JOURNAL OF BACTERIOLOGY, May 1991, p. 3265–3268
0021-9193/91/103265-04$02.00/0

Vol. 173, No. 10

$endA_{FS}$, a Novel Family E Endoglucanase Gene from *Fibrobacter succinogenes* AR1

R. CAVICCHIOLI,[1]* P. D. EAST,[2] AND K. WATSON[1]

Department of Biochemistry, Microbiology and Nutrition, University of New England, Armidale, New South Wales 2351,[1] and Division of Entomology, Commonwealth Scientific and Industrial Research Organization, Canberra, Australian Capital Territory 2601,[2] Australia

Received 19 November 1990/Accepted 15 March 1991

The complete nucleotide sequence of $endA_{FS}$, an endoglucanase gene isolated from the ruminal anaerobe *Fibrobacter succinogenes* AR1, was determined. $endA_{FS}$ encodes two overlapping open reading frames (ORF1 and ORF2), and it was proposed that a −1 ribosomal frameshift was required to allow contiguous synthesis of a 453-amino-acid endoglucanase. A proline- and threonine-rich region at the C terminus of ORF1 and rare codons for arginine and threonine were coincident with the proposed frameshift site. $ENDA_{FS}$ is proposed to be a member of subgroup 1 of family E endoglucanases, of which endoglucanases from *Thermomonospora fusca* and *Persea americana* (avocado) are also members. Endoglucanases from *Clostridium thermocellum* and *Pseudomonas fluorescens* form subgroup 2.

Fibrobacter succinogenes is a gram-negative obligate anaerobe that is highly cellulolytic and proliferates in the rumen of animals fed high-fiber diets (16). Recently, three separate cellulase genes have been isolated from strain AR1, which is proposed to be a member of *F. succinogenes* subsp. *elongata* (2). In total, at least nine cellulase or cellulase-related genes from *F. succinogenes* have been cloned in *Escherichia coli* (1, 2, 17). DNA sequences for rumen bacterial glucanase genes, including those of *F. succinogenes* (13, 17; this paper) have only recently been published. This paper reports the DNA sequence of a novel endoglucanase gene from *F. succinogenes* AR1. The gene, $endA_{FS}$, encodes two overlapping open reading frames (ORFs), and a −1 ribosomal frameshift is proposed for contiguous translation. The translation product, $ENDA_{FS}$, was shown to exhibit significant homology to family E endoglucanases (8), and two new subgroups within family E are proposed.

Overlapping deletions of $endA_{FS}$ were generated from $pRCZ^+$ and $pRCZ^-$ (2) by using the Erase-a-base kit (Promega). DNA base composition was determined by the dideoxy-chain termination method (14) with double-stranded and single-stranded DNA templates and *Taq* DNA polymerase. Primer extension was performed essentially by the method of Hartz et al. (7). Oligonucleotide TTGTCATCCA CGCAGCT was complementary to positions 476 to 492, and oligonucleotide CTCCGTCTAGACCGCAC was complementary to positions 446 to 463 of the reverse strand.

A 2.1-kb *Eco*RI-*Hin*dIII fragment containing $endA_{FS}$ was subcloned from pRCO93 (2) into Gemini series vectors and sequenced completely on both strands. Analysis of potential translated regions indicated three ORFs, two encoded in the *Eco*RI-to-*Hin*dIII direction (ORF1 and ORF2) and the third, ORF3, on the complementary strand (Fig. 1). Oligonucleotides were synthesized that were complementary to regions towards the 5′ end of both ORF1 and ORF3. Primer extension analysis revealed a single defined band (Fig. 2, lane 6), which was indicative of mRNA synthesis in the *Eco*RI-to-*Hin*dIII direction originating at nucleotide 355 (Fig. 3). The absence of any cDNA synthesis from the heptadecamer synthesized for ORF3 (Fig. 2, lane 5) indicated that this strand was not transcribed. This was confirmed by RNA dot blot hybridization (results not shown).

Promoter regions typical of recognition sites for *E. coli* σ^{70} were found upstream of the transcription start site (Fig. 3). The −35 and −10 regions were separated by 17 bp. The A+T content upstream of the −35 region is 67%, and 14 out of the first 15 nucleotides immediately 5′ were either A or T. By comparison, the region encoding $ENDA_{FS}$ was 48% A+T. At nucleotide position 381, there is a proposed ATG translation start site. Initiation at this site seems likely in view of the fact that 6 bp upstream there is a purine-rich region which strongly resembles a Shine-Dalgarno sequence found in gram-negative bacteria. A sequence typical of bacterial signal peptides was predicted for the first 26 amino acids (21).

Sequences from the cellulases EGD of *Clostridium thermocellum* (11), endoglucanase (EG) of *Persea americana* (avocado) (19), EGA of *Pseudomonas fluorescens* subsp. *cellulosa* (6), and E4 of *Thermomonospora fusca* (partial sequence; 24) have shown significant homology with predicted amino acid sequences from both ORF1 and ORF2 (Fig. 4). A distinct family grouping, family E with two subgroups, is proposed for these five cellulases. The sequences of cellulase subgroup 1, EG of *P. americana* and E4 of *T. fusca*, were 40% identical and 59% similar, and the sequences of *P. fluorescens* EGA and *C. thermocellum* EGD in subgroup 2 were 29% identical and 51% similar. By comparison, these members of subgroup 1 and 2 were 20 to 22% identical to each other. *F. succinogenes* ENDA shared significant homology to both subgroups but has been grouped in subgroup 1 because of its higher overall identity, 25 to 26%, compared with 22 to 23% with subgroup 2.

A codon usage chart was constructed for $ENDA_{FS}$ (Table 1). Codons for minor tRNA species (10) for arginine (AGG) and threonine (ACA) are underlined. The rare AGG codon and four of the rare ACA codons are clustered at the C terminus of ORF1 at nucleotide 738 and nucleotides 726, 735, 747, and 753, respectively.

Flanking the ATG start site of ORF2 is a proline- and threonine-rich region at the C terminus of ORF1. Approximately half of the cellulase and xylanase sequences so far

* Corresponding author.

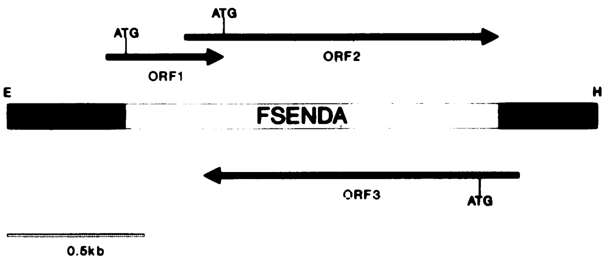

FIG. 1. A 2,075-bp *Eco*RI-*Hin*dIII fragment encoding endoglucanase $ENDA_{FS}$ (FSENDA) from *F. succinogenes* AR1. ORF1 and ORF2 encoded in the *Eco*RI-to-*Hin*dIII direction and ORF3 encoded on the complementary strand are shown.

published contain regions rich in serine, proline, or threonine (SPT), including the two published *Fibrobacter* sequences (13, 17). Discrete catalytic domains (CDs) and cellulose-binding domains (CBDs) are often separated by SPT-rich regions. The separation of distinct functional domains and the potential for glycosylation of SPT regions have led to the hypothesis that degradation of soluble and insoluble substrates can be optimized in cellulolytic microorganisms by balancing the proportions of intact cellulases with CDs through the control of precise proteolysis regulated by posttranslational glycosylation (25). Alternative roles for SPT regions have been suggested, and it has been shown that removing the SPT region in *F. succinogenes* has no effect on activity or binding to oat glucan; a role for the SPT region in protein stabilization has been suggested (17, 18). Ferreira et al. (5) have postulated that the DNA sequences encoding SPT regions may provide a role analogous to introns, whereby cleavage at these sites would allow synthesis of novel enzymes from newly arranged genes.

SPT regions are found in both *T. fusca* E4 and *P. fluorescens* EGA as well as in $ENDA_{FS}$, although position within each sequence is not conserved. In EGA, the N-terminal 600 amino acid residues are separated from the C-terminal region by an extensive SPT region. Henrisatt et al. (8)

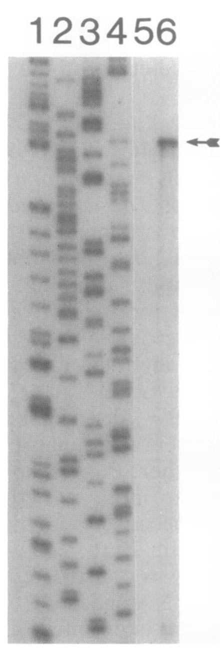

FIG. 2. Mapping of the 5′ end of $endA_{FS}$ mRNA. Primer extension analysis for ORF3 (lane 5) and ORF1 (lane 6) was performed as outlined in Materials and Methods. The molecular weight of cDNA (band in lane 6 indicated by arrow) was determined by comparing it with an M13 DNA sequencing ladder (lanes 1 to 4) and the mRNA start site assigned to position 355 on the DNA sequence (Fig. 3).

```
GAATTCCAAGAATTTTGCCCGAGAAGCCGAAACCTTCGAACGCGTCAAATGCTTTGGCTCGGCCGTTATC   70

GAGGGCTGTTTTACCGCCTGTGGGCGCCTGGACTGCTTCGTGATGACCATGAGCTACCCCTGGGATATTG  140
 Δ1
CGGCTATTGCCCTTCTGGTAGAAGAAGCAGGCGGAAAATCGACCCATATCGATGGCACGCCCATGCAGTT  210

TGTGGACGCCGAACAAGTCATTTTTAGTAACGGAATTTTACATAATACATTGATTTCGACGCTTCAATAG  280

AGTCGTACCATCCCAGTGTAAATATTCTTTTTAATGAAAATTTGCCAAAATAAGCTATATTGGCATTGGG  350
    ↓
TTTGCGGTGTGGTCATTTTTGGAGAAAGTCATGAATTGTAGAAAATACCTTTTGTCGGGCCTTGCCGTTT  420
                                M  N  C  R  K  Y  L  L  S  G  L  A  V  F

TTGGTTTGGCTGCGACGAGTGCTGTTGCAGCACTTAGCACCGACGACTATGTGGAAGCTGCGTGGATGAC  490
  G  L  A  A  T  S  A  V  A  A  L  S  T  D  D  Y  V  E  A  A  W  M  T
 Δ2
AACCCGATTCTTCGGTGCGCAGCGTTCTGGCCAGGGCCCCAACTGGATTTTGGATGGCACGAGCAACCCG  560
 T  R  F  F  G  A  Q  R  S  G  Q  G  P  N  W  I  L  D  G  T  S  N  P

ACTAGCTTTACCAAGGATAGTTATAACGGTAAAGACGTGAGCGGTGGTTGGTTTGACTGCGGTGACCACG  630
T  S  F  T  K  D  S  Y  N  G  K  D  V  S  G  G  W  F  D  C  G  D  H  V
                                                                ***P  R

TGATGTATGGTCAGTCCCAGGGTTACGCCTCTTATGTGCTCGCCTTGGCTTATGCCGAATTTACCGAAGT  700
  M  Y  G  Q  S  Q  G  Y  A  S  Y  V  L  A  L  A  Y  A  E  F  T  E  V
 D  V  W  S  V  P  G  L  R  L  L  C  A  R  L  G  L  C  R  I  Y  R  S

TTCTACGACCTTTATACTGGTGACTACACCGACTACAAGGAAGCCAACAACTACACCATGAAGAGCGGTA  770
 S  T  T  F  I  L  V  T  T  P  T  T  R  K  P  T  T  T  P  ***
F  Y  D  L  Y  T  G  D  Y  T  D  Y  K  E  A  N  N  Y  T  M  K  S  G  K

AGCCCAACAAGGTGCGTGACCTGCTCGAAGAACTCCGCTACGAAGCAGATTTCTGGGTAAAGGCTGCCAT  840
  P  N  K  V  R  D  L  L  E  E  L  R  Y  E  A  D  F  W  V  K  A  A  I

CGATGGAAACAACTTTGTGACGGTTAAGGGCGATGGTAACGCAGACCACCAGAAATGGGTGACTGCTGGT  910
 D  G  N  N  F  V  T  V  K  G  D  G  N  A  D  H  Q  K  W  V  T  A  G

GCCATGAGCAAGCTCGGCTCGGGCGAAGGTGGTGAACCGCGTTGCATTACGGGTAACGCAAACGATGGCT  980
A  M  S  K  L  G  S  G  E  G  G  E  P  R  C  I  T  G  N  A  N  D  G  F

TTACTTCGGGCCTTGCCGCCGCTATGCTTGCCGTGATGGCTCGTGTTGACCCTGATACGGCAAATCAGGC 1050
  T  S  G  L  A  A  A  M  L  A  V  M  A  R  V  D  P  D  T  A  N  Q  A

CAAGTACCTGAAGGCTGCAAAGACTGCCTATTCTTACGCCAAGTCTCACAAGGGCGTTACGAACTCTCAG 1120
 K  Y  L  K  A  A  K  T  A  Y  S  Y  A  K  S  H  K  G  V  T  N  S  Q

GGCTTCTATGAATCTAGCTGGTGGGATGGTCGTTGGGAAGACGGCCCGTTTCTTGCTGAACTTGAACTTT 1190
G  F  Y  E  S  S  W  W  D  G  R  W  E  D  G  P  F  L  A  E  L  E  L  Y

ACCGCACTACCGGTGAAAATTCTTACAAGACAGCCGCTATTGACCGTTATGATAACTTGAAGTTCAGCCT 1260
  R  T  T  G  E  N  S  Y  K  T  A  A  I  D  R  Y  D  N  L  K  F  S  L

GGGCGAAGGCACGCACTTTATGTACAGTAACGTGGTGCCGCTGTCTGCCGTGATGGCCGAAGCCGTGTTC 1330
 G  E  G  T  H  F  M  Y  S  N  V  V  P  L  S  A  V  M  A  E  A  V  F

GAAGAAACTCCGCATGGCATGCGTAAGGAAGCCATCGGCGTGTTGGACTTGATCTACGAAGAAAAGGCCA 1400
E  E  T  P  H  G  M  R  K  E  A  I  G  V  L  D  L  I  Y  E  E  K  A  K

AGGACAAGATTTTCCAGAATCCCAATGGCATGGGTTCGGGCAAGTTCCCGGTACGTGTTCCGTCGGGTGG 1470
  D  K  I  F  Q  N  P  N  G  M  G  S  G  K  F  P  V  R  V  P  S  G  G

TGCTTTCCTGTACGCATTGTCTGACAAGTTCAATAACACGAACGAACACATGGAAATGATCGAAAAGAAC 1540
 A  F  L  Y  A  L  S  D  K  F  N  N  T  N  E  H  M  E  M  I  E  K  N

GTGAGCTACTTGCTTGGCGATAACGGCAGCAAGAAGTCTTACGTGGTTGGTTTCTCCAAGAACGGCGCGA 1610
V  S  Y  L  L  G  D  N  G  S  K  K  S  Y  V  V  G  F  S  K  N  G  A  N

ACGCTCCGTCTAGACCGCACCATCGTGGCTACTATGCTAACGAAAAACGCTGGCGTCGAAGTCGGCGGTG 1680
  A  P  S  R  P  H  H  R  G  Y  Y  A  N  E  K  R  W  R  R  S  R  R  C

CTCCGAATCCTCCCGAAAAGAACAAGCTCTTGGGCGGTATGATTGCTGGCGACTTTACTAGCGGAAACCA 1750
 S  E  S  S  R  K  E  Q  A  L  G  R  Y  D  C  W  R  L  Y  ***
                                                 Δ3
TGACGGCAATACGTCCAACTGGCAGACGAACGAAGTTTGCGTTGACTTGAACGCTCCGCTCGTTGGCGCT 1820

CTCGGCTATATCTTGAGCAAGAAGGCTCCGAAGACCGACGAAGATCTCGGCATCAAGTCTATTGTCAAGA 1890

AGGACACGACGAAGAAGGATACGGTTGTCAAGGATACGACGAAGAAGGATACCATCGAAGCTATCGTTCC 1960
      Δ4
GCGTCTTGCCTTGGCCAAGAGTTTCAACCTGACTTCTAACGGTTCCTTGGTGAGTGTTTCCCAGGTTGCA 2030

CGCAAGCCCTTCAAGGTGCAGGTGTTCGACCTCACAGGTAAGCTT   2075
```

FIG. 3. Nucleotide sequence of $endA_{FS}$ and flanking sequences. The derived amino acid sequence of $ENDA_{FS}$ is shown as single-letter codes for overlapping ORFs. ORF2 is delimited by translation stop codons at the N and C termini, indicated by ***. A proline- and threonine-rich region delimiting the C terminus of ORF1 is underlined, and potential Shine-Dalgarno regions for both ORFs are underlined. The transcription start site is marked with an arrow, and promoters are overlined. A section of 50 nucleotides is underlined in the 5′ and 3′ flanking sequences, and these have been identified as potential rho-independent termination regions. TGT-X_9-ACA, possible upstream activator sequences, are underlined in the 5′ leader region, and direct repeats are underlined in the 3′ trailing region. Exonuclease deletion mutants constructed from the 5′ (Δ_1 and Δ_2) and 3′ (Δ_3 and Δ_4) ends are indicated. Δ_1 and Δ_4 direct endoglucanase synthesis and Δ_2 and Δ_3 do not.

```
 51   NYAEALQKSMFFYEAQRSGKLPENNRVSWRGDSGLNDGADVGL   E4
 29   HYSDALEKSILFFEGQRSGKLPTNQRLTWRGDSGLSDGSSYHV   EG
 29   DYVEAAWMTTRFFGAQRSGQGPNWILDGTSNPTSFTKDSYNGK   FSENDA
      .* .*  ..  *...****. *.    . .. .......

 94   DLTGGWYDAGDHVKFGFPMAFTATMLAWGAS    E4
 72   DLVGGYYDAGDNLKFGLPMAFTTTMLAWGII    EG
 72   DVSGGWFDCGDHVMYGQSQGYASYVLALAYA    FSENDA
193   NVTKGWYDAGDHGKYVVNGGISVWTLLNLYE    EGA
191   DSTKGWHDAGDYNKYVVNAGITVGSMFLAWE    EGD
      .  *. * **  ..   . .  .

131   RSGQMPYLKDNLRWVNDYFIKAH-PSPNVLYVQVGDGDADHKWWGPAEVMP   E4
108   MPEQVENARAALRWSTDYLLKASTATSNSLYVQVGEPNADHRCWERPEDM-   EG
131   KPNKVRDLLEELRYEADFWVKAAIDGNNFVTVK-GDGNADHQKWVTAGAMS   FSENDA
245   SGNGVADILDEARWQMEFMLAMQVPQG-----QAKAGMAHHKIHDVGWTGL   EGA
237   KNNSIPDFLDELKYEIDWILTMQYPDG--------SGRVAHKVSTRNFGGF   EGD
        . .   .. ..  .. .              . . *.

392   RQINYALGDNPRNSSYVVGF---GNNPPRNPHHRT   E4
383   KQVDYILGQNPAKMSYMVGF---GERYPQHVHHRG   EG
386   KNVSYLLGDNGSKKSYVVGFSKNGANAPSRPHHRG   FSENDA
495   KGINYLFGSNVLSTSFITGL---GTNTVAQPHHRF   EGA
487   DAISHVFGRNYYNRSYVTGL---GINPPMNPHDRR   EGD
        ... .* *  . *...*.   * .   . *.*
```

FIG. 4. Alignment of family E cellulases. Predicted amino acid sequence alignments were determined for *F. succinogenes* ENDA, *T. fusca* E4, *P. americana* EG, *P. fluorescens* EGA, and *C. thermocellum* EGD. The amino acid residue number for each sequence is shown to the left. Sequences having identical residues are indicated by *, and those with conserved sequence substitutions (Higgins and Sharp [9]) are indicated by dots.

have indicated that the N-terminal region represents the core, catalytic domain, having homology with *C. thermocellum* EGD, whereas the C-terminal region has homology with CBDs of *Cellulomonas fimi* cellulases. The conserved regions of EGA that have homology with family E cellulases are in the N-terminal CD. By inference, regions of homology for all five cellulases in family E arise in CDs. The presence of SPT regions within E4 and $ENDA_{FS}$ and the observation that these sequences are dispensable in other cellulases suggest that the spacer model may not be generally applicable. Consideration of other potential functions for SPT regions therefore seems appropriate.

We propose that a −1 frameshift may occur in $endA_{FS}$ between ORF1 and ORF2 to translate a single contiguous protein of 453 amino acids (Fig. 3). Evidence to support a single contiguous protein includes the observation that the predicted amino acid sequence resulting from the frameshift (ORF1 and ORF2) has homology with other endoglucanases in family E, which is observed over the entire length of the sequence in a continuous fashion. If each ORF encoded a separate endoglucanase, it would seem unlikely that the encoded proteins would have homology to distinct but consecutive regions of the same endoglucanase. Furthermore, the molecular weight prediction for $ENDA_{FS}$ from ORF1 and ORF2 is in agreement with preliminary data on the molecular weight (46,500) of periplasmic endoglucanase from *E. coli* (2). The mature endoglucanase has a predicted molecular weight of 47,000 after cleavage of the putative signal peptide. Further evidence supporting continuous translation of both ORFs is indicated by expression of deletion mutants. Exonuclease deletions from the 5′ and 3′ ends of the 2.1-kb *Eco*RI-*Hin*dIII fragment encoding $endA_{FS}$ demonstrated that the entire sequence predicted for the gene, including regulatory regions, is required for expression (Fig. 3).

Characteristics of the primary structure for the proposed frameshift site have some similarity with elements that control high-efficiency frameshifting at the frameshift site of RF2 (23). A Shine-Dalgarno region upstream of the stop site in *prfB* has been shown to interact with the complementary region of 16S rRNA. It is suggested that the 3′ end of 16S rRNA continually scans the mRNA during elongation until it binds at the complementary region, causing a pause in translation (4). It is possible that unique interaction of 16S rRNA with both the ribosome-binding site and the termination codon results in frameshifting.

It is noteworthy that, immediately upstream of the proposed frameshift, the arginine codon and four threonine codons, AGG and ACA, respectively, are read by minor tRNA species (10). Tandem rare codons AGG-AGG preceding a TGA stop site mediate high-efficiency frameshifting by the depletion of cognate tRNA (15). Although rare codons per se may not be sufficient to cause frameshifting, they can increase pausing, which enhances frameshifting (22). The role of rare codons at the proposed frameshift site is un-

TABLE 1. Codon usage chart for nucleotide sequence encoding $ENDA_{FS}$

Amino acid	Codon	Residue no.	Amino acid	Codon	Residue no.	Amino acid	Codon	Residue no.	Amino acid	Codon	Residue no.
G	GGG	1	E	GAG	0	V	GTG	16	A	GCG	4
G	GGA	1	E	GAA	26	V	GTA	2	A	GCA	8
G	GGT	20	D	GAT	11	V	GTT	8	A	GCT	15
G	GGC	22	D	GAC	13	V	GTC	0	A	GCC	18
R[a]	AGG	1	K	AAG	25	M	ATG	13	T	ACG	9
R	AGA	2	K	AAA	5	I	ATA	1	T	ACA	6
S	AGT	4	N	AAT	6	I	ATT	4	T	ACT	9
S	AGC	10	N	AAC	19	I	ATC	4	T	ACC	6
W	TGG	10	***	TAG	1	L	TTG	9	S	TCG	5
***[b]	TGA	1	***	TAA	0	L	TTA	0	S	TCA	0
C	TGT	1	Y	TAT	10	F	TTT	9	S	TCT	12
C	TGC	4	Y	TAC	14	F	TTC	11	S	TCC	5
R	CGG	3	Q	CAG	8	L	CTG	6	P	CCG	10
R	CGA	4	Q	CAA	1	L	CTA	0	P	CCA	2
R	CGT	10	H	CAT	2	L	CTT	11	P	CCT	1
R	CGC	3	H	CAC	6	L	CTC	4	P	CCC	3

[a] Underlining denotes minor tRNA species.
[b] Translation stop codon.

known, but the occurrence of consecutive rare codons for threonine and arginine may facilitate ribosomal pausing.

Two similar examples of ribosomal frameshifting in endoglucanase genes are known from two separate isolates of ruminal strains of *Bacteroides ruminicola* (12, 20). Insufficient cellulase sequence data are available from *Bacteroides* spp. or *Fibrobacter* spp. to speculate on a general role for a frameshifting phenomenon. However, the fact that an SPT region occurs upstream of and ends precisely at the frameshift may indicate that the $endA_{FS}$ gene has arisen from an insertion event derived from an evolutionarily primitive cellulase, consistent with the suggestions of Ferreira et al. (5) that SPT regions may have fulfilled a role in transferring cellulases between organisms. Furthermore, the regions may mark junctions not only between CBDs and CDs but also between other functionally important but as yet unidentified domains. We are currently undertaking a detailed analysis of transcriptional events involved in the expression of $endA_{FS}$, and there is evidence for a second transcript that originates from promoters within ORF1 and which exhibits characteristics different from that encoding the functional endoglucanase gene (3).

Nucleotide sequence accession number. The sequence reported in this paper has been assigned GenBank accession number M58520.

We are grateful to John Pemberton and coworkers, Ken Reed, John Watson, Alan Richardson, Ifor Beacham, John Argyle, Athol Klieve, Robert Learmonth, Keith Gregg, and Cheryl Ware for valuable advice and discussion. We are also indebted to Philip Vercoe, David Wilson, Ronald Teather, Geoff Hazlewood, Bryan White, and Chung-Ming Huang for invaluable preprints, sequence data, and discussion. Thanks to Andrew Gooley for providing oligonucleotides and to Jean Hansford for the typing of the manuscript.

REFERENCES

1. **Cavicchioli, R., D. H.-L. Lai, and K. Watson.** 1989. Cloning and expression of cellulase genes from *Bacteroides succinogenes*, p. 153–156. *In* M. Sleigh (ed.), Eighth Australian Biotechnology Conference. Australian Biotechnology Association, Sydney, Australia.
2. **Cavicchioli, R., and K. Watson.** 1991. Molecular cloning, characterization, and expression of endoglucanase genes from the ruminal bacterium *Fibrobacter succinogenes* AR1. Appl. Environ. Microbiol. **57**:359–365.
3. **Cavicchioli, R., and K. Watson.** Unpublished data.
4. **Craigen, W. J., C. C. Lee, and C. T. Caskey.** 1990. Recent advances in peptide chain termination. Mol. Microbiol. **4**:861–865.
5. **Ferreira, L. M. A., A. J. Durrant, J. Hall, G. P. Hazlewood, and H. J. Gilbert.** 1990. Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. Biochem. J. **269**:261–264.
6. **Hall, J., and H. J. Gilbert.** 1988. The nucleotide sequence of a carboxymethylcellulase gene from *Pseudomonas fluorescens* subsp. *cellulosa*. Mol. Gen. Genet. **213**:112–117.
7. **Hartz, D., D. S. McPheeters, R. Traut, and L. Gold.** 1988. Extension inhibition analysis of translation initiation complexes. Methods Enzymol. **164**:419–425.
8. **Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene **81**:83–95.
9. **Higgins, D. G., and P. M. Sharp.** 1988. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73**:237–244.
10. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151**:389–409.
11. **Joliff, G., P. Beguin, and J.-P. Aubert.** 1986. Nucleotide sequence of the cellulase gene *celD* encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. **14**:8605–8613.
12. **Matsushita, O., J. B. Russell, and D. B. Wilson.** Unpublished data.
13. **McGavin, M. J., C. W. Forsberg, B. Crosby, A. W. Bell, D. Dignard, and D. Y. Thomas.** 1989. Structure of the *cel-3* gene from *Fibrobacter succinogenes* S85 and characteristics of the encoded gene product, endoglucanase 3. J. Bacteriol. **171**:5587–5595.
14. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
15. **Spanjaard, R. A., and J. Van Duin.** 1988. Translation of the sequence AGG-AGG yields 50% ribosomal frameshift. Proc. Natl. Acad. Sci. USA **85**:7967–7971.
16. **Stewart, C. S., and H. J. Flint.** 1989. *Bacteroides* (*Fibrobacter*) *succinogenes*, a cellulolytic anaerobic bacterium from the gastrointestinal tract. Appl. Microbiol. Biotechnol. **30**:433–439.
17. **Teather, R. M., and J. D. Erfle.** 1990. DNA sequence of a *Fibrobacter succinogenes* mixed linkage β-glucanase (1,3-1,4-β-D-glucan 4-glucanohydrolase) gene. J. Bacteriol. **172**:3837–3841.
18. **Teather, R. M., H. J. Gilbert, and G. P. Hazlewood.** *In* Genetics and molecular biology of anaerobes, in press.
19. **Tucker, M. L., M. L. Durbin, M. T. Clegg, and L. N. Lewis.** 1987. Avocado cellulase: nucleotide sequence of a putative full length cDNA clone and evidence for a small gene family. Plant Mol. Biol. **9**:197–203.
20. **Vercoe, P.** Personal communication.
21. **Von Heijne, G.** 1988. Transcending the impenetrable: how proteins come to terms with membranes. Biochim. Biophys. Acta **947**:307–333.
22. **Weiss, R. B., and J. A. Gallant.** 1986. Frameshift suppression in aminoacyl tRNA limited cells. Genetics **112**:727–739.
23. **Weiss, R. B., D. M. Dunn, A. E. Dahlberg, J. F. Atkins, and R. F. Gesteland.** 1988. Reading frame switch caused by base pair formation between the 3′ end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. EMBO J. **7**:1503–1507.
24. **Wilson, D. B.** Personal communication.
25. **Yablonsky, M. D., K. O. Elliston, and D. E. Eveleigh.** 1989. The relationship between the endoglucanase *MBcelA* of *Microbispora bispora* and the cellulases of *Cellulomonas fimi*, p. 112–133. *In* M. P. Coughlan (ed.), Enzyme systems for lignocellulose degradation. Elsevier Applied Science Publications, London.