Sequencing and Expression of the *Butyrivibrio fibrisolvens xylB* Gene Encoding a Novel Bifunctional Protein with β -D-Xylosidase and α -L-Arabinofuranosidase Activities[†]

E. A. UTT, C. K. EDDY, K. F. KESHAV, AND L. O. INGRAM*

Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611

Received 4 October 1990/Accepted 28 January 1991

A single gene (xylB) encoding both β -D-xylosidase (EC 3.2.1.37) and α -L-arabinofuranosidase (EC 3.2.1.55) activities was identified and sequenced from the ruminal bacterium *Butyrivibrio fibrisolvens*. The *xylB* gene consists of a 1,551-bp open reading frame (ORF) encoding 517 amino acids. A subclone containing a 1,843-bp DNA fragment retained both enzymatic activities. Insertion of a 10-bp *NotI* linker into the *Eco*RV site within the central region of this ORF abolished both activities. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis of cytoplasmic proteins from recombinant *Escherichia coli* confirmed the presence of a 60,000-molecular-weight protein in active subclones and the absence of this protein in subclones lacking activity. With *p*-nitrophenyl- β -D-xylopyranoside and *p*-nitrophenyl- α -L-arabinofuranoside as substrates, the specific activity of arabinosidase was found to be approximately 1.6-fold higher than that of xylosidase. The deduced amino acid sequence of the *xylB* gene was located between two incomplete ORFs within the 4,200-bp region which was sequenced. No sequences resembling terminators were found within this region, and these three genes are proposed to be part of a single operon. Based on comparison with other glycosidases, a conserved region was identified in the carboxyl end of the translated *xylB* gene which is similar to that of glucoamylase from *Aspergillus niger*.

Xylan is a major component in the cell walls of monocots and hardwoods, representing up to 30% of the dry weight of these plants (39). This polymer is second only to cellulose in natural abundance and represents a major reserve of reduced carbon in the environment. Unlike cellulose, xylan is a complex polymer consisting of a β -D-1,4-linked xylopyranoside backbone substituted with arabinosyl, acetyl, uronyl, mannosyl, and glucosyl side chains.

Many bacteria and fungi are able to degrade xylan in the environment (8, 9, 18, 38). Xylan degradation is a multistep process involving multiple enzymatic activities. Xylanases (1,4- β -D-xylan xylanohydrolase; EC 3.2.1.8) are extracellular enzymes which hydrolyze the internal β -1,4xylosidic linkages of the xylan backbone structure. Typically, small oligoxyloside fragments are transported into microbial cells where xylosidases (1,4-β-D-xylan xylohydrolase; EC 3.2.1.37) complete the hydrolysis by releasing free sugars for glycolysis (10). Removal of side-chain substituents requires additional enzymatic activities such as arabinofuranosidase, uronidase, glucosidase, mannosidase, and acetyl esterase. These enzymes are important for the efficient utilization of plant materials in animal feed. A mixture of xylanases is currently being added to feed to increase digestion of swine feeds in some countries. The genetic manipulation of anaerobic bacteria and ruminal organisms to increase the production of xylan-degrading enzymes is currently being investigated as a complementary means of improving the digestion of plant materials for increased milk and beef production (2, 12, 29).

Butyrivibrio fibrisolvens is an obligate anaerobe which is particularly abundant in the rumen and other anaerobic environments in which plant cell wall material serves as a primary substrate (7, 17). Strains of *B. fibrisolvens* have been isolated that produce the key enzymes necessary to degrade xylan (16, 34). In three such isolates from an anaerobic digester, we have previously shown that the expression of xylanase and xylosidase is repressed by glucose and induced by xylan (34). The gene encoding β -D-xylosidase activity was subsequently cloned and proposed to lie within the central region of a 4.2 kbp fragment (35).

In this report, we present the complete nucleotide sequence of the 4.2-kbp fragment from strain GS113 containing the xylB gene. This gene encodes a bifunctional protein which exhibits both β -D-xylopyranosidase and α -L-arabinofuranosidase activities.

MATERIALS AND METHODS

Medium and growth conditions. Escherichia coli DH5 α was propagated at 37°C in Luria broth or on Luria agar supplemented with 50 mg of ampicillin per liter (23).

Genetic methods. Plasmid pUC18 was used as a cloning vector. Xylosidase-positive constructs pLOI1001 and pLOI 1005 were reported previously (35). Restriction enzymes (Bethesda Research Laboratories) were used according to the manufacturer's instructions. All DNA ligations, transformations, isolations, and other manipulations were conducted by standard methods (24). Transformed colonies were screened on agar plates containing 20 μ g of the fluorogenic substrates 4-methylumbelliferyl- β -D-xylopyranoside or 4-methylumbelliferyl- α -L-arabinofuranoside (Sigma Chemical Co., St. Louis, Mo.) per ml.

DNA sequencing. Double-stranded DNA was sequenced in both directions by using the dideoxy-chain termination method and Sequenase (United States Biochemical Corp.) according to the manufacturer's instructions. Additional

^{*} Corresponding author.

[†] Florida Agricultural Experiment Station Publication no. R-01081.

sequencing primers were synthesized by the University of Florida ICBR Oligonucleotide Core and the Department of Microbiology and Cell Science Nucleotide Core.

DNA sequences were assembled and analyzed by using the GENEPRO software package (Hoefer Scientific Instruments, San Francisco, Calif.) and the University of Wisconsin Genetics Computer Group GCG package (version 6.1).

Preparation of cell extracts. Recombinant cultures were harvested from the mid-exponential phase by centrifugation $(10,000 \times g, 10 \text{ min, 4}^{\circ}\text{C})$ and washed twice with cold 5 mM sodium phosphate buffer (pH 6.8). Cell pellets were stored at -70°C until needed. Pellets were resuspended in an equal volume of 5 mM sodium phosphate buffer (pH 6.8) containing 10 mM β -mercaptoethanol and lysed by two passes through a French pressure cell at 20,000 lb/in². Membranes and debris were removed by centrifugation (100,000 $\times g$, 1 h, 4°C). The supernatants containing cytoplasmic proteins were stored at -70°C .

Enzyme assays. β -D-Xylopyranosidase and α -L-arabinofuranosidase activities were measured by determining the rate of hydrolysis of *p*-nitrophenol- β -D-xylopyranoside and *p*-nitrophenol- α -L-arabinofuranoside (1 mM final concentration), respectively, in 50 mM phosphate buffer (pH 6.8) at 37°C. Other nitrophenol derivatives of sugars were also tested as substrates under the same conditions. Assays were conducted in 1-ml volumes and terminated by the addition of 2 ml of 500 mM sodium carbonate. The cleavage of 1 nmol of substrate resulted in an increase in absorbance of 0.070 at 405 nm. Specific activities are expressed as nanomoles of nitrophenol released per minute per milligram of protein. Sugar derivatives were purchased from Sigma. Protein concentration was estimated by the method of Bradford (6).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Cell proteins were separated in denaturing gels by the method of Laemmli (21) and stained with Coomassie blue.

Nucleotide sequence accession number. The nucleotide sequence reported in this article has been assigned GenBank accession number M55537.

RESULTS

DNA sequence of xylosidase gene (xylB). The xylB gene was originally cloned as pLOI1001 (35) on a 4.2-kbp fragment of B. fibrisolvens DNA. The strategy used to sequence this fragment in both directions is shown in Fig. 1A. The sequence of this entire fragment is summarized in Fig. 2. Analysis of this sequence revealed three open reading frames (ORFs) (Fig. 1B). The first ORF (ORF1) is 1,340 bp long. ORF1 is incomplete and lacks a Shine-Dalgarno sequence and an initiation codon. A second ORF was found 15 bp downstream from the translational terminator of ORF1 and defines a complete gene with a putative Shine-Dalgarno sequence 6 bp upstream from the initiation codon. ORF2 is 1,551 bp long and spans the predicted xylB coding region (35). The deduced 517-amino-acid polypeptide from ORF2 encodes a protein with a calculated molecular weight of 62,040. A third incomplete ORF was found 123 bp downstream and also includes a putative Shine-Dalgarno sequence 5 bp upstream from the initiation codon. ORF3 continued for 1,173 bp to the end of the cloned fragment and was incomplete. No sequences were found with strong homology to rho-independent terminators.

Codon usage. Table 1 summarizes codon usage in the three *B. fibrisolvens* GS113 ORFs. Codon usage for *B. fibrisolvens* 49 xylA (25) and averaged codon usage for *E. coli* (1) are



NotI linker insertion

FIG. 1. Sequencing strategy and map of the 4.2-kbp DNA fragment from *B. fibrisolvens.* (A) Sequencing strategy of insert in pLO11001. Arrows indicate direction of sequencing. Vertical bars at the front of sequencing arrows indicate sequencing from a subclone with universal primers for pUC18 as opposed to sequencing from within *B. fibrisolvens* DNA with oligonucleotide primers (bars absent). Abbreviations: E, *Eco*RI; H, *Hind*III; A, *AccI*; X, *XbaI*; D, *DraI*; S, *SspI*; P, *PstI*; ERV, *Eco*RV. (B) Diagram showing ORFs and selected subclones with a qualitative evaluation of xylosidase and arabinosidase activity by using fluorogenic indicator plates. The double vertical bars in pLO11040 indicate the site at which a 10-bp *NotI* linker was inserted. Abbreviations are the same as those for panel A.

included for comparison. The patterns of codon usage among the three GS113 ORFs are very similar to each other and to strain 49 xylA. The lower GC content of B. fibrisolvens compared with E. coli is clearly evident in the preferred usage of codons with A or T. Dominant codons used by these two organisms differed for 12 of the 20 amino acids. With the exception of CAG for glutamine, AAG for lysine, and GAG for glutamic acid, A or T was present in the wobble position of all dominant codons in B. fibrisolvens.

Identification of ORF2 as xylB by insertional inactivation and subclone analysis. A variety of subclones were made during sequencing, and all were tested for β -D-xylosidase and α -L-arabinofuranosidase activities on umbelliferyl-glycoside plates (Fig. 1B). Both activities were expressed or lost coordinately. Only constructs which contained the complete xylB coding region (Fig. 1B) exhibited these activities. Both activities were retained by pLOI1043, a 1,843-bp SspI subclone which contained only 17 bp 5' and 274 bp 3' in addition to xylB.

The xylB gene was expressed only when inserted in the

AAT TGT GGA TGC ACA TAT GAA AAG CTG ATT TAT GCT TAT AAG GCA GGT CTT GTC AAG GAA GAG ACC ATC GAT GAG GCT GTT ACT CGA CTT 90 G L V K E E T I D E A V T R L N C G C T Y E K L I T A T K A ATG GAA ATC AGA CTT CGT CTA GGT ACT ATT CCA GAG AGA AAG AGT AAG TAT GAT GAT ATC CCA TAT GAA GTG GTC GAA TGC AAA GAG CAT 180 M E I R L R L G T I P E R K S K Y D D I P Y E V V E C K E H ATC AAA CTT GCT CTT GAC GCT GCA AAG GAT AGC TTT GTC CTT TTG AAG AAT GAT GGT TTA CTT CCA CTG AAT AAA AAG GAT TAT AAA TCT 270 I K L A L D A A K D S F V L L K N D G L L P L N K K D Y K S ATT GCT GTT ATT GGA CCC AAC TCT GAT TCA AGA AGA GCT TTA ATT GGA AAT TAT GAG GGC CTT TCT TCA GAG TAT ATT ACA GTT TTA GAG 360 ATT GET GTI ATT GUC LLC AND THE MEN HOW GUT THE ATT GUE WITH AND ADDITING AND GUT THE SECOND TO THE ADDITING AND ADDITING ADDITING AND ADDITING ADDITALIA ADDITING ADDITING ADDITING ADDITING ADDITING ADDIT P K D G F A E A K I V A E H S D L V V M C L G L D A S I E G GAA GAA GGA GAC GAG GGT AAT CAG TTC GGT AGC GGA GAC AAG CCT GGA TTA AAG CTT ACA GGT TGT CAG CAA GAG CTA CTT GAG GAA ATT 630 E E G D E G N Q F G S G D K P G L K L T G C Q Q E L L E E I GCC AMA ATC GGC AMG CCT GTT GTA CTT CTT GTG GTG TCT GGC TTT ACA GGC GGG GAG GAA TCT AAT AAC GTA ATT GGG 720 A K I G K P V V L L V I S G S A T GUE ANA ALC GUE ANG CET GIT GIT CIT GIT CIT GIT CIT GET TET GET CIT GAT ITA CA TGG GEG CAG GAA TET ANI ANG GTA ANT GEG ZAG A K I G K P V V L L V L S G S A L D L S W A Q E S N N V N Å ATA ATG CAG TGC TGG TAT CCA GGC GCA GAA GGT GGA CGT GCT ATT GCA GAG GTT TTAT TIT GGC AGA GGC AGT CCA GGC GGT AAA ATG CCT 810 I M Q C W Y P G A R G G R A I A E V L F G K A S P G G K N P CIT ACA TIT TAT GCC TCA GAT GAT GAC CIT CCT GAT TIT TCT GAT TAT TCA ATG GAA ATA GG ACA TAC AGA TAT TTC AGG GGC ACA CCA 900 L T F Y A S D D L P F S A Y B F S A Y S A CT ATT GAT TAT CA TTAT CAT TTG ATT ATT GAT AAA GAT AAG GCA ATT GGC GAT ACA 990 I T AT CCA TIT GAT TAT GGA CAT GGT GAT TT TCT AAA ATT AGT ATT GAT TAT GAT TAT GAT AGA GGA ACT AGG GCA ATT GGT GAT ACA 990 GAT TIT GCC ATT ATT GAT GAA AGG GGA AAA TGT ATC ATA GAG CCA GGC AAG TTT AG ATT TCT ATT GGG GGA CAA CAG CCA GAC GAT AGA 1260 D F G I I D E R G K C I I E P G K F K I S I G G O O P D D R AGT AAA GAA CTT ATG GGC AGA GAG TGT GAT ATT TTT GAA ATT GAA ATT GAA TTA ACA GGC TCT GTT ACA GAA GTT GAA TAT TAA TTG A<u>GA GGT</u> GCA 1350 S K E L M G R E C D I F E I E L T G S V T E V E Y $^{\circ}$ S.D TC <u>ATG</u> GT ATA GCT AGC AAT CCA ATT TA ATT CCA ACT CCT TCT ATC TGG AGA AAA GGG GAT GAT TTT ATT GTA GTT TGT 1439 TC <u>ATG</u> GT ATA GCT AGC AAT CCA ATT TTA AGA GGT TTT ATT CCA ACT CCT TCT ATC TGG AGA AAA GGG GAT GAT TTT TAT CTA GTT TGT 1439 TC A<u>TG</u> GT ATT GCT ACC GGA GTA CCG ATT TTT CCA CAC CCT TCT ATC TGC AGC AAATT GGA AATT GGA ATT AG CAG AGA 1529 S S F V Y A P G V P I L K G F Y P D P S I C R K G D D F Y L V C S S F V Y A P G V P I F H T K D L A H F E Q I G N I L D R E AGT CAA CTT CCA ATT GC GGA GAT ATA TCT AGA GGA TAT TGT AGA GAG GAC CCT ATT AGA AT GA ATC ATT TAC ATG ATA ACA ACT 1619 S C P N M I S C D S C V M I S C C C C A CCA ATA TA GGA GAG CCT ATT GCA CCC OCA ACA ATA GGA ATT GGA ATT GGA ATA GAA ATA GA ACA ACT 1619 S Q L P L S G D I S R G I F A P T I R E H N G I F Y M I T T AAT GTA AGC TCT GGC GGC AAC TTT ATT GTT ACT GCA AAA GAT CCA GCT GGT TCT TGG TCA GAG GCCA TAT TAT TTA GGT GAA GAT GAG GCG 1709 N V S S G G N F I V T A K D P A G P W S E P Y Y L G E D E A N V S G G N F I V T A K D P A G P W S E P Y Y L G E D E A CCA GAT AGT CAT GAT AGA CAT AT CT CAT GAT GAT GAT AAC 1799 P G I D P S L F F D D D G K C Y Y V G T R P N P D G V R Y N GGT GAT TGG GAG ATA TGG GTI CAA GAG CTG GAT TTA GAG CAA ATG AAA CTT GTA GGT CCT TCG ATG GCA ATT TGG AAG GGC GCT CTT AAG 1889 G D W E I W V G E L D L E G W K L V G P S M A I W K G A L K G A L K G A L TAT TAT CTT TAT CAT GCA GAA GCA GCA CCA CAA GCA TAT AAG AAA GAT GGA TAT TAT CTT TAT CAT GCA GAA GCT GCC ACA AGC TIT GAA GAT (AT 1977) O V I W P E G P H L Y K K D G Y Y L L H A E A G T S F E H GCT ATT TCT GTA GCT CGC TCA AAG GAG CTA TTC AAA TGG TIT GAG GGA TGT CCT AGA AAT CCT ATA TIT ACC CAT AGA AAT TTA GGC AAG 2069 A I S V A R S K E L F K W F E G C P R N P I F T H R N L G K GAT TAT CCA GTA TGC AAT GTT GGA CAT GCT GAT TTA GTT GAT GAT ATAT GGC AAC TGG TAT ATG GTG ATG CTG GCA TCT AGA CCA TGC 2159 D Y P V C N V G H A D L V D D I N G N W Y N V N L A S R P C AAG GGA AAG TGC AGC TTG GGA CGA GAG ACA TTC CTT GCA AAA GTA ATT TGG GAA GAC GGA TGG CCA GTG GTT AAT CCG GGA GTT GGT CGT 2249 K G K C S L G R E T F L A K V I W E D G W P V V N P G V G R TIG ACT GAT GAG GIG GAG ATG GAC CTT CCT GAA TAT AGA TTC TCA AMA GAG ATT ACT ACA AMG GAT AMA ATG ACC TTT GAA GAG ACA GTC 2339 L T D E V E N D L P E Y R F S K E I T T K D K N T F E E T V CTA GAT GAT AGA ATT GAT AGA ATG AAGC AGT GAG GAC TTT TAT TCC CTT ACT GAC AAT CCT GGA TTC TTA AGA TTA AAG CTT CGT 2429 LIA GAT AGA THE GIT GAA AGA AGA AGA AGA AGA GAG GAC THI TAT ICE CIT ACA AC CIT GGA THE THA AGA THA AGA CIT CE 24.24 L D R F V G I E R R S E D F Y S L T D N P G F L R L K L R CCT GAG GCC ATA GAA AAT ACT GGC AAT CCA TCT TAC TTA GGA ATT GGT CAA AGA CAT CAT TCG TTA AGA GCA AGG CAG CTT AGA TT 2519 P E A I E N T G N P S Y L G I R G K T H S F R A S C G L K F ACA CCA GCA AAA GAT AAT GAA TGT GCA GGA ATG GTG TTA TTC CAG AAT AAT GAA AAT CAC TTG GAT CAT TTA GAT GAT ATA GAA AGA AGA AGA T P A K D N E C A G M V L F Q N N E N H L E L V V K K K D AGG CTA CAG GTT AAA GTA GGA GCA GTT ATT AAA GGA ACC AAA ATC AGA CTT GCT TTT GAT ATT TCA TCA GGA GTT TTA GAA ATT ATT 2649 K I Q F K V G P V I K G C K K I A T GAA ATT AAA GGA ACC AAA ATC AGA CTT GCT ATT TCA TCA TCA GG GTT TTA GAA ATT ATT 2649 K L Q F K V G P V I K G T K I R L A T F D I S S G D L E I I CTT GAG GCA GCA AAT CAG CTG GCT AAT ATC TAT ATT AAA AAG AAT AAT GAA AAG ATT CTT GTG GCA GAA TGT ATT GAT TTG AGC CCA TAC 2789 LE A A N G L A N I Y I K K N N E K I L V A E C I D L S P Y ACT ACC GAA GAA TCA GGC GGA TTC GTA GGA TGT ACC ATT GGA CTA TAT GCT TCT CA ATT GGA CAG AGC AGC GGT GAT AAC TAT TGC GAT TAT 2879 T T E E S G G F V G C T I G L Y A S S N G K T S D N Y C D Y TCC TAC TIT ACA GTA GAA GAA GAT AGA CTA TTT CAA TGA GCG AAT TTG CAT TAT ACC GGA TTA ATT GTA CGT AAA AAC CAT ACC 2969 S Y F T V E E V * S Y F T V E E V * GGT GTA AMA TAG TIT CCA GAG AMA GTT TIT TCT CTG GAA TIT TIT ATT ATG GAG GGG ATT ATG CTT CAG GAA AGT ATT AAG AAG TTG GTA 3059 CAG TAC GGT ATT GAT ATG GGG CTT ACA CCA GAA TGT GAT GAG AGA ATA TAT ACT ACA UND ATT ATT AGA GAG TGT GTA 3059 S.D. $H \ C = S \ I \ K \ K \ V$ Q Y G I D M G L T P E C E R I Y T T A L L L E C M K E D E Y ATA GAT CCA GAC TGT GAT TTA AGA GAT ATT ATA CT TTA AGA GAA CAT GAT GAT GAT 329 I D P D C D L S N I I L E D V L K E L L D E A V N R G I I E GAT TCA GTT ACA CAT AGG GAT TTG TTT GAT ACA AMG CTA ATG AAT CAG CTA TGC CCA CGT CCT AMA CAG GTT ATA GAT GAT TTT AAC CGT 3329 D S V T H R D L F D T K L M N Q L C P R P K Q V I D D F N R ATA TAC GAT AAC CAT GGT CCA ATA GCT GCA ACA GAT TAT TTT TAC AGG TTA AGC AMA GCC TCT GAC TAT ATC CGT ACT TAC AGG GTA AMA 3419 I Y D N H G P I A A T D Y F Y K L S K A S D Y I R T Y R V K AG GAC CTA ANA TGG ACA TGG GAT ACA GAG TAT GGC ACT CTT GAC ATA ACA ATT AAT CTC TCT AAG CCA GAA AAA GAC CCA AAG GCA ATT 3509 ANG GAC CTA AMA TGG ACA TGG GAT ACA GAG TAT GGC ACT CTT GAC ATA ACA ATT AN LET CIC ICI AMG UA WAA GAC CLA ANG UA ATI 530° K D L K W T C D T E Y G T L D I T I N L S K P E K D P K A I GCT GCA GCT AMG AAT GCA AMA CAA TCC ACA TAT CCG AMG TGC CAA TTA TCT ATG AMA ATT GAA GGC TAT GCT GGT GGC ATT AMT TAT CCT 3590 A A K W A K Q S T Y P K C Q L C W E N E G Y A G R I N H P GCT AGA GAG AMT CAT CGC ATA ATT CCT ATA ACT ATA TAAT CAC GC AMC TGG GGA TTT ATAT AGT CGC TATA GTT TAT CAAT GAG CAT 3589 A R E N H R I I P I T I N S N W G F Q Y S P Y V Y N E H TGC ATA GTC TTT AAC GGA GAG CAT ACT CCT ATG AMA ATA GAG GGA GCT ACT TTT AGT TAAC CAT TTT CAA TAT ACC CTA TTT ATT TAC AMA CTA TTT CCA CAC 3779 F C ATA GTC TTT AAC GGA GAG CAT ACT CCT ATG AMA ATA GAG CGA GCT ACT TTT GAT AGC CTA TTT ACT AMA CTA TTT CCA CAC 3779 TGC ATA GTC TIT AAC GGA GAG CAT ACT CCT ATG AMA ATA GAG GGA GGC TACT TIT GTT AAG CTA TIT GAT TIT CAC AAC CTA TIT CCC AAC CTA TT CAC AAC CTA TIT CAC AAC ATT CAC ACTA TIT CAC AAC ATT CAC ACTA TIT CAC AAC ATT CAC ACTA TIT CAC AAC ATT CAC ACA TIT GAC GAC AGC AGC CAT TAC ACA TIT GAC GAC AGC CAC TIT CAC ACA TIT GAC GAC AGC CAC ATT ATT CAG GAA TIT ACC GAC AGC CAT TAC ACA TIT GAC GAC AGC CAC ATT ATT CAG GAA ATT GAC AAA AGG GAC GAT GAC AGC CAC ATT ATT CAG GAA TIT ACT GAC AAA AGG GAC GAT ATT AAG GCT GGT ATA GTT AAA GGC CAC CT TAC ACA TIT GAC GAC ATT ATT GAC AAA AGG GAC TAT ATT ATT GAC AAA TGG AGA AAT TAC ACC GAT GAC GAC GAC ATT ATT GAC AAA TGG AGA AAT TAC ACC GAT GAA GAG GCA TAT 4049 R L Q C K D E T R L I D L A T N I L D K W R N Y T D E E A Y ATT TIT CA GAC AAAC CT AAT GAC CT AAT GAC AAA AGG GG GAT TAC ACC GAT GAA GAG GCA TAT 4049 R L Q C K D E T R L I D L A T N I L D K W R N Y T D E E A Y ATT TIT GAT GAA GAG GCC TAC AACC ATT ACG ATT ACG CT AAC GAT GAC CT ATT GAC CT AAT GAC CT AAT GAC GAA ACAC CT ATT GAC AAA AGG GG GAT TAC ACC GAT GAC GAC TAT AGG ATT AC ACC GAT GAC GAC TAT AGG ATT ACC ACC GAT GAC GAC TAT CAC GAT CAC CT ACG ACC TAT GAC AAA CAG GC GAT TAC ACC GAT CAC GAC CT ATT GAC AGA CT AAT AGC AAA CAG GC GAT TAC ACC GAT CAC GAC CT ACC ATT GAC GAC CT ATT GAC AAA CAG GC GAT TAC ACC CT ACG AACC TAT GAC GAT CAC CT ACT GAC CT ACC GAC CT ACC ACT ACC GAC CT ACC ACT A

FIG. 2. Nucleotide sequence and translated amino acid sequence for the 4.2-kbp insert in pLOI1001. Putative Shine-Dalgarno (S.D.) regions and initiation codons are underlined. Translational termination is indicated by an asterisk (*).

direction of transcription from the *lac* promoter, indicating a dependence on this promoter in *E. coli*. A frameshift mutation was created within the central region of this gene by the insertion of a 10-bp *Not*I linker into the *Eco*RV site (pLOI1040). This mutation resulted in the loss of both enzymatic activities (Table 2).

Comparison of specific activities for xylosidase and arabinosidase in cell extracts was consistent with their origin from a single gene (Table 2). Using *p*-nitrophenol sugar substrates, arabinosidase activity was 1.6-fold higher than xylosidase activity. The ratio of these activities was essentially the same for the three active subclones. The highest enzymatic

 TABLE 1. Comparison of codon usage frequency for the three

 B. fibrisolvens ORFs

acid D. Jordsovens E. B. fibrisolvens ORF1 ORF2 ORF3 ColP ^a B. fibrisolvens ⁴ xylA Phe TTC 0.9 1.2 0.3 2.2 1.5 Leu TTA 2.0 2.5 2.8 0.7 0.7 TTG 0.2 1.2 1.0 0.9 0.5 CTC 0 0 0.3 8.0 0 CTA 1.4 1.0 1.8 0.2 0.2 CTG 0.5 0.8 0.3 6.8 1.0 Ile ATT 4.7 3.9 4.3 2.2 2.7 ATC 2.0 1.0 0.8 .7 1.7 ATA 0.5 1.7 3.8 0.2 0.5 Met ATG 1.4 1.7 2.0 2.8 2.7 Val GTT 2.9 2.7 2.3 2.9 2.0 GTC 0.9 0.2	Amino	Calar	Frequency (mol%) of codo	usage	
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	acid	Codon		. fibrisolve		E. coli ^a	B. fibrisolvens ^b xylA	
Prife TIT 2.7 4.1 5.6 1.3 2.9 Leu TTA 2.0 2.5 2.8 0.7 0.7 TG 0.2 1.2 1.0 0.9 0.5 CTT 5.6 2.7 2.5 0.8 4.4 CTC 0 0 0.3 0.8 0 CTG 0.5 0.8 0.3 6.8 1.0 Ile ATT 4.7 3.9 4.3 2.2 2.7 ATA 0.5 1.7 3.8 0.2 0.5 Met ATG 1.4 1.7 2.0 2.8 2.7 Val GTT 2.9 2.7 2.3 2.9 2.0 GTC 0.9 0.2 0.5 1.2 0.5 GTA 1.4 2.1 1.3 1.8 4.2 GTC 0.9 0.8 0.3 0.3 0.5 CCT 1.4	Dha		27	4 1	2 4	1 2	2.0	
	rne	TTC	0.9	4.1 1.2	0.3	2.2	1.5	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Leu	TTA	2.0	2.5	2.8	0.7	0.7	
$\begin{array}{ccccccc} C11 & 5.6 & 2.7 & 2.5 & 0.8 & 4.4 \\ CTC & 0 & 0 & 0.3 & 0.8 & 0.2 \\ CTG & 0.5 & 0.8 & 0.3 & 6.8 & 1.0 \\ \hline \\ CTG & 0.5 & 0.8 & 0.3 & 6.8 & 1.0 \\ \hline \\ Ile & ATT & 4.7 & 3.9 & 4.3 & 2.2 & 2.7 \\ ATC & 2.0 & 1.0 & 0.8 & 3.7 & 1.7 \\ ATC & 0.5 & 1.7 & 3.8 & 0.2 & 0.5 \\ \hline \\ Met & ATG & 1.4 & 1.7 & 2.0 & 2.8 & 2.7 \\ \hline \\ Val & GTT & 2.9 & 2.7 & 2.3 & 2.9 & 2.0 \\ GTC & 0.9 & 0.2 & 0.5 & 1.2 & 0.5 \\ GTA & 1.4 & 2.1 & 1.3 & 1.8 & 4.2 \\ GTG & 1.1 & 1.2 & 0 & 2.2 & 1.0 \\ \hline \\ Ser & TCT & 2.3 & 1.5 & 0.8 & 1.3 & 1.2 \\ TCC & 0 & 0.4 & 0.3 & 1.5 & 0.5 \\ TCA & 1.8 & 1.5 & 0.5 & 0.4 & 2.9 \\ TCG & 0 & 0.6 & 0.3 & 0.6 & 0.2 \\ AGT & 0.9 & 0.8 & 0.3 & 0.3 & 0 \\ AGC & 0.7 & 1.0 & 1.8 & 1.4 & 1.2 \\ \hline \\ Pro & CCT & 1.4 & 1.7 & 1.8 & 0.5 & 1.2 \\ CCC & 0.2 & 0 & 0 & 0.3 & 0 \\ AGC & 0.7 & 1.0 & 1.8 & 1.4 & 1.2 \\ \hline \\ Pro & CCT & 1.4 & 1.7 & 1.8 & 0.5 & 1.2 \\ CCG & 0 & 0.6 & 0.3 & 2.5 & 0 \\ \hline \\ Thr & ACT & 0.7 & 1.9 & 2.3 & 1.1 & 0.7 \\ ACC & 0.7 & 1.4 & 0.5 & 2.4 & 1.0 \\ ACC & 0.7 & 1.4 & 0.5 & 2.4 & 1.0 \\ ACC & 0.7 & 1.4 & 0.8 & 2.2 & 1.2 \\ GCG & 0.5 & 0 & 0.3 & 0.8 & 0.5 \\ \hline \\ Ala & GCT & 2.7 & 2.1 & 2.8 & 2.6 & 1.5 \\ GCC & 1.1 & 0.4 & 0.8 & 2.2 & 1.2 \\ GCG & 0.5 & 0.2 & 0.2 & 3.2 & 0 \\ \hline \\ Tyr & TAT & 4.3 & 3.5 & 2.8 & 1.0 & 3.9 \\ TAC & 0.2 & 1.2 & 2.5 & 1.5 & 3.4 \\ \hline \\ His & CAT & 0.9 & 1.4 & 2.5 & 0.7 & 1.5 \\ CAG & 1.6 & 0.6 & 1.5 & 3.2 & 2.2 \\ \hline \\ Gln & CAA & 0.7 & 1.0 & 1.3 & 1.0 & 0.5 \\ CAG & 1.6 & 0.6 & 1.5 & 3.2 & 2.2 \\ \hline \\ Asn & AAT & 2.0 & 4.2 & 3.8 & 1.0 & 3.9 \\ Asp & GAT & 5.8 & 5.6 & 5.9 & 2.5 & 3.9 \\ GAC & 1.6 & 1.4 & 2.0 & 3.0 & 0.2 \\ \hline \\ Glu & GAA & 4.3 & 3.7 & 3.7 & 4.9 & 1.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ GCG & 1.7 & 1.0 & 1.5 & 0.5 & 0.5 \\ \hline \end{array}$		TTG	0.2	1.2	1.0	0.9	0.5	
$\begin{array}{cccccc} CTA & 1.4 & 1.0 & 1.8 & 0.2 & 0.2 \\ CTG & 0.5 & 0.8 & 0.3 & 6.8 & 1.0 \\ \hline CTG & 0.5 & 0.8 & 0.3 & 6.8 & 1.0 \\ \hline Ile & ATT & 4.7 & 3.9 & 4.3 & 2.2 & 2.7 \\ ATC & 2.0 & 1.0 & 0.8 & 3.7 & 1.7 \\ ATA & 0.5 & 1.7 & 3.8 & 0.2 & 0.5 \\ \hline Met & ATG & 1.4 & 1.7 & 2.0 & 2.8 & 2.7 \\ \hline Val & GTT & 2.9 & 2.7 & 2.3 & 2.9 & 2.0 \\ GTC & 0.9 & 0.2 & 0.5 & 1.2 & 0.5 \\ GTG & 1.1 & 1.2 & 0 & 2.2 & 1.0 \\ \hline Ser & TCT & 2.3 & 1.5 & 0.8 & 1.3 & 1.2 \\ TCC & 0 & 0.4 & 0.3 & 1.5 & 0.5 \\ TCA & 1.8 & 1.5 & 0.5 & 0.4 & 2.9 \\ TCG & 0 & 0.6 & 0.3 & 0.6 & 0.2 \\ AGT & 0.9 & 0.8 & 0.3 & 0.3 & 0.6 \\ AGT & 0.9 & 0.8 & 0.3 & 0.3 & 0.6 \\ AGT & 0.9 & 0.8 & 0.3 & 0.3 & 0.6 \\ AGT & 0.9 & 0.8 & 0.3 & 0.3 & 0.6 \\ AGT & 0.9 & 0.8 & 0.3 & 0.3 & 0.6 \\ CCC & 0.2 & 0 & 0 & 0.03 & 0.6 \\ CCC & 0.2 & 0 & 0 & 0.3 & 0.5 \\ CCC & 0.2 & 0 & 0 & 0.3 & 0.5 \\ CCC & 0.2 & 0 & 0 & 0.3 & 0.5 \\ CCC & 0.2 & 0 & 0 & 0.3 & 0.5 \\ CCC & 0.2 & 0 & 0 & 0.3 & 0.5 \\ \end{array}$		CIT	5.6	2.7	2.5	0.8	4.4	
$\begin{array}{cccccc} \mbox{CrG} & 0.7 & 0.8 & 0.3 & 6.8 & 0.1 \\ \mbox{ATC} & 2.0 & 1.0 & 0.8 & 3.7 & 1.7 \\ \mbox{ATA} & 0.5 & 1.7 & 3.8 & 0.2 & 0.5 \\ \mbox{Met} & ATG & 1.4 & 1.7 & 2.0 & 2.8 & 2.7 \\ \mbox{Val} & GTT & 2.9 & 2.7 & 2.3 & 2.9 & 2.0 \\ \mbox{GTC} & 0.9 & 0.2 & 0.5 & 1.2 & 0.5 \\ \mbox{GTA} & 1.4 & 2.1 & 1.3 & 1.8 & 4.2 \\ \mbox{GTG} & 1.1 & 1.2 & 0 & 2.2 & 1.0 \\ \mbox{Ser} & TCT & 2.3 & 1.5 & 0.8 & 1.3 & 1.2 \\ \mbox{TCC} & 0 & 0.4 & 0.3 & 1.5 & 0.5 \\ \mbox{TCA} & 1.8 & 1.5 & 0.5 & 0.4 & 2.9 \\ \mbox{TCG} & 0.6 & 0.3 & 0.6 & 0.2 \\ \mbox{AGT} & 0.9 & 0.8 & 0.3 & 0.3 & 0.6 \\ \mbox{AGC} & 0.7 & 1.0 & 1.8 & 1.4 & 1.2 \\ \mbox{Pro} & CCT & 1.4 & 1.7 & 1.8 & 0.5 & 1.2 \\ \mbox{CCG} & 0 & 0.6 & 0.3 & 0.6 & 0.2 \\ \mbox{AGC} & 0.7 & 1.9 & 2.3 & 1.1 & 0.7 \\ \mbox{ACC} & 0.7 & 1.4 & 0.5 & 2.4 & 1.0 \\ \mbox{ACC} & 0.5 & 0 & 0.3 & 0.8 & 0.5 \\ \mbox{ACC} & 0.7 & 1.4 & 0.5 & 2.4 & 1.0 \\ \mbox{ACC} & 0.7 & 1.4 & 0.5 & 2.4 & 1.0 \\ \mbox{ACC} & 0.7 & 1.4 & 0.5 & 2.4 & 1.0 \\ \mbox{ACC} & 0.5 & 0 & 0.3 & 0.8 & 0.5 \\ \mbox{ACC} & 0.5 & 0 & 0.3 & 0.8 & 0.5 \\ \mbox{Alex} & 0.7 & 1.9 & 2.3 & 1.1 & 0.7 \\ \mbox{ACC} & 0.5 & 0.2 & 0.2 & 3.2 & 0 \\ \mbox{Thr} & ACT & 0.7 & 1.9 & 2.3 & 1.1 & 0.7 \\ \mbox{ACC} & 0.5 & 0.2 & 0.2 & 3.2 & 0 \\ \mbox{Tyr} & TAT & 4.3 & 3.5 & 2.8 & 1.0 & 3.9 \\ \mbox{ACC} & 0.5 & 0.2 & 0.2 & 3.2 & 0 \\ \mbox{Tyr} & TAT & 4.3 & 3.5 & 2.8 & 1.0 & 3.9 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 1.0 & 0.5 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 2.8 & 2.7 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 2.8 & 2.7 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 2.8 & 2.7 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 2.8 & 2.7 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 2.8 & 2.7 \\ \mbox{AAA} & 3.4 & 3.1 & 3.6 & 4.1 & 3.4 \\ \mbox{AAC} & 0.5 & 1.1 & 1.3 & 2.8 & 2.7 \\ \mbox{AAA} & 3.4 & 3.7 & 3.7 & 4.9 & 2.7 \\ \mbox{CYs} & TGT & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ \mbox{GAG} & 4.9 & 3.5 & 3.8 & 1.8 & 2.7 \\ \mbox{Cys} & TGT & 1.4 & 1.4 & 1.0 & 0.4 & 0.7 \\ \mbox{TCG} & 0.7 & 1.0 & 1.5 & 0.5 & 0.5 \\ \mbox{TCS} & 0.7 & 1.0 & 1.5 & 0.5 \\ \mbox{TCS} & 0.7 & 1.0 & 1.5 & 0.5 \\ \mbox{TCS} & 0.7 & 1.0 & 1.5 & 0.5 \\ $		СТА	14	10	1.8	0.8	02	
lie ATT ATC 4.7 2.0 3.9 1.0 4.3 8.8 2.2 3.7 2.7 1.7 Met ATG 1.4 1.7 2.0 2.8 2.7 Val GTT GTC 2.9 0.9 2.7 0.2 2.3 0.5 2.9 1.2 2.0 0.5 Ser TCT 		CTG	0.5	0.8	0.3	6.8	1.0	
ATC 2.0 1.0 0.8 3.7 1.7 ATA 0.5 1.7 3.8 0.2 0.5 Met ATG 1.4 1.7 2.0 2.8 2.7 Val GTT 2.9 2.7 2.3 2.9 2.0 GTC 0.9 0.2 0.5 1.2 0.5 GTG 1.1 1.2 0 2.2 1.0 Ser TCT 2.3 1.5 0.8 1.3 1.2 TCC 0 0.4 0.3 1.5 0.5 0.4 2.9 TCG 0 0.4 0.3 0.5 0.2 0.6 0.2 0.6 0.2 0.6 0.2 0.6 0.2 0.6 0.2 0.6 0.2 0.6 0.2 0.6 0.2 0.2 0.6 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 <td>Île</td> <td>ATT</td> <td>4.7</td> <td>3.9</td> <td>4.3</td> <td>2.2</td> <td>2.7</td>	Île	ATT	4.7	3.9	4.3	2.2	2.7	
ATA 0.5 1.7 3.8 0.2 0.5 Met ATG 1.4 1.7 2.0 2.8 2.7 Val GTT 2.9 2.7 2.3 2.9 2.0 GTC 0.9 0.2 0.5 1.2 0.5 GTA 1.4 2.1 1.3 1.8 4.2 GTG 1.1 1.2 0 2.2 1.0 Ser TCT 2.3 1.5 0.8 1.3 1.2 TCC 0 0.4 0.3 0.6 0.2 1.0 Ser TCT 2.3 1.1 1.2 0 2.2 TCG 0 0.6 0.3 0.6 0.2 0 AGT 0.9 0.8 0.3 0.5 1.2 CCC 1.4 1.7 1.8 0.5 1.2 CCC 0.7 1.4 0.5 2.4 0 CCC 0.		ATC	2.0	1.0	0.8	3.7	1.7	
Met ATG 1.4 1.7 2.0 2.8 2.7 Val GTT 2.9 2.7 2.3 2.9 2.0 GTC 0.9 0.2 0.5 1.2 0.5 GTG 1.1 1.2 0 2.2 1.0 Ser TCT 2.3 1.5 0.8 1.3 1.2 TCC 0 0.4 0.3 1.5 0.5 1.4 2.2 Ser TCC 0 0.4 0.3 1.5 0.5 0.4 2.9 TCG 0 0.6 0.3 0.6 0.2 0.4 1.7 1.8 1.5 0.5 0.4 2.9 TCG 0 0.6 0.3 0.6 0.2 0.2 0 0 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.5 0 0.3 0.8 0.5 0 0.5 <t< td=""><td></td><td>ΑΤΑ</td><td>0.5</td><td>1.7</td><td>3.8</td><td>0.2</td><td>0.5</td></t<>		ΑΤΑ	0.5	1.7	3.8	0.2	0.5	
Val GTT 2.9 2.7 2.3 2.9 2.0 GTC 0.9 0.2 0.5 1.2 0.5 GTA 1.4 2.1 1.3 1.8 4.2 GTG 1.1 1.2 0 2.2 1.0 Ser TCT 2.3 1.5 0.8 1.3 1.2 TCC 0 0.4 0.3 1.5 0.5 1.4 AGT 0.9 0.8 0.3 0.6 0.2 0 AGT 0.9 0.8 0.3 0.3 0 0 0 AGC 0.7 1.0 1.8 1.4 1.2 0 0 0 Pro CCT 1.4 1.7 1.8 0.5 1.2 0 Thr ACT 0.7 1.9 2.3 1.1 0.7 ACC 0.7 1.4 0.5 2.4 1.0 ACC 0.7 1.4 0.5 2.4 1.0 ACC 0.7 1.4 0.5 2	Met	ATG	1.4	1.7	2.0	2.8	2.7	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	Val	GTT	2.9	2.7	2.3	2.9	2.0	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		GTC	0.9	0.2	0.5	1.2	0.5	
SerTCT2.31.50.81.31.2TCC00.40.31.50.5TCA1.81.50.50.42.9TCG00.60.30.60.2AGT0.90.80.30.30AGC0.71.01.81.41.2ProCCT1.41.71.80.51.2CCC0.2000.30CCA2.33.52.80.72.4CCG00.60.30.50ThrACT0.71.92.31.10.7ACC0.71.40.52.41.0ACG0.500.30.80.5AlaGCT2.72.12.82.61.5GCA2.02.32.02.34.1GCG0.50.20.23.20TyrTAT4.33.52.81.03.9TyrTAT0.50.60.51.20.2GlnCAA0.71.01.31.00.5CAG1.60.61.53.22.2AsnAAT2.04.23.81.05.4AAC0.51.11.33.64.13.4HisCAG0.51.11.33.00.2GlnCAA0.51.11.33.6 <td></td> <td>GTA</td> <td>1.4</td> <td>2.1</td> <td>1.3</td> <td>1.8</td> <td>4.2</td>		GTA	1.4	2.1	1.3	1.8	4.2	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$		GIG	1.1	1.2	0	2.2	1.0	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Ser	TCT	2.3	1.5	0.8	1.3	1.2	
TCA 1.8 1.5 0.5 0.4 2.9 TCG 0 0.6 0.3 0.6 0.2 AGT 0.9 0.8 0.3 0.3 0 AGC 0.7 1.0 1.8 1.4 1.2 Pro CCT 1.4 1.7 1.8 0.5 1.2 CCC 0.2 0 0 0.3 0 CCC 0.2 0 0 0.3 0 CCC 0.2 0 0 0.3 0 CCG 0 0.6 0.3 2.5 0 Thr ACT 0.7 1.9 2.3 1.1 0.7 ACG 0.7 1.4 0.5 2.4 1.0 3.5 ACG 0.5 0 0.3 0.8 0.5 Ala GCT 2.7 2.1 2.8 2.6 1.5 GCC 1.1 0.4 0.8 <		TCC	0	0.4	0.3	1.5	0.5	
AGT0.90.80.30.30.3AGC0.71.01.81.41.2ProCCT1.41.71.80.51.2CCC0.2000.32.4CCG0.60.32.50ThrACT0.71.92.31.1ACC0.71.40.52.4ACG0.500.30.5ACC0.71.40.52.4ACC0.71.40.52.4ACC0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.500.30.8ACG0.50.20.23.2AlaGCG0.50.20.23.2TyrTAT4.33.52.81.0TAC0.21.22.51.53.4HisCAT0.91.42.50.7TAC0.51.11.31.00.5CAC0.51.11.31.05.4LysAAA3.43.13.64.13.4AAG4.7		TCA	1.8	1.5	0.5	0.4	2.9	
AG1 0.3 0.3 0.3 0.3 0.3 0.3 0.3 AGC 0.7 1.0 1.8 1.4 1.2 ProCCT 1.4 1.7 1.8 0.5 1.2 CCC 0.2 0 0 0.3 0 CCG 0.2 0 0 0.3 0 ThrACT 0.7 1.9 2.3 1.1 0.7 ACC 0.7 1.4 0.5 2.4 1.0 ACG 0.5 0 0.3 0.8 0.5 AlaGCT 2.7 1.7 3.1 0.3 5.1 AlaGCT 2.7 2.1 2.8 2.6 1.5 AlaGCT 2.7 2.1 2.8 2.6 1.5 GCG 0.5 0.2 0.2 3.2 0.5 AlaGCT 2.7 2.1 2.8 2.6 1.5 AlaGCT 2.7 2.1 2.8 2.6 1.5 GCG 0.5 0.2 0.2 3.2 0.7 3.4 HisCAT 0.9 1.4 2.5 0.7 1.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 AsnAAT 2.0 4.2 3.8 1.0 5.4 AsnAAT 2.0 4.2 3.8 1.0 5.4 LysAAA 3.4 3.1 3.6 4.1 3.4 AspGAT 5.8 <th< td=""><td></td><td>AGT</td><td>0</td><td>0.0</td><td>0.3</td><td>0.6</td><td>0.2</td></th<>		AGT	0	0.0	0.3	0.6	0.2	
ProCCT1.41.71.80.51.2CCG0.2000.32.4CCG00.60.32.50ThrACT0.71.92.31.10.7ACC0.71.40.52.41.0ACA2.71.73.10.35.1ACG0.500.30.80.5AlaGCT2.72.12.82.61.5GCA2.02.32.02.32.02.3GCG0.50.20.23.20TyrTAT4.33.52.81.03.9TAC0.21.22.51.53.4HisCAT0.91.42.50.71.5CAG0.50.60.51.20.2GlnCAA0.71.01.31.00.5AsnAAT2.04.23.81.05.4AAG1.60.61.53.22.2AspGAT5.85.65.92.53.9AspGAT5.85.65.92.53.9GluGAA4.33.73.74.91.7CysTGT1.41.41.00.40.7TGC0.71.01.50.50.50.5		AGC	0.9	0.8 1.0	1.8	0.3 1.4	1.2	
$\begin{array}{cccccc} CCC & 0.2 & 0 & 0 & 0.3 & 0 \\ CCA & 2.3 & 3.5 & 2.8 & 0.7 & 2.4 \\ CCG & 0 & 0.6 & 0.3 & 2.5 & 0 \\ \end{array}$ $\begin{array}{cccccccccccccccccccccccccccccccccccc$	Pro	ССТ	1.4	1.7	1.8	0.5	1.2	
$\begin{array}{ccccccc} CCA & 2.3 & 3.5 & 2.8 & 0.7 & 2.4 \\ CCG & 0 & 0.6 & 0.3 & 2.5 & 0 \\ \end{array}$ $\begin{array}{cccccccccccccccccccccccccccccccccccc$		CCC	0.2	0	0	0.3	0	
CCG00.60.32.50ThrACT0.71.92.31.10.7ACC0.71.40.52.41.0ACA2.71.73.10.35.1ACG0.500.30.80.5AlaGCT2.72.12.82.61.5GCA2.02.32.02.34.1GCG0.50.20.23.20TyrTAT4.33.52.81.03.9TAC0.21.22.51.53.4HisCAT0.91.42.50.71.5CAC0.50.60.51.20.2GlnCAA0.71.01.31.00.5AsnAAT2.04.23.81.05.4AAG4.73.53.31.33.9AspGAT5.85.65.92.53.9GluGAA4.33.73.74.91.7GAG1.61.42.03.00.2GluGAA4.33.73.74.91.7CysTGT1.41.41.00.40.7TGC0.71.01.50.50.50.5		CCA	2.3	3.5	2.8	0.7	2.4	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$		CCG	0	0.6	0.3	2.5	0	
ACC 0.7 1.4 0.5 2.4 1.0 ACA 2.7 1.7 3.1 0.3 5.1 ACG 0.5 0 0.3 0.8 0.5 AlaGCT 2.7 2.1 2.8 2.6 1.5 GCA 2.0 2.3 2.0 2.3 4.1 GCG 0.5 0.2 0.2 3.2 0 TyrTAT 4.3 3.5 2.8 1.0 3.9 TyrTAT 4.3 3.5 2.8 1.0 3.9 HisCAT 0.9 1.4 2.5 0.7 1.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 CAG 1.6 0.6 1.5 3.2 2.2 AsnAAT 2.0 4.2 3.8 1.0 5.4 AAC 0.5 1.1 1.3 2.8 2.7 LysAAA 3.4 3.1 3.6 4.1 3.4 AspGAT 5.8 5.6 5.9 2.5 3.9 GluGAA 4.3 3.7 3.7 4.9 1.7 CysTGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	Thr	ACT	0.7	1.9	2.3	1.1	0.7	
ACA 2.7 1.7 3.1 0.3 5.1 ACG 0.5 0 0.3 0.8 0.5 AlaGCT 2.7 2.1 2.8 2.6 1.5 GCA 2.0 2.3 2.0 2.3 2.0 2.3 GCG 0.5 0.2 0.2 3.2 0 TyrTAT 4.3 3.5 2.8 1.0 3.9 TyrTAT 4.3 3.5 2.8 1.0 3.9 HisCAT 0.9 1.4 2.5 0.7 1.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 CAG 1.6 0.6 1.5 3.2 2.2 AsnAAT 2.0 4.2 3.8 1.0 5.4 LysAAA 3.4 3.1 3.6 4.1 3.4 AspGAT 5.8 5.6 5.9 2.5 3.9 GluGAA 4.3 3.7 3.7 4.9 1.7 CysTGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5		ACC	0.7	1.4	0.5	2.4	1.0	
AlaGCT GCC GCA 2.7 1.1 0.4 0.8 2.3 2.3 2.0 		ACA ACG	0.5	1.7 0	3.1 0.3	0.3	5.1 0.5	
GCC1.1 0.4 0.8 2.2 1.2 GCA2.02.32.02.3 4.1 GCG 0.5 0.2 0.2 3.2 0 TyrTAT 4.3 3.5 2.8 1.0 3.9 HisCAT 0.9 1.4 2.5 0.7 1.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 AsnAAT 2.0 4.2 3.8 1.0 5.4 LysAAA 3.4 3.1 3.6 4.1 3.4 AspGAT 5.8 5.6 5.9 2.5 3.9 AspGAT 5.8 5.6 5.9 2.5 3.9 GluGAA 4.3 3.7 3.7 4.9 1.7 CysTGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	Ala	GCT	2.7	2.1	2.8	2.6	1.5	
GCA 2.0 2.3 2.0 2.3 4.1 GCG 0.5 0.2 0.2 3.2 0 Tyr TAT 4.3 3.5 2.8 1.0 3.9 His CAT 0.9 1.4 2.5 0.7 1.5 Gln CAA 0.7 1.0 1.3 1.0 0.5 Gln CAA 0.7 1.0 1.3 1.0 0.5 Asn AAT 2.0 4.2 3.8 1.0 5.4 Lys AAA 3.4 3.1 3.6 4.1 3.4 Asp GAT 5.8 5.6 5.9 2.5 3.9 Asp GAT 5.8 5.6 5.9 2.5 3.9 Glu GAA 4.3 3.7 3.7 4.9 1.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 I.0 1.5 0.5 0.5 0.5 0.5 0.5		GCC	1.1	0.4	0.8	2.2	1.2	
GCG 0.5 0.2 0.2 3.2 0 Tyr TAT 4.3 3.5 2.8 1.0 3.9 His CAT 0.2 1.2 2.5 1.5 3.4 His CAT 0.9 1.4 2.5 0.7 1.5 Gln CAA 0.7 1.0 1.3 1.0 0.5 Gln CAA 0.7 1.0 1.3 1.0 0.5 Asn AAT 2.0 4.2 3.8 1.0 5.4 Lys AAA 3.4 3.1 3.6 4.1 3.4 Asp GAT 5.8 5.6 5.9 2.5 3.9 Asp GAT 5.8 5.6 5.9 2.5 3.9 Glu GAA 4.3 3.7 3.7 4.9 1.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 ToC 0.7 1.0 1.5 0.5 0.5 0.5		GCA	2.0	2.3	2.0	2.3	4.1	
TyrTAT TAC4.3 0.2 3.5 1.2 2.8 2.5 1.0 1.5 3.9 3.4 HisCAT CAC0.9 0.5 1.4 0.6 2.5 0.5 0.7 1.2 1.5 0.2 GlnCAA CAG0.7 1.6 1.0 0.6 1.3 1.5 1.0 		GCG	0.5	0.2	0.2	3.2	0	
TAC 0.2 1.2 2.5 1.5 3.4 HisCAT 0.9 1.4 2.5 0.7 1.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 GlnCAA 0.7 1.0 1.3 1.0 0.5 AsnAAT 2.0 4.2 3.8 1.0 5.4 LysAAA 3.4 3.1 3.6 4.1 3.4 AspGAT 5.8 5.6 5.9 2.5 3.9 GluGAA 4.3 3.7 3.7 4.9 1.7 CysTGT 1.4 1.4 1.0 0.4 0.7	Tyr	TAT	4.3	3.5	2.8	1.0	3.9	
HisCAT CAC 0.9 0.5 1.4 0.6 2.5 0.5 0.7 1.2 1.5 0.2 GlnCAA CAG 0.7 1.6 1.0 0.6 1.3 1.5 1.0 3.2 0.5 2.2 AsnAAT AAC 2.0 0.5 4.2 1.1 3.8 1.3 1.0 2.8 5.4 2.7 LysAAA AAG 3.4 4.7 3.1 3.5 3.6 3.3 4.1 3.3 3.4 3.9 AspGAT GAC 5.8 1.6 5.6 1.4 5.9 2.0 2.5 3.0 3.9 0.2 GluGAA GAG 4.3 4.9 3.7 3.5 3.7 3.8 4.9 1.8 1.7 2.7 CysTGT TGC 1.4 0.7 1.4 1.0 1.4 1.5 0.5 0.5		TAC	0.2	1.2	2.5	1.5	3.4	
Gln CAA 0.7 1.0 1.3 1.0 0.5 Asn AAT 2.0 4.2 3.8 1.0 5.4 Lys AAA 3.4 3.1 3.6 4.1 3.4 Asp GAT 5.8 5.6 5.9 2.5 3.9 Glu GAA 4.3 3.7 3.7 4.9 1.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 Lys AAA 3.4 3.1 3.6 4.1 3.4 Asp GAT 5.8 5.6 5.9 2.5 3.9 GAC 1.6 1.4 2.0 3.0 0.2 Glu GAA 4.3 3.7 3.7 4.9 1.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	His	CAT	0.9	1.4	2.5	0.7	1.5	
GinCAA CAG 0.7 1.6 1.0 1.3 1.0 1.5 0.5 2.2 AsnAAT AAC 2.0 0.5 4.2 1.1 3.8 1.3 1.0 2.8 5.4 2.7 LysAAA AAG 3.4 4.7 3.1 3.5 3.6 3.3 4.1 3.4 3.4 3.9 AspGAT GAC 5.8 1.6 5.6 1.4 5.9 2.0 2.5 3.0 3.9 AspGAT GAC 5.8 1.6 5.6 1.4 2.0 3.0 3.0 0.2 GluGAA GAG 4.3 4.9 3.7 3.5 3.8 3.8 1.8 2.7 CysTGT TGC 1.4 0.7 1.4 1.0 1.6 0.5 0.5		C.A.C.	0.5	0.0	0.5	1.2	0.2	
Asn AAT 2.0 4.2 3.8 1.0 5.4 Lys AAA 3.4 3.1 1.3 2.8 2.7 Lys AAA 3.4 3.1 3.6 4.1 3.4 Asp GAT 5.8 5.6 5.9 2.5 3.9 Glu GAA 4.3 3.7 3.7 4.9 1.7 Gus GAG 4.9 3.5 3.8 1.8 2.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 Gold GAA 4.3 3.7 3.7 4.9 1.7 GAG 4.9 3.5 3.8 1.8 2.7	Gin	CAA CAG	0.7 1.6	1.0 0.6	1.3 1.5	1.0 3.2	0.5 2.2	
AshAA1 2.0 4.2 3.8 1.0 5.4 AAC 0.5 1.1 1.3 2.8 2.7 LysAAA 3.4 3.1 3.6 4.1 3.4 AAG 4.7 3.5 3.3 1.3 3.9 AspGAT 5.8 5.6 5.9 2.5 3.9 GaC 1.6 1.4 2.0 3.0 0.2 GluGAA 4.3 3.7 3.7 4.9 1.7 GAG 4.9 3.5 3.8 1.8 2.7 CysTGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	A am	A A T	2.0	4.2	2.0	1.0	 c .	
Lys AAA AAG 3.4 4.7 3.1 3.5 3.6 3.3 4.1 1.3 3.4 3.9 Asp GAT GAC 5.8 1.6 5.6 1.4 5.9 2.0 2.5 3.0 3.9 0.2 Glu GAA GAG 4.3 4.9 3.7 3.5 3.7 3.8 1.8 2.7 Cys TGT TGC 1.4 1.4 1.0 1.5 0.5 0.5	Asn	AAT	2.0 0.5	4.2 1.1	3.8 1.3	1.0 2.8	5.4 2.7	
LysAAA 3.4 3.1 3.6 4.1 3.4 AAG 4.7 3.5 3.3 1.3 3.9 AspGAT 5.8 5.6 5.9 2.5 3.9 GAC 1.6 1.4 2.0 3.0 0.2 GluGAA 4.3 3.7 3.7 4.9 1.7 GAG 4.9 3.5 3.8 1.8 2.7 CysTGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	Luc		2.4	2 1	26	4.1	2.4	
Asp GAT GAC 5.8 1.6 5.6 1.4 5.9 2.0 2.5 3.0 3.9 0.2 Glu GAA GAG 4.3 4.9 3.7 3.5 3.7 3.8 1.8 2.7 Cys TGT TGC 1.4 1.4 1.0 0.4 0.7	Lys	AAA AAG	3.4 4.7	3.1 3.5	3.6 3.3	4.1 1.3	3.4 3.9	
Alsp GAT 5.6 5.6 5.7 2.5 3.7 GAC 1.6 1.4 2.0 3.0 0.2 Glu GAA 4.3 3.7 3.7 4.9 1.7 GAG 4.9 3.5 3.8 1.8 2.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	Asn	GAT	58	56	5 0	25	2.0	
Glu GAA 4.3 3.7 3.7 4.9 1.7 GAG 4.9 3.5 3.8 1.8 2.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5		GAC	1.6	1.4	2.0	3.0	0.2	
GAG 4.9 3.5 3.8 1.8 2.7 Cys TGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5	Glu	GAA	4.3	3.7	3.7	4.9	17	
Cys TGT 1.4 1.4 1.0 0.4 0.7 TGC 0.7 1.0 1.5 0.5 0.5		GAG	4.9	3.5	3.8	1.8	2.7	
TGC 0.7 1.0 1.5 0.5 0.5	Cys	TGT	1.4	1.4	1.0	0.4	0.7	
		TGC	0.7	1.0	1.5	0.5	0.5	

Continued ^{co}

TABLE 1-Continued

		Frequency (mol%) of codon usage					
Amino acid	Codon	B. fibrisolvens			Е.	B. fibrisolvens ^b	
		ORF1	ORF2	ORF3	colia	xylA	
Тгр	TGG	1.3	1.7	1.3	0.7	2.0	
Arg	CGT	0.7	0.8	1.3	3.1	1.0	
-	CGC	0	0.2	0.5	2.0	0.2	
	CGA	0.2	0.2	0.2	0.2	0	
	CGG	0	0	0	0.2	0	
	AGA	3.1	2.9	2.0	0.1	2.0	
	AGG	0.5	0	0.8	0.1	0.2	
Gly	GGT	3.4	1.7	1.5	3.8	2.4	
	GGC	2.3	2.1	1.0	3.1	1.5	
	GGA	3.1	4.4	1.5	0.4	2.7	
	GGG	0.5	0.2	0.8	0.6	0.5	

^a Reference 1.

^b Reference 25.

activity was observed with pLOI1005, although pLOI1043 contained a smaller fragment of *B. fibrisolvens* DNA which included the complete xylB gene. During subcloning, a clone was found in which two *SspI* fragments each containing xylB had been inserted (pLOI1050). Both fragments were oriented to allow transcription from the *lac* promoter. This double construct was more active than the single insertion (pLOI1043) but less active than the original construct (pLOI1005).

Utilization of other nitrophenyl substrates. The activities of the xylB-encoded protein with various para- and orthonitrophenol-glycosidic substrates was tested (Table 3). A 19-fold-higher activity was observed with o-nitrophenyl- β -D-xylopyranoside than with the p-nitrophenyl derivative. No activity above the control level was detected with the other substrates. The low level of activity observed with o-nitrophenyl- β -D-fucopyranoside may not be significant. Activity with this substrate was 0.5% of that with the analogous xyloside derivative.

SDS-PAGE analysis of cloned proteins. A new protein band of approximately 60,000 molecular weight was present in cells harboring pLOI1005, the most active clone, and was absent in the pUC18 control (Fig. 3). Extracts from cells harboring pLOI1040 in which a *Not*I linker was used to disrupt the *xylB* gene also lacked this new protein. The amount of this protein was reduced in the *SspI* clone (pLOI1043) compared with pLOI1005, consistent with mea-

TABLE 2. Expression of xylB activities in recombinant E. coli: β -D-xylopyranosidase and α -L-arabinofuranosidase activities

Dloomid		Sp act ^a	
Plasmid	Xylosidase	Arabinosidase	Ratio ara/xyl ^b
pLOI1005	8.9	15.5	1.76
pLOI1040	0.2	0.2	
pLOI1043	1.9	3.0	1.65
pLOI1050	6.8	10.2	1.52
pUC18	0.2	0.2	

^a Nanomoles per minute per milligram of cell protein.

 b Ratio calculated after subtraction of background values from the pUC18 control.

 TABLE 3. Hydrolysis of different nitrophenyl-glycosides by the xylB gene product

Substrate	Sp act		
Substrate	pLOI1005	pUC18	
<i>p</i> -Nitrophenyl-β-D-xylopyranoside	8.9	0.20	
<i>p</i> -Nitrophenyl-α-L-arabinofuranosidase	15.5	0.20	
p -Nitrophenyl- α -L-arabinopyranoside	0.14	0.21	
p -Nitrophenyl- α -D-galactopyranoside	0.17	0.16	
p-Nitrophenyl- α -D-glucopyranoside	3.45	3.21	
<i>p</i> -Nitrophenyl-β-D-mannopyranoside	0.10	0.12	
p -Nitrophenyl- α -D-mannopyranoside	0.15	0.14	
p -Nitrophenyl- α -L-fucopyranoside	0.19	0.21	
p -Nitrophenyl- β -D-fucopyranoside	0.22	0.24	
p -Nitrophenyl- β -L-fucopyranoside	0.42	0.41	
<i>p</i> -Nitrophenyl-α-L-rhamnopyranoside	0.28	0.20	
o-Nitrophenyl-β-D-xylopyranoside	195	3.1	
o-Nitrophenyl- β -D-fucopyranoside	1.44	0.98	
o-Nitrophenyl-α-D-galactopyranoside	0.78	1.14	
o-Nitrophenyl-β-D-galactopyranoside	1.00	1.05	

surements of specific activity (Table 2). The recombinant containing a double insertion of the *xylB* gene (pLOI1050) contained higher levels of this protein than the single-insertion recombinant (pLOI1043).

Comparison of *B. fibrisolvens* ORFs with other related genes. The three ORFs from *B. fibrisolvens* exhibited 42 to 45% amino acid similarity to each other with 14 to 19% amino acid identity (Table 4). Comparison of *xylB* (ORF2) with GenBank and EMBL sequence libraries identified the β -glucosidase gene from *Kluyveromyces fragilis* as being most similar, with 44% similarity and 20% identity. Additional comparisons were also made to other xylan-degrading enzymes and selected glycosidases. No genes were found which showed a high percentage of amino acid identity to the *B. fibrisolvens xylB* gene. The *Bacillus pumilus* β -D-xylosidase gene exhibited higher identity in primary structure (21%), although this translated sequence exhibited poor overall similarity, only 28%. The identity between these proteins was more prevalent in the amino-terminal region.

The translated sequences for ORF1 and ORF3 did not share strong identity with previously sequenced genes. Both were also most similar to the gene encoding *K. fragilis* β -glucosidase. Deduced amino acid sequences for the two



FIG. 3. SDS-PAGE of cytoplasmic cell extracts from recombinant *E. coli* harboring selected constructs. Approximately 15 μ g of soluble protein was loaded in each lane. Lanes 1 and 7, molecular weight markers (×10³); lane 2, DH5 α (pLOI1005); lane 3, DH5 α (pUC18); lane 4, DH5 α (pLOI1040); lane 5, DH5 α (pLOI1043); lane 6, DH5 α (pLOI1050). The band corresponding to the xylosidasearabinosidase enzyme is marked with an arrowhead.

incomplete genes exhibited similarity to other xylan-degrading enzymes from *Bacillus subtilis* and *Bacillus pumilus* and to cellulases from *Clostridium thermocellum*.

Comparison of the translated protein sequence of ORF2 (xylB) with those of hen egg white lysozyme (HEWL), *Aspergillus niger* glucoamylase, xylanases, and cellulases revealed a region homologous to the HEWL active site (Table 5). Glucoamylase provided the best match in this region, with identity between 8 of the 21 amino acids shown. The glutamate residues, the asparagine, and spacing were all conserved. The conserved regions of other carbohydrate hydrolases are also presented for comparison. In the *B. fibrisolvens xylB* protein and in glucoamylase, one glutamate in HEWL is replaced by aspartate.

DISCUSSION

The xylB gene was initially cloned into pUC18 on a 4.2-kbp segment of B. fibrisolvens DNA (35). This gene is present as a single chromosomal copy (35). Based on comparisons of deduced amino acid sequence, xylB does not appear to be closely related to the two xylosidase genes from Caldocellum saccharolyticum (22), the Bacillus pumilus xylosidase (26), or other glycohydrolases. It does, however, retain a conserved region which has been previously implicated in catalytic function. Two incomplete ORFs were found on either side of xylB which do not appear to exhibit enzymatic activity. Although the identities of ORF1 and ORF3 are unknown, the similarity in amino acid sequences to other glycohydrolases is consistent with their involvement in xylan degradation.

The xylanase and xylosidase in B. fibrisolvens GS113 have been shown to be under coordinate control, induced by xylan and repressed by glucose (34). The control of arabinosidase activity was not investigated. A cadre of enzymes are needed to hydrolyze the backbone and substituents of plant xylans including xylanase, xylosidase, xylobiase, acetyl-esterase, arabinosidase, glucosidase, glucuronidase, and others which may be under similar control. The three ORFs described share sequence similarity and may all be involved in hemicellulose hydrolysis. The proximity of ORF1, ORF2 (xylB), and ORF3 on the cloned DNA is suggestive of organization for common control. It is unlikely that a terminator is present between ORF1 and ORF2 since these are separated by only 15 bases. No terminator functions in E. coli between these two ORFS. The ORF2 protein is expressed in large amounts in E. coli from the lac promoter in constructs which include ORF1. It is also unlikely that a terminator is present between ORF2 and ORF3. No terminator or prominent stem-loop was identified by computer analysis. This region of DNA does not function as a terminator in E. coli. Dual insertions of consecutive SspI fragments containing the xylB coding region through part of ORF3 (pLOI1050) expressed over twice the enzyme activity of single-insertion constructs (pLOI1043) in E. coli. These data are consistent with the presence of an operon which includes these three ORFs from *B. fibrisolvens*.

The gene encoding xylosidase and arabinosidase activities, xylB, is the first of its kind to be sequenced. The bifunctionality of this gene product has been correlated with the presence of a single new protein in SDS-PAGE. Further evidence for the bifunctional nature of this enzyme includes the subcloning of a minimal DNA segment from which both activities are expressed, the establishment that both activities require the same direction of transcription, and the loss

1232 UTT ET AL.

TABLE 4. Comparison of the translated amino acid sequence	es of the three B . fibrisolven	as ORFs in pLOI1005 with the	ose of other genes
---	--	------------------------------	--------------------

	% Similarity (% identity)			Deferrer
Organism (gene)	ORF1	ORF2 (xylB)	ORF3	Reference
Butyrivibrio fibrisolvens (xynA)	40 (16)	41 (16)	41 (13)	25
Butyrivibrio fibrisolvens (end1)	16 (8)	14 (6)	20 (11)	4
Bacillus pumilus (xynA)	46 (21)	44 (16)	12 (8)	13
Bacillus pumilus (xynB)	39 (16)	28 (21)	38 (17)	26
Bacillus subtilis (xynA)	48 (21)	44 (16)	12 (6)	28
Caldocellum saccharolyticum (xynB)	20 (14)	25 (12)	21 (12)	22
Caldocellum saccharolyticum (xynA/xynB)	16 (8)	20 (9)	19 (10)	22
Caldocellum saccharolyticum (xynC)	16 (8)	22 (11)	31 (3)	22
Clostridium thermocellum (xynZ)	18 (7)	40 (16)	15 (8)	15
Clostridium thermocellum (celA)	45 (22)	41 (18)	39 (15)	3
Clostridium thermocellum (celB)	44 (20)	44 (18)	43 (19)	14
Clostridium thermocellum (celD)	45 (16)	39 (15)	43 (18)	19
Aspergillus niger (glucoamylase)	40 (16)	41 (18)	43 (19)	5
Kluyveromyces fragilis (β-glucosidase)	52 (31)	44 (20)	46 (22)	32
ORF1	100	44 (19)	45 (19)	This study
ORF2	44 (19)	100	42 (14)	This study
ORF3	45 (19)	42 (14)	100	This study

of both activities after disruption of the predicted reading frame by the insertion of a 10-bp linker.

A recent study with *Bacteriodes ovatus* has reported the cloning and expression of xylanosidase and arabinosidase activities (40). These activities were both encoded by DNA within a cluster of genes involved in hemicellulose degradation. All subclones with active xylosidase also exhibited arabinosidase activity. Both activities copurified and were proposed to reside on a single gene product. A bifunctional

xylanase-xylosidase has been cloned and sequenced from *Caldocellum saccharolyticum* (22) but lacks arabinosidase activity.

It has been suggested that the hydrolytic mechanism of lysozyme (36) and taka-amylase can serve as a model for other carbohydrate hydrolases (20). Studies of HEWL revealed a general acid-base catalysis involving Glu-35 and Asp-52 as important catalytic residues (31). Subsequent studies have demonstrated that this catalytic region is con-

Organism or protein"	Sequence	Reference
	35 44 52	<u> </u>
HEWL	F <u>E</u> S N F N T Q A T . <u>N</u> R N T D G S T <u>D</u> Y 331	41
A.n. (glu)	PEDTY.YNGNPWFLCTLAAAEQ 342 350 362	5
B.f. xylB	SEDFYSLTDNPGFLRLKLRPEA 320 328 337	This study
B.f. endl	GETSATNRN. NTAERVKWA. DY 355 362 369	4
B.f. xynA	N <u>E</u> K P L I W S <u>N</u> I G V A K P A Y <u>D</u> E 325 335 343	25
B.p. xynB	I E C T R L A Q L N W N T C S M Q F V E E 458 466 475	26
C.t. xynZ	GEALLRADV. NRSGKVDS. TDY 130 138 149	15
C.t. celA	Q E V V N Y C I D. NKMYVILNTHHDV 418 427 436	3
C.t. celB	TEGGHPLLDL.NLKYLRCMR.DF 376 384 395	14
C.t. celD	DEEYLRDFE. NRAAQFSKKEADF 408 418 427	19
C.s. xynB	REVFVERIDEYNANPKRVWL.EM 244 254 263	22
T.r. CBH II	L <u>E</u> CINYAVTQL <u>N</u> LPNVAMYL <u>D</u> A 586 596 605	33
K.f. (β-glucosidase)	GEWETEGYDRENMDLPKRTN. EL 33 42 50	32
S.c. EG I	N <u>E</u> SCAEFGNQ. <u>N</u> IPGVKNT <u>D</u> Y	41

TABLE 5. Amino acid sequence alignment of conserved regions

^a Abbreviations: A.n., Aspergillus niger; glu, glucoamylase; B.f., B. fibrisolvens endoglucanase 1, xylosidase; B.p., Bacillus pumilus xylosidase; C.t., Clostridium thermocellum endoglucanases; C.s., Caldocellum saccharolyticum xylosidase; T.r., Trichoderma reesei cellobiohydrolase II; K.f., Kluyveromyces fragilis; S.c., Schizophyllum commune endoglucanase I.

served in cellulase enzymes and amylases (20, 33). On the basis of the similarity of sequence in this region, Morosoli et al. (27) proposed a similar catalytic mechanism for the xylanase from *Cryptococcus albidus*. The *B. fibrisolvens* xylB gene product also contains elements of this conserved region. Thus, it seems likely that the *B. fibrisolvens xylB* product also shares a similar catalytic mechanism.

Substrate ambiguity between xylanase and carboxymethyl cellulases is well known (11). Substrate ambiguity among xylosidases also occurs and has recently been reported for cloned genes from Bacteroides ovatus (40) and Caldocellum saccharolyticum (22). By using enzyme kinetic methods, Uziie et al. (37) demonstrated that a bifunctional β -Dxylosidase-B-glucosidase from Chaetomium trilaterale contained a single active site with dual binding regions (37). Purified β -xylosidase from *Trichoderma reesei* was reported to also exhibit both xylosidase and arabinosidase activities (30). Since rotation about the α -1,3 glycosidic bond that links the arabinofuranosyl residue to xylopyranose produces a bond conformation that resembles the β -1,4 linkage between xylose residues, a single catalytic site could be responsible for both enzymatic activities in the B. fibrisolvens xylB gene product.

It seems reasonable to speculate that bifunctionality among cellulases and hemicellulases is common in the microbial world. The structural similarities between the various cellulosic and hemicellulosic substrates which occur together in lignocellulose may be responsible for the apparent evolution of bifunctionality. Since many activities are required for the complete degradation of xylans and cellulose, the evolution of bifunctional enzymes could be of selective advantage in the rumen and other environments.

ACKNOWLEDGMENTS

We gratefully acknowledge the helpful discussions and assistance provided by R. B. Hespell.

These studies were supported in part by the Florida Agricultural Experimental Station and by a Gas Research Institute/University of Florida Cooperative Research Program.

REFERENCES

- 1. Allf-Steinberger, C. 1984. Evidence for coding pattern on the non-coding strand of the *Escherichia coli* genome. Nucleic Acids Res. 12:2235–2241.
- 2. Armstrong, D. G., and H. J. Gilbert. 1985. Biotechnology and the rumen: a mini-review. J. Sci. Food Agric. 36:1039-1046.
- Benguin, P., P. Cornet, and J. P. Aubert. 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium ther*mocellum. J. Bacteriol. 162:102-105.
- Berger, E., W. A. Jones, D. T. Jones, and D. R. Woods. 1989. Cloning and sequencing of an endoglucanase (endl) gene from Butyrivibrio fibrisolvens H17c. Mol. Gen. Genet. 219:193–198.
- Boel, E., M. T. Hansen, I. Hjort, and N. P. Fiil. 1984. Two different types of intervening sequences in the glucoamylase gene from Aspergillus niger. EMBO J. 3:1581–1585.
- 6. Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of proteins utilizing the principle of protein-dye binding. Anal. Biochem. 72:248-254.
- Cotta, M. A., and R. B. Hespell. 1986. Proteolytic activity of the ruminal bacteria *Butyrivibrio fibrisolvens*. Appl. Environ. Microbiol. 52:51-58.
- Dehority, B. A. 1966. Characterization of several bovine rumen bacteria isolated with a xylan medium. J. Bacteriol. 91:1724– 1729.
- 9. Dehority, B. A. 1968. Mechanism of isolated hemicellulose and xylan degradation by cellulolytic rumen bacteria. Appl. Microbiol. 16:781-786.
- 10. Dekker, R. F. H., and G. N. Richards. 1976. Hemicellulases: their occurrence, purification, properties, and mode of action.

Adv. Carbohydr. Chem. Biochem. 32:277-352.

- Flint, H. J., C. A. McPherson, and J. Bisset. 1989. Molecular cloning of genes from *Ruminococcus flavefaciens* encoding xylanase and β(1-3, 1-4)glucanase activities. Appl. Environ. Microbiol. 55:1230-1233.
- 12. Forsberg, C. W., B. Crosby, and D. Y. Thomas. 1986. Potential for manipulation of the rumen fermentation through the use of recombinant DNA techniques. J. Anim. Sci. 63:310–325.
- 13. Fukusaki, E., W. Panbangred, A. Shinmyo, and H. Okada. 1984. The complete nucleotide sequence of the xylanase gene (xylA) of *Bacillus pumilus*. FEBS Lett. 171:197–201.
- Grepinet, O., and P. Benguin. 1986. Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B. Nucleic Acids Res. 14:1791–1799.
- 15. Grepinet, O., M.-C. Chebrou, and P. Beguin. 1988. Nucleotide sequence and deletion analysis of the xylanase gene (xylZ) of Clostridium thermocellum. J. Bacteriol. 170:4582-4588.
- Hespell, R. B., and P. J. O'Bryan-Shah. 1988. Esterase activities in *Butyrivibrio fibrisolvens* strains. Appl. Environ. Microbiol. 54:1917-1922.
- Hespell, R. B., R. Wolf, and R. J. Bothast. 1987. Fermentation of xylans by *Butyrivibrio fibrisolvens* and other ruminal bacterial species. Appl. Environ. Microbiol. 53:2849–2853.
- Hobson, P. N., and M. R. Purdom. 1961. Two types of xylan fermenting bacteria from the sheep rumen. J. Appl. Bacteriol. 24:188–193.
- Joliff, G., P. Benguin, and J. P. Aubert. 1986. Nucleotide sequence of the cellulase gene *celD* encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. 14:8605–8613.
- 20. Knowles, J., P. Lehtovaara, and T. Teeri. 1987. Cellulase families and their genes. Trends Biotechnol. 5:255–261.
- 21. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) 227:680-685.
- Luthi, E., D. R. Love, J. McAnulty, C. Wallace, P. A. Caughey, D. Saul, and P. Bergquist. 1990. Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile *Caldocellum saccharolyticum*. Appl. Environ. Microbiol. 56:1017-1024.
- Lynch, J. M. 1987. Utilization of lignocellulosic wastes. J. Appl. Bacteriol. Symp. Suppl. 16:71S-83S.
- 24. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 25. Mannarelli, B. M., S. Evans, and D. Lee. 1990. Cloning, sequencing, and expression of a xylanase gene from the anaerobic rumen bacterium *Butyrivibrio fibrisolvens*. J. Bacteriol. 172: 4247-4254.
- Moriyama, H., E. Fufusaki, J. Cabrera Crespo, A. Shinmyo, and H. Okada. 1987. Structure and expression of genes coding for xylan-degrading enzymes of *Bacillus pumilus*. Eur. J. Biochem. 166:539-545.
- 27. Morosoli, R., C. Roy, and M. Yaguchi. 1986. Isolation and partial primary sequence of a xylanase from the yeast *Cryptococcus albidus*. Biochim. Biophys. Acta 870:473–478.
- 28. Paice, M. G., R. Bourbonnais, M. Desrochers, L. Jurasek, and M. Yaguchi. 1986. A xylanase gene from *Bacillus subtilis*: nucleotide sequence and comparison with *B. pumilus* gene. Arch. Microbiol. 144:201-206.
- Patterson, J. A. 1989. Prospects for establishment of genetically engineered microorganisms in the rumen. Enzyme Microb. Technol. 11:187-189.
- Poutanen, K., and J. Puls. 1988. Characteristics of *Trichoderma* reesei β-xylosidase and its use in the hydrolysis of solubilized xylans. Appl. Microbiol. Biotechnol. 28:425–432.
- Quiocho, F. A. 1986. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. Annu. Rev. Biochem. 55:287-315.
- 32. Raynal, A., C. Gerbaud, M. C. Francingues, and M. Guerineau. 1987. Sequence and transcription of the β-glucosidase gene of *Kluyveromyces fragilis* cloned in *Saccharomyces cerevisiae*. Curr. Genet. 12:175–184.
- 33. Rouvinen, J., T. Bergfors, T. Teeri, J. K. C. Knowles, and T. A.

Jones. 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249**:380–386.

- 34. Sewell, G. W., H. C. Aldrich, D. Williams, B. Mannarelli, A. Wilkie, R. B. Hespell, P. H. Smith, and L. O. Ingram. 1988. Isolation and characterization of xylan-degrading strains of *Butyrivibrio fibrisolvens* from a Napier grass-fed anaerobic digester. Appl. Environ. Microbiol. 54:1085–1090.
- 35. Sewell, G. W., E. A. Utt, R. B. Hespell, K. F. MacKenzie, and L. O. Ingram. 1989. Identification of the *Butyrivibrio fibrisol*vens xylosidase gene (xylB) coding region. Appl. Environ. Microbiol. 55:306-311.
- Teeri, T. T., P. Lehtovaara, S. Kauppinen, I. Salovouri, and J. Knowles. 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase. Gene 51:43-52.
- Uzie, M., M. Matsuo, and T. Yasui. 1985. Possible identity of β-xylosidase and β-glucosidase of *Chaetomium trilaterale*. Agric. Biol. Chem. 49:1167-1173.
- Ward, O. P., and M. Moo-Young. 1989. Enzymatic degradation of cell wall and related plant polysaccharides. Crit. Rev. Biotechnol. 8:237-274.
- 39. Weinstein, L., and P. Albersheim. 1979. Structure of plant cell walls. Plant Physiol. 63:425-432.
- 40. Whitehead, T. R., and R. B. Hespell. 1990. The genes for xylan-degrading activities from *Bacteroides ovatus* are clustered in a 3.8-kb region. J. Bacteriol. 172:2408-2412.
- 41. Yaguchi, M., C. Roy, C. F. Rollin, M. G. Paice, and L. Jurasek. 1983. A fungal cellulase shows sequence homology with the active site of hen egg-white lysozyme. Biochem. Biophys. Res. Commun. 116:408-411.