JOURNAL OF BACTERIOLOGY, Sept. 1993, p. 5585–5594
0021-9193/93/175585-10$02.00/0

Vol. 175, No. 17

Sequencing and Characterization of a Gene Cluster Encoding the Enzymes for L-Rhamnose Metabolism in *Escherichia coli*

PILAR MORALEJO,[1] SUSAN M. EGAN,[2] ELENA HIDALGO,[1] AND JUAN AGUILAR[1]*

Department of Biochemistry, School of Pharmacy, University of Barcelona, Diagonal 643, 08028 Barcelona, Spain,[1] and Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218[2]

Received 9 March 1993/Accepted 22 June 1993

The sequencing of the *Eco*RI-*Hin*dIII fragment complementing mutations in the structural genes of the L-rhamnose regulon of *Escherichia coli* has permitted identification of the open reading frames corresponding to *rhaB*, *rhaA*, and *rhaD*. The deduced amino acid sequences gave a 425-amino-acid polypeptide corresponding to rhamnulose kinase for *rhaB*, a 400-amino-acid polypeptide corresponding to rhamnose isomerase for *rhaA*, and a 274-amino-acid polypeptide corresponding to rhamnulose-1-phosphate aldolase for *rhaD*. Transcriptional fusions of the three putative promoter regions to *lacZ* showed that only the *rhaB* leader region acted as a promoter, as indicated by the high β-galactosidase activity induced by rhamnose, while no significant activity from the *rhaA* and *rhaD* constructions was detected. The *rhaB* transcription start site was mapped to −24 relative to the start of translation. Mutations in the catabolic genes were used to show that L-rhamnose may directly induce *rhaBAD* transcription.

L-Rhamnose, a methylpentose, is metabolized in *Escherichia coli* by a set of enzymes encoded by genes constituting the rhamnose regulon, which maps at 88.4 min in the chromosome (2). Four structural genes have been described: *rhaA*, encoding rhamnose isomerase; *rhaB*, encoding rhamnulose kinase; *rhaD*, encoding rhamnulose-1-phosphate aldolase (32); and *rhaT*, encoding the rhamnose transport system (17). The *rhaT* gene has been mapped in the *rha* locus, separated from *rhaA*, *rhaB*, and *rhaD* by the regulatory operon *rhaC*, which has been found to be formed by two partially overlapping genes, *rhaR* and *rhaS* (40). The gene order of the region, counterclockwise, is *glpK*. . .*sodA-rhaT-rhaR-rhaS-rhaB-rhaA-rhaD*.

In *E. coli*, *rhaT*, encoding the transporter (17, 39), and *rhaR* and *rhaS*, governing expression (40, 41), have been sequenced and extensively analyzed. In this species, another methylpentose, L-fucose, is metabolized by a parallel metabolic pathway integrated by a set of specific enzymes encoded by the *fuc* gene cluster, which has been located at 60.2 min (23) and completely sequenced (11, 25). In *Salmonella typhimurium* LT2, *rhaB*, for rhamnulose kinase; *rhaC2*, one of the regulatory genes (31); and *rhaT*, encoding the transporter (39) have also been sequenced.

Here we present a sequence analysis of three structural genes of the rhamnose pathway, some experiments involving their expression, and a comparison with the corresponding gene sequences of the L-fucose system.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The bacterial strains used in this study are listed in Table 1. Cells were grown aerobically as described previously (7) on LB or minimal medium. For growth on minimal medium, L-rhamnose, glucose, or L-fucose was added to 0.2%. When indicated, X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was added to 40 μg/ml. Ampicillin was used at 100 μg/ml, kanamycin was used at 30 μg/ml, and streptomycin was used at 25 μg/ml. For primer extension analysis, the strains were grown in M10 medium (33) containing 0.4% glycerol, 0.2% Casamino Acids, and 50 μM thiamine, in the presence or absence of 0.2% rhamnose.

Preparation of cell extracts and enzyme assay. For enzyme assay, the cells were harvested at the end of the exponential phase and the cell extract was prepared as described previously (7) with 10 mM Tris-HCl buffer (pH 7.0). The β-galactosidase activity in strains grown under specified conditions was assayed as described by Miller, and the values are reported in the units defined by Miller (28).

The protein concentration in cell extracts was determined by the method of Lowry et al. (24) with bovine serum albumin as the standard.

DNA manipulation. Plasmid DNA was routinely prepared by the boiling method (34). For large-scale preparation, a crude DNA sample was subjected to purification by cesium chloride-ethidium bromide density gradient centrifugation or on a column (Qiagen GmbH, Düsseldorf, Germany). DNA manipulations were performed essentially as described by Sambrook et al. (34). The DNA sequence was determined by using the dideoxy-chain termination procedure of Sanger et al. (36). Double-stranded plasmid DNA was used as the template. Plasmid purified with a Qiagen column was used for the construction of ordered deletions with the Erase-a-Base system (Promega Biotec, Madison, Wis.). We resolved the numerous sequencing gel compressions as described previously (20).

Transcriptional fusions were constructed by inserting the DNA fragments into plasmid pRS550 of Simons et al. (37). Plasmid pRS550 carries a cryptic *lac* operon and genes that confer resistance to both kanamycin and ampicillin. After introduction of the recombinant plasmid into the streptomycin-resistant strain MC1061, blue colonies on X-Gal plates containing ampicillin, kanamycin, and streptomycin were isolated, and plasmid DNA was sequenced by using the M13 primer to ensure that the desired fragment was inserted in the correct orientation.

For the *rhaB* promoter, a fragment starting with a *Bgl*II site upstream of *Eco*RI (position 560 in the *rhaR-rhaS* sequence in reference 40) and ending with a *Cla*I site (position 392 in Fig. 2) was prepared by digestion of pJB3.1

* Corresponding author.

TABLE 1. *E. coli* strains used in this work

Strain	Genotype	Source or reference
ECL1	HfrC *phoA8 relA1 tonA22* $T2^r$ (lambda)	13
DH5α	*supE44 ΔlacU169(ϕ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	BRL
MC1061	*hsdR mcrB araD139 Δ(araABC-leu) 7679 lacX74 galU galK rpsL thi*	37
JA121	MC1061(pRS550)	This study
JA123	MC1061(pRS550-P*rhaB-lacZ*)	This study
JA124	MC1061(pRS550-P*rhaA-lacZ*)	This study
JA125	MC1061(pRS550-P*rhaD-lacZ*)	This study
JA126	MC1061(pRS550-P*rhaBA-lacZ*)	This study
ECL116	F^- *ΔlacU169 endA hsdR thi*	3
ECL339	As ECL116 but *Δ(rha-pfkA)15 zig-1*::Tn*10*	10
ECL714	As ECL116 but *rhaB101*	12
ECL715	As ECL116 but *rhaA502*	12
ECL716	As ECL116 but *rhaD701*	12
ECL717	As ECL116 but *rhaR702*	12

(see Fig. 1), blunted, and ligated to vector pRS550 (37). Analogously, the *rhaA* fusion contained a *Pvu*I-*Nhe*I fragment (positions 1178 to 1666 in Fig. 2), and the *rhaD* fusion contained a *Sac*II-*Eco*RV fragment (positions 2888 to 3538 in Fig. 2) (strains JA123, JA124, and JA125, respectively). Another fusion contained the *Bgl*II-*Nhe*I fragment encompassing the whole *rhaB* gene and the *rhaB-rhaA* intergenic space up to the 5′ end of *rhaA* (strain JA126). A control with the vector pRS550 containing no insertion was strain JA121.

RNA preparation and Northern (RNA) blot hybridization. For the preparation of total RNA, cells of a 25-ml culture grown up to an A_{650} of 0.5 or 0.2 as indicated were collected by centrifugation and resuspended in 125 μl of ice-cold 0.3 M sucrose–0.01 M sodium acetate solution (pH 4.5). To this suspension, 125 μl of 0.01 M sodium acetate (pH 4.5) with 2% (wt/vol) sodium dodecyl sulfate (SDS) was added. The extract was then heated at 70°C for 3 min. The DNA and proteins were extracted with 1 volume of phenol at 70°C for 3 min and then cooled for 15 s in a −80°C bath by means of a dry ice-acetone freezing mixture. After centrifugation at 15,000 × *g* for 5 min, the supernatant was extracted with phenol at 70°C two times more. The RNA was precipitated by addition of 1 ml of 100 mM sodium acetate in ethanol and then washed with 70% ethanol. After centrifugation, the total cellular RNA was resuspended in 50 μl of 20 mM sodium phosphate buffer (pH 6.5) containing 1 mM EDTA.

For the Northern blot hybridization, each RNA sample (10 μg) was electrophoresed on a 1% agarose-formaldehyde gel and transferred to a nylon membrane filter (Schleicher & Schuell) in 10× SSPE (34). Prehybridization and hybridization were carried out in 50% formamide–5× Denhardt reagent–50 mM sodium phosphate (pH 6.5)–10 mM NaCl–0.1% SDS–125 μg of sonicated salmon sperm DNA per ml at 42°C. The restriction nuclease fragments used as probes were 3′ end labeled with the random-primed DNA labeling kit (Boehringer, Manheim, Germany). Filters were washed two times at room temperature and two times at 65°C in 2× SSC (34)–0.5% SDS and then washed once at 50°C and once at 55°C in 0.1× SSC. The filters were exposed to X-ray films at −70°C with an intensifying screen.

Primer extension analysis. RNA was isolated from cells harvested at an A_{600} of 0.2 as previously described (33), except that 10 ml of cells was used and an ethanol precipitation was performed following the isopropanol precipitation. One microgram of RNA was mixed with 2.5 ng of ^{32}P-labeled primer (5′-AAAACGATGGATTTCGCGCAGC GTCAGGCT-3′), and primer extension reactions were performed as described previously (33).

Nucleotide sequence accession number. The nucleotide sequence of the *Eco*RI-*Hin*dIII fragment encompassing *rhaB*, *rhaA*, and *rhaD* has been deposited in the GenBank data library under accession no. X60472.

RESULTS AND DISCUSSION

Nucleotide sequence analysis and identification of *rhaB*, *rhaA*, and *rhaD*. In a previous study we used plasmid pJB4.1 (Fig. 1) to analyze the products of genes *rhaB*, *rhaA*, and *rhaD* (5). The *Eco*RI-*Hin*dIII fragment was inserted in Bluescript and designated plasmid pPM.2. This insert was subcloned in plasmids pPM.2.1 and pPM.2.2, containing fragments *Eco*RI-*Bam*HI and *Bam*HI-*Hin*dIII, respectively. Serial deletions of the three plasmids were obtained and sequenced by the strategy presented in Fig. 1. The DNA was sequenced at least once on each strand. Figure 2 depicts the 5,677 bp of DNA sequenced between the *Eco*RI and *Hin*dIII restriction sites as well as the proteins encoded by the five open reading frames (ORFs) found in the sequence.

According to previous mapping employing complementation (5), the first ORF downstream of the *Eco*RI site corresponded to *rhaB*, encoding rhamnulose kinase. The second one, which has the *Bam*HI site, corresponded to *rhaA*, encoding rhamnose isomerase, and the third one, including the *Sal*I restriction site, corresponded to *rhaD*, encoding rhamnulose-1-phosphate aldolase. Two additional unidentified ORFs designated URF1 and URF2 were found between the end of *rhaD* and the *Hin*dIII restriction site.

rhaB corresponds to a 1,276-nucleotide ORF which can encode a 425-amino-acid polypeptide with a calculated molecular mass of 47,708 Da. No −10 or −35 boxes were apparent upstream from *rhaB*, but a catabolite repression protein recognition inverted repeat could be identified between positions 50 and 65. *rhaA* corresponds to a 1,200-nucleotide ORF which can encode a polypeptide of 400 amino acids with a calculated molecular mass of 44,246 Da, and *rhaD* corresponds to an 822-nucleotide ORF which can encode a polypeptide of 274 amino acids with a calculated molecular mass of 30,149 Da. No evident transcription initiation regulatory signal was apparent in the *rhaB-rhaA* and *rhaA-rhaD* intergenic spaces. The N-terminal amino acid sequences of rhamnose isomerase and rhamnulose-1-phosphate aldolase (4) were in agreement with the assigned ORFs. Unfortunately, the N-terminal end of rhamnulose kinase was blocked, and the sequence is not known.

Two sequences seem to fulfill the requirements of a transcription terminator (dashed arrows in Fig. 2) (14, 29). The first one has a 14-bp inverted repeat separated by 4 bp that could form a stable stem-loop structure with a calculated free energy of stabilization of 24 kcal (ca. 100.4 kJ) and is followed by two thymidine residues. The second one, downstream of *rhaD* and more likely to be functional, has a 10-bp inverted repeat separated by 4 bp, which would correspond to a stem-loop with a calculated free energy of stabilization of 14 kcal (ca. 58.5 kJ), followed by four thymidine residues. This terminator is located 330 bp downstream from the TAA stop codon, an unusually long distance between translation and transcription termination; perhaps there is posttranscriptional processing, as is the case for the

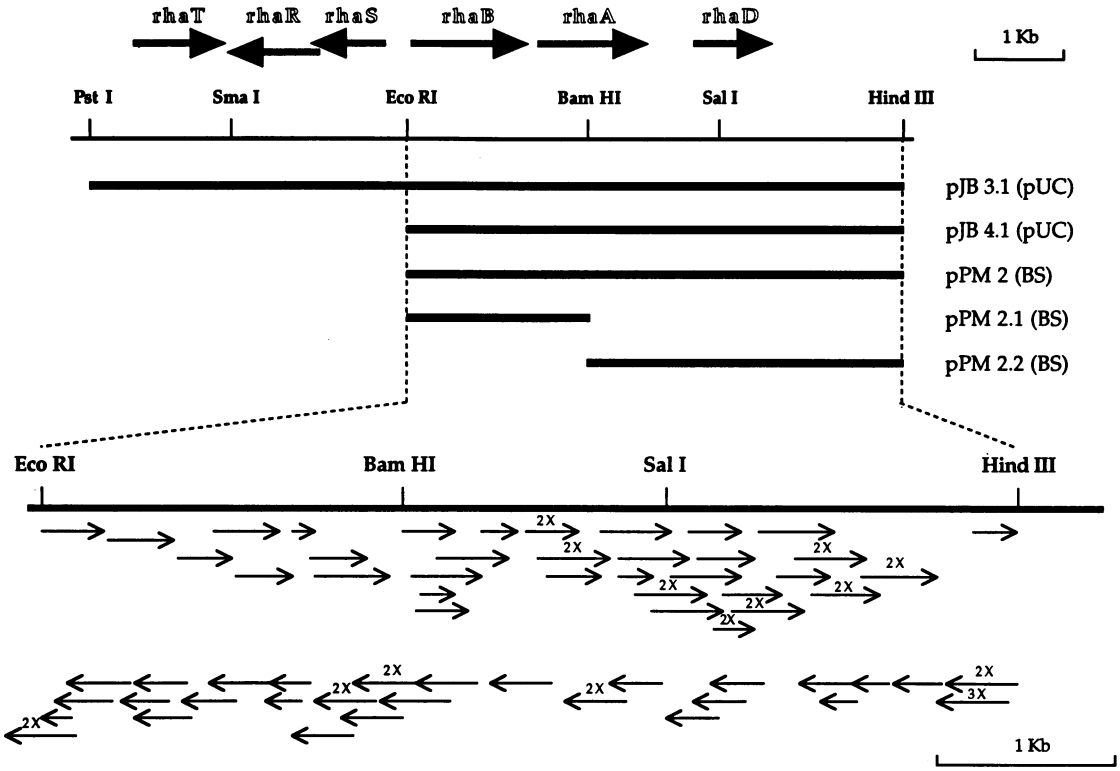

FIG. 1. Restriction map of the genomic region containing the rhamnose system. The thin line represents the genomic DNA, while the thick lines correspond to the fragments subcloned into plasmid pUC18 (pUC) or Bluescript (BS). Thick arrows at top show the directions of transcription and the positions of *rha* genes. The sequencing strategy for the *Eco*RI-*Hin*dIII fragment encompassing the *rhaB*, *rhaA*, and *rhaD* genes is also presented. Thin arrows indicate the start point, direction, and extent of sequence determined from each subclone. If a particular sequence was obtained more than once, this is indicated by a number over the arrow.

lactose operon of *E. coli* (27), or rho-dependent termination (16, 29).

Analysis of DNA sequences of the *rhaA-rhaD* intergenic fragment showed five 88- to 92-bp repetitions plus an incomplete sixth one (Fig. 2). Each complete element was a combination of the motifs described by Gilson et al. (18, 19) which appeared sequentially as follows: an REP (PU) palindromic unit of 34 bp (Y motif) highly homologous to the REP consensus previously described (15, 38), a right internal segment (S motif), a second PU sequence in the opposite orientation (Z^2 motif), and a left internal segment (s motif). The sixth repetition was a Y motif followed by a B-like external fragment. This combination of motifs is in agreement with what is generally known as the BIME (bacterial interspersed mosaic elements) family (18, 19), although the number of repetitions found in other intergenic regions is usually lower than that in the one described here. The function of these short, interspersed repetitive DNA sequences in the regulation of gene expression has been widely discussed (26) but remains uncertain. No repetitive sequences were found in the other intergenic fragments of the rhamnose regulon.

Transcription. Total RNA of cells of strain ECL1 grown aerobically on L-rhamnose was prepared as indicated in Materials and Methods. Northern blot hybridizations showed a major RNA of ca. 1.4 kb for an *rhaB* probe and transcripts of ca. 1.6 and 2.5 kb for both *rhaA* and *rhaD* probes. A larger probe containing *rhaB*, *rhaA*, and part of *rhaD* also gave (but only for some mRNA preparations from early exponential-phase growth) a minor band which might correspond to a 3.8-kb transcript (Fig. 3). RNA prepared from cells of the same strain grown on L-fucose or glucose gave no band of hybridization with any of the probes used (not shown), indicating the specificity for L-rhamnose of the RNA analysis performed.

Strains carrying transcriptional fusions (see Materials and Methods) were grown under different conditions, and β-galactosidase activity in their extracts was determined. As shown in Table 2, the promoter of *rhaB* yielded high activity in growth on L-rhamnose, while the promoters of *rhaA* and *rhaD* yielded very low activities. Thus, only the *rhaB* leader region appears to contain an L-rhamnose-inducible promoter. The strain containing the *rhaBA-lacZ* fusion also displayed high activities, indicating that no strict termination occurs between these two genes. Growth on glucose gave very low activities, close to basal levels. Similarly, growth on the isomer sugar L-fucose, which differs only in the stereoconfiguration at carbons 2 and 4, yielded undetectable activity, indicating no cross induction with this sugar.

In view of these results, we conclude that *rhaBAD* is probably transcribed as a single transcription unit and that the smaller RNAs observed (Fig. 3) may result from degradation or in vivo processing. The role of the differential mRNA stability in the regulation of gene expression of *rhaBAD*, as described for other systems (9, 30, 42), deserves more study. Moreover, intercistronic transcription terminators, such as the one proposed to exist between *rhaA* and *rhaD*, could also be acting as gene expression regulators (30). Alternatively, this stem-loop structure could simply

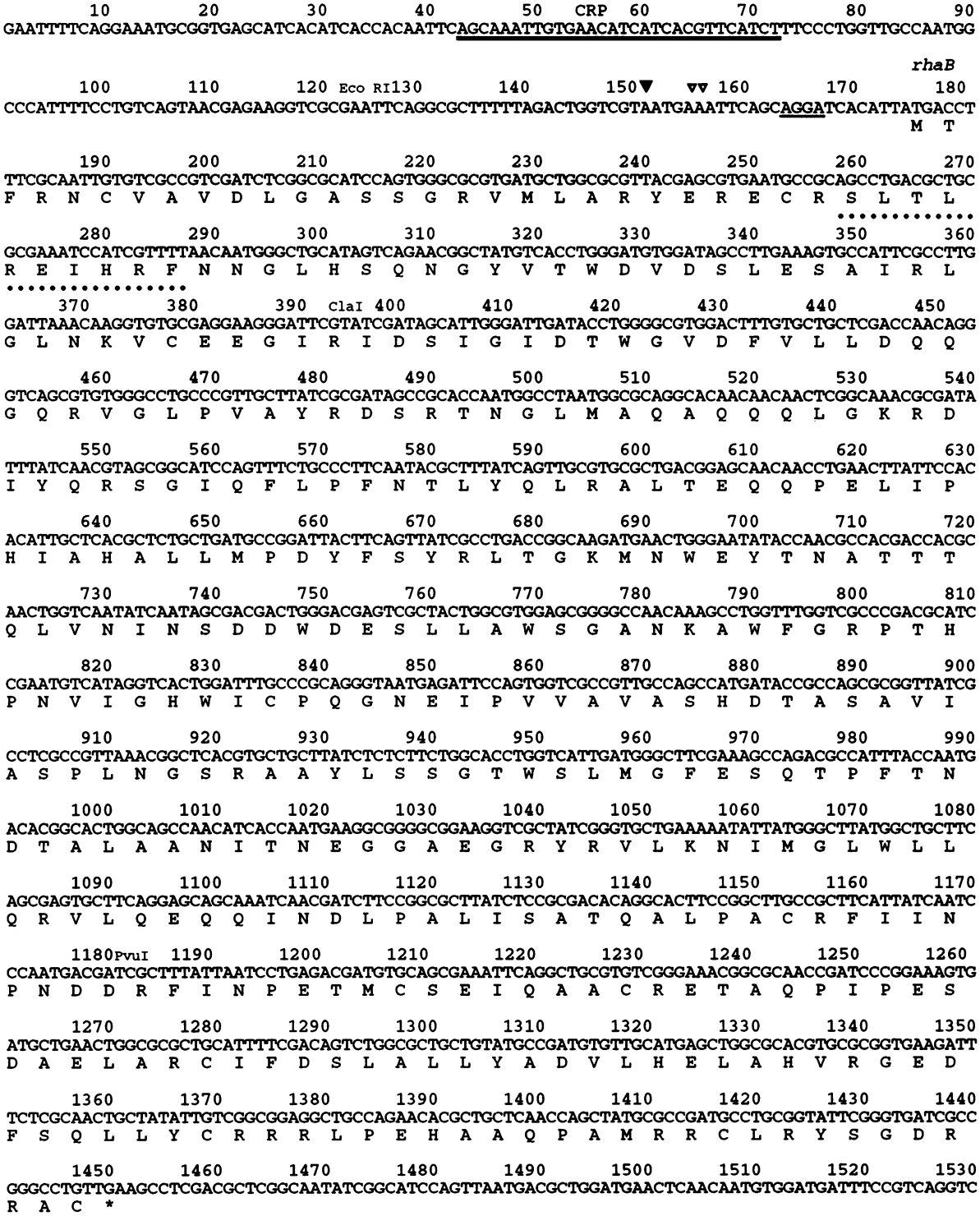

FIG. 2. Nucleotide sequence of the *Eco*RI-*Hin*dIII fragment. The coding regions of the ORFs present in the fragment have been translated and are indicated by the single-letter amino acid code. The restriction sites used in this work are indicated. The inverted repeat constituting a good catabolite repression protein (CRP) consensus binding site is indicated by heavy underlining. DNA sequences predicted to form hairpin loop structures are shown by dashed arrows. The putative Shine-Dalgarno sequences are underlined. Stop codons are indicated by asterisks. The major transcription start site is indicated by a closed triangle, while open triangles show the two minor or processing transcription start sites found. Dots indicate the position and length of the primer used in primer extension analysis. The locations of motifs in repetitive DNA sequences as described in the text are as follows: Y (>), Z^2 (<), S (#), s (⋆), and B-like (+).

```
      1540      1550      1560      1570      1580      1590      1600      1610      1620
GTCAGCACCACCGCGAATCTGACCACCTTTACCCCTAATCCTGACAGTGAAATTGCCCACTATGTGCGCAGATTCACTCTACACGACAGA

                rhaA
      1630      1640      1650      1660 NheI 1670      1680      1690      1700      1710
CAAAGGAGCTTTGCGCATGACCACTCAACTGGAACAGGCCTGGGAGCTAGCGAAACAGCGTTTCGCGGCGGTGGGATTGATGTCGAGGA
                  M  T  T  Q  L  E  Q  A  W  E  L  A  K  Q  R  F  A  A  V  G  I  D  V  E  E

      1720      1730      1740      1750      1760      1770      1780      1790      1800
GGCGCTGCGCCAACTTGATCGTTTACCCGTTTCAATGCACTGCTGGCAGGGCGATGATGTTTCCGGTTTTGAAAACCCGGAAGGTTCGCT
  A  L  R  Q  L  D  R  L  P  V  S  M  H  C  W  Q  G  D  D  V  S  G  F  E  N  P  E  G  S  L

      1810      1820      1830      1840      1850      1860      1870      1880      1890
GACCGGGGGGATTCAGGCCACAGGCAATTATCCGGGCAAAGCGCGTAATGCCAGTGAGCTACGTGCCGATCTGGAACAGGCTATGCGGCT
  T  G  G  I  Q  A  T  G  N  Y  P  G  K  A  R  N  A  S  E  L  R  A  D  L  E  Q  A  M  R  L

      1900      1910      1920      1930      1940      1950      1960      1970      1980
GATTCCGGGGCCGAAACGGCTTAATTTACATGCCATCTATCTGGAATCAGATACGCCAGTCTCGCGCGACCAGATCAAACCAGAGCACTT
  I  P  G  P  K  R  L  N  L  H  A  I  Y  L  E  S  D  T  P  V  S  R  D  Q  I  K  P  E  H  F

      1990      2000      2010      2020      2030      2040      2050      2060      2070
CAAAAACTGGGTTGAATGGGCGAAAGCCAATCAGCTCGGTCTGGATTTTAACCCCTCCTGCTTTTCGCATCCGCTAAGCGCCGATGGCTT
  K  N  W  V  E  W  A  K  A  N  Q  L  G  L  D  F  N  P  S  C  F  S  H  P  L  S  A  D  G  F

      2080      2090      2100      2110      2120      2130      2140      2150      2160
TACGCTTTCCCATGCCGACGACAGCATTCGCCAGTTCTGGATTGATCACTGCAAAGCCAGCCGTCGCGTTTCGGCCTATTTTGGCGAGCA
  T  L  S  H  A  D  D  S  I  R  Q  F  W  I  D  H  C  K  A  S  R  R  V  S  A  Y  F  G  E  Q

      2170      2180      2190BamHI 2200      2210      2220      2230      2240      2250
ACTCGGCACACCATCGGTGATGAACATCTGGATCCCGGATGGTATGAAAGATATCACCGTTGACCGTCTCGCCCCGCGTCAGCGTCTGCT
  L  G  T  P  S  V  M  N  I  W  I  P  D  G  M  K  D  I  T  V  D  R  L  A  P  R  Q  R  L  L

      2260      2270      2280      2290      2300      2310      2320      2330      2340
GGCAGCACTGGATGAGGTGATCAGCGAGAAGCTAAACCCTGCGCACCATATCGACGCCGTTGAGAGCAAATTGTTTGGCATTGGCGCAGA
  A  A  L  D  E  V  I  S  E  K  L  N  P  A  H  H  I  D  A  V  E  S  K  L  F  G  I  G  A  E

      2350      2360      2370      2380      2390      2400      2410      2420      2430
GAGCTACACGGTTGGCTCCAATGAGTTTTACATGGGGTATGCCACCAGCCGCCAGACTGCGCTGTGCCTGGACGCCGGGCACTTCCACCC
  S  Y  T  V  G  S  N  E  F  Y  M  G  Y  A  T  S  R  Q  T  A  L  C  L  D  A  G  H  F  H  P

      2440      2450      2460      2470      2480      2490      2500      2510      2520
GACTGAAGTGATTTCCGACAAGATTTCCGCCGCCATGCTGTATGTGCCGCAGTTGCTGCTGCACGTCAGCCGTCCGGTTCGCTGGGACAG
  T  E  V  I  S  D  K  I  S  A  A  M  L  Y  V  P  Q  L  L  L  H  V  S  R  P  V  R  W  D  S

      2530      2540      2550      2560      2570      2580      2590      2600      2610
CGATCACGTAGTGCTGCTGGATGATGAAACCCAGGCAATTGCCAGTGAGATTGTGCGTCACGATCTGTTTGACCGGGTGCATATCGGCCT
  D  H  V  V  L  L  D  D  E  T  Q  A  I  A  S  E  I  V  R  H  D  L  F  D  R  V  H  I  G  L

      2620      2630      2640      2650      2660      2670      2680      2690      2700
TGACTTCTTCGATGCCTCTATCAACCGCATTGCCGCGTGGGTCATTGGTACACGCAATATGAAAAAAGCCCTGCTGCGTGCGTTGCTGGA
  D  F  F  D  A  S  I  N  R  I  A  A  W  V  I  G  T  R  N  M  K  K  A  L  L  R  A  L  L  E

      2710      2720      2730      2740      2750      2760      2770      2780      2790
ACCTACCGCTGAGCTGCGCAAGCTGGAAGCGGCGGGCGATTACACTGGCGCGTCTGGCACTGCTGGAAGAGCAGAAATCGTTGCCGTGGC
  P  T  A  E  L  R  K  L  E  A  A  G  D  Y  T  G  A  S  G  T  A  G  R  A  E  I  V  A  V  A

      2800      2810      2820      2830      2840      2850      2860      2870      2880
AGGCGGTCTGGGAAATGTATTGCCAACGTCACGATACGCCAGCAGGTAGCGAATGGCTGGAGAGCGTGCGGGCTTATGAGAAAGAAATTT
  G  G  L  G  N  V  L  P  T  S  R  Y  A  S  R  *

      2890SacII 2900      2910      2920      2930      2940      2950      2960      2970
TGAGTCGCCGCGGGTAAACACTGCCGGATGCGGCGCGAGCGCCTTATCCGGCCTACGGGTCGGCAACAGTTGTAGGCCTGATAAGACGCG
                                                                         ------------->
                 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>############<<<<<<<<<<<<<<<<<<

      2980      2990      3000      3010      3020      3030      3040      3050      3060
ACAGCGTCGCATCAGGCATTGATTGCCGGATGCGGCACAAGTGCCTTATCCGGCCTACAGGTCGGCAATAGTTGTAGGCCTGATAAGACG
  <-------------
<<<<<<<<<<<<<<<<<<★>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>############<<<<<<<<<<<<<<
```

FIG. 2—*Continued.*

function in protection from degradation by 3′ exoribonucleases (6).

The *rhaB* transcription start site was mapped by using primer extension analysis. We found in strain ECL116 a major rhamnose-inducible transcription start that mapped to 24 bp upstream of the *rhaB* translational start site (Fig. 4, lanes 1 and 2). Two minor RNA 5′ ends were mapped 18 and 19 bp upstream of the *rhaB* translational start site. It is not clear whether these represented independent transcription start sites or degradation products of the larger transcript.

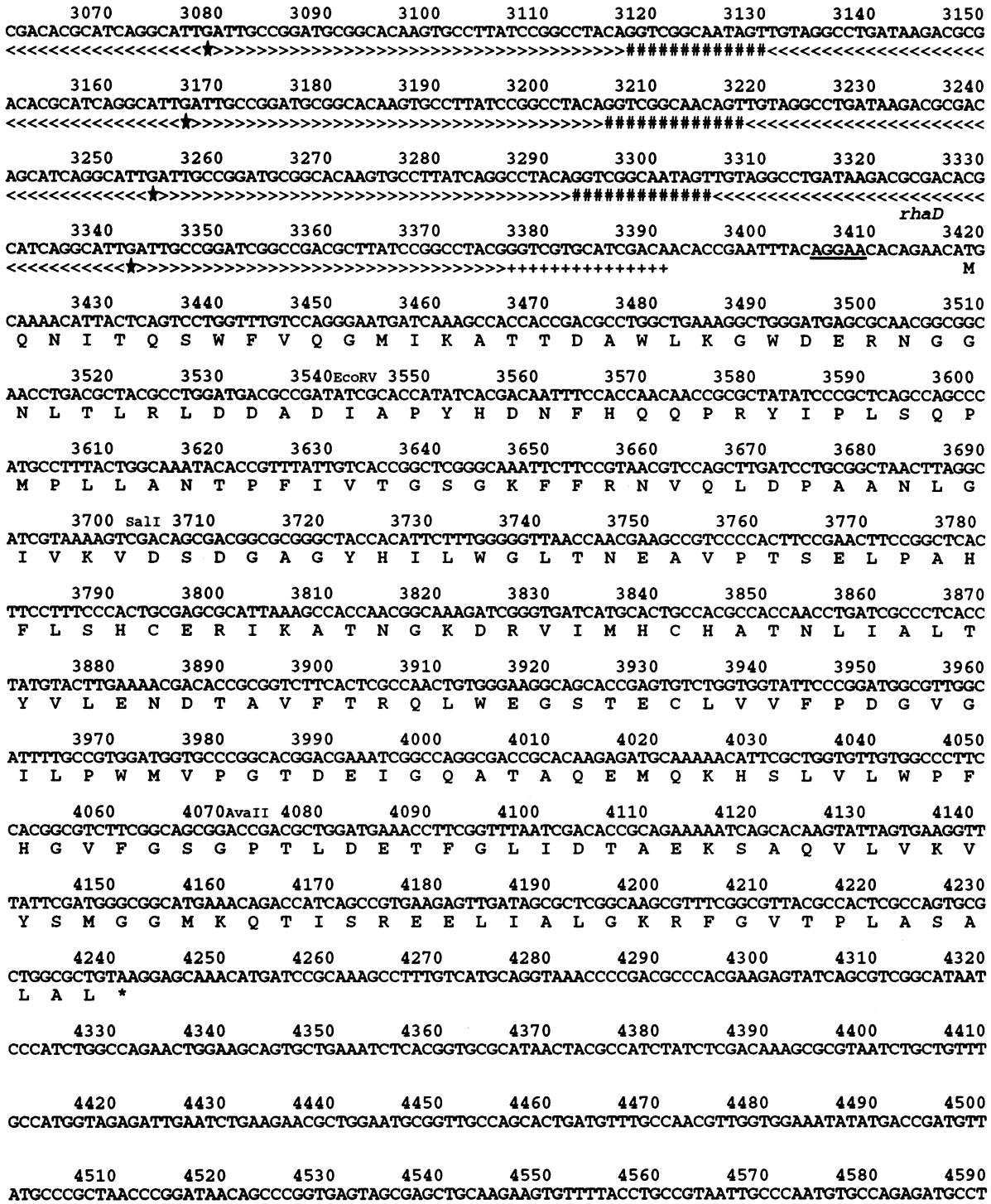

FIG. 2—*Continued.*

Neither transcript was detected in strain ECL339, which carries a deletion of the *rha* region (Fig. 4, lane 3). Prior to this sequencing of *rhaBAD*, Tobin and Schleif used S1 mapping to identify three transcription start points (p1, p2, and p3) in the *rhaBAD* region (40). The transcription start point upstream of the *rhaB* open reading frame identified here (pB) is distinct from all of the previously identified start points. It seems likely that, at least under most conditions, pB defines the major promoter for *rhaBAD* transcription. The p1 and p2 promoters defined by Tobin and Schleif (40) may correspond to the 5′ ends of the same breakdown products identified in the Northern blot experiments in this work. The locations of p1 and p2 suggest that *rhaBAD* transcripts may have been processed between the *rhaB* and

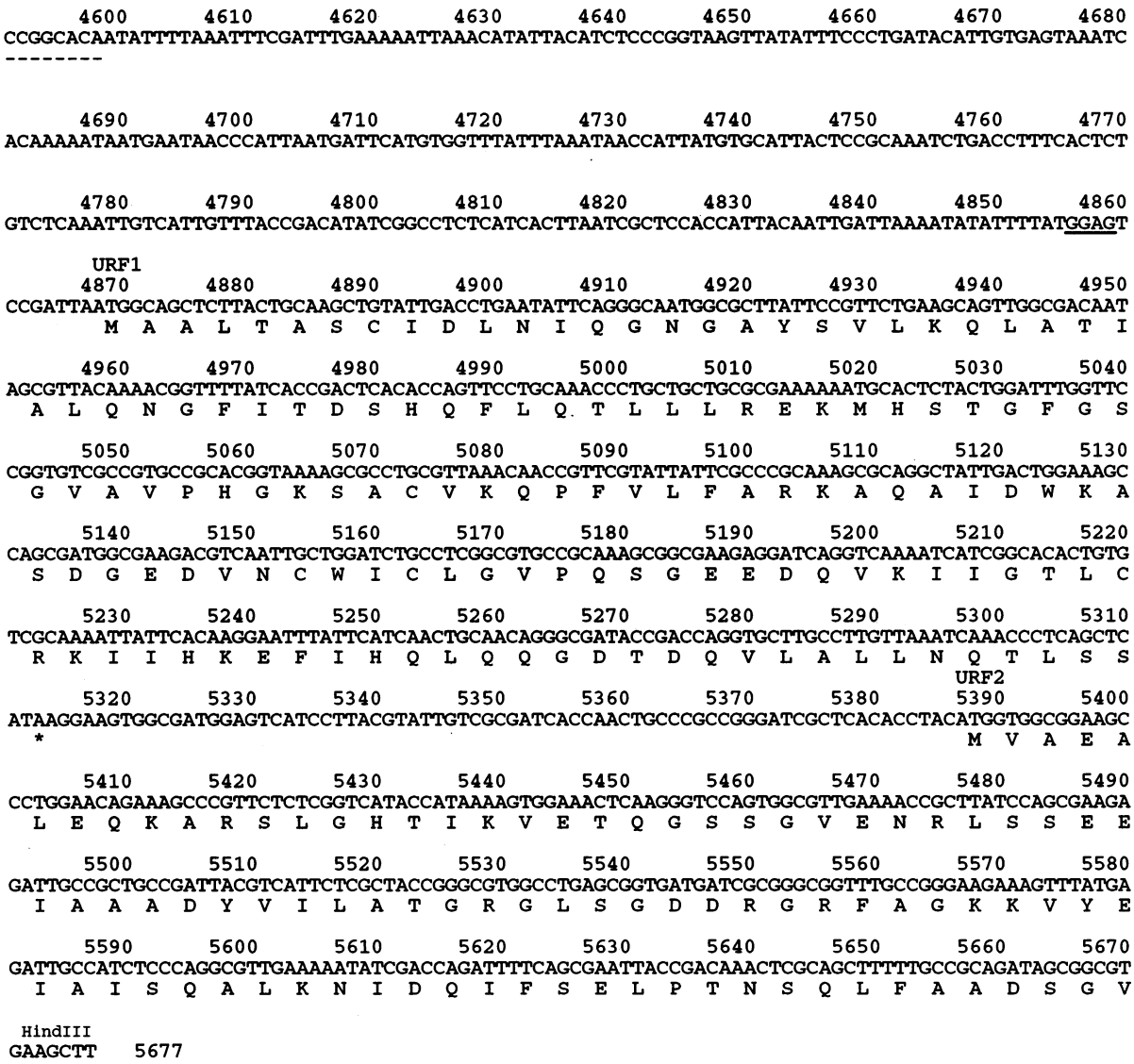

FIG. 2—*Continued.*

the *rhaA* genes. The origin of the p3 transcript (40), which would begin after the 3′ end of *rhaD* but with opposite transcriptional polarity, is unknown at this time.

To determine whether L-rhamnose or a metabolite of L-rhamnose was the direct inducer of the *rhaBAD* operon, we also determined whether L-rhamnose induced *rhaB* transcription in strains carrying point mutations in *rhaA*, *rhaB*, *rhaD*, or *rhaR* (Fig. 4, lanes 4 to 7). Mutations in *rhaA* are expected to block any catabolism of L-rhamnose, while *rhaB* mutations prevent formation of L-rhamnulose-1-phosphate and *rhaD* mutations prevent its further catabolism to dihydroxy acetone phosphate and L-lactaldehyde. *rhaB* transcription was detected in strains carrying mutations in each of the structural genes, indicating that L-rhamnose may be the direct inducer of *rhaBAD* transcription. The lower level of transcription in the *rhaD* mutant strain is likely due to the fact that this strain grew very poorly in the presence of L-rhamnose. This poor growth is presumably a consequence of the accumulation of the phosphorylated intermediate L-rhamnulose-1-phosphate (1). As expected (12), a point mutation in *rhaR* abolished *rhaBAD* transcription (Fig. 4, lane 7).

Sequence similarity with other proteins. The reported sequences of the L-fucose regulon gene cluster (11, 25) permitted a comparison with the functionally analogous enzymes encoded by the L-rhamnose regulon gene cluster whose sequences are presented here. Rhamnulose kinase was found to have 25% identity with *E. coli* fuculose kinase, 18% identity with *E. coli* xylulose kinase, and 66% identity with the rhamnulose kinase of *S. typhimurium* (31). No other kinase in the EMBL-GenBank data bank showed any homology with the *rhaB* product of *E. coli*, according to the TFASTA program.

In the cases of rhamnose isomerase and fucose isomerase of *E. coli*, homology was lower, with the alignment displaying only 15% identity. Likewise, in the cases of rhamnulose-1-phosphate aldolase and fuculose-1-phosphate aldolase, homology was very low, with an identity of 18%. As for

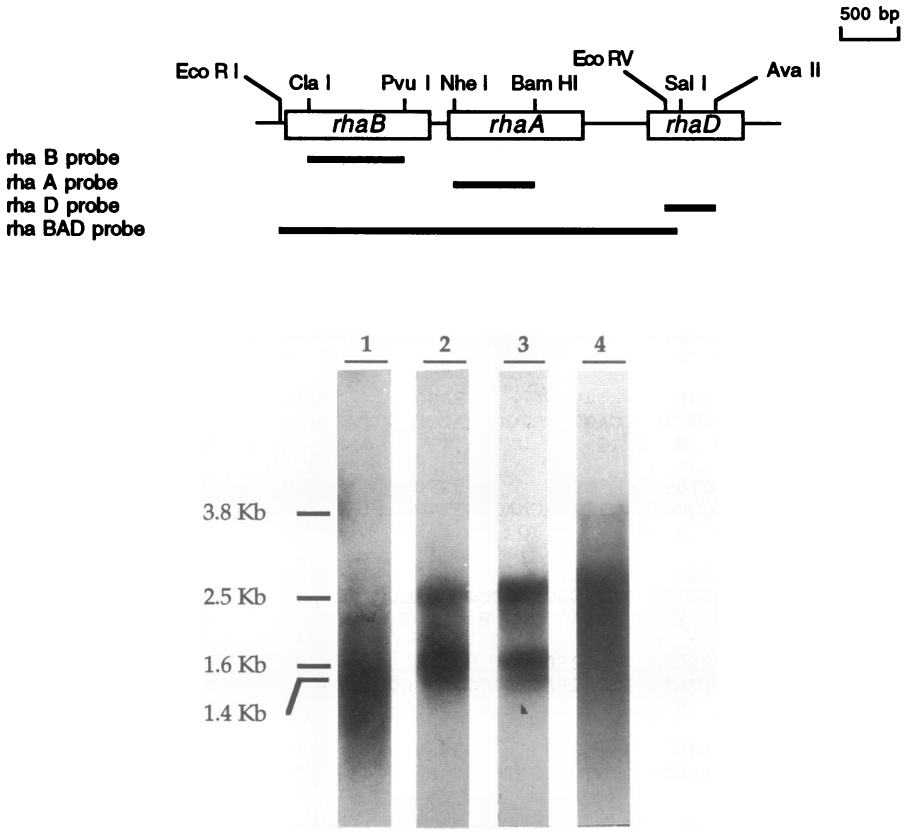

FIG. 3. Northern blots of mRNA from strain ECL1. RNA was isolated from cells grown on rhamnose to an A_{650} of 0.5 (lanes 1 to 3) or 0.2 (lane 4) and hybridized with the probes shown as thick lines in the upper part. A major transcript of 1.4 kb when the *rhaB* gene probe was used is apparent (lane 1). Two transcripts of 2.5 and 1.6 kb are present with the *rhaA* gene probe (lane 2) and with the *rhaD* gene probe (lane 3). A full-length transcript of 3.8 kb together with other RNA species appears when the probe used encompasses the three *rhaBAD* genes (lane 4).

rhamnulose kinase, no other isomerase or aldolase with significant homology to rhamnose isomerase or rhamnulose-1-phosphate aldolase was found.

In spite of the similarity between the reactions catalyzed by the corresponding enzymes in the parallel metabolic pathways for rhamnose and fucose, homologies between the sequences are rather low. Conservation is more stringent in specific short fragments, which are presumably involved in the active center, for the kinases but not for the isomerases or the aldolases. According to Sander and Schneider (35), the 25% identity between rhamnulose kinase and fuculose kinase is at the lower limit at which structural homology could be inferred, while sequence identities for the corresponding isomerases or aldolases are below the threshold permitting inference of structure homology.

TABLE 2. β-Galactosidase activities in strains containing transcriptional fusions of the *rhaB*, *rhaA*, and *rhaD* genes and grown under different conditions

Strain	β-Galactosidase activity[a] of cells grown on:		
	L-Rhamnose	L-Fucose	Glucose
JA121 (control)	<100	<100	<100
JA123 (*rhaB-lacZ*)	32,600	<100	<100
JA124 (*rhaA-lacZ*)	1,330	<100	<100
JA125 (*rhaD-lacZ*)	130	<100	<100
JA126 (*rhaBA-lacZ*)	24,900	<100	<100

[a] Enzyme activities are given in Miller units (28).

The degree of conservation between the genes of the two systems would be very low if one accepts a divergent evolution for the fucose and rhamnose genetic systems. However, the high homology between rhamnulose kinases of two different species, *E. coli* and *S. typhimurium*, seems to point to a convergent rather than to a divergent evolution.

G+C content and codon usage. The G+C content of *rhaB*, *rhaA*, and *rhaD* (55.3, 56.3, and 56%, respectively) is significantly higher than the approximately 50% G+C content which is the average of the whole genomes of *E. coli* K-12 and *S. typhimurium*. This could be interpreted to indicate that these genes were transferred to the enteric bacteria from an ancestor with a genome which was G+C rich, as has been proposed for the A+T-rich *rfb* region (8, 22). This possible horizontal acquisition of the rhamnose system would also hold true for the *rhaB* gene of *S. typhimurium* (31), which has a G+C content of 55%. The 50% G+C content of the fucose genes is interesting in view of the lack of homology between the corresponding proteins of the rhamnose and fucose systems presented above. If the high G+C content is indicative of an ancestral transfer of genes, the differences in G+C content suggest that each system was of a different origin.

The high G+C content of the three genes is reflected in the codon usage. At all codon positions, but particularly at the third one, there is a strong preference for G or C over A or

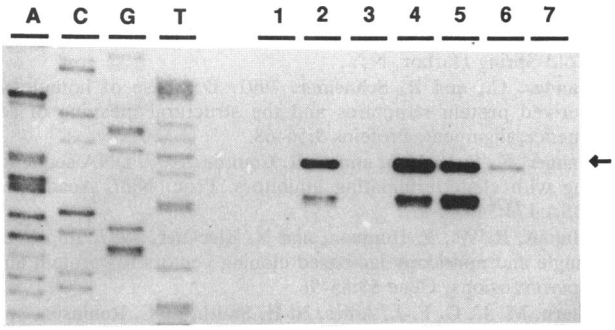

FIG. 4. Primer extension analysis of *rhaBAD* transcription. Primer extension reactions were performed as described in Materials and Methods. Sequencing reactions were performed by using the same ^{32}P-labeled oligonucleotide as used for the primer extension reactions (lanes A, C, G, and T). The wild-type strain was grown in the absence or presence of L-rhamnose, while all other strains were grown in the presence of L-rhamnose (see Materials and Methods). Lanes: 1, ECL116 (wild type) without rhamnose; 2, ECL116 with rhamnose; 3, ECL339 [Δ(*rha-pfk*)]; 4, ECL714 (*rhaB101*); 5, ECL715 (*rhaA502*); 6, ECL716 (*rhaD701*); 7, ECL717 (*rhaR702*). The arrow indicates nucleotide position at which major transcription start takes place.

U. The scores for the frequency of optimal codon usage (21) of 0.64 for *rhaB*, 0.61 for *rhaA*, and 0.65 for *rhaD*, close to the 0.59 reported for *rhaT* (39), are significantly lower than the scores of highly expressed proteins from genes such as *ompA* or *lpp*, which have scores of 0.92 and 0.98, respectively (21). These differences in codon usage could indicate that *rhaBAD* encodes proteins which are not highly expressed in *E. coli*. The genes *rhaR* and *rhaS* for the rhamnose regulatory proteins (40) have scores of 0.56 and 0.55, respectively, also corresponding to proteins with low expression levels, as do other regulatory proteins encoded by genes such as *trpR* (0.56) or *araC* (0.54) (21).

ACKNOWLEDGMENTS

This work was supported by grant PB91-437 from the DGICYT of Spain and by NIH grant GM18277 to R. Schleif. P.M. and E.H. were recipients of predoctoral fellowships (FPI) from the Ministerio de Educación y Ciencia of Spain, and S.M.E. was the recipient of NIH postdoctoral fellowship GM14364.

We thank M. Aldea for the gift of bacterial strains and plasmids and for helpful discussion and R. Schleif for critical reading of the manuscript.

ADDENDUM IN PROOF

While this article was being reviewed, we found that a nucleotide was erroneously inserted at position 2747. Deletion of this nucleotide changes the carboxyl-terminal part of rhamnose isomerase and adds 19 amino acids, yielding a molecular mass of 47,231 Da for the complete protein.

REFERENCES

1. **Al-Zarban, S., L. Hefferman, J. Nishitani, L. Ransone, and G. Wilcox.** 1984. Positive control of the L-rhamnose genetic system in *Salmonella typhimurium* LT2. J. Bacteriol. **158**:603–608.
2. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
3. **Backman, K., Y.-M. Chen, and B. Magasanik.** 1981. Physical and genetic characterization of the *glnA-glnG* region of the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **78**:3743–3747.
4. **Badía, J. (University of Barcelona).** 1991. Personal communication.
5. **Badia, J., L. Baldoma, J. Aguilar, and A. Boronat.** 1989. Identification of the *rhaA*, *rhaB*, and *rhaD* gene products from *Escherichia coli* K12. FEMS Microbiol. Lett. **65**:253–258.
6. **Belasco, J. G., T. Beatty, C. W. Adams, A. von Gabain, and S. N. Cohen.** 1985. Differential expression of photosynthesis genes in *R. capsulata* results from segmental differences in stability within the polycistronic *rxcA* transcript. Cell **40**:171–181.
7. **Boronat, A., and J. Aguilar.** 1979. Rhamnose-induced propanediol oxidoreductase in *Escherichia coli*: purification, properties, and comparison with the fucose-induced enzyme. J. Bacteriol. **140**:320–326.
8. **Brown, P. K., L. K. Romana, and P. R. Reeves.** 1992. Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): genetic basis of the polymorphism between groups C2 and B. Mol. Microbiol. **6**:1385–1394.
9. **Chen, C.-Y. A., J. T. Beatty, S. N. Cohen, and J. G. Belasco.** 1988. An intercistronic stem-loop structure functions as an mRNA decay terminator necessary but insufficient for puf mRNA stability. Cell **52**:609–619.
10. **Chen, Y.-M., and E. C. C. Lin.** 1984. Dual control of a common L-1,2-propanediol oxidoreductase by L-fucose and L-rhamnose in *Escherichia coli*. J. Bacteriol. **157**:828–832.
11. **Chen, Y.-M., Z. Lu, and E. C. C. Lin.** 1989. Constitutive activation of the *fucAO* operon and silencing of the divergently transcribed *fucPIK* operon by an IS*5* element in *Escherichia coli* mutants selected for growth on L-1,2-propanediol. J. Bacteriol. **171**:6097–6105.
12. **Chen, Y.-M., J. F. Tobin, Y. Zhu, R. F. Schleif, and E. C. C. Lin.** 1987. Cross-induction of the L-fucose system by L-rhamnose in *Escherichia coli*. J. Bacteriol. **169**:3712–3719.
13. **Chen, Y.-M., Y. Zhu, and E. C. C. Lin.** 1987. The organization of the *fuc* regulon specifying L-fucose dissimilation in *Escherichia coli* K12 as determined by gene cloning. Mol. Gen. Genet. **210**:331–337.
14. **d'Aubenton Carafa, Y., E. Brody, and C. Thermes.** 1990. Prediction of Rho-independent *Escherichia coli* transcription terminators. J. Mol. Biol. **216**:835–858.
15. **Dimri, G. P., K. E. Rudd, M. K. Morgan, H. Bayat, and G. F.-L. Ames.** 1992. Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria. J. Bacteriol. **174**:4583–4593.
16. **Galloway, J. L., and T. Platt.** 1988. Signals sufficient for rho-dependent transcription termination at *trp t'* span a region centered 60 base pairs upstream of the earliest 3' end point. J. Biol. Chem. **263**:1761–1767.
17. **García-Martín, C., L. Baldomá, J. Badía, and J. Aguilar.** 1992. Nucleotide sequence of the *rhaR-sodA* interval specifying *rhaT* in *Escherichia coli*. J. Gen. Microbiol. **138**:1109–1116.
18. **Gilson, E., W. Saurin, D. Perrin; S. Bachellier, and M. Hofnung.** 1991. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. **19**:1375–1383.
19. **Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung.** 1991. The BIME family of bacterial highly repetitive sequences. Res. Microbiol. **142**:217–222.
20. **Hidalgo, E., Y. M. Chen, E. C. C. Lin, and J. Aguilar.** 1991. Molecular cloning and DNA sequencing of the *Escherichia coli* K-12 *ald* gene encoding aldehyde dehydrogenase. J. Bacteriol. **173**:6118–6123.
21. **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.
22. **Jiang, X.-M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5**:695–713.

23. **Lin, E. C. C., and T. T. Wu.** 1984. Functional divergence of the L-fucose system in mutants of *Escherichia coli*, p. 135–163. *In* R. P. Mortlock (ed.), Microorganisms as model systems for studying evolution. Plenum, New York.
24. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193:**265–275.
25. **Lu, Z., and E. C. C. Lin.** 1989. The nucleotide sequence of *Escherichia coli* genes for L-fucose dissimilation. Nucleic Acids Res. **17:**4883–4884.
26. **Lupski, J. R., and G. M. Weinstock.** 1992. Short, interspersed repetitive DNA sequences in prokaryotic genomes. J. Bacteriol. **174:**4525–4529.
27. **McCormick, J. R., J. M. Zengel, and L. Lindahl.** 1991. Intermediates in the degradation of mRNA from the lactose operon of *Escherichia coli*. Nucleic Acids Res. **19:**2767–2776.
28. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
29. **Morgan, W. D., D. G. Bear, B. L. Litchman, and P. H. von Hippel.** 1985. A sequence and secondary structure requirements for rho-dependent transcription termination. Nucleic Acids Res. **13:**3739–3754.
30. **Newbury, S. F., N. H. Smith, and C. F. Higgins.** 1987. Differential mRNA stability controls relative gene expression within a polycistronic operon. Cell **51:**1131–1143.
31. **Nishitani, J., and G. Wilcox.** 1991. Cloning and characterization of the L-rhamnose regulon in *Salmonella typhimurium* LT2. Gene **105:**37–42.
32. **Power, J.** 1967. The L-rhamnose genetic system in *Escherichia coli* K12. Genetics **55:**557–568.
33. **Reeder, T., and R. Schleif.** 1993. AraC protein can activate transcription from only one position and when pointed in only one direction. J. Mol. Biol. **231:**205–218.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Sander, C., and R. Schneider.** 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins **9:**56–68.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **174:**5463–5467.
37. **Simons, R. W., F. Houman, and N. Kleckner.** 1987. Improved single and multicopy lac-based cloning vectors for protein and operon fusions. Gene **53:**85–96.
38. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic (REP) sequences: a major component of the bacterial genome. Cell **37:**1015–1026.
39. **Tate, C. G., J. A. R. Muiry, and P. J. F. Henderson.** 1992. Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H^+ transport protein in *Salmonella typhimurium* and *Escherichia coli*. J. Biol. Chem. **267:**6923–6932.
40. **Tobin, J. F., and R. F. Schleif.** 1987. Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J. Mol. Biol. **196:**789–799.
41. **Tobin, J. F., and R. F. Schleif.** 1990. Purification and properties of *rhaR*, the positive regulator of the L-rhamnose operons of *Escherichia coli*. J. Mol. Biol. **211:**75–89.
42. **Ziemke, P., and J. E. G. McCarthy.** 1992. The control of mRNA stability in *Escherichia coli*: manipulation of the degradation pathway of the polycistronic *atp* mRNA. Biochim. Biophys. Acta **1130:**297–306.