Biochem. J. (1990) 272, 369–376 (Printed in Great Britain)

Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. *cellulosa* contain identical cellulose-binding domains and are encoded by adjacent genes

Louise E. KELLETT,* Debbie M. POOLE,* Luís M. A. FERREIRA,* Alastair J. DURRANT,* Geoffrey P. HAZLEWOOD† and Harry J. GILBERT*‡

*Department of Agricultural Biochemistry and Nutrition, 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.

The complete nucleotide sequence of the *Pseudomonas fluorescens* subsp. *cellulosa xynB* gene, encoding an endo-β-1,4-xylanase (xylanase B; XYLB) has been determined. The structural gene consists of an open reading frame (ORF) of 1775 bp coding for a protein of M_r 61000. A second ORF (*xynC*) of 1712 bp, which starts 148 bp downstream of *xynB*, encodes a protein, designated xylanase C (XYLC), of M_r 59000. XYLB hydrolyses oat spelt xylan to xylobiose and xylose, whereas XYLC releases only arabinose from the same substrate. Thus XYLB is a typical xylanase and XYLC is an arabinofuranosidase. Both enzymes bind to crystalline cellulose (Avicel), but not to xylan. The nucleotide sequences between residues 114 and 931 of *xynB* and *xynC* were identical, as were amino acid residues 39–311 of XYLB and XYLC. This conserved sequence is reiterated elsewhere in the *P. fluorescens* subsp. *cellulosa* genome. Truncated derivatives of XYLB and XYLC, in which the conserved sequence had been deleted, retained catalytic activity, but did not exhibit cellulose binding. A hybrid gene in which the 5′ end of *xynC*, encoding residues 1–110 of XYLC, was fused to the *Escherichia coli phoA′* gene (encodes mature alkaline phosphatase) directed the synthesis of a fusion protein which exhibited alkaline phosphatase activity and bound to cellulose.

INTRODUCTION

Xylan, a major component of plant hemicellulose, is hydrolysed by the co-operative actions of endo-β-1,4-xylanase (EC 3.2.1.8), β-xylosidase (EC 3.2.1.37) and a series of enzymes which cleave side-chain sugars or remove acetyl groups from the xylan backbone. These enzymes, described by Dekker & Richards (1976), include α-L-arabinofuranosidase (EC 3.2.1.55), β-glucuronidase (EC 3.2.1.31) and xylan acetylesterase (EC 3.1.1.6). Xylanolytic micro-organisms often synthesize isoenzymic forms of xylanases which are encoded by multiple genes (Gilbert *et al.*, 1988). Some xylanases only hydrolyse xylan (Hall *et al.*, 1989), whereas others exhibit endoglucanase activity (Hall *et al.*, 1988; Wong *et al.*, 1988).

Many micro-organisms that hydrolyse xylan also attack other plant cell-wall polysaccharides such as cellulose. In view of the similarity of the substrates cleaved (both contain β-1,4-glycosidic linkages) and the cross-specificity of some endoglucanases and xylanases, the properties and structural relationships of those enzymes which attack only cellulose or hemicellulose are worthy of examination. A common feature of cellulases from *Fibrobacter succinogenes* (McGavin & Forsberg, 1989), *Cellulomonas fimi* (Ong *et al.*, 1989), *Thermomonospora fusca* (Ghangas & Wilson, 1988), *Trichoderma reesei* (Van Tilbeurgh *et al.*, 1986) and *Pseudomonas fluorescens* subsp. *cellulosa* (Gilbert *et al.*, 1990) is the presence of two distinct structural domains; a cellulose-binding domain (CBD) and a catalytic domain (CD). Recent studies have shown that xylanase A (XYLA) from *P. fluorescens* subsp. *cellulosa* also contains a CBD which is distinct from the active site (Ferreira *et al.*, 1990). It remains to be established whether such multidomain proteins are prevalent among the other xylanases of *P. fluorescens* subsp. *cellulosa*, or if the molecular architecture of XYLA is unique, and a consequence of the evolution of the XYLA gene (*xynA*) from an adjacent endoglucanase gene (*celA*; Gilbert *et al.*, 1988).

With the object of investigating the molecular architecture of further *P. fluorescens* subsp. *cellulosa* xylanases, we have determined the nucleotide sequence of a second xylanase gene (*xynB*) and have analysed the capacity of the encoded enzyme (xylanase B; XYLB) to bind cellulose. Results showed that downstream of *xynB* is a second gene (*xynC*), which encodes an arabinofuranosidase (xylanase C; XYLC). The 5′ 817 bp of the two genes were identical and encoded a CBD.

MATERIALS AND METHODS

Strains, vectors and culture media

Escherichia coli strains employed in this study were JM83 *ara*Δ(*lac-pro*), *Sm*®, *thi* (φ80d *lacZ*Δ*M15*) and JM101 Δ(*lac-pro*), *SupE*, *thi* [F′*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*Δ*M15*]. The vectors used were pUC18, pUC19, M13mp8, M13mp18 and M13mp19 (Norrander *et al.*, 1983). Genes *xynB* and *xynC* were subcloned from pGHJ5 (Gilbert *et al.*, 1988), and the *phoA′* gene was derived from pMTL221. Recombinant strains harbouring plasmid and bacteriophage were cultured as described by Hall *et al.* (1988).

General recombinant DNA procedures

Agarose-gel electrophoresis, transformation of *E. coli* and the modification of DNA using restriction enzymes and T4 DNA ligase, extraction of *P. fluorescens* subsp. *cellulosa* chromosomal

Abbreviations used: CBD, cellulose-binding domain; CD, catalytic domain; ORF, open reading frame; XYLB, xylanase B; XYLC, arabinofuranosidase; *xynB* and *xynC*, genes encoding XYLB and XYLC.

‡ To whom correspondence and reprint requests should be sent.

The nucleotide sequence data reported will appear in the EMBL, Genbank and DDBJ Nucleotide Sequence databases under the accession number X54523.

DNA and Southern hybridization were as described by Gilbert *et al.* (1988). Plasmid DNA was extracted from *E. coli* by 'Brij lysis' and subsequent CsCl density-gradient centrifugation (Clewell & Helinski, 1969). The rapid-boiling method of Holmes & Quigley (1981) was used for small-scale plasmid isolation.

Nucleotide sequencing

DNA from recombinant M13 clones was sequenced 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 described by Staden (1980). The complete sequences of *xynB* and *xynC* were determined in both strands.

Assays

Xylanase activity was measured by determining the reducing sugar released from oat spelt xylan (Miller, 1959), and was expressed as xylose equivalents; one unit of enzyme activity released 1 μmol of pentose/min. Protein was assayed as described by Lowry *et al.* (1951), with BSA as standard. Products released from xylan by the action of XYLB or XYLC were identified by h.p.l.c. Periplasmic proteins (in a final volume of 5 ml) were prepared from 100 ml cultures of *E. coli*, harbouring either pLE7 or pLE9, as described by Hsiung *et al.* (1986). Reaction mixtures (2 ml final volume) contained periplasmic protein (0.1 ml) and oat spelt xylan (0.2% for XYLB, 1.0% for XYLC) in 50 mM-potassium phosphate/12 mM-citric acid, pH 6.5, and were incubated at 37 °C until there was no further release of reducing sugar. Insoluble material was removed by centrifuging and aliquots (20 μl) of supernatant were applied to a CarboPac PA1 column (0.4 cm × 25 cm). Material was eluted from the column with 15 ml of 1.75 mM-NaOH, a 5 ml gradient of 1.75–17.5 mM-NaOH, 10 ml of 17.5 mM-NaOH and finally 10 ml of 25 mM-NaOH. Alkaline phosphatase was assayed as described by Minton *et al.* (1983).

Cellulose-binding studies

Periplasmic fraction (2 ml), prepared as described above from *E. coli* cells harbouring *xynB*, *xynC* or truncated derivatives of the two genes, was mixed with 2 ml of Avicel (PH105; 5%, w/v) suspended in 100 mM-Tris/HCl buffer, pH 7.5, as described by Gilbert *et al.* (1990). Cellulose-bound and unbound protein was analysed by SDS/PAGE (Laemmli, 1970).

Protein sequencing

XYLB and XYLC were purified from total periplasmic proteins of *E. coli* harbouring pLE9 and pLE7 respectively by affinity chromatography with Avicel as described previously for endoglucanase B of *P. fluorescens* subsp. *cellulosa* (Gilbert *et al.*, 1990). The *N*-terminal sequences of XYLB and XYLC were determined by automated Edman sequencing using a 470A gas-phase sequenator equipped with a 120A on-line phenylthiohydantoin analyser (Applied Biosystems; Hunkapillar *et al.*, 1983).

RESULTS

Nucleotide sequences of *xynB* and *xynC*

Previous studies showed that pGHJ5 contained a 6.2 kb *Eco*RI restriction fragment of *Pseudomonas* genomic DNA cloned into the *Eco*RI site of pUC18. The plasmid directed the synthesis of a functional xylanase (Gilbert *et al.*, 1988). A restriction map of the xylanase-coding region of pGHJ5 is shown in Fig. 1.

Examination of the nucleotide sequence of the pseudomonad insert in pGHJ5 revealed two open reading frames of 1775 bp (*xynB*) and 1712 bp (*xynC*) which were transcribed from the same strand and separated by 148 bp (Fig. 2). The nucleotide sequences between residues 114 and 931 of *xynB* and *xynC* were identical. Translation of the two genes revealed that *xynB* and *xynC* encoded proteins with M_r values of 61000 and 59000 respectively. The first 11 amino acids of purified XYLB and

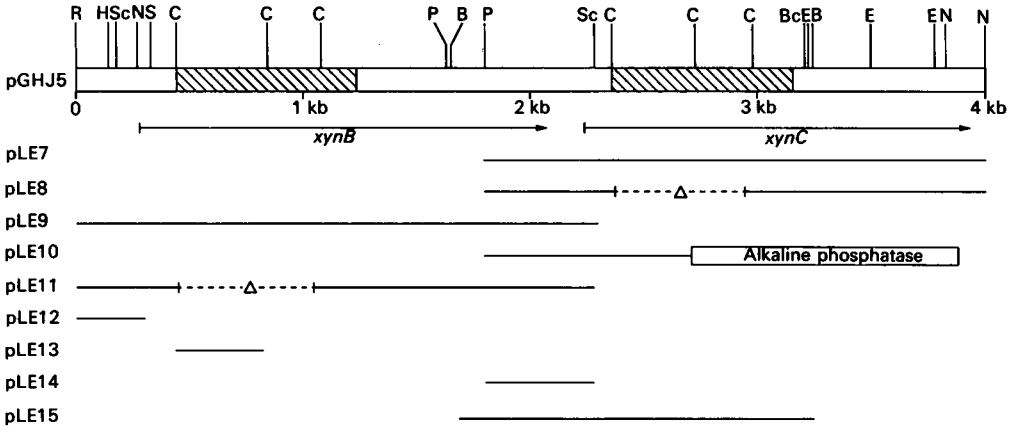


Fig. 1. Restriction maps of the recombinant pUC plasmids containing *xynB* and *xynC*

The positions of the cleavage sites of *Bam*HI (B), *Sca*I (Sc), *Bcl*I (Bc), *Nco*I (N), *Eco*RI (R), *Eco*RV (E), *Cla*I (C), *Pst*I (P), *Sma*I (S) and *Hpa*I (H) are shown. The solid arrows show the extent and orientation of *xynB* and *xynC*. For sequencing purposes the 1.6 kb *Bam*HI restriction fragments of *P. fluorescens* subsp. *cellulosa* genomic DNA were isolated from pLE15 and pLE9, and random subfragments, generated by sonication (Minton *et al.*, 1986), were cloned into the *Sma*I site of M13mp18. The resultant clones were sequenced as described in Materials and methods section. The nucleotide sequence over the *Bam*HI site was determined by cloning the 0.25 kb *Pst*I fragment from pLE9 into M13mp18. The sequence of DNA upstream from the *Bam*HI site was determined by cloning appropriate restriction fragments into M13 vectors. The plasmid pLE10 was generated as follows: pLE7 was digested with *Eco*RV and subsequently with BAL31 exonuclease (Hall *et al.*, 1988). After blunt-ending with T4 DNA polymerase, the DNA was digested with *Hin*dIII and ligated to the *phoA'* gene, which had been isolated on a 1.25 kb *Stu*I–*Hin*dIII fragment, coding for mature alkaline phosphatase (residue 8 to the *C*-terminus). Recombinants were selected in a *phoA*$^-$ mutant of *E. coli* by expression of the PhoA$^+$ phenotype using a chromogenic substrate (X-phosphate). The extent of deletion of *xynC* was determined by sequencing over the junction of *xynC* and *phoA'*. Regions of deleted sequence are shown by -----Δ-----.

Table 1. Reducing sugar released from oat spelt xylan by XYLB and XYLC contained in a periplasmic fraction prepared from *E. coli* JM83 harbouring pLE9 and pLE7 respectively

The conditions for the enzymic reactions are described in the Materials and methods section.

Enzyme	Time (h)	Reducing sugar released (μg)
XYLB	0	0
	0.25	450
	0.5	740
	1	970
	3	1100
	5	1130
XYLC	0	0
	0.1	270
	0.2	360
	1	430
	3	500
	5	510

XYLC, determined by *N*-terminal sequencing, were identical and exhibited a perfect match with amino acid residues 38–46 (inclusive) of the translated sequence. It is therefore likely that the 37 residues preceding the mature *N*-terminus represent the signal peptides of the respective proteins. The deduced *N*-terminal sequences of each of the encoded polypeptides consisted of predominantly hydrophobic residues characteristic of bacterial signal peptides, but the basic hydrophilic region, normally present at the *N*-terminus of secreted bacterial proteins, was less obvious in both XYLB and XYLC. This agrees with our previous observation that two extracellular endoglucanases from the same organism lacked a basic hydrophilic *N*-terminus, but contained a hydrophobic region within their respective signal peptides (Hall & Gilbert 1988; Gilbert *et al.*, 1990). Proposed translational start codons for *xynB* and *xynC* (Fig. 2) were each preceded (7 bp) by the sequence TGGAGA, which is characteristic of the ribosome-binding site of Gram-negative bacteria (Hall & Gilbert, 1988). Codon utilization in *xynB* and *xynC* was similar to that observed in genes *celA*, *celB* and *xynA* from the same organism (Hall & Gilbert, 1988; Hall *et al.*, 1989; Gilbert *et al.*, 1990).

Substrate specificity of XYLB and XYLC

The substrate specificities of XYLB and XYLC, expressed by *E. coli* cells harbouring pLE9 and pLE7 respectively, were determined. The data (Table 1) showed that XYLB and XYLC released 283 mg and 26 mg of reducing sugar per g of xylan respectively.

Analysis of hydrolysis products (Fig. 3) revealed that XYLB liberated predominantly xylobiose with some xylose, and may therefore be regarded as a typical bacterial endo-β-1,4-xylanase (EC 3.2.1.8). No disaccharides containing arabinose, the major side-chain sugar of oat spelt xylan, were detected among the products, suggesting that the enzyme does not cleave glycosidic linkages involving xylose residues which are linked to side-chain sugars. Hydrolysis of xylan by XYLC resulted in the production of arabinose, indicating that the enzyme is an arabinofuranosidase (EC 3.2.1.55). Furthermore, its failure to hydrolyse the synthetic substrate *p*-nitrophenyl α-L-arabinofuranoside suggests that XYLC belongs to a minority group of such enzymes which have a narrow substrate range and are active only against high-molecular-mass substrates in which arabinose is linked to a polymeric backbone (Kaji, 1984).

The conserved sequence in *xynB* and *xynC*

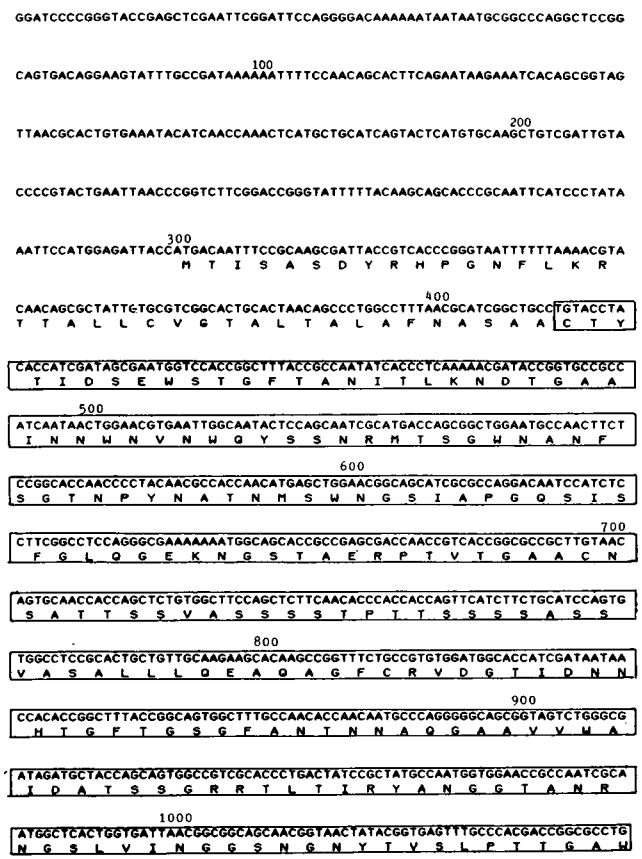
Previous studies showed that the 6.2 kb fragment of genomic DNA cloned in pGHJ5 contained sequences which hybridized to other regions of the *P. fluorescens* subsp. *cellulosa* chromosome (Gilbert *et al.*, 1988). To locate the repeated region more precisely, three DNA probes generated from pLE12, pLE13 and pLE14 (Fig. 1) were hybridized to *P. fluorescens* subsp. *cellulosa* genomic DNA which had been restricted with *Bam*HI, *Pst*I or *Eco*RI, electrophoresed in an agarose gel and blotted on to transfer membrane (Hybond N). The results (Fig. 4) showed that sequences upstream and downstream of the 817 bp region, which was identical in *xynB* and *xynC*, hybridized to single restriction fragments. The conserved sequence (pLE13) hybridized with four *Bam*HI, two *Eco*RI and three *Pst*I restriction fragments, suggesting that the sequence conserved in *xynB* and *xynC* is reiterated elsewhere in the *P. fluorescens* subsp. *cellulosa* genome. The presence of multiple genomic copies of a specific sequence is indicative of either a transposon or an insertion sequence. However, inspection of the 817 bp region, conserved in *xynB* and *xynC*, revealed no direct or inverted repeats at the ends of these sequences, suggesting that the duplicated region does not constitute an insertion sequence.

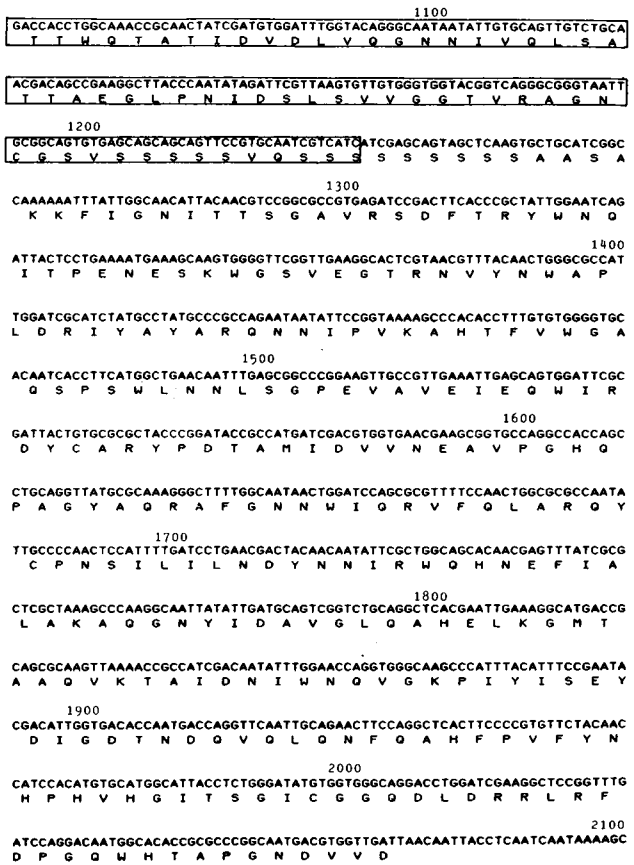
An alternative explanation is that the repeated sequence codes for a specific protein domain which is common to a family of related proteins. Thus either loci within the *P. fluorescens* subsp. *cellulosa* genome at which the conserved sequence occurs may be the sites of genes that, like *xynB* and *xynC*, code for enzymes involved in the hydrolysis of plant structural polysaccharides.

Role of the conserved regions in XYLB and XYLC

The mature forms of the enzymes XYLB and XYLC were identical between amino acid residues 39 and 311. Within this region, features of primary structure common to other cellulases and xylanases of *P. fluorescens* subsp. *cellulosa* were discernible. Firstly, two regions rich in hydroxy amino acids were apparent (Hall & Gilbert, 1988; Hall *et al.*, 1989; Gilbert *et al.*, 1990). Secondly, the sequence spanning residues 39–136 of XYLB and XYLC exhibited similarity with the CBD previously shown to be present in one xylanase (XYLA) and two endoglucanases (EGA and EGB) from *P. fluorescens* subsp. *cellulosa* (Gilbert *et al.*, 1990). To investigate whether XYLB and XYLC contain CBDs, periplasmic fractions prepared from *E. coli* JM83 harbouring pLE9 and pLE7 respectively were incubated with crystalline cellulose (Avicel). Results (Fig. 5) showed that both recombinant strains expressed a protein with an M_r of 58000 which bound tightly to cellulose and could be eluted with SDS. *N*-Terminal sequence analysis of these two proteins confirmed their identity as XYLB and XYLC respectively. To determine the location of the CBDs in XYLB and XYLC, the conserved sequence was deleted from *xynB and xynC* and the catalytic activities and cellulose-binding capacities of the enzymes encoded by the truncated genes were assessed. The data (Table 2) showed that the truncated enzymes retained catalytic activity, but did not bind cellulose. From this result we concluded that the conserved sequence in XYLB and XYLC encodes a CBD and that the catalytic centres of both enzymes are located in the *C*-terminal portion of the proteins.

To determine whether cellulose-binding capacity in XYLB and XYLC depends on the presence of the whole conserved sequence, the *N*-terminal region of *xynC* (coding for residues 1–110) was fused to the *E. coli phoA'* gene, which encodes mature alkaline phosphatase. In *E. coli* JM83, the hybrid gene directed the synthesis of a fusion protein which exhibited alkaline


GGATCCCCGGGTACCGAGCTCGAATTCGGATTCCAGGGGACAAAAAATAATAATGCGGCCCAGGCTCCGG
100
CAGTGACAGGAAGTATTTGCCGATAAAAAATTTTCCAACAGCACTTCAGAATAAGAAATCACAGCGGTAG
200
TTAACGCACTGTGAAATACATCAACCAAACTCATGCTGCATCAGTACTCATGTGCAAGCTGTCGATTGTA
CCCCGTACTGAATTAACCCGGTCTTCGGACCGGGTATTTTTACAAGCAGCACCCGCAATTCATCCCTATA
300
AATTCCATGGAGATTACCATGACAATTTCCGCAAGCGATTACCGTCACCCGGGTAATTTTTTAAAACGTA
M T I S A S D Y R H P G N F L K R
400
CAACAGCGCTATTCTGCGTCGGCACTGCACTAACAGCCCTGGCCTTTAACGCATCGGCTGCCTGTACCTA
T T A L L C V G T A L T A L A F N A S A A C T Y
CACCATCGATAGCGAATGGTCCACCGGCTTTACCGCCAATATCACCCTCAAAAACGATACCGGTGCCGCC
T I D S E W S T G F T A N I T L K N D T G A A
500
ATCAATAACTGGAACGTGAATTGGCAATACTCCAGCAATCGCATGACCAGCGGCTGGAATGCCAACTTCT
I N N W N V N W Q Y S S N R M T S G W N A N F
600
CCGGCACCAACCCCTACAACGCCACCAACATGAGCTGGAACGGCAGCATCGCGCCAGGACAATCCATCTC
S G T N P Y N A T N M S W N G S I A P G Q S I S
700
CTTCGGCCTCCAGGGCGAAAAAAATGGCAGCACCGCCGAGCGACCAACCGTCACCGGCGCCGCTTGTAAC
F G L Q G E K N G S T A E R P T V T G A A C N
AGTGCAACCACCAGCTCTGTGGCTTCCAGCTCTTCAACACCCACCACCAGTTCATCTTCTGCATCCAGTG
S A T T S S V A S S S S T P T T S S S S A S S
800
TGGCCTCCGCACTGCTGTTGCAAGAAGCACAAGCCGGTTTCTGCCGTGTGGATGGCACCATCGATAATAA
V A S A L L L Q E A Q A G F C R V D G T I D N N
900
CCACACCGGCTTTACCGGCAGTGGCTTTGCCAACACCAACAATGCCCAGGGGGCAGCGGTAGTCTGGGCG
H T G F T G S G F A N T N N A Q G A A V V W A
ATAGATGCTACCAGCAGTGGCCGTCGCACCCTGACTATCCGCTATGCCAATGGTGGAACCGCCAATCGCA
I D A T S S G R R T L T I R Y A N G G T A N R
1000
ATGGCTCACTGGTGATTAACGGCGGCAGCAACGGTAACTATACGGTGAGTTTGCCCACGACCGGCGCCTG
N G S L V I N G G S N G N Y T V S L P T I G A W


1100
GACCACCTGGCAAACCGCAACTATCGATGTGGATTTGGTACAGGGCAATAATATTGTGCAGTTGTCTGCA
T T W Q T A T I D V D L V Q G N N I V Q L S A
ACGACAGCCGAAGGCTTACCCAATATAGATTCGTTAAGTGTTGTGGGTGGTACGGTCAGGGCGGGTAATT
T T A E G L P N I D S L S V V G G T V R A G N
1200
GCGGCAGTGTGAGCAGCAGCAGTTCCGTGCAATCGTCATCATCGAGCAGTAGCTCAAGTGCTGCATCGGC
C G S V S S S S S S V Q S S S S S S S S S S A A S A
1300
CAAAAAATTTATTGGCAACATTACAACGTCCGGCGCCGTGAGATCCGACTTCACCCGCTATTGGAATCAG
K K F I G N I T T S G A V R S D F T R Y W N Q
1400
ATTACTCCTGAAAATGAAAGCAAGTGGGGTTCGGTTGAAGGCACTCGTAACGTTTACAACTGGGCGCCAT
I T P E N E S K W G S V E G T R N V Y N W A P
TGGATCGCATCTATGCCTATGCCCGCCAGAATAATATTCCGGTAAAAGCCCACACCTTTGTGTGGGGTGC
L D R I Y A Y A R Q N N I P V K A H T F V W G A
1500
ACAATCACCTTCATGGCTGAACAATTTGAGCGGCCCGGAAGTTGCCGTTGAAATTGAGCAGTGGATTCGC
Q S P S W L N N L S G P E V A V E I E Q W I R
1600
GATTACTGTGCGCGCTACCCGGATACCGCCATGATCGACGTGGTGAACGAAGCGGTGCCAGGCCACCAGC
D Y C A R Y P D T A M I D V V N E A V P G H Q
CTGCAGGTTATGCGCAAAGGGCTTTTGGCAATAACTGGATCCAGCGCGTTTTCCAACTGGCGCGCCAATA
P A G Y A Q R A F G N N W I Q R V F Q L A R Q Y
1700
TTGCCCCAACTCCATTTTGATCCTGAACGACTACAACAATATTCGCTGGCAGCACAACGAGTTTATCGCG
C P N S I L I L N D Y N N I R W Q H N E F I A
1800
CTCGCTAAAGCCCAAGGCAATTATATTGATGCAGTCGGTCTGCAGGCTCACGAATTGAAAGGCATGACCG
L A K A Q G N Y I D A V G L Q A H E L K G M T
CAGCGCAAGTTAAAACCGCCATCGACAATATTTGGAACCAGGTGGGCAAGCCCATTTACATTTCCGAATA
A A Q V K T A I D N I W N Q V G K P I Y I S E Y
1900
CGACATTGGTGACACCAATGACCAGGTTCAATTGCAGAACTTCCAGGCTCACTTCCCCGTGTTCTACAAC
D I G D T N D Q V Q L Q N F Q A H F P V F Y N
2000
CATCCACATGTGCATGGCATTACCTCTGGGATATGTGGTGGGCAGGACCTGGATCGAAGGCTCCGGTTTG
H P H V H G I T S G I C G G Q D L D R R L R F
2100
ATCCAGGACAATGGCACACCGCGCCCGGCAATGACGTGGTTGATTAACAATTACCTCAATCAATAAAAGC
D P G Q W H T A P G N D V V D

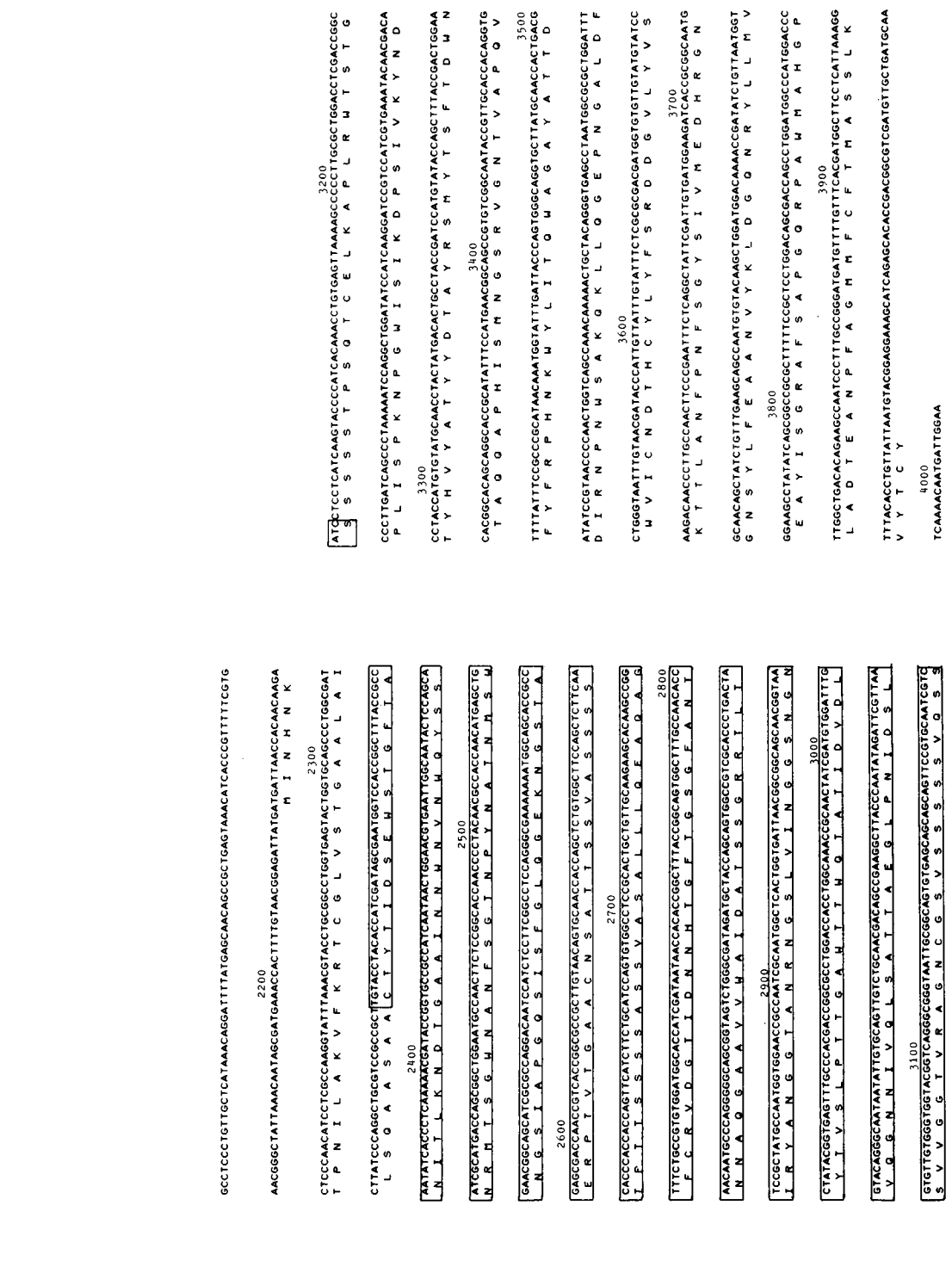


Fig. 2. Nucleotide sequence of *xynB* [nucleotides (nts) 229–2074], *xynC* (nts 2222–3934) and the derived amino acid sequence of their respective proteins, XYLB and XYLC

The conserved regions are boxed.

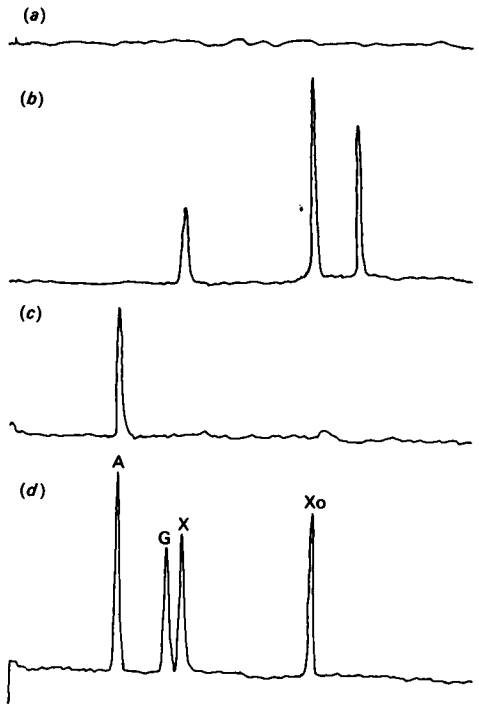


Fig. 3. H.p.l.c. analysis of sugars released after incubating periplasmic proteins of *E. coli* cells harbouring pLE7 (*c*) and pLE9 (*b*) with xylan

The retention times of the sugar standards arabinose (A), glucose (G), xylose (X) and xylobiose (Xo) were 6.6, 12.2, 13.3 and 21.9 min respectively (*d*). A blank consisting of xylan incubated for zero time with the periplasmic proteins is shown in (*a*). The conditions of the enzymic reactions and subsequent sugar analyses are described in the Materials and methods section.

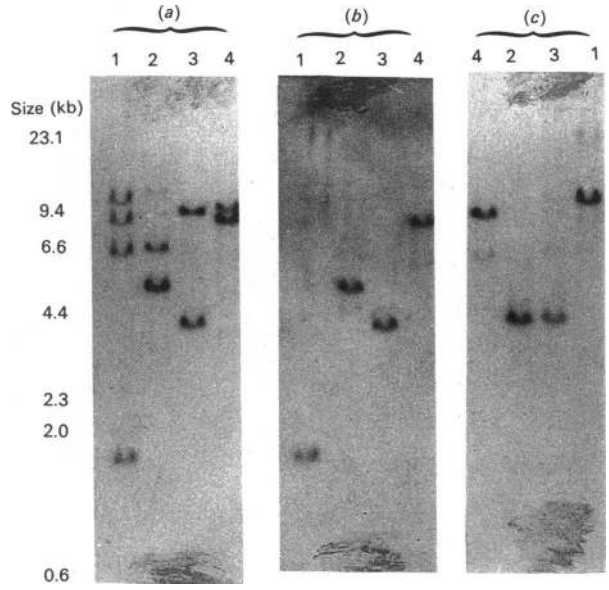


Fig. 4. Hybridization of restriction fragments encoding *xynB* and *xynC* with *P. fluorescens* subsp. *cellulosa* genomic DNA

The three probes consisted of the pseudomonad insert in pLE13 (*a*), pLE14 (*b*) and pLE12 (*c*). *Pseudomonas* genomic DNA was digested with *Bam*HI (1), *Sma*I (2), *Eco*RV (3) and *Eco*RI (4). The conditions used for radiolabelling the probes and subsequent Southern hybridization were as described by Gilbert *et al.* (1988). The filters were autoradiographed on X-ray film at −70 °C for 2 h. The position and sizes of restriction fragments generated by digesting λ- DNA with *Hin*dIII are shown.

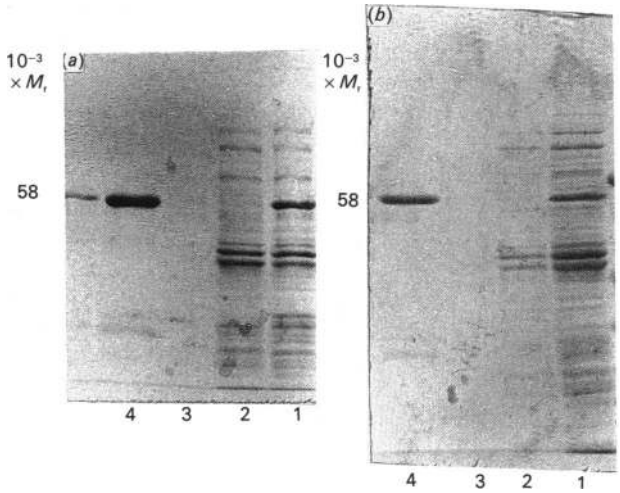


Fig. 5. SDS/PAGE of proteins expressed by *E. coli* strains harbouring pLE7 and pLE9, which bind to cellulose

Periplasmic proteins prepared from recombinant strain LE7 (*a*) and LE9 (*b*) were mixed and incubated with Avicel, and the unbound and bound proteins were analysed by SDS/PAGE as described in the Materials and methods section. Lane 1 contained periplasmic proteins; lane 2, filtrate from Avicel/periplasmic protein mixture containing unbound protein; lane 3, final Tris/HCl wash of Avicel; lane 4, bound protein eluted from Avicel by boiling for 5 min with 10% (w/v) SDS. A 10%-(w/v)-polyacrylamide gel was used. The M_r values of the Avicel-bound proteins are shown.

Table 2. Xylanase, arabinofuranosidase and cellulose (Avicel)-binding capacity of proteins expressed by *E. coli* strains harbouring derivatives of *xynB* and *xynC*

Plasmid	Xylanase	Arabinofuranosidase	Avicel binding
pLE7	−	+	+
pLE8	−	+	−
pLE9	+	−	+
pLE11	+	−	−

phosphatase activity and bound tightly to cellulose; alkaline phosphatase alone was unable to bind to cellulose (Table 3). These data show that the CBD is located in the *N*-terminal portion of XYLC (and XYLB) and, more specifically, occupies the region which displays sequence identity with other cellulases and xylanases of *P. fluorescens* subsp. *cellulosa*, known to contain CBDs. It has been argued that the presence of a CBD may confer cellulase activity (West *et al.*, 1989). This is clearly not the case, since it is now apparent that two xylanases [XYLA (Ferreira *et al.*, 1990) and XYLB] and one arabinofuranosidase (XYLC) from *P. fluorescens* subsp. *cellulosa* also contain CBDs.

DISCUSSION

The results described here show that two genes which encode enzymes involved in hemicellulose hydrolysis are located in close proximity on the chromosome of *P. fluorescens* subsp. *cellulosa*. A remarkable feature of the genes is that they are identical over approximately half of their primary sequences, even though the encoded enzymes have widely differing substrate specificities. In addition, the conserved region in the respective proteins is not essential for catalytic activity but, in part, constitutes a CBD.

Table 3. Binding to Avicel and xylan of alkaline phosphatase and an XYLC–alkaline phosphatase fusion protein

Protein	Polysaccharide used in binding	Alkaline phosphatase activity (total munits)*			
		Periplasmic fraction	Filtrate	Tris/HCl washes	Avicel-bound enzyme
XYLC–alkaline phosphatase†	Avicel	6.9	0	0	3.4
Alkaline phosphatase	Avicel	9.5	6.9	1.3	0
XYLC–alkaline phosphatase	Xylan	9.5	6.1	1.4	0.04

* The preparation of periplasmic proteins and the cellulose-binding assay were described in the Materials and methods section.
† The hybrid encoding the XYLC–alkaline phosphatase fusion protein was constructed as described in the legend to Fig. 1.

The conserved DNA sequence of the genes *xynB* and *xynC* is reiterated elsewhere in the genome of *P. fluorescens* subsp. *cellulosa*. The data presented pose some interesting questions with respect to the phylogeny of *xynB* and *xynC*.

Perhaps the most obvious explanation for the conservation of sequence is that *xynB* was duplicated to produce a second gene (*xynC*), which subsequently underwent substantial mutation, resulting in the change in substrate specificity. The lack of alteration at the 5′ end of *xynC* could reflect a strong selective pressure for retention of cellulose-binding capacity by the encoded enzyme. However, other cellulases and xylanases from *P. fluorescens* subsp. *cellulosa*, which contain a CBD, do not exhibit 100% identity such as occurs between *xynB* and *xynC* (Fig. 6), although it could be argued that the duplication of *xynB* was a recent event which left little time for the accumulation of mutations within the conserved sequence. The lack of overall sequence identity between *xynB* and *xynC* tends to argue against this theory. An alternative possible explanation for the observed sequence conservation is that the two genes did not evolve by duplication. Instead, the conserved sequence could have constituted a mobile genetic element encoding a CBD, which, in the course of domain shuffling, has come to be associated with two different catalytic domains. The absence of repeated sequences at the extremities of the conserved sequence is, however, contrary to the accepted structure for insertion sequences.

Although the CBDs of XYLB and XYLC do not bind xylan or play an important role in xylan hydrolysis *in vitro*, the conservation of CBDs in a range of plant-cell-wall-degrading enzymes suggests they are required for efficient hydrolysis of xylan and cellulose. It is our view that, in the normal environment of soil saprophytes, where efficient degradation of plant structural polysaccharides demands prolonged association of active enzymes with relatively recalcitrant substrates, CBDs would confer a selective advantage on plant-cell-wall-degrading enzymes by promoting intimate contact between enzyme and substrate. The diversity of xylan structure coupled with the high concentration of cellulose in different plant species could be the basis for the evolution of a single CBD, rather than protein domains which bind to each plant cell-wall polysaccharide.

The present results confirm earlier studies (Ferreira *et al.*, 1990; Gilbert *et al.*, 1990) showing that only part of the conserved sequences of *Pseudomonas* plant-cell-wall-degrading enzymes constitute CBDs. Within the conserved regions of cellulases and xylanases from a variety of micro-organisms, and also mammalian sucrase-isomaltase, are sequences rich in hydroxy amino acids. These structures are thought to be linker sequences between functional domains of proteins. *O*-Glycosylation of the threonine and serine residues confers protection against proteolysis (Fagerstam *et al.*, 1984). From the foregoing discussion it is apparent that the primary structure between the regions rich in hydroxy amino acids, in XYLB and XYLC, is not essential for catalysis or cellulose binding. However, the fact that this sequence is highly conserved is indicative of an important role in enzyme function, and it is our belief that it could be involved in secretion of these extracellular proteins; alternatively it could constitute a docking sequence which interacts with other enzymes to form a plant-cell-wall-degrading multienzyme complex.

The data presented here, in conjunction with previous studies (Hall & Gilbert, 1988; Hall *et al.*, 1989; Gilbert *et al.*, 1990) have shown that a range of different plant-cell-wall-degrading enzymes from *P. fluorescens* subsp. *cellulosa* contain CBDs. Whether such molecular architecture is a common feature of other enzyme systems which attack hemicellulose remains to be established.

```
XYLB/C 38 ACTYTIDSEWSTGFTANITLKNDTGAAINNWNVNWQYSSNRMTSGWNANFSG
XYLA   30 TCSYNITNEWNTGYTGDITITNRGSSAINGWSVNWQYATNRLSSSWNANVSG
EGB    31 ACEYRVTNEWGSGFTASIRIKNNGSSTINGWSVSWNYTDSRVTSSWNAGLSG
EGA   864 NCQYVVTNQWNNGFTAVIRVRNNGSSAINRWSVNWSYSDSRITNSWNANVTG
             * * *   *  ***  * * *     ** * * * *   ***  ***  **

XYLB/C 90 TNPYNATNMSWNGSIAPGQSISFGLQGEKNGSTAERPTVTGAACN
XYLA   82 SNPYSASNLSWNGNIQPGQSVSFGFQVNKNGGSAERPSVGGSICS
EGB    83 ANPYSATPVGWNTSIPIGSSVEFGVQGNNGSSRAQVPAVTGAIC
EGA   916 NNPYAASALGWNANIQPGQTAEFGFQGTKGAGSRQVPAVTGSVCQ
             *** ** * **  *  * *  ** *          * * *  *
```

Fig. 6. Amino acid sequence alignments for the CBDs of endoglucanases and xylanases from *P. fluorescens* subsp. *cellulosa*

The enzymes compared were as follows: endoglucanase A (EGA; Hall & Gilbert, 1988), endoglucanase B (EGB; Gilbert *et al.*, 1990), XYLA (Hall *et al.*, 1989), XYLB and XYLC. The residue at the start of each sequence is numbered.

We thank the Agricultural and Food Research Council (Grant LRG 138) for supporting this work.

REFERENCES

Clewell, D. B. & Helinski, D. R. (1969) Proc. Natl. Acad. Sci. U.S.A. **62**, 1159–1166
Dekker, R. F. H. & Richards, G. N. (1976) Adv. Carbohydr. Chem. Biochem. **32**, 277–352
Fägerstam, L. G., Pettersson, L. G. & Engström, J. Å. (1984) FEBS Lett. **167**, 309–315
Ferreira, L. M. A., Durrant, A. J., Hall, J., Hazlewood, G. P. & Gilbert, H. J. (1990) Biochem. J. **269**, 261–264
Ghangas, G. S. & Wilson, D. B. (1988) Appl. Environ. Microbiol. **54**, 2521–2526

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
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
Holmes, D. S. & Quigley, M. (1981) Anal. Biochem. **114**, 193–197
Hsiung, H. M., Mayne, N. G. & Becker, G. W. (1986) Bio/Technology **4**, 991–995
Hunkapillar, M. W., Hewick, R. E., Dreyer, W. J. & Hood, L. E. (1983) Methods Enzymol. **91**, 399–413
Kaji, A. (1984) Adv. Carbohydr. Chem. Biochem. **42**, 383–394
Laemmli, U. K. (1970) Nature (London) **277**, 680–685
Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. (1951) J. Biol. Chem. **193**, 265–275
McGavin, M. J. & Forsberg, C. W. (1989) J. Bacteriol. **171**, 3310–3315
Miller, G. L. (1959) Anal. Chem. **31**, 426–428
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 **49**, 25–35
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., Jr. & Warren, R. A. J. (1989) Trends Biotechnol. **7**, 239–243
Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) J. Mol. Biol. **143**, 161–178
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
Van Tilbeurgh, H., Tomme, P., Claeyssens, M., Bhikhabhai, R. & Petterson, G. (1986) FEBS Lett. **204**, 223–227
West, C. A., Elzanowski, A., Yeh, L.-S. & Barker, W. (1989) FEMS Microbiol. Lett. **59**, 167–172
Wong, K. K. Y., Tan, L. U. L. & Saddler, J. N. (1988) Microbiol. Rev. **52**, 305–317

Received 24 May 1990/12 July 1990; accepted 20 July 1990