subsp. *cellulosa* grown in the presence of different carbon sources

Carbon source	MUGase activity* (units/mg of protein)
Glucose	4.5×10^{-2}
Cellobiose	3.6×10^{-2}
Carboxymethylcellulose	6.2×10^{-2}

* Assays were conducted with unfractionated sonicated cells.

Fig. 7. Nucleotide sequence of *celD*.

The derived amino acid sequence of CELD is shown. The horizontal arrow shows the experimentally determined N-terminal sequence of purified recombinant CELD. The small hydroxyamino acid-rich region is boxed. The 5' nucleotide of each truncated derivative of *celD* is indicated by a vertical arrow, and the designation of the plasmid encoding the derivative is also shown.

CTB 170	V	D	E	R	T	L	R	E	I	Y	F	A	S	F	E	N	A	V	K	K
CELD 281	G	D	E	K	V	L	R	D	I	H	G	A	G	Y	F	S	A	I	E	S
CTB	A	R	P	W	V	V	M	C	A	Y	N	K	L	N	G	E	Y	C	S	E
CELD	G	V	Q	V	V	M	A	S	P	T	S	W	E	G	T	R	M	H	G	
CTB	N	R	Y	L	L	T	E	V	L	K	N	E	W	M	H	D	G	F	V	V
CELD	H	K	Y	L	L	T	D	V	L	K	D	R	M	G	F	D	G	L	V	V
CTB	S	D	W									G	A	V	N	D	R	V	S	G
CELD	G	D	W	S	G	H	S	F	I	P	G	C	T	A	L	N	C	P	Q	S
CTB	L	D	A	G	L	D	L	E	M	P	T	S	H	G	I	T	D	K		
CELD	L	M	A	G	L	D	I	Y	M	V	P	E	P	D	W	K	E	L	Y	K
CTB	K	I	V	E	A	V	K	S	G	K	L	S	E	N	I	L	N	R	A	V
CELD	N	L	L	A	Q	A	K	T	G	E	L	P	M	A	R	V	D	D	A	V
CTB	E	R	I	L	K	V	I	I	M	A	L	E	N	K	K	E	N	A		
CELD	R	I	L	R	V	K	I	R	A	G	L	F	E	K	G	A	P	S	T	
CTB	Q	Y	E	Q			D		A	H		H	R			L	A	R		
CELD	R	P	L	A	G	K	K	D	V	L	G	A	P	E	H	R	E	V	A	R
CTB	Q	A	A	A	E	S	M	V	L	L	K	N	E	D	D	V	L	P	L	K
CELD	Q	A	V	R	E	S	L	V	L	L	K	N	K	N	N	L	L	P	L	A
CTB	K	S	G	T	I	A	L	I												337
CELD	R	Q	Q	T	V		L	V												465

Fig. 8. Alignment of similar or identical regions of CELD and β -glucosidase B from *C. thermocellum* (CTB)

Residues which show identity or similarity of structure are boxed. The positions of the first and last residues of the similar or identical regions, in their respective primary sequence, are shown.

observed in cellulases from *Pseudomonas* and other micro-organisms (Hall & Gilbert, 1988; Gilbert *et al.*, 1990; Ong *et al.*, 1989), constitute the 'hinge regions' which link together the discrete domains found in these multidomain enzymes. Interestingly CELD does not contain the highly conserved 100-amino-acid CBD present in other *P. fluorescens* subsp. *cellulosa* plant-cell-wall hydrolases (Kellett *et al.*, 1990; Gilbert *et al.*, 1990; Ferreira *et al.*, 1990). Furthermore, the conserved 40-amino-acid sequence, located between the serine-rich regions of cellulases and xylanases from the pseudomonad, was also absent from CELD. The significance of this remains to be elucidated.

Sequence comparisons

The comparison of sequences revealed that residues 281–465 of CELD exhibited 34% sequence identity with residues 170–337 of *Clostridium thermocellum* β -glucosidase B (Fig. 8). If amino acid replacement by conservative substitution is taken into account, overall identity is increased to 44%. No significant sequence identity was observed between CELD and other microbial cellulases and xylanases.

Structure of CELD

Previous studies have shown that conserved sequences in the *N*- and *C*-terminal domains of *P. fluorescens* subsp. *cellulosa* cellulases and xylanases do not appear to be essential for catalytic activity when using purified substrates (Hall & Gibert, 1988; Hall *et al.*, 1989; Gilbert *et al.*, 1990). To investigate whether the *N*-terminal region of CELD is required for the enzyme to retain catalytic activity, BAL31 and PCR were used to generate 5'

deletions of *celD*. The resultant derivatives of the gene were cloned in-phase with *lacZ'* of pUC19 (Figs. 1 and 7; Table 2). Removal of up to 141 bp of 5' sequence encoding the putative signal peptide and hydroxyamino-acid-rich region, resulted in a truncated form of *celD* which directed the synthesis of catalytically active CELD (CELD'). Cell-fractionation studies revealed that CELD' was located exclusively in the cytoplasmic fraction of *E. coli* (Table 2), suggesting that the signal peptide plays a direct role in the attachment of CELD to the cell envelope.

The 3' 148 bp of *celD* were deleted by cloning the 4.5 kb *EcoRI* restriction fragment into pUC18. The resultant plasmid pMC2 (Fig. 1), did not direct the synthesis of a functional MUGase.

DISCUSSION

The present paper describes the cloning and sequencing of *celD*, and the characterization of its encoded product, CELD. The selection strategy for isolating 1,4- β -D-glucan glucohydrolase and β -glucosidase genes was based on the capacity of recombinants to hydrolyse MUG. Only one clone expressing 1,4- β -D-glucan glucohydrolase was isolated. Although this indicates that *P. fluorescens* subsp. *cellulosa* only synthesizes a single exoglucanase, other interpretations of the data are possible. Thus: (i) other cellobiose genes may be present in the pseudomonad but not functional in *E. coli*; (ii) the number of recombinants screened may not have been sufficient to encompass the whole *P. fluorescens* subsp. *cellulosa* genome; (iii) other β -glucosidases or 1,4- β -D-glucan glucohydrolases derived from the pseudomonad, may be lethal when synthesized by *E. coli*. Support for this latter notion is provided by the fact that *xynA* (Hall *et al.*, 1989) and other *P. fluorescens* subsp. *cellulosa* *cel* and *xyn* genes (Ferreira *et al.*, 1991) can cause cell death when highly expressed in the enteric bacterium.

The recombinant 1,4- β -D-glucan glucohydrolase described in the present study contains a signal peptide similar to that of extracellular cellulases and xylanases expressed by *P. fluorescens* subsp. *cellulosa*. Although CELD is not exported into the culture medium when synthesized by the pseudomonad, the presence of an *N*-terminal signal peptide indicates that the enzyme is normally secreted out of the cytoplasm. In *E. coli* the 1,4- β -D-glucan glucohydrolase is clearly associated with the envelope fraction. The cellular location of CELD in its endogenous host, however, is less certain. Although CELD seemed to fractionate with NADH oxidoreductase in the pseudomonad, there was considerable contamination of the soluble fraction with the membrane marker enzyme. Furthermore, attempts to isolate periplasmic proteins from *P. fluorescens* subsp. *cellulosa* using either osmotic-shock or lysozyme treatment resulted in extensive cell lysis (results not shown) and thus it was not possible to determine whether CELD was exported to the periplasm. Nevertheless, it does appear that some, if not all, of CELD is associated with the cell envelope of *P. fluorescens* subsp. *cellulosa*.

Deletion studies indicate that CELD consists of two discrete domains; a hydrophobic *N*-terminal region, responsible for association with the cell envelope, and a large catalytic domain, linked to the former by a short sequence rich in hydroxyamino acids. The molecular architecture of the bacterial 1,4- β -D-glucan glucohydrolase is similar to that of sucrase-isomaltase, which associates with the brush-border membrane of intestinal enterocytes (Semenza, 1986). The eukaryotic enzyme also contains an *N*-terminal membrane anchor which is linked to an extensive catalytic domain by a sequence rich in hydroxyamino acids.

All *P. fluorescens* subsp. *cellulosa* plant-cell-wall hydrolases that we have characterized to date contain a CBD which is discrete from the catalytic domain (Hall & Gilbert, 1988; Ferreira

Fig. 9. Alignment of CELD sequences with the *A. wentii* β -glucosidase A₃ sequence in the vicinity of the aspartate active-site residue

The enzymes compared were as follows: *P. fluorescens* subsp. *cellulosa* 1,4- β -D-glucan glucohydrolase D (CELD), *C. thermocellum* β -glucosidase B (CTB; Gräbnitz *et al.*, 1989), *Candida pelliculosa* β -glucosidase (CPE; Kohchi & Toh-e, 1985), *Kluyveromyces fragilis* β -glucosidase (KFG; Raynal *et al.*, 1987) and *A. wentii* β -glucosidase A₃ (AAW; Bause & Legler, 1980). Boxes indicate regions of identity or similarity. The arrow indicates the aspartate residue located at the active site of AAW.

et al., 1990; Gilbert *et al.*, 1990; Kellett *et al.*, 1990). Here we have shown that CELD does not contain a sequence which resembles that of a CBD. Since the enzyme is intracellular and would not normally come into contact with the structural polysaccharides of plant cell walls, there is no selection pressure for the enzyme to acquire a CBD distinct from the active site.

CELD exhibits some sequence identity with β -glucosidase B from *C. thermocellum*. Recently, the phylogeny of cellulases and xylanases has been evaluated using hydrophobic cluster analysis. According to this classification CELD is a member of the subfamily BGB which contains β -glucosidases from *C. thermocellum*, *Kluyveromyces fragilis*, *Candida pelliculosa* and *Schizophyllum commune* (Béguin, 1990; Henrissat *et al.*, 1989). It is likely, therefore, that CELD and other genes from group BGB have evolved from a common ancestral gene. There is no evidence to suggest that *celD* is related to other plant-cell-wall hydrolases from its endogenous host.

Legler *et al.*, (1979), using an active-site affinity label, have shown that an aspartic acid residue is located in the active site of *Aspergillus wentii* β -glucosidase A₃. The structure of CELD and other subfamily-BGB enzymes exhibit considerable sequence identity with the *A. wentii* β -glucosidase in the vicinity of the active-site acidic residue (Fig. 9). Thus it is possible that aspartic acid-341 and proximal residues could constitute the active site of CELD. Further studies using active site ligands and *in vitro* site-directed mutagenesis are needed to verify whether aspartic acid-341 is essential for catalytic activity.

We thank the Ministry of Agriculture, Fisheries and Food, ICI and the Agricultural and Food Research Council (grant LRG 138) for supporting this work.

REFERENCES

Bause, E. & Legler, G. (1980) *Biochim. Biophys. Acta* **626**, 459–465

- Béguin, P. (1990) *Annu. Rev. Microbiol.* **44**, 219–248
 Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523
 Chambers, S. P., Prior, S. E., Barstow, D. A. & Minton, N. P. (1988) *Gene* **68**, 139–149
 Enari, T.-M. (1983) in *Microbial Enzymes and Biotechnology* (Fogarty, W. M., ed.), pp. 183–223, Applied Science Publishers, London and New York
 Ferreira, L. M. A., Durrant, A. J., Hall, J., Hazlewood, G. P. & Gilbert, H. J. (1990) *Biochem. J.* **269**, 261–264
 Ferreira, L. M. A., Hazlewood, G. P., Barker, P. J. & Gilbert, H. J. (1991) *Biochem. J.* **279**, 793–799
 Gilbert, H. J., Blazek, R., Bullman, H. M. S. & Minton, N. P. (1986) *J. Gen. Microbiol.* **132**, 151–160
 Gilbert, H. J., Jenkins, G., Sullivan, D. A. & Hall, J. (1987) *Mol. Gen. Genet.* **210**, 551–556
 Gilbert, H. J., Sullivan, D. A., Jenkins, G., Kellett, L. E., Minton, N. P. & Hall, J. (1988) *J. Gen. Microbiol.* **134**, 3239–3247
 Gilbert, H. J., Hall, J., Hazlewood, G. P. & Ferreira, L. M. A. (1990) *Mol. Microbiol.* **4**, 759–767
 Gräbnitz, F., Rücknagel, K. P., Seiss, M. & Staudenbauer, W. L. (1989) *Mol. Gen. Genet.* **217**, 70–76
 Hall, J. & Gilbert, H. J. (1988) *Mol. Gen. Genet.* **213**, 112–117
 Hall, J., Hazlewood, G. P., Barker, P. J. & Gilbert, H. J. (1988) *Gene* **69**, 29–38
 Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J. & Gilbert, H. J. (1989) *Mol. Microbiol.* **3**, 1211–1219
 Hazlewood, G. P., Davidson, K., Clarke, J. H., Durrant, A. J., Hall, J. & Gilbert, H. J. (1990a) *Enzyme Microb. Technol.* **12**, 656–662
 Hazlewood, G. P., Davidson, K., Laurie, J. I., Romaniec, M. P. M. & Gilbert, H. J. (1990b) *J. Gen. Microbiol.* **136**, 2089–2097
 Henrissat, B., Claeysens, M., Tomme, P., Lemesle, L. & Mornon, J.-P. (1989) *Gene* **81**, 83–95
 Hunkapillar, M. W., Hewick, R. E., Dreyer, W. J. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 399–413
 Kellett, L. E., Poole, D. M., Ferreira, L. M. A., Durrant, A. J., Hazlewood, G. P. & Gilbert, H. J. (1990) *Biochem. J.* **272**, 369–376
 Kohchi, C. & Toh-e, A. (1985) *Nucleic Acids Res.* **13**, 6273–6282
 Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
 Legler, G., Roeser, K. R. & Illig, H. K. (1979) *Eur. J. Biochem.* **101**, 85–92
 Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
 Minton, N. P., Atkinson, T. & Sherwood, R. F. (1983) *J. Bacteriol.* **156**, 1222–1227
 Minton, N. P., Bullman, H. M. S., Scawen, M. D., Atkinson, T. & Gilbert, H. J. (1986) *Gene* **46**, 25–33
 Norrander, J., Kemp, T. & Messing, J. (1983) *Gene* **26**, 101–116
 Ong, E., Greenwood, J. M., Gilkes, N. R., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1989) *Trends Biotechnol.* **7**, 239–243
 Raynal, A., Gerbaud, C., Francinques, M. C. & Guérineau, M. (1987) *Curr. Genet.* **12**, 175–184
 Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178
 Semenza, G. (1986) *Annu. Rev. Cell. Biol.* **2**, 255–313
 Staden, R. (1980) *Nucleic Acids Res.* **16**, 3673–3694
 Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4767–4771
 Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350–4354
 Wakarchuk, W. W., Greenberg, N. M., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1988) *J. Bacteriol.* **170**, 301–307
 Yamane, K., Suzuki, H., Hirotsu, M., Ozawa, H. & Nisizawa, K. (1970) *J. Biochem. (Tokyo)* **67**, 9–18

Received 20 August 1991/27 January 1992; accepted 5 February 1992