A modular esterase from *Pseudomonas fluorescens* subsp. *cellulosa* contains a non-catalytic cellulose-binding domain

Lúis M. A. FERREIRA,* Thomas M. WOOD,† Gary WILLIAMSON,‡ Craig FAULDS,‡ Geoffrey P. HAZLEWOOD,§ Gary W. BLACK* and Harry J. GILBERT*||

*Department of Biological and Nutritional Sciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU

†The Rowett Research Institute, Greenburn Road, Bucksburn, Aberdeen AB2 9SB, Scotland

‡AFRC Institute of Food Research, Norwich Laboratory, Norwich Research Park, Colney Lane, Norwich NR4 7UA

§Department of Biochemistry, AFRC Institute of Animal Physiology and Genetics Research, Babraham, Cambridge CB2 4AT, U.K.

The 5' regions of genes *xynB* and *xynC*, coding for a xylanase and arabinofuranosidase respectively, are identical and are reiterated four times within the *Pseudomonas fluorescens* subsp. *cellulosa* genome. To isolate further copies of the reiterated *xynB/C* 5' region, a genomic library of *Ps. fluorescens* subsp. *cellulosa* DNA was screened with a probe constructed from the conserved region of *xynB*. DNA from one phage which hybridized to the probe, but not to sequences upstream or downstream of the reiterated *xynB/C* locus, was subcloned into pMTL22p to construct pFG1. The recombinant plasmid expressed a protein in *Escherichia coli*, designated esterase XYLD, of M_r 58 500 which bound to cellulose but not to xylan. XYLD hydrolysed aryl esters, released acetate groups from acetylxylan and liberated 4-hydroxy-3-methoxycinnamic acid from de-

starched wheat bran. The nucleotide sequence of the XYLD-encoding gene, *xynD*, revealed an open reading frame of 1752 bp which directed the synthesis of a protein of M_r 60 589. The 5' 817 bp of *xynD* and the amino acid sequence between residues 37 and 311 of XYLD were almost identical with the corresponding regions of *xynB* and *xynC* and their encoded proteins XYLB and XYLC. Truncated derivatives of XYLD lacking the N-terminal conserved sequence retained the capacity to hydrolyse ester linkages, but did not bind cellulose. Expression of truncated derivatives of *xynD*, comprising the 5' 817 bp sequence, encoded a non-catalytic polypeptide that bound cellulose. These data indicate that XYLD has a modular structure comprising of a N-terminal cellulose-binding domain and a C-terminal catalytic domain.

INTRODUCTION

Accessory enzymes that cleave bonds other than the β -1,4-glycosidic linkages of the xylan backbone are an essential part of the consortium of enzymes that hydrolyse the hemicellulose fraction of plant polysaccharides. They include phenolic acid esterases, acetyl esterases (EC 3.1.1.6), α -glucuronidases and α -L-arabinofuranosidases (EC 3.2.1.55). The role of these enzymes in saccharification is to cleave the esterified phenolic groups which mediate cross-linking and the attachment of xylan to lignin, and to promote unrestricted access of the depolymerizing enzyme, endo-1,4- β -D-xylanase (EC 3.2.1.8), to the xylan backbone by removing acetyl, glucuronosyl and arabinosyl residues. A number of the accessory enzymes have been purified (see, for example, Poutanen et al., 1991; Faulds and Williamson, 1991; Borneman et al., 1992), and their synergistic action with endoxylanase is well established (Biely et al., 1986; Poutanen and Sundberg, 1988; Poutanen et al., 1991), but, with few exceptions, little is known about their molecular architecture. As with other components of microbial xylanolytic systems, biochemical analysis of the accessory enzymes has been hampered by the difficulty of purifying discrete forms of these enzymes, free of other plant-cell-wall-hydrolysing activities. The expression of genes encoding these enzymes in a non-plant-cell-wall-degrading host, such as *Escherichia coli*, provides an attractive alternative approach for characterizing individual xylanolytic enzymes.

Numerous endoxylanase genes have been isolated, and the primary structures of the encoded proteins have been deduced (Gilkes et al., 1991; Hazlewood and Gilbert, 1992). The enzymes fall into two distinct families, suggesting that the xylanases evolved from two ancestral sequences (Gilkes et al., 1991; Hazlewood and Gilbert, 1992). Xylanases from prokaryotic and eukaryotic micro-organisms are found in each of the enzyme families, indicating that significant transfer of xylanase genes, between bacteria and fungi, has occurred during the course of evolution (see, for example, Gilbert et al., 1992). No convincing sequence similarity has been detected between cellulases and endoxylanases, suggesting that the two functionally related enzyme species (both cleave β -1,4-glycosidic bonds in the plant cell wall) have evolved from distinct genes.

The molecular analysis of plant-cell-wall hydrolases has shown that many cellulases, and some xylanases, consist of a catalytic domain linked by sequences rich in serine, threonine and proline residues to a non-catalytic cellulose-binding domain (CBD; Gilkes et al., 1991). There is increasing evidence indicating that CBDs play a pivotal role in the hydrolysis of insoluble cellulose by endoglucanases and cellobiohydrolases (Van Tilbeurgh et al., 1986; Tomme et al., 1988; Ong et al., 1989). The precise role of these domains in hemicellulose hydrolysis is less certain. The role of linker sequences in multidomain cellulases and xylanases is also unclear. Ferreira et al. (1990) showed that removal of the linker sequences of xylanase A from *Pseudomonas fluorescens*

Abbreviations used: CBD, cellulose-binding domain; LB, Luria broth.

|| To whom correspondence should be sent.

The nucleotide sequence of the *xynD* gene has been submitted to the EMBL, GenBank and DDBH Nucleotide Sequence Databases under the accession no. X58956.

subsp. *cellulosa* did not affect the enzyme's catalytic or cellulose-binding capacities. In contrast, deletion of the corresponding region from *Cellulomonas fimi* CenA altered the enzyme's catalytic activity (Ong et al., 1989).

Recent work in our laboratories has focused on the plant-cell-wall hydrolases of *Ps. fluorescens* subsp. *cellulosa*. Two endoglucanases, two endoxylanases, a cellodextrinase, an arabinofuranosidase and a β -glucosidase have been characterized (Hazlewood and Gilbert, 1992). Apart from the β -glucosidase, each enzyme comprises a catalytic domain linked to a CBD via serine-rich sequences. An 817 bp sequence located in the 5' region of the XYLB- and XYLC-encoding genes (*xynB* and *xynC*) is apparently reiterated four times within the pseudomonad genome (Kellett et al., 1990). Although a large number of endo- β -1,4-xylanases have been sequenced (Gilkes et al., 1991), those accessory enzymes which hydrolyse xylan side chains have received little attention. To date only one acetylxytan esterase (Lüthi et al., 1990) and two arabinofuranosidase (Kellett et al., 1990; Utt et al., 1991) genes have been described. To increase our understanding of this neglected class of enzymes, we have used the conserved 817 bp sequence of *xynB/C* as a probe to isolate genes from the *Ps. fluorescens* subsp. *cellulosa* genome that code for plant-cell-wall hydrolases other than endoglucanases and endoxylanases. Here we describe the isolation and sequencing of a gene, designated *xynD*, which codes for an esterase that binds tightly to cellulose. The role of the enzyme in plant-cell-wall hydrolysis is discussed.

MATERIALS AND METHODS

Strains, vectors and media

The *E. coli* strains employed in this study were LE392 and JM83. The vectors used were pUC18 (Norrande et al., 1983), pMTL22p and pMTL23p (Chambers et al., 1988). *E. coli* strains were cultured in Luria broth (LB) supplemented with ampicillin (100 μ g/ml) and 5-bromo-4-chloroindol-3-yl β -D-galactoside (2 μ g/ml), to select for transformants and recombinants respectively. Bacteriophage were plated out on top agar comprising of LB supplemented with 10 mM $MgSO_4$ and 0.7% agar.

Enzyme and protein assays

Cell-free extracts from *E. coli* clones were prepared by sonication. Xylanase- and cellulase-related activities were assayed as described by Hazlewood et al. (1990) and Kellett et al. (1990). The specificity of XYLD arylesterase activity was determined by the method of Hespell and O'Bryan-Shah (1988) with the 1-naphthyl esters of acetic, butyric, octanoic and dodecanoic acids as substrates. Acylesterase activity was routinely determined using 4-nitrophenyl acetate (1 mM) as substrate in a 1 ml reaction volume, containing 100 mM Mops, pH 6.0. The reaction mixture was incubated at 37 °C for appropriate lengths of time and the release of 4-nitrophenol was detected at 420 nm. The release of acetate from acetylxytan (prepared as described by Johnson et al., 1988) and of 4-hydroxy-3-methoxycinnamic acid from destarched wheat bran, were detected using the Boehringer acetate kit and h.p.l.c. (Faulds and Williamson, 1991) respectively. The methyl esters of 4-hydroxycinnamate, 3,5-dimethoxy-4-hydroxycinnamate and 3,4-dihydroxycinnamate were prepared by the method of Borneman et al. (1990). The esters (1 mM) were prepared in 100 mM Mops buffer, pH 6.0, and the release of the free acid due to enzyme hydrolysis was measured as described by Faulds and Williamson (1991). Activity is expressed as μ mol of

acid released/min per mg of protein at pH 6.0 and 37 °C. The free acids of 4-hydroxy-3-methoxycinnamate, 4-hydroxycinnamate, 3,5-dimethoxy-4-hydroxycinnamate and 3,4-dihydroxycinnamate were used as reference standards for the h.p.l.c. Protein was determined by the method of Lowry et al. (1951), with BSA as standard. The capacity of protein, present in cell-free extracts of *E. coli* strains harbouring full-length or truncated derivatives of *xynD*, to bind cellulose was evaluated as follows: cell-free extracts (2 ml derived from 100 ml of culture) were mixed with an equal volume of crystalline cellulose (Avicel PH105; 5%, w/v) suspended in 100 mM Tris/HCl buffer, pH 7.5. After washing the cellulose five times with the Tris/HCl buffer, bound protein was either assayed for esterase activity using 4-nitrophenyl acetate as substrate, or was eluted with 10% (w/v) SDS by boiling for 5 min. Protein eluted from Avicel was subjected to SDS/PAGE using a 10% (w/v) polyacrylamide gel (Laemmli, 1970).

Purification of XYLD and determination of N-terminal sequence

E. coli JM83 harbouring pFG1 was cultured for 16 h in LB broth (up to 6 litres) containing ampicillin (100 μ g/ml). A cell-free extract was prepared by sonicating the harvested cells and recovering the soluble fraction. $(NH_4)_2SO_4$ (0.243 g/ml) was added to precipitate acetyl esterase activity, and the pellet obtained by centrifuging was dissolved in 10 mM Tris/HCl buffer, pH 8.0, and dialysed at 4 °C against three changes of the same buffer. Acetyl esterase activity was purified to apparent homogeneity by anion-exchange chromatography on a column of DEAE-Trisacryl M (Rixon et al., 1992). Active fractions, eluted by a NaCl gradient extending from 0 to 400 mM, contained a single protein with an M_r of 58500 that bound quantitatively to Avicel (Figure 1).

The first 15 N-terminal residues of recombinant XYLD were determined, by automated Edman sequencing, on enzyme that had been purified from the cell-free extract by allowing it to adsorb on to crystalline cellulose as described above. Cellulose-bound XYLD, released by boiling with SDS (10% w/v), was applied directly to a 470A gas-phase sequencer equipped with a 120A on-line phenylthiohydantoin-derivative analyser (Applied Biosystems; Hunkapillar et al., 1983).

Construction and screening of gene library

A gene library of *Ps. fluorescens* subsp. *cellulosa* DNA, constructed in λ 47.1 (Gilbert et al., 1987), was plated on to LB-top agarose (0.7%) at a density of three plaques/cm². Phage DNA from 5000 plaques was transferred to nitrocellulose, denatured, neutralized and then hybridized with a radiolabelled probe comprising the 5' region of *xynB* as described by Sambrook et al. (1989). Positive plaques were rescreened with other sequences from within the *xynB/C* locus. Those phages which hybridized with only the 5' region of *xynB* were selected for further studies.

Recombinant DNA techniques

Agarose-gel electrophoresis, Southern hybridization, transformation of *E. coli*, plasmid DNA preparation and the use of restriction enzymes and T4 DNA ligase were as described by Gilbert et al. (1987). To determine the nucleotide sequence of *xynD*, the pseudomonad insert in pFG1 was subjected to BAL-31 digestion as described by Hall et al. (1988), and the truncated derivatives excised and cloned into suitably restricted pMTL22p and pMTL23p. The recombinant plasmids generated were pre-

Figure 1 SDS/PAGE of purified XYLD

XYLD was purified as described in the Materials and methods section. The proteins present in each fraction were subjected to SDS/PAGE on 10%-(w/v)-polyacrylamide gels. (a) Lane 1, Sigma high- M_r markers; lane 2, *E. coli* cell-free extract; lane 3, purified XYLD. (b) Lane 1, SDS/PAGE of Sigma high- M_r markers; lane 2, purified XYLD; lane 3, soluble material from purified XYLD/cellulose incubation; lane 4, material released from cellulose, which had been incubated with XYLD, by boiling in 10% (w/v) SDS for 10 min.

pared on a small scale (Birnbom and Doly, 1976), denatured with alkali, neutralized and further purified by spin dialysis (Murphy and Kavanagh, 1988). Sequencing of the resultant DNA was performed using Sequenase according to the manufacturer's instructions. Sequences were compiled and ordered using the computer programs described by Staden (1980). The complete *xynD* sequence was determined in both strands.

RESULTS

Isolation of *xynD*

The *Ps. fluorescens* subsp. *cellulosa* genome contains four regions that exhibit extensive sequence similarity. Previous studies showed that two of these sequences constituted the 5' 817 bp of

Figure 2 Physical map of the *xynB/xynC* locus

The restriction-endonuclease cleavage sites were as follows: Ba, *Bam*HI; P, *Pst*I; S, *Sma*I; Ev, *Eco*RV; E, *Eco*RI; C, *Cla*I; and H, *Hpa*I. The solid arrows show the orientation and location of *xynB* and *xynC*. Regions encompassed by probes A, B and C are indicated. The hatched boxes represent the conserved 5' regions of *xynB* and *xynC*.

two closely-linked genes designated *xynB* and *xynC*, which code for an endo-1,4- β -xylanase and an arabinofuranosidase respectively (Gilbert et al., 1988; Kellett et al., 1990). To isolate additional DNA sequences which exhibit identity with the 5' region of *xynB/C*, a *Ps. fluorescens* subsp. *cellulosa* gene library, constructed in λ 47.1, was screened with probe B which consisted of the 5' region of *xynB* (Figure 2). Clones which hybridized to probe B, 20 in all, were rescreened with probes consisting of DNA sequences upstream (probe A) and downstream (probe C) of *xynB* and *xynC* respectively (Figure 2). Phage which hybridized to probe B, but not to probes A and C, were assumed to contain a region of the pseudomonad genome which, although exhibiting sequence similarity to the 5' region of the two xylanase genes, did not constitute the *xynB/xynC* locus. Phage DNA that hybridized only to probe B was digested with several restriction enzymes and subjected to Southern hybridization using the 5' region of *xynB* (probe B) as the probe. A 10 kb *Eco*RI restriction fragment, which hybridized to probe B, was isolated from one of the phages and cloned into the *Eco*RI site of pUC18 to generate the recombinant plasmid pFG1. A restriction map of the pseudomonad insert in pFG1 is shown in Figure 3.

As pFG1 contains DNA that exhibits sequence identity with the CBD-encoding region of *xynB*, the capacity of polypeptides expressed by *E. coli* harbouring pFG1 to bind cellulose was assessed. Results presented in Figure 4 show that *E. coli* cells containing the recombinant plasmid synthesize a 58.5 kDa protein that binds Avicel. The N-terminal amino acid sequence of the polypeptide was ACTYITDSEWSTGFT, which is identical with the N-terminal regions of XYLB and XYLC (Kellett et al., 1990). These data indicate that pFG1 contains a gene, designated *xynD*, that encodes a protein of M_r 58 500 which binds cellulose, and exhibits sequence identity with XYLB and XYLC from the same organism.

Catalytic activity of the polypeptide encoded by *xynD*

The *xynD* encoded protein, designated XYLD, was purified as described in the Materials and methods section and its catalytic activity assessed. Although the enzyme did not hydrolyse glycosidic bonds in a range of plant structural polysaccharides or aryl β -D-glycosides, XYLD cleaved the ester linkages in 4-nitrophenyl acetate, 4-methylumbelliferyl acetate and 1-naphthyl butyrate (Table 1); failure to hydrolyse 1-naphthyl esters containing longer acyl chains was indicative of a lack of lipase activity.

In view of the cellulose-binding capacity of XYLD, its primary substrate is likely to be ester linkages contained in the plant cell

Figure 3 A restriction map of the *Ps. fluorescens* subsp. *cellulosa* insert in pFG1

The location of the cleavage sites of *ClaI* (C), *EcoRI* (E), *EcoRV* (EV), *HindIII* (H), *PstI* (P), *StuI* (St), *SmaI* (Sm) and *SphI* (Sp) are shown. The orientation and position of *xynD* is shown by an arrow. Truncated derivatives of pFG1 were generated using BAL-31, and the extent of the pseudomonad inserts in the resultant plasmids are shown. The positions of the 5' (pFG2 and pFG3) and 3' (pFG4) nucleotides of the pseudomonad inserts, within the *xynD* sequence depicted in Figure 4, are shown. Derivatives of *xynD* lacking their respective translational start codons were inserted into appropriate pMTL- and pUC-based vectors, in-phase with the plasmids' *lacZ* gene. The hatched area denotes λ DNA present in the insert in pFG1. The capacity of proteins, encoded by full-length and truncated forms of *xynD*, to cleave 4-nitrophenyl acetate and to bind to cellulose are shown.

Figure 4 Binding of XYLD to cellulose

Cell-free extracts prepared from *E. coli* JM83 harbouring pFG1 were incubated with Avicel and bound protein was eluted with 10% (w/v) SDS by boiling as described in the Material and methods section. Proteins in the various fractions were subjected to SDS/PAGE on a 10% (w/v) polyacrylamide gel. Lane 1, Sigma high- M_r markers; lane 2, cell-free extract proteins; lane 3, proteins recovered in the filtrate after binding; lane 4, final Tris/HCl Avicel wash; lane 5, proteins eluted from the cellulose with SDS. The polyacrylamide gel was stained with Coomassie Blue, and the major polypeptide bound to Avicel (M_r 58 500) is indicated by the arrow.

wall. To evaluate the activity of the enzyme against plant-cell-wall material, XYLD was incubated with acetylxylan and destarched wheat bran, and the release of organic acids was monitored. When the two substrates were pretreated with XYLD in the presence of endo- β -1,4-xylanase, acetate and 4-hydroxy-3-methoxycinnamate were released from acetylxylan and wheat bran respectively. No acetic or 4-hydroxy-3-methoxycinnamic acid was released from these substrates by XYLD in the absence of endoxylanase (Table 2). The enzyme also exhibited activity against methyl esters of cinnamic acids. These data indicate that XYLD is capable of cleaving ester linkages in the plant cell wall, although the enzyme exhibits greatest activity against 4-nitrophenyl acetate.

Table 1 Catalytic activity of the polypeptide encoded by *xynD* (XYLD)

XYLD was purified as described in the Materials and methods section. The enzyme assays were performed in 100 mM Mops buffer, pH 6.0 at 37 °C using a substrate concentration of 1 mM. A unit of activity is 1 μ mol of product liberated/min. The following substrates were not hydrolysed: carboxymethylcellulose, barley β -glucan, oat-spelt xylan, larchwood xylan, lichenan, laminarin, 4-methylumbelliferyl β -D-cellobioside, 4-methylumbelliferyl β -D-glucoside, 4-nitrophenyl β -D-cellobioside, 4-nitrophenyl β -D-glucoside and 4-nitrophenyl β -D-xyloside.

Substrate	Catalytic activity (units/mg of protein)
4-Nitrophenyl acetate	5.8
4-Methylumbelliferyl acetate	2.7
1-Naphthyl acetate	0.31
1-Naphthyl butyrate	0.58
1-Naphthyl octanoate	0.01
1-Naphthyl dodecanoate	0.01
Methyl 4-hydroxy-3-methoxycinnamate	0.86
Methyl 3,5-dimethoxy-4-hydroxycinnamate	0.33
Methyl 4-hydroxycinnamate	0.65
Methyl 3,4-dihydroxycinnamate	0.27

Table 2 Hydrolysis of acetylxylan and destarched wheat bran by XYLD

Substrate	Xylanase (units)*	Esterase activity (units/mg of protein)†
Destarched wheat bran	0	ND‡
	3	2.5
Acetylxylan	0	ND
	3	4.1

* The xylanase added to reaction mixtures was purchased from Megazyme (Australia), and was derived from *Trichoderma viride*; this preparation is electrophoretically homogeneous and contains no esterase activity.

† 4-Hydroxy-3-methoxycinnamate and acetate were released from destarched wheat bran (50 mg/ml) and acetylxylan (30 mg/ml) respectively.

‡ ND, not detected.

Nucleotide sequence of *xynD* and the deduced primary structure of XYLD

Examination of the nucleotide sequence of the pseudomonad insert in pFG1 (Figure 5) revealed a single open reading frame of

CGGTAGCAATACGATTGAGTTCCGTGCAACAGCAGCAAGGTCGAGCATTATTTACTTCGA

CAATATCGTTATTACGAGGTAATACTTCGGTTTGCCCGGCCACAGCCGGCATTTT

TATTAATACAGCACTACATAACTTTATGTGTAAGTCAGGAGTTCTAACAATGAATAATG
M N N

60
GAAATACCCACTCACTCATTATGAAATACTCAAGTACGGCTCGATGGTGCCGGCC
G K Y P L N S L H K I L K Y G S H G A G

120
TGGCTTTAATGCTTATCCACACGCGAGCGTCCGCGCGCTTGTACCTACACCATCGATA
L A L I A L S H T Q R P P A C T Y T I D

180
CGGAATGGTCCACCGGCTTACCGCAATATACCCCTCAAAAACGATACCGGTGCCGCCA
S E W S T G F T A N I T L K N D T G A A

240
TCAATAACTGGAACGCGTAATGGCAATCTCCAGCAATCGCATGACGAGCGGTGGAATG
I N N W N V N W Q Y S S N R M T S G W N

300
CCAACTTCTCCGGCACCAACCCCTACAACGCCACCAACATGAGCTGGAACCGGACATCG
A N F S G T N P Y N A T T N M S W H G S I

360
CGCCAGGACAATCCATCTCTCGGCTCCAGGCGCAAAAAATGGCAGCACCGCCGAGC
A P G Q S I S F G L Q G E K N G S T A E

420
GACCAACCGTCAACCGCGCGCTTGAACAGTGAACCAACGCTCTGTGGCTTCAGCT
R P T V T G A A C N S A T T T S S V A S S

480
CTTCAACACCCACCACTGTCATCTTGCATCCAGTGTGGCTCGGCACTGCTGTTGC
S S T P T T S S S A S S V A S A L L L

540
AAGAAGCACAAGCCGTTTCTGCGCTGTGGATGGCACCATCGATAAATACCAACCGGCT
G E A Q A G A F C R V D G T I D N N H T G

600
TTACCGGCACTGCTTTGCAACCAACAATGCGCAGGCGGCGAGTGTGGGCGA
F T G S G F A N T N N A Q G A A V V W A

660
TAGATGCTACCAAGTGGCGTCCGCACTGATATCGCTATGCAATGTGGAAACCG
I D A T S S G R R T L T I R Y A N G G T

720
CCAAATCGCAATGGCTCACTGTTGATTAACGGGGGCGAGCAACGTAATACCGTGAAT
A N R N G S L V I N G G S N G N Y T V S

780
TGCCACGACCGCGCTGGACACCTGCGCAACCGCAATATCGATGTGGATTGGTAC
L P T T G A T T G A T I D V D L V

840
AGGGCAATAATATTGTGCACTGTGCAACAACAGCGAAGGCTTACCAATATAGATT
Q G N N I V Q L S A T T A E G L P N I D

900
CGTTAAGTGTGTGGGTGGTACCGTCAAGGCGAGTAAATGCGGCGAGTGTGAGCAGCA
S L S V V G G T V R A G N C G S V S S S

960
GTTCCGTGCAATCGTCACTGTCATCGAAGCAGTGCAGTGCAGCAGTACCCAGTGCAG
S S V Q S S S S S S S A G A A V P S A

1020
GTCTGGGTAAACCGCGCACACTGCAAAATGGTACCGTCAATATTCAAAGTGGCGAACAC
G L G K P R T L Q N G T V N I Q S G G T

1080
AACGTTCTTACATTCGAGCGCGCGGTAACATAAACAAGCGTTATCGTCTGG
Q R S Y I L D A P G N Y N N N K P Y R L

1140
TTATTGGTTATCACTGCGGGGAGGACACGCGCAAGATGTACATACCGGACAAACCGTTC
V I G Y H W R G G T A Q D V H T G Q T V

1200
AACCGGATGCAATGGGCACATTACCGCATGAAAAATGGCAGGTGACTCCAGATTTTGT
Q R D A W A H Y G H K K L A G D S T I F

1260
TGGCGCCACAGGGTATAGGTAACGGGTGGGCAATGGTGGCGGTGACATGTCACTGTTA
V A P Q G I G N G W G N G G G A D V T F

1320
CGGATAATATTCTTGGCAATTAAGGAGAAATGGTATCGACGAGAGCCGCGTCTTTTG
T D N I L A Q L K T E L C I D E S R V F

1380
CCAATGGATTAGCTTTCGGCGGTGGCATGTCTACGCTATGCGTGTGCGCGGCAATG
A N G F S F G G G H S Y A I A C A R A N

1440
TTTTCAAAGCGGTAGCAGTCTATTCCGCGAGGGTTATCAGCGGTTGTAGCGAGCAACA
V F K A V A V Y S A G V I S G C S G A N

1500
GCCCAATTGCTACTTCGGTGACATGGATTGGGAGATTCACTGCTGGGCTTTTCGGGTG
S P I A Y F G A H G L G D S V L G I S G

1560
GCCGAGCATGCGCATCGCTTTGTGAGAAATAACGCGTGTACCGCAATGACACCCCG
G R S M R D R F R N N A C T A H T P P

1620
AACCCCTGTGGGTAGCGGTTGACACACTTGCACTTCTTACGTTTGGCTTAACAGGCAAT
E P S V G S R L H T C T S Y V C P N R Q

1680
ACCGGTCGGTTGGTGTGCAATTTGATGGGGACATGATGCGTCACCACTGGATAATGGTG
Y P V R W C A F D G G H D A S P L D N G

1740
TGGGCCAGAAAGGCTACTCTGCTGCGGATGAGGCGCTGGAAGTTTTCACAGATT
V G Q K G Y T W L A D E A W K F F T Q F

1800
AATACAGCTGTGCTTATCTTGGAGTTCTGCAAGTAACAACCTCAATTGAAATAACGGTT

1752 bp encoding a polypeptide of M_r 60589. The open reading frame contained 53% G+C nucleotides and had a codon utilization similar to other *Ps. fluorescens* subsp. *cellulosa* plant-cell-wall hydrolases (Ferreira et al., 1991). The codon utilization and relatively low G+C content was atypical of *Ps. fluorescens*, suggesting that *Ps. fluorescens* subsp. *cellulosa* has been incorrectly classified, or alternatively that the organism has only recently acquired the genes coding for plant-cell-wall hydrolases. The proposed translational start codon of the open reading frame was preceded (8 bp) by the sequence GGAGT, which is similar to ribosome-binding sequences of Gram-negative bacteria (Hall and Gilbert, 1988). Residues 37–51 of the translated sequence exhibited a perfect match with the N-terminal 15 residues of mature XYLD. These data indicate that the 1752 bp open reading frame constitutes *xynD*, and the encoded protein contains an N-terminal signal peptide of 36 residues which is cleaved in *E. coli*. The leader sequence contains a stretch of 14 small hydrophobic residues preceded by a 16-amino-acid sequence containing three basic residues. The signal peptide of XYLD is similar to the secretion signals of other *Ps. fluorescens* subsp. *cellulosa* plant-cell-wall hydrolases (Kellett et al., 1990; Gilbert et al., 1990; Ferreira et al., 1991; Hall et al., 1989).

Inspection of the nucleotide sequence of *xynD* revealed that residues 112–928 exhibited almost complete identity (only one mismatch in the 817 bp sequence) with the corresponding regions in *xynB* and *xynC*. Similarly the primary sequences of the regions of XYLB, C and D encoded by the conserved 817 bp sequence were identical. In a previous study, Kellett et al. (1990) showed that the N-terminal 100 residues of the conserved domain constituted a CBD. It would appear likely, therefore, that the corresponding region in XYLD also mediates attachment of the esterase to cellulose. To assess whether the conserved N-terminal region of XYLD is essential for catalytic activity, the 3' region of *xynD* was fused in-frame to the *lacZ'* translational start codon supplied by the vector. *E. coli* cells containing the resultant plasmid expressed elevated esterase activity (Figure 3). A truncated derivative of *xynD* lacking the 5' 1152 bp sequence did not encode a functional esterase (Figure 3). Expression of the 5' region of the esterase gene, encoding the conserved domain, generated a polypeptide that bound Avicel but did not exhibit esterase activity (Figure 3). These data indicate that the catalytic domain of XYLD is located in the C-terminal portion of the enzyme, and the conserved N-terminal 273-residue domain, as in XYLB and C, contains a non-catalytic CBD.

Database searches (SWISSPROT) did not reveal any proteins which exhibited sequence similarity to the catalytic domain of XYLD. However, the N-terminal 100 residues, in addition to exhibiting identity with XYLB and C, also showed strong sequence similarity to other *Pseudomonas* and *Cellulomonas* CBDs.

DISCUSSION

Results presented here describe the cloning and sequencing of *xynD* and the characterization of its encoded product, XYLD. The 5' region of *xynD* and the N-terminal domain of XYLD

Figure 5 Nucleotide sequence of *xynD* and the deduced primary structure of XYLD

The *xynD* gene was sequenced as described in the Materials and methods section. The experimentally determined N-terminal sequence of the esterase is indicated by a broken line. The region of *xynD* and XYLD that exhibits identity with *xynB/xynC* and XYLB/XYLC respectively is boxed.

exhibit almost complete sequence similarity with the corresponding regions of *xynB/C* and *XYLB/C* respectively. The mechanism for the evolution of this reiterated sequence in the three genes is currently unknown. It is possible that the conserved sequence is contained within a mobile genetic element. However, the lack of repeated sequences at the extremities of the homologous DNA argues against this possibility. An alternative explanation is that the encoded CBD, when fused to plant-cell-wall hydrolases, confers a significant improvement in the enzymes' capacity to cleave their substrates. Thus there is a strong selection pressure for the dissemination of the 817 bp sequence among the plant-cell-wall-hydrolase genes contained in the pseudomonad genome. A similar argument has been proposed to explain the presence of CBDs in *Ps. fluorescens* subsp. *cellulosa* XYLA, CELA, CELB and CELC (Ferreira et al., 1991). However, these CBDs, although clearly related (37% of residues are conserved in all seven CBDs) both to each other and to the XYLB, XYLC and XYLD CBDs, are not identical. It is, however, difficult to reconcile the view that there is such an extreme selection pressure on the 817 bp sequence, that even silent mutations have not accumulated. It is likely, therefore, that the reiteration of this sequence is a very recent event, and that there has been insufficient time for the individual copies of this sequence to acquire mutations.

In addition to containing a CBD, XYLD displays esterase activity. The presence of the CBD suggests that the enzyme's primary substrate is the plant cell wall. In the presence of endo-1,4- β -D-xylanase, XYLD releases acetate and ferulate from plant structural polysaccharides. In addition, the enzyme also cleaves 4-hydroxy-3-methoxycinnamic acid and 4-hydroxycinnamic acid methyl esters. The esterase does not hydrolyse ester linkages in the plant cell wall in the absence of xylanase. Thus XYLD's primary substrates are likely to be acetylated xylo-oligosaccharides, or xylo-oligosaccharides containing arabinose side chains substituted with 4-hydroxy-3-methoxycinnamate groups. It should be noted, however, that XYLD also hydrolyses aryl esters and displays maximum activity against 4-nitrophenyl acetate. Although the classification of esterases in terms of their primary substrates is far from clear; in general, enzymes which have been classified as xylan esterases display 1.5–5.0 times more activity against aryl esters than against acetylated plant structural polysaccharides (Biely et al., 1986). By contrast, arylesterases expressed by orange peel and porcine liver display respectively 67 and 4518 times more activity against 4-nitrophenyl acetate than against acetylated hemicellulose. In view of the presence of a CBD, and the relatively low ratio of activity towards 4-nitrophenyl acetate to that towards acetylxylan, we tentatively classify XYLD as a xylan esterase. However, in order to define which of the potential substrates contained in the structural matrix of the plant cell wall are preferred by XYLD, it will be necessary to generate acetylated and 4-hydroxy-3-methoxycinnamated xylo-oligosaccharides of known structure.

The role of CBDs in plant-cell-wall hydrolysis is controversial. Fungal cellobiohydrolase CBDs appear to channel single cellulose chains into the active site, substantially increasing the activity of the enzyme against crystalline substrate (Tomme et al., 1988). The corresponding domain in prokaryotes has also been reported to enhance cellulase action against insoluble substrates (Ong et al., 1989). However, the location of CBDs in several plant-cell-wall hydrolases which do not attack cellulose argues against a direct role for this domain in cellulose hydrolysis (Kellett et al., 1990; Ferreira et al., 1990; Clarke et al., 1991). We have previously suggested that the CBDs, at least in hemicellulases, promote intimate contact with the plant cell wall, thus increasing the efficiency of the enzyme through prolonged

association with its recalcitrant substrate. A recent study by Din et al. (1991) showed that a CBD from *Cellulomonas fimi* CenA, although incapable of cleaving covalent bonds, disrupted the crystalline structure of cellulose, which could result in the plant cell wall becoming more accessible to a range of plant-cell-wall hydrolases. It is possible that the XYLD CBD exerts a similar effect on cellulose to that described for the corresponding domain from *C. fimi* CenA.

Analysis of cellulases and xylanases through primary-sequence comparisons and hydrophobic-cluster analysis has revealed nine enzyme families in which each member of a family is thought to have evolved from a common ancestral sequence (Béguin, 1990; Henrissat et al., 1989; Gilkes et al., 1991). Comparison of the catalytic domain of XYLD with those of other esterases, including the xylan esterase from *Caldocellum saccharolyticum* (Lüthi et al., 1990), failed to reveal any significant sequence similarities. This may reflect the great variety of substrates cleaved by esterases, the diversity of the type of ester bond cleaved and the differing organisms which express these enzymes. It is not surprising, therefore, that the extensive sequence identity observed between cellulases and xylanases is not apparent between XYLD and other esterases. Until the primary structures of several microbial xylan esterases have been determined, it remains to be established whether these enzymes have evolved through convergent or divergent evolution.

We thank J. Allen for his assistance with the nucleotide sequencing. We also wish to thank the Agricultural and Food Research Council for their support (grant no. LE13/138).

REFERENCES

- Béguin, P. (1990) Annu. Rev. Microbiol. **44**, 219–248
- Biely, P., Mackenzie, C. R., Puls, J. and Schneider, H. (1986) Bio/Technology **4**, 731–733
- Birnboim, H. C. and Doly, P. (1979) Nucleic Acids Res. **7**, 1513–1523
- Borneman, W. S., Hartley, R. D., Morrison, W. H., Akin, D. E. and Ljungdahl, L. G. (1990) Appl. Microbiol. Biotechnol. **33**, 345–351
- Borneman, W. S., Ljungdahl, L. G., Hartley, R. D. and Akin, D. E. (1992) Appl. Environ. Microbiol. **58**, 3762–3766
- Chambers, S. P., Prior, S. E., Barstow, D. A. and Minton, N. P. (1988) Gene **68**, 139–149
- Clarke, J. H., Laurie, J. I., Gilbert, H. J. and Hazlewood, G. P. (1991) FEMS Microbiol. Lett. **83**, 305–310
- Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, R. A. J. and Kilburn, D. G. (1991) Bio/Technology **9**, 1096–1099
- Faulds, C. B. and Williamson, G. (1991) J. Gen. Microbiol. **137**, 2339–2345
- Ferreira, L. M. A., Durrant, A. J., Hall, J., Hazlewood, G. P. and Gilbert, H. J. (1990) Biochem. J. **269**, 261–264
- Ferreira, L. M. A., Hazlewood, G. P. and Gilbert, H. J. (1991) Biochem. J. **279**, 793–799
- Gilbert, H. J., Jenkins, G., Sullivan, D. A. and Hall, J. (1987) Mol. Gen. Genet. **210**, 551–556
- Gilbert, H. J., Sullivan, D. A., Jenkins, G., Kellett, L. E., Minton, N. P. and Hall, J. (1988) J. Gen. Microbiol. **134**, 3239–3247
- Gilbert, H. J., Hall, J., Hazlewood, G. P. and Ferreira, L. M. A. (1990) Mol. Microbiol. **4**, 759–767
- Gilbert, H. J., Hazlewood, G. P., Laurie, J. I., Orpin, C. G. and Xue, G. P. (1992) Mol. Microbiol. **6**, 2065–2072
- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1991) Microbiol. Rev. **55**, 303–315
- Hall, J. and Gilbert, H. J. (1988) Mol. Gen. Genet. **213**, 112–117
- Hall, J., Hazlewood, G. P., Barker, P. J. and Gilbert, H. J. (1988) Gene **69**, 29–38
- Hall, J., Hazlewood, G. P., Huskisson, N. S., Durrant, A. J. and Gilbert, H. J. (1989) Mol. Microbiol. **3**, 1211–1219
- Hazlewood, G. P. and Gilbert, H. J. (1992) in Xylans and Xylanases (Visser, J., ed.), pp. 259–273. Elsevier Science Publishers, Amsterdam
- Hazlewood, G. P., Davidson, K., Laurie, J. I., Romaniec, M. P. M. and Gilbert, H. J. (1990) J. Gen. Microbiol. **136**, 2089–2097
- Henrissat, B., Claessens, M., Tomme, P., Lemesle, L. and Mornon, J.-P. (1989) Gene **81**, 83–95
- Hespell, R. B. and O'Bryan-Shah, P. J. (1988) Appl. Environ. Microbiol. **54**, 1917–1922
- Hunkapillar, M. W., Hewick, R. E., Dreyer, W. J. and Hood, L. E. (1983) Methods Enzymol. **91**, 399–413

- Johnson, K. G., Fontana, J. D. and MacKenzie, C. R. (1988) *Methods Enzymol.* **160**, 551–560
- Kellett, L. E., Poole, D. M., Ferreira, L. M. A., Durrant, A. J., Hazlewood, G. P. and Gilbert, H. J. (1990) *Biochem. J.* **272**, 369–376
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- Lüthi, E., Love, D. R., McAnulty, J., Wallace, C., Caughey, P. A., Saul, D. and Bergquist, P. L. (1990) *Appl. Environ. Microbiol.* **56**, 1017–1024
- Murphy, G. and Kavanagh (1988) *Nucleic Acids Res.* **16**, 5918
- Norlander, J., Kemp, T. and Messing, T. (1983) *Gene* **26**, 101–116
- Ong, E., Greenwood, J. M., Gilkes, N. R., Kilburn, D. G., Miller, R. C., Jr. and Warren, R. A. J. (1989) *Trends Biotechnol.* **7**, 239–243
- Poutanen, K. and Sundberg, M. (1988) *Appl. Microbiol. Biotechnol.* **28**, 419–424
- Poutanen, K., Tenkanen, M., Korte, H. and Puls, J. (1991) *ACS Symp. Ser.* **460**, 426–436
- Rixon, J. E., Ferreira, L. M. A., Durrant, A. J., Laurie, J. I., Hazlewood, G. P. and Gilbert, H. J. (1992) *Biochem. J.* **285**, 947–955
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Staden, R. (1980) *Nucleic Acids Res.* **16**, 3673–3694
- Tomme, P., van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeysens, M. (1988) *Eur. J. Biochem.* **170**, 575–581
- Utt, E. A., Eddy, C. K., Kershav, K. F. and Ingram, L. O. (1991) *Appl. Environ. Microbiol.* **57**, 1227–1234
- Van Tilbeurgh, H., Tomme, P., Claeysens, M., Bhikhabhai, R. and Pettersson, G. (1986) *FEBS Lett.* **204**, 223–227

Received 22 February 1993/7 April 1993; accepted 21 April 1993