APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 1991, p. 1227–1234
0099-2240/91/041227-08$02.00/0

Vol. 57, No. 4

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 *Not*I 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 product did not exhibit a high degree of identity with other xylan-degrading enzymes or glycosidases. 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-1,4-xylosidic 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* DH5α 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 μg 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 × *g*, 10 min, 4°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/in^2. Membranes and debris were removed by centrifugation (100,000 × *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 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*.

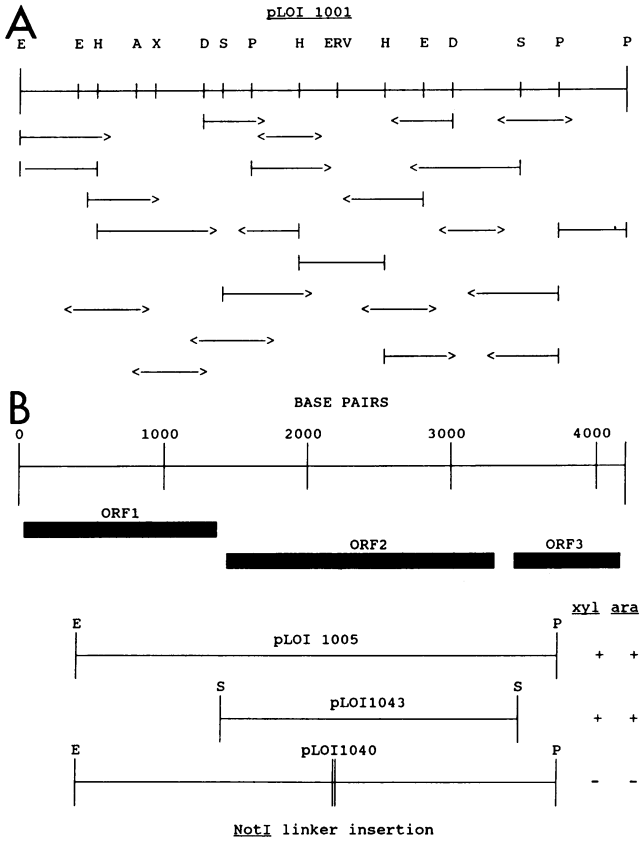

FIG. 1. Sequencing strategy and map of the 4.2-kbp DNA fragment from *B. fibrisolvens*. (A) Sequencing strategy of insert in pLOI1001. Arrows indicate direction of sequencing. Vertical bars at the front of sequencing arrows indicate sequencing from a subclone with universal primers for pUC18 as opposed to sequencing from within *B. fibrisolvens* DNA with oligonucleotide primers (bars absent). Abbreviations: E, *Eco*RI; H, *Hin*dIII; A, *Acc*I; X, *Xba*I; D, *Dra*I; S, *Ssp*I; P, *Pst*I; 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 pLOI1040 indicate the site at which a 10-bp *Not*I linker was inserted. Abbreviations are the same as those for panel A.

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 *Ssp*I 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

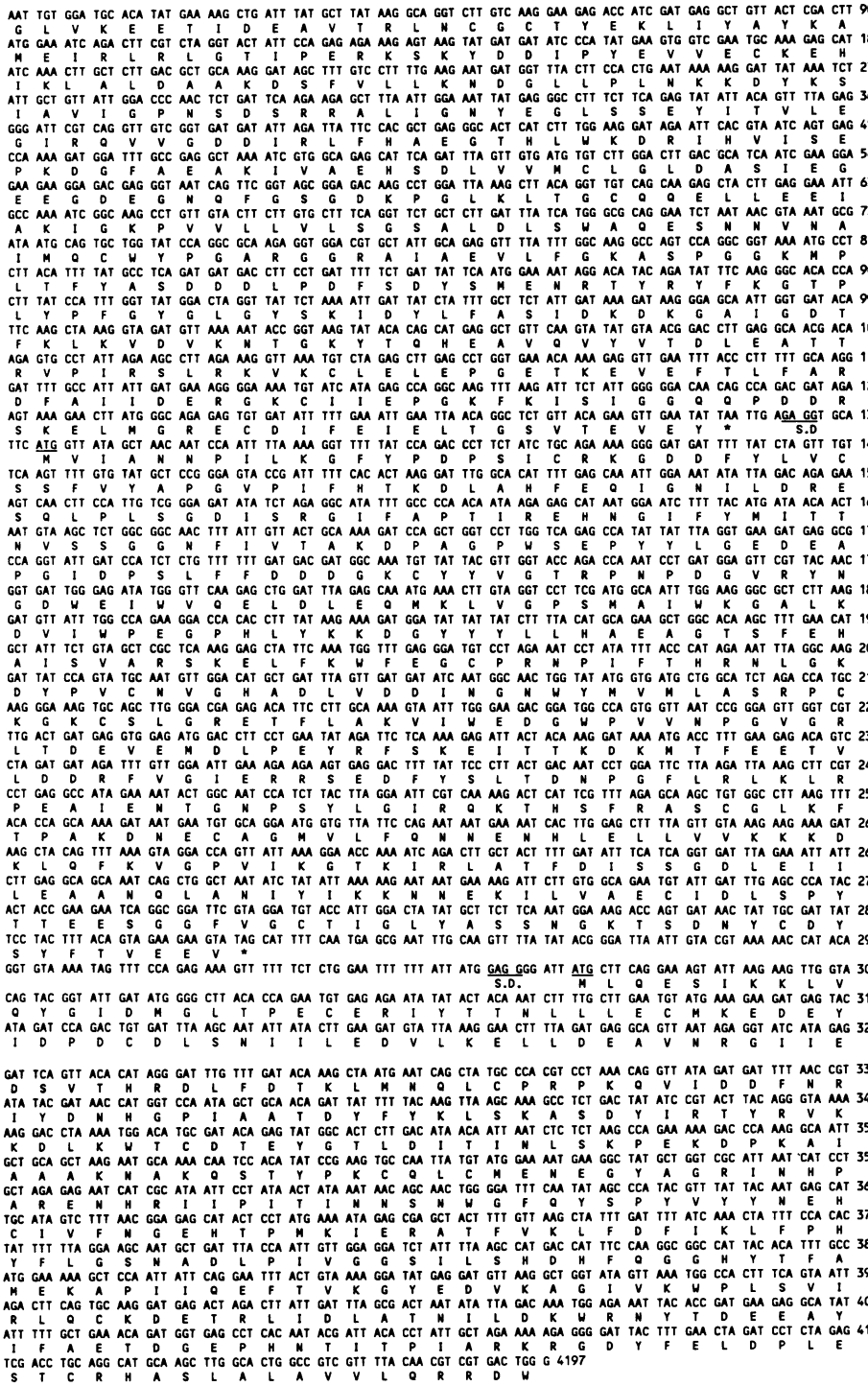

FIG. 2. Nucleotide sequence and translated amino acid sequence for the 4.2-kbp insert in pLOI1001. Putative Shine-Dalgarno (S.D.) regions and initiation codons are underlined. Translational termination is indicated by an asterisk (*).

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

Amino acid	Codon	Frequency (mol%) of codon usage				
		B. fibrisolvens			*E. coli*[a]	*B. fibrisolvens*[b] *xylA*
		ORF1	ORF2	ORF3		
Phe	TTT	2.7	4.1	3.6	1.3	2.9
	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
	CTT	5.6	2.7	2.5	0.8	4.4
	CTC	0	0	0.3	0.8	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	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
	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
	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
	CCA	2.3	3.5	2.8	0.7	2.4
	CCG	0	0.6	0.3	2.5	0
Thr	ACT	0.7	1.9	2.3	1.1	0.7
	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
Ala	GCT	2.7	2.1	2.8	2.6	1.5
	GCC	1.1	0.4	0.8	2.2	1.2
	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
	TAC	0.2	1.2	2.5	1.5	3.4
His	CAT	0.9	1.4	2.5	0.7	1.5
	CAC	0.5	0.6	0.5	1.2	0.2
Gln	CAA	0.7	1.0	1.3	1.0	0.5
	CAG	1.6	0.6	1.5	3.2	2.2
Asn	AAT	2.0	4.2	3.8	1.0	5.4
	AAC	0.5	1.1	1.3	2.8	2.7
Lys	AAA	3.4	3.1	3.6	4.1	3.4
	AAG	4.7	3.5	3.3	1.3	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
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
Trp	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 *Ssp*I 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 *ortho*-nitrophenol-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 *Ssp*I 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

Plasmid	Sp act[a]		
	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	
	pLOI1005	pUC18
p-Nitrophenyl-β-D-xylopyranoside	8.9	0.20
p-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
p-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
p-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 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.

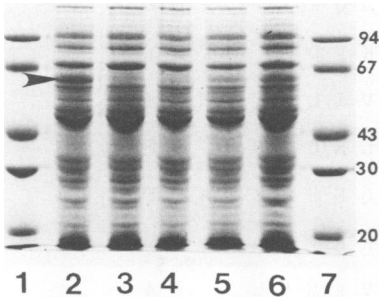

FIG. 3. SDS-PAGE of cytoplasmic cell extracts from recombinant *E. coli* harboring selected constructs. Approximately 15 μg of soluble protein was loaded in each lane. Lanes 1 and 7, molecular weight markers (×10³); lane 2, DH5α(pLOI1005); lane 3, DH5α (pUC18); lane 4, DH5α(pLOI1040); lane 5, DH5α(pLOI1043); lane 6, DH5α(pLOI1050). The band corresponding to the xylosidase-arabinosidase enzyme is marked with an arrowhead.

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 *Ssp*I 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

TABLE 4. Comparison of the translated amino acid sequences of the three *B. fibrisolvens* ORFs in pLOI1005 with those of other genes

Organism (gene)	% Similarity (% identity)			Reference
	ORF1	ORF2 (*xylB*)	ORF3	
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-

TABLE 5. Amino acid sequence alignment of conserved regions

Organism or protein[a]	Sequence	Reference
	35 44 52	
HEWL	F E S N F N T Q A T . N R N T D G S . . . T D Y	41
	331 338 349	
A.n. (glu)	P E D T Y . Y N G . . N P W F L C T L A A A E Q	5
	342 350 362	
B.f. *xylB*	S E D F Y S L T D . . N P G F L R L K L R P E A	This study
	320 328 337	
B.f. *end1*	G E T S A T N R N . . N T A E R V K W A . . D Y	4
	355 362 369	
B.f. *xynA*	N E K P L I W S . . . N I G V A K P A Y . . D E	25
	325 335 343	
B.p. *xynB*	I E C T R L A Q L N W N T C S M Q F V . . . E E	26
	458 466 475	
C.t. *xynZ*	G E A L L R A D V . . N R S G K V D S . . T D Y	15
	130 138 149	
C.t. *celA*	Q E V V N Y C I D . . N K M Y V I L N T H H D V	3
	418 427 436	
C.t. *celB*	T E G G H P L L D L . N L K Y L R C M R . . D F	14
	376 384 395	
C.t. *celD*	D E E Y L R D F E . . N R A A Q F S K K E A D F	19
	408 418 427	
C.s. *xynB*	R E V F V E R I D E Y N A N P K R V W L . . E M	22
	244 254 263	
T.r. CBH II	L E C I N Y A V T Q L N L P N V A M Y L . . D A	33
	586 596 605	
K.f. (β-glucosidase)	G E W E T E G Y D R E N M D L P K R T N . . E L	32
	33 42 50	
S.c. EG I	N E S C A E F G N Q . N I P G V K N . . . T D Y	41

[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-D-xylosidase–β-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.
3. **Benguin, P., P. Cornet, and J. P. Aubert.** 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. J. Bacteriol. **162**:102–105.
4. **Berger, E., W. A. Jones, D. T. Jones, and D. R. Woods.** 1989. Cloning and sequencing of an endoglucanase (*endI*) gene from *Butyrivibrio fibrisolvens* H17c. Mol. Gen. Genet. **219**:193–198.
5. **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.
7. **Cotta, M. A., and R. B. Hespell.** 1986. Proteolytic activity of the ruminal bacteria *Butyrivibrio fibrisolvens*. Appl. Environ. Microbiol. **52**:51–58.
8. **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.
11. **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.
14. **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.
16. **Hespell, R. B., and P. J. O'Bryan-Shah.** 1988. Esterase activities in *Butyrivibrio fibrisolvens* strains. Appl. Environ. Microbiol. **54**:1917–1922.
17. **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.
18. **Hobson, P. N., and M. R. Purdom.** 1961. Two types of xylan fermenting bacteria from the sheep rumen. J. Appl. Bacteriol. **24**:188–193.
19. **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.
22. **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.
23. **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.
26. **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.
29. **Patterson, J. A.** 1989. Prospects for establishment of genetically engineered microorganisms in the rumen. Enzyme Microb. Technol. **11**:187–189.
30. **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.
31. **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 fibrisolvens* xylosidase gene (*xylB*) coding region. Appl. Environ. Microbiol. **55**:306–311.
36. **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.
37. **Uziie, M., M. Matsuo, and T. Yasui.** 1985. Possible identity of β-xylosidase and β-glucosidase of *Chaetomium trilaterale*. Agric. Biol. Chem. **49**:1167–1173.
38. **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.