JOURNAL OF BACTERIOLOGY, May 1988, p. 2070–2077
0021-9193/88/052070-08$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 5

Nucleotide Sequence and Regulation of a Gene Involved in Bile Acid 7-Dehydroxylation by *Eubacterium* sp. Strain VPI 12708

JAMES P. COLEMAN,† W. BRUCE WHITE, MARK LIJEWSKI, AND PHILLIP B. HYLEMON*

Department of Microbiology and Immunology, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia 23298-0678

Received 16 November 1987/Accepted 11 February 1988

***Eubacterium* sp. strain VPI 12708 is an anaerobic intestinal bacterium that has inducible bile acid 7-dehydroxylation activity. At least four new polypeptides were synthesized after addition of primary bile acids to the growth medium. One of these, of molecular weight 27,000 (P-27), was shown to be involved in the 7-dehydroxylation reaction sequence. The gene coding for P-27 was cloned, and the entire DNA sequence for the protein-coding region was determined. In addition, sequence information was obtained for 294 bases upstream from the translational start codon and 329 bases downstream from the stop codon. Induction studies with a synthetic oligonucleotide probe (16-mer) revealed the presence of a cholic acid-inducible mRNA species approximately 900 bases long. A 5′ terminus of this mRNA was detected by primer extension analysis, and the location of the 3′ terminus of the mRNA was estimated by using S1 nuclease mapping. The 3′ terminus of the mRNA contained a large element with dyad symmetry of unknown function. The open reading frame contained 249 codons, and the corresponding polypeptide had a calculated molecular weight of 26,745. The amino acid sequence of P-27 showed significant homology to several previously described alcohol-polyol dehydrogenases ("nonzinc" dehydrogenases), especially in the region believed to contain a pyridine nucleotide-binding domain. The implications of this homology and the possible function of P-27 in bile acid 7-dehydroxylation are discussed.**

Cholic acid and chenodeoxycholic acid are the major bile acids synthesized by humans. During their enterohepatic circulation, these primary bile acids are 7α-dehydroxylated by intestinal bacteria, producing deoxycholic acid and lithocholic acid, respectively. These microbial products constitute 20 to 25% of the total biliary bile acid pool of humans (31). The secondary bile acids have markedly different physical-chemical properties (3), which are manifested in a variety of physiological effects (1, 12, 13, 21, 24, 27). Because 7α-dehydroxylation of primary bile acids is an important physiological reaction in the intestinal ecosystem, there is considerable interest in the mechanism and regulation of this microbial transformation.

Early work on the mechanism of 7-dehydroxylation by Samuelsson (29) led to the proposal that the 7α-hydroxy group is removed by dehydration from carbons 6 and 7 (7α-hydroxy, 6β-hydrogen), yielding a Δ^6-bile acid intermediate, which is then reduced. *Eubacterium* sp. strain VPI 12708 is a human intestinal bacterium that has bile acid 7α-dehydroxylation activity that is induced by culturing in the presence of a C_{24} bile acid containing an unhindered 7α-hydroxy group (37). Also inducible under these conditions are an NADH:flavin oxidoreductase (16) and a unique adenosine nucleotide:bile acid conjugating activity (4). Work in this laboratory with crude extracts of *Eubacterium* sp. strain VPI 12708 supported Samuelsson's proposed mechanism by showing that exogenous Δ^6-bile acids could be reduced to the corresponding saturated bile acids (36). Surprisingly, this reaction and 7α-dehydroxylation in general, both of which represent net reductions, were found to be strongly stimulated by NAD^+ and inhibited by NADH (36, 38). The reduction of Δ^6-bile acid was further stimulated by addition of reduced free flavins (38). These anomalous observations could not be explained until it was discovered that bile acid ring modifications occurred while the bile acid was covalently linked to an adenosine nucleotide (4). The major bile acid metabolite found linked to the adenosine nucleotide with deoxycholic acid as a substrate was 3-keto-12α-hydroxy-4-cholenoic acid. This finding, in conjunction with data suggesting that a 3-keto-7α-hydroxy-Δ^4-steroid configuration easily loses the 7α-hydroxy group (2), has led to the hypothesis that 7-dehydroxylation by *Eubacterium* sp. strain VPI 12708 occurs after two sequential oxidations at the 3-hydroxy group and at the C4-C5 bond (Fig. 1). Loss of the 7α-hydroxy group then yields 3-keto-12-hydroxy-4,6-cholenoic acid. This intermediate is then sequentially reduced, ultimately yielding the 7-dehydroxylated bile acid. It should be pointed out that this mechanism is entirely consistent with the data of Samuelsson (29), who showed loss of a tritium from the 6β position after 7α-dehydroxylation of cholic acid. Analysis of culture media from bile acid transformations by aerobic bacteria has also shown the presence of several of these proposed intermediates, including bile acids with a 3-keto group, with a 3-keto-Δ^4 structure, and with a 3-keto-$\Delta^{4,6}$ structure (6, 8, 9, 23). No bile acids with the Δ^6 structure alone were found, suggesting that oxidation at carbons 3 and 4 is required before loss of the 7-hydroxy group can occur.

Induction of 7-dehydroxylation activity in *Eubacterium* sp. strain VPI 12708 coincides with the appearance of at least four new polypeptides, with molecular weights of 23,500, 27,000, 45,000, and 77,000, as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (25, 36). The 27,000- and 45,000-molecular-weight polypeptides have been shown to coelute with 7-dehydroxylating activity from a

* Corresponding author.
† Present address: Department of Microbiology and Immunology, East Carolina University School of Medicine, Greenville, NC 27858-4354.

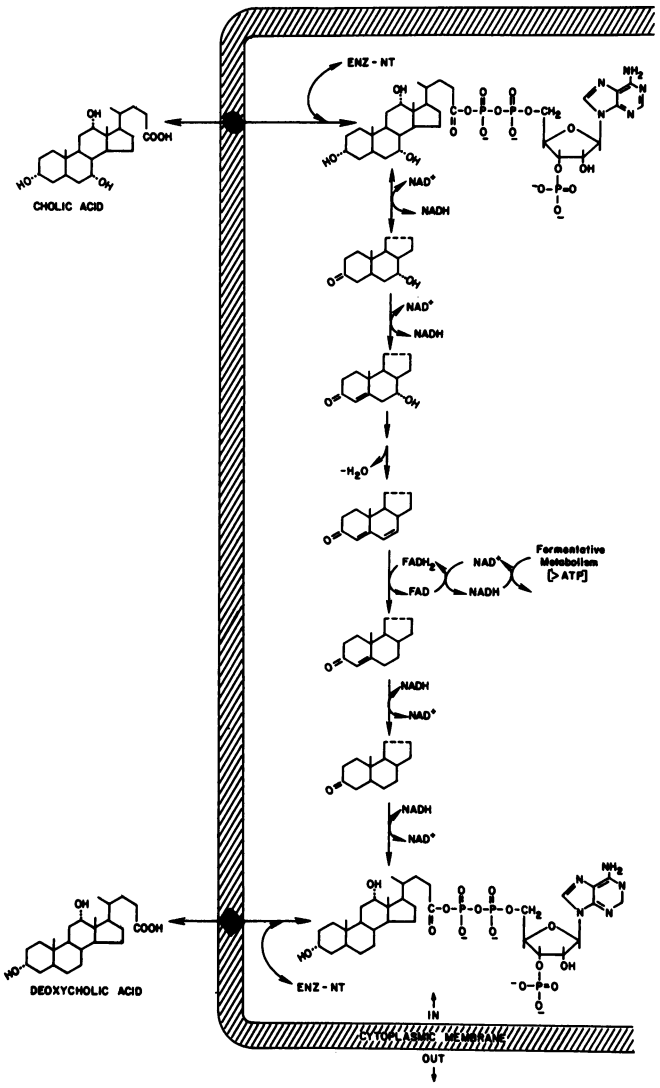

FIG. 1. Current proposal for bile acid 7-dehydroxylation in *Eubacterium* sp. VPI 12708 based on recent observations. ENZ-NT, Enzyme-bound nucleotide.

high-performance gel filtration column (25). Immunoinhibition experiments have implicated the 27,000-molecular-weight polypeptide (P-27) as a major catalytic component of the bile acid 7α-dehydroxylation pathway (25). To gain further knowledge of the reactions involved in 7-dehydroxylation by *Eubacterium* sp. strain VPI 12708 and the regulation of its synthesis, studies were undertaken to characterize these inducible polypeptides. In a previous study, we described the cloning of the gene coding for P-27 from *Eubacterium* sp. strain VPI 12708 into *Escherichia coli* (5). This report describes the characterization of the P-27 gene and its protein product and the possible role of P-27 in the dehydroxylation reaction.

MATERIALS AND METHODS

Materials. T4 polynucleotide kinase and T4 DNA ligase were obtained from Bethesda Research Laboratories, Inc. Restriction enzymes were obtained from Pharmacia, Inc., and International Biotechnologies, Inc. S1 nuclease, dideoxynucleotide sequencing reagents, and M13mp19 were obtained from Pharmacia. Reverse transcriptase, [^{32}P] dATP, [^{35}S]dATP analog, and GeneScreen hybridization membranes were obtained from New England Nuclear Corp. All other chemicals and materials were of the highest grade commercially available.

Organisms and growth conditions. *Eubacterium* sp. strain VPI 12708 was grown as described previously (37). *E. coli* HB101 containing recombinant plasmids was grown on Luria broth containing 50 μg of ampicillin per ml (17). Recombinant plasmids pPBH12 (pUC8-12) and pPBH-22 (pUC8-22) were described previously (5).

Nucleic acid isolation. Plasmid DNA was isolated by the Ish-Horowitz modification of the method of Birnboim and Doly (17) and further purified by CsCl gradient centrifugation. M13 DNA was prepared for sequencing by the procedure described by Pharmacia. RNA was isolated from *Eubacterium* sp. VPI 12708 by a procedure developed for RNA isolation from clostridia (15). Cells in the exponential phase of growth were mixed with 1/5 volume of 200 mM Tris hydrochloride (pH 8.0)–60% (wt/vol) sucrose containing chloramphenicol and lysozyme at 0.6 and 1.2 mg/ml, respectively. After incubation at room temperature for 15 min, the protoplasts were recovered by centrifugation at 6,000 × *g* for 15 min at room temperature. The cell pellet was resuspended in 3/100 of the original culture volume in 10 mM Tris hydrochloride (pH 8.0)–10 mM NaCl–5 mM EDTA–0.5% (wt/vol) sodium dodecyl sulfate. The resulting suspension was swirled gently in a 60°C bath until clearing was observed. The lysate was then centrifuged at 10,000 × *g* for 10 min. A volume of 5 M sodium acetate (pH 5.0) was added to the supernatant for a final sodium acetate concentration of 25 mM. After buffer-saturated phenol and chloroform-isoamyl alcohol (24:1 vol/vol) extractions, the nucleic acids were precipitated with 2.5 volumes of cold 95% ethanol and then stored as suspensions in 80% ethanol at −20°C

Oligonucleotide synthesis and purification. Oligonucleotides were synthesized and purified as described previously (5).

Recombinant DNA methods. Restriction endonuclease digestions were performed as recommended by the suppliers. Electrophoresis of DNA and RNA, ligation reactions, and other nucleic acid manipulations were performed as described by Maniatis et al. (17).

DNA sequencing. Sequence information was initially obtained by using double-stranded inserts in pUC8 or pUC19 by the general procedure of Sanger et al. (30) as modified by Wallace et al. (35) for use with a double-stranded template. Commercially available oligonucleotide primers complementary to vector DNA on either side of the insert were used to sequence both strands of the insert proximal to the vector. Oligonucleotide primers were then chemically synthesized to sequence internal areas of the inserts. Regions of ambiguous DNA sequences were further analyzed after subcloning of smaller fragments into M13mp19 and sequencing by the dideoxynucleotide procedure of Sanger et al. (30). [^{35}S] dATP (>600 Ci/mmol; New England Nuclear Corp.) was used for all sequencing reactions.

Analysis of sequence data. Analyses of nucleic acid and protein sequence data were performed with either the Microgenie sequence analysis program (Beckman Instruments, Inc.) or the IBI/Pustell DNA and protein sequence analysis system (International Biotechnologies, Inc.).

Primer extension reactions. A 5′ terminus of the mRNA synthesized by *Eubacterium* sp. strain VPI 12708 was

mapped by the primer extension technique (14). *Eubacterium* total RNA (10 μg) from a culture containing cholic acid was mixed with 2.5×10^5 cpm of P-27 gene-specific primer (labeled with [^{32}P]ATP and T4 polynucleotide kinase), and the mixture was ethanol precipitated and dried briefly. It was then dissolved in 10 μl of 0.1 M sodium chloride–0.01 M Tris hydrochloride (pH 7.9)–0.1 mM EDTA. The mixture was drawn into a glass capillary tube and, after the ends were sealed, heated in a beaker of boiling water for 1 min. The beaker was then removed and allowed to cool slowly to 60°C, after which it was maintained at that temperature for 4 to 5 h. The contents of the tube were then expelled into an equal volume of 80 mM Tris hydrochloride (pH 7.9)–10 mM $MgCl_2$–4 mM dithiothreitol–0.4 mM each dATP, dCTP, dGTP, and dTTP–5 U of avian myeloblastosis virus reverse transcriptase–10 U of RNasin (Promega Biotech). The solution was kept at 4°C for 5 min and then incubated at 37°C for 30 min. The nucleic acids were then ethanol precipitated, dried briefly, and dissolved in 10 μl of formamide-dye mix before being run on an 8% polyacrylamide–urea sequencing gel (2.5 μl per lane). Primer extension reactions were run parallel to standard DNA dideoxy sequencing reactions with the same primer and plasmid pPBH22, which contains the nucleic acid sequence upstream of the gene in addition to the first 290 base pairs (bp) of the protein-coding region.

S1 nuclease mapping. The location of the 3′ terminus of P-27 mRNA was estimated by the procedure of Murray (20). RNA (10 μg) from either induced or uninduced *Eubacterium* sp. strain VPI 12708 was mixed with 0.2 μg of a double-stranded DNA restriction fragment containing the region expected to encompass the termination site (see Fig. 7A). The mixture was lyophilized in a 1.5-ml microcentrifuge tube and then suspended in 10 μl of 3.0 M sodium trichloroacetate–50 mM piperazine-*N*,*N*′-bis(ethanesulfonic acid) (PIPES, pH 7.0)–5 mM EDTA. After being overlaid with approximately 50 μl of mineral oil, the tubes were heated at 60°C for 5 min and then cooled to 50°C and maintained at that temperature for 4 h. The reassociation was terminated by addition of 200 μl of cold 200 mM NaCl–40 mM sodium acetate (pH 5.5)–1 mM $ZnCl_2$–20 μg of denatured calf thymus DNA per ml–2,000 U of S1 nuclease per ml. The samples were incubated at 37°C for 30 min, followed by addition of sodium dodecyl sulfate and EDTA to 0.2% and 20 mM, respectively. The samples were extracted once with 200 μl of chloroform-isoamyl alcohol (24:1) before precipitation with 1 volume of isopropanol. After washing with 50% isopropanol in 0.3 M ammonium acetate, the samples were dried briefly, suspended in formamide-dye sequencing sample buffer, and run on a 5% acrylamide–7 M urea gel.

RESULTS

DNA sequence analysis. Previously reported sequence data indicated that pPBH22 (pUC8-22) contained an *Eco*RI insert with a terminal 349-bp *Taq*I-*Eco*RI fragment which contained the N terminus of P-27 (5). Furthermore, pPBH12 (pUC8-12) contained a *Taq*I insert beginning with this same *Taq*I-*Eco*RI fragment and was of a size sufficient to code for P-27. A partial restriction map of the insert in pPBH12, as well as the adjoining upstream sequence from pPBH22, is shown in Fig. 2, along with the sequencing strategy used to determine the nucleotide sequences of both strands. The complete DNA sequence is shown in Fig. 3. Only one open reading frame was present, and it contained start and stop codons bounded by residues 295 and 1044.

A putative ribosome binding site (AAGGAAGG) was present 8 bases upstream from the initiation codon. Provided

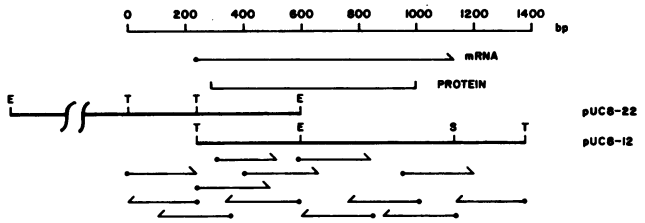

FIG. 2. Sequencing strategy for P-27 gene. The two separate inserts into pUC8, pPBH12 (pUC8-12), and pPBH22 (pUC8-22) are indicated as heavy solid lines, with restriction enzyme cleavage sites marked as follows: E, *Eco*RI; T, *Taq*I; S, *Sma*I. The scale at the top indicates the span of nucleotides whose sequence was determined (total length, 1,371 bp), and the lines designated mRNA and protein represent the regions containing the entire transcript and amino acid sequence information, respectively. The closed circles and arrows below the inserts indicate the priming areas and regions sequenced, respectively, for each set of dideoxy sequencing reactions. Closed circles located at restriction sites indicate those reactions for which commercially available primers were used to sequence fragments of the gene subcloned into either pUC19 or M13mp19, whereas closed circles located between restriction sites correspond to oligonucleotide primers specifically synthesized for sequencing the appropriate region.

the sequence of *Eubacterium* 16S RNA is homologous to that from the eubacterial 16S RNAs (39), the free energy of binding between the mRNA and the 16S RNA has been calculated to be −14.8 kcal (1 cal = 4.184 J)/mol (34).

Once a putative 5′ terminus of the P-27 mRNA was detected (see below), the sequence immediately upstream from it was examined for the presence of sequences resembling promoter consensus sequences (28). One possible region was present between nucleotides 195 and 227. A sequence in the −10 region (TAaAgT) matched the *E. coli* −10 consensus sequence (TAtAaT) in four of six bases. Moreover, these four bases correspond to the most highly conserved bases of the *E. coli* consensus −10 sequence (28). The end of this sequence was separated from the 5′ end of the predominant P-27 mRNA species by only six bases. The region upstream from the putative −10 sequence contained two six-base sequences with reasonable similarity to the *E. coli* consensus −35 sequence (TTGACa). One (TTtcCa) matched at four of six bases, and the other (TTccaa) matched at three of six bases. However, these sequences were separated from the start of the putative −10 sequence by 21 and 20 bases, respectively.

A large element of dyad symmetry was present 43 bases downstream from the stop codon (Fig. 4). It had a stem length of 23 bases (including mismatches) with a 6-base hairpin loop. The free energy of formation was calculated to be −48.8 kcal/mol (34).

Transcriptional activation of the P-27 gene by cholic acid. The bile acid inducibility of the P-27 gene was demonstrated by monitoring the presence of P-27-specific RNA before and after addition of cholic acid (0.1 mM final concentration) to *Eubacterium* sp. strain VPI 12708 in the exponential phase of growth. Northern (RNA) blot analysis of RNA isolated under these conditions showed the appearance of P-27-specific mRNA at 10 min, with none detectable before cholic acid addition (Fig. 5). The size of the RNA band was estimated to be in the range of 850 to 950 bases.

Detection of 5′ and 3′ ends of P-27 mRNA. A 5′ end of the predominant P-27 mRNA was detected by the primer extension technique. Comparison of the product of the reverse transcriptase reaction by using a primer hybridized to total

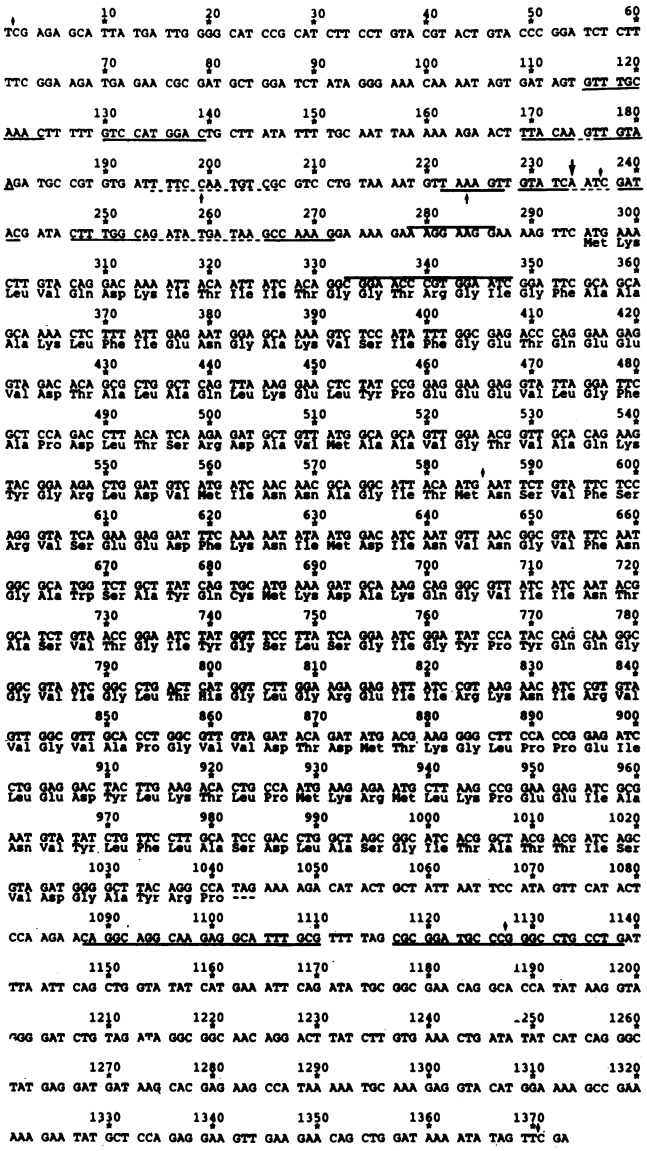

FIG. 3. DNA sequence of the P-27 gene and adjacent regions. The large arrow pointing downward at nucleotide 234 indicates the transcriptional start site. The putative −10 region is underlined, and the base at the position 10 bases from the start site is marked with an arrow pointing upward. The bases from positions 195 to 206, underlined with the broken line, indicate the area containing seven possible 6-base sequences with a spacing from the −10 region that varies from 15 to 21 bases. The nucleotide 35 bases from the start site is indicated by an arrow pointing upward. The overlined region at nucleotides 279 to 286 is the most likely ribosome-binding site. The large dyad symmetry element between nucleotides 1086 and 1136 is underlined. Several other elements with dyad symmetry in the region upstream from the protein-coding area are also underlined (nucleotides 115 to 124, 130 to 139, 169 to 181, 229 to 242, and 247 to 271). Restriction enzyme sites are marked by small arrows pointing downward. *Taq*I cut sites are present after nucleotides 1, 236, and 1368. The *Eco*RI site is present after nucleotide 585, and the *Sma*I cut site is after nucleotide 1125. The bases overlined at positions 333 to 348 indicate the sequence to which a complementary synthetic oligonucleotide was made to assay for the presence of a P-27-specific transcript and to use as a primer in 5′ terminus mapping.

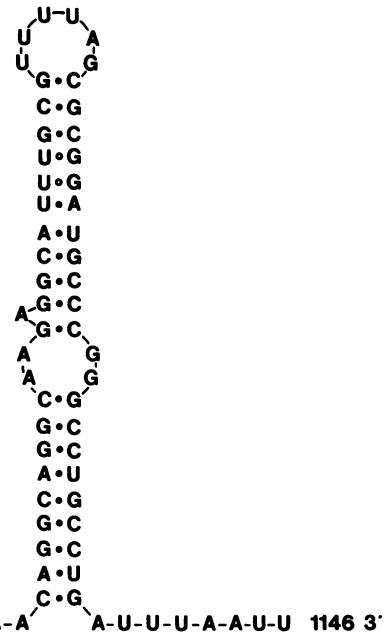

FIG. 4. Potential secondary structure at the 3′ end of the P-27 gene. The stem-and-loop structure shown begins at a nucleotide corresponding to DNA sequence residue 1088 and ends at nucleotide 1138. The RNA sequence is shown here and includes G-U base pairing. The free energy of formation for this structure (25°C) is −48.8 kcal/mol (34). It is possible for a slightly higher free energy to be calculated if the interior loop consisting of G-A residues is paired as described by Woese et al. (39).

Eubacterium RNA with DNA sequencing reactions containing the same primer showed a distinct band corresponding to base 234 (Fig. 6). No other bands were evident within the range of separation of the gel (300 bases). Interestingly, we also observed (data not shown) a 5′ terminus corresponding to nucleotide positions 234 and 235 when RNA was isolated from *E. coli* transformed with pPBH22. S1 nuclease mapping revealed a 3′ end of the RNA from *Eubacterium* sp. about 600 bases downstream from the *Eco*RI site (Fig. 7).

Amino acid sequence. The derived amino acid sequence of P-27 is shown below the nucleotide sequence in Fig. 3. The polypeptide contains 249 amino acids and has a calculated molecular weight of 26,745. A computer-aided protein sequence homology search indicated significant homology between P-27 and several previously sequenced alcohol-polyol dehydrogenases, specifically the ribitol dehydrogenase from *Klebsiella aerogenes* (18), the alcohol dehydrogenase from *Drosophila melanogaster* (33), and the glucose dehydrogenase from *Bacillus megaterium* (10). These enzymes are all of the same approximate size (247 to 262 amino acids) and belong to a group of pyridine nucleotide-linked dehydrogenases referred to as "short" or "nonzinc" dehydrogenases (11). The area of highest sequence similarity between P-27 and these other proteins is shown in Fig. 8. This conserved amino acid sequence has previously been postulated to contain two β-sheet areas responsible for pyridine nucleotide coenzyme binding (11).

DISCUSSION

Two separate genomic DNA fragments from *Eubacterium* sp. strain VPI 12708 were cloned into *E. coli* with the plasmid vector pUC8. One insert consisted of a 2,200-bp

*Eco*RI fragment, and the other was an 1,154-bp *Taq*I fragment. The two inserts contained a common 349-bp *Taq*I-*Eco*RI fragment. *E. coli* strains containing the plasmid with the 1,150-bp insert produced a 27,000-molecular-weight polypeptide which was immunologically cross-reactive with P-27 purified from *Eubacterium* sp. strain VPI 12708. In addition, preliminary nucleic acid sequence data analysis indicated an open reading frame coding for an amino acid sequence identical to that found for the N terminus of P-27 (5). Together, the two inserts constitute a continuous 3-kilobase stretch of *Eubacterium* DNA containing the entire coding region for P-27, as well as approximately 2 kilobases of upstream DNA and 329 bp downstream from the P-27 stop codon.

The genus *Eubacterium* has been classified, on the basis of 16S rRNA fingerprint analyses, with the clostridia in the subdivision of "gram-positive eubacteria with low G+C content" of the gram-positive group (40). Because of the absence of literature on gene structures in *Eubacterium* species, the DNA sequence analyses in this report were analyzed and compared with bacillus and clostridium sequence information (7, 19), as well as with information about *E. coli* systems (28). Reports from other laboratories on several putative bacillus and clostridial vegetative promoters (7, 19) have indicated that the −10 and −35 regions of these promoters are very similar to those seen in *E. coli* promoters. The data presented here indicate that the *Eubacterium* system contains upstream sequence information similar to previously defined promoter structures. The 5′ terminus of the major P-27 mRNA species follows by seven bases a nucleotide sequence which shows some similarity to *E. coli* consensus promoter sequences. However, the two six-base sequences with the highest degree of similarity to consensus −35 sequences occurred at 21 and 20 bases, respectively, from the start of the only good –10 sequence. Most previously described promoters have a −35 sequence separated from the −10 sequence by 17 ± 2 bp. Expression of P-27 is regulated by bile acid, although the nature of the control is not known. In view of the observed 5′ terminus of the P-27 mRNA and sequences which appear similar to previously described promoters, this region of the P-27 gene represents an obvious region to test for the presence of regulatory elements. We recognize that the 5′ and 3′ termini observed in this work could very well represent the ends of processed RNA. Therefore, further experiments with DNA footprinting and deletion analysis approaches will be necessary to elucidate the exact location and function of this putative promoter element.

We also observed an extended dyad symmetry element beginning at 43 bp downstream from the stop codon (Fig. 4). An S1 nuclease protection experiment indicated that the in vivo synthesized mRNA exhibits a 3′ terminus in the vicinity of the element. It is known that some terminators exhibit G-C-rich hairpin loops (26). However, this particular element has a longer stem structure (23 bp) than most of these

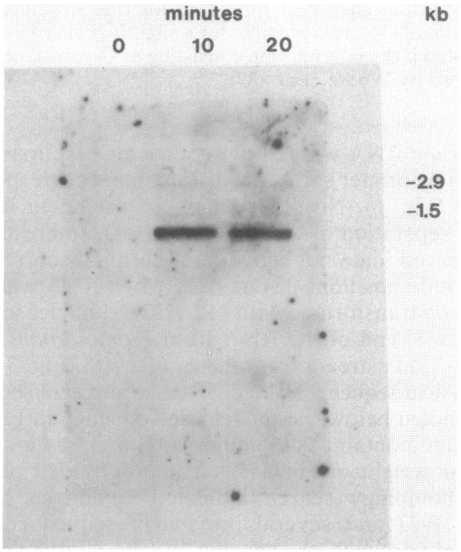

FIG. 5. Induction of the P-27 message by cholic acid. A 500-ml culture of *Eubacterium* sp. strain VPI 12708 was grown at 37°C in BHI medium to the early exponential phase of growth. A 100-ml volume was removed, and the RNA was extracted as described in Materials and Methods. Cholic acid (sodium salt) was then added to a final concentration of 0.1 mM, and incubation was continued. Further 100-ml samples were removed at 10 and 20 min, and RNA was extracted from each. The RNA samples (20 μg each) were run on a 1% agarose gel containing 2.2 M formaldehyde, 0.05 M morpholinepropanesulfonic acid (pH 7.0), and 0.001 M EDTA. The gel was run in a horizontal apparatus with 0.05 M morpholinepropanesulfonic acid (pH 7.0)–0.001 M EDTA as the electrophoresis buffer. Samples were electrophoresed at 60 V until the bromophenol blue dye marker migrated to the end of the gel. After electrophoretic transfer of the RNA to a Nylon membrane (GeneScreen; New England Nuclear), the presence of P-27-specific mRNA was detected with the 16-mer oligonucleotide probe shown in Fig.3, labeled at the 5′ end with ^{32}P (T4 kinase reaction). The positions of the 16S and 23S rRNAs are also indicated. kb, Kilobases.

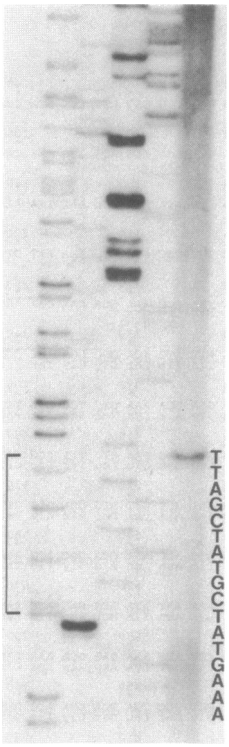

FIG. 6. Transcriptional start sites of P-27 mRNA. The 5′ terminus of P-27 mRNA was determined by the primer extension technique (see Materials and Methods). The first four lanes (A, C, G, and T, respectively) are dideoxy nucleotide sequencing reactions with double-stranded pPBH22 DNA and the synthetic primer shown in Fig. 3. Lane 5 is the RNA template reaction with the same primer. The sequence of the newly synthesized strand is shown on the right.

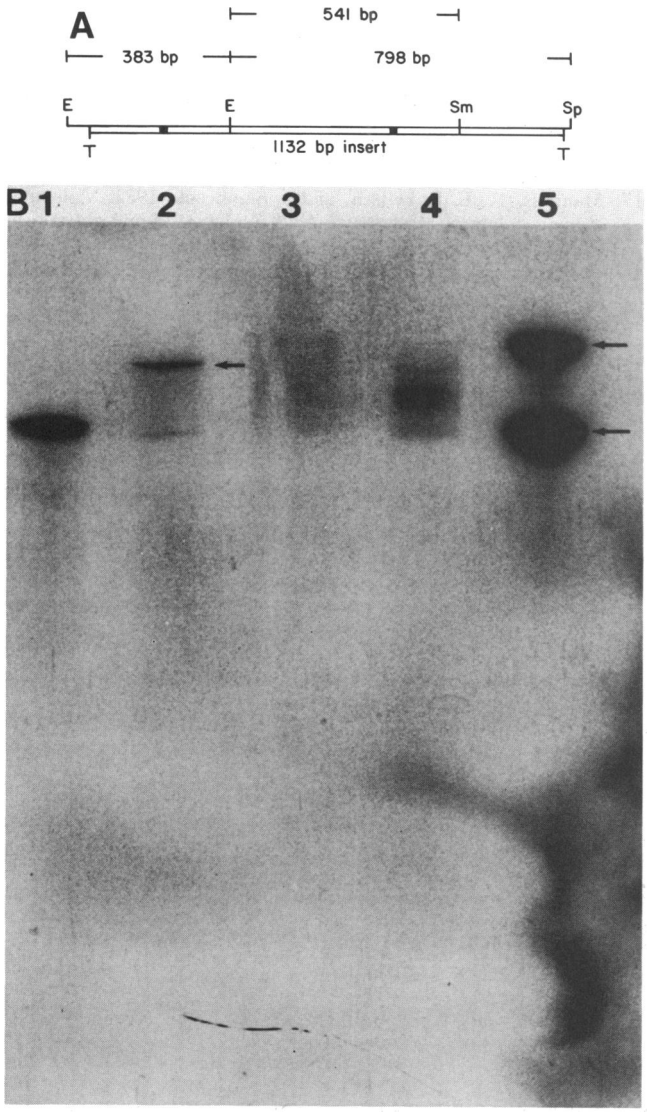

FIG. 7. S1 nuclease mapping of 3′ end of RNA. Total RNA from either cholic acid-induced or uninduced *Eubacterium* sp. strain VPI 12708 was hybridized to a 798-bp *Eco*RI (E)-*Sph*I (Sp) fragment containing 786 bp of the *Eubacterium* sequence plus 12 bp of the pUC19 multiple cloning site (A). The hybrid was digested with S1 nuclease (see Materials and Methods), and the product was run on a polyacrylamide gel containing 8 M urea. Standards were the 798-bp piece and a 383-bp *Eco*RI-*Eco*RI fragment (A). After electrophoresis, the protected species were denatured in dilute NaOH, electroblotted onto a Nylon membrane, and hybridized to 5′-end-labeled ([^{32}P]T4 kinase reaction) synthetic oligonucleotides (dark blocks on the 383- and 798-bp fragments in A). The oligonucleotides were originally synthesized for use as sequencing primers. Sm, *Sma*I; T, *Taq*I. (B) Lanes: 1, 383-bp fragment; 2, total induced RNA plus the 798-bp fragment; lane 3, total uninduced RNA plus the 798 bp fragment; 4, 798-bp fragment (no RNA); 5, 383- and 798-bp fragments (arrows). The arrow in lane 2 corresponds to approximately 600 bp. The samples in lanes 2, 3, and 4 were subjected to hybridization and S1 nuclease treatment, whereas those in lanes 1 and 5 were not treated before being loaded onto the gel.

previously described terminators and, although the region on the 3′ side of the stem is T rich, there is not an extended run of T's as is often observed in rho-independent terminators (26). It is possible that this structure functions as a regulatory element to inhibit 3′→5′ degradation of the mRNA. These types of RNA sequences have been described for certain genes in *E. coli* (22, 32), *Salmonella typhimurium* (22, 32), and *Bacillus thuringiensis* (41). A function of this type is supported by the observation of an uncharacteristically sharp hybridizing band after Northern blotting of *Eubacterium* RNA. Whether the element reported here acts as a terminator, as a stabilizing structure, or as a recognition site for some other processing system remains to be determined. In the case of the P-27 gene, no other potential terminator structures are apparent in the remaining 235 bp downstream from the stem structure, nor are there any discernible Shine-Dalgarno sequences followed by initiation codons. This indicates that the mRNA is monocistronic or, if it is polycistronic, the next coding sequence is further downstream.

The finding that regions of P-27 show high similarity to sequences believed to encompass the pyridine nucleotide-binding domain of several alcohol-polyol dehydrogenases (11) is noteworthy in light of new evidence concerning the

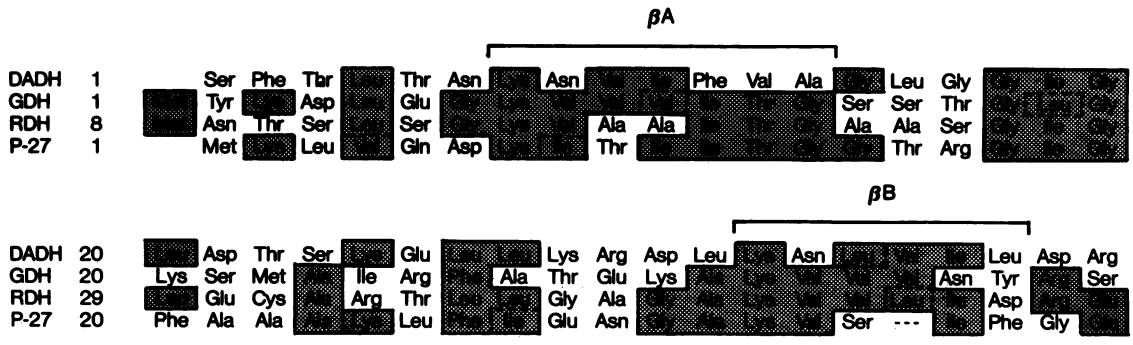

FIG. 8. Alignment of P-27 with other dehydrogenases. Abbreviations: DADH, *Drosophila* alcohol dehydrogenase (33); GDH, *Bacillus megaterium* glucose dehydrogenase (10); RDH, *Klebsiella aerogenes* ribitol dehydrogenase (18); P-27, *Eubacterium* sp. strain VPI 12708 bile acid-inducible polypeptide. The number after each protein designation is the residue number of the first amino acid shown in that line. Shaded areas indicate regions of identity (or those containing Ile-Val-Leu residues). The Leu-Phe residues below DADH position 26 are shaded, since there is an even distribution between these two possibilities.

mechanism of 7-dehydroxylation (Fig. 1). In the *Eubacterium* system, P-27 is believed to play a key role in the dehydroxylation pathway. Immunoinhibition experiments have shown that P-27 is associated with 7-dehydroxylation activity (25), although no enzymatic activity has been demonstrated in assays with purified P-27. It may be, however, that the actual substrate for P-27 is the bile acid linked to the adenosine nucleotide. The sequence homology of P-27 with pyridine nucleotide-linked dehydrogenases supports the idea that P-27 is involved in at least one of the oxidation-reduction reactions postulated for the 7-dehydroxylation pathway. Research is under way to determine the exact function of P-27 in this 7-dehydroxylation pathway.

ACKNOWLEDGMENTS

We thank M. Holmes, F. Macrina, and G. Christie for helpful discussions and criticism and M. Cousins for preparing the manuscript. The DNA and protein sequence analysis programs were used courtesy of F. Macrina.

This work was supported by Public Health Service grants CA 17747 from the National Cancer Institute and P01 DK38030 from the National Institutes of Health.

LITERATURE CITED

1. **Bagheri, S. A., M. G. Bolt, J. L. Boyers, and R. H. Palmer.** 1978. Stimulation of thymidine incorporation in mouse liver and biliary tract epithelium by lithocholate and deoxycholate. Gastroenterology **74**:188–192.
2. **Björkhem, I., and S. Skrede.** 1982. Biosynthesis of cholestanol from intestinal 7α-hydroxy-4-cholesten-3-one. J. Biol. Chem. **257**:8363–8367.
3. **Carey, M. C.** 1985. Physical-chemical properties of bile acids and their salts, p. 345–403. *In* H. Danielsson and J. Sjövall (ed.), Sterols and bile acids: new comprehensive biochemistry, vol. 12. Elsevier Science Publishing, Inc., New York.
4. **Coleman, J. P., W. B. White, B. Egestad, J. Sjövall, and P. B. Hylemon.** 1987. Biosynthesis of a novel bile acid nucleotide and mechanism of 7α-dehydroxylation by an intestinal *Eubacterium* species. J. Biol. Chem. **262**:4701–4707.
5. **Coleman, J. P., W. B. White, and P. B. Hylemon.** 1987. Molecular cloning of bile acid 7-dehydroxylase from *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **169**:1516–1521.
6. **Ferrari, A., and L. Beretta.** 1977. Activity on bile acids of a *Clostridium bifermentans* cell-free extract. FEBS Lett. **75**:163–165.
7. **Graves, M. C., and J. C. Rabinowitz.** 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene. J. Biol. Chem. **261**:11409–11415.
8. **Hayakawa, S., Y. Kanematsu, and T. Fujiwara.** 1969. Microbial degradation of bile acids. Ring A cleavage and 7α,12α-dehydroxylation of cholic acid by *Arthrobacter simplex*. Biochem. J. **115**:249–256.
9. **Hayakawa, S., and B. Samuelsson.** 1964. Transformation of cholic acid *in vitro* by *Corynebacterium simplex*. Bile acids and steroids 132. J. Biol. Chem. **239**:94–97.
10. **Jany, K.-D., W. Ulmer, M. Fröschle, and G. Pfleiderer.** 1983. Complete amino acid sequence of glucose dehydrogenase from *Bacillus megaterium*. FEBS Lett. **165**:6–10.
11. **Jörnvall, H., H. von Bahr-Lindström, K. D. Jany, W. Ulmer, and M. Fröschle.** 1984. Extended superfamily of short alcohol-polyol-sugar dehydrogenases: structural similarities between glucose and ribitol dehydrogenases. FEBS Lett. **165**:190–196.
12. **Kelsey, M. I., and R. J. Pienta.** 1979. Transformation of hamster embryo cells by cholesterol-α-epoxide and lithocholic acid. Cancer Lett. **9**:143–149.
13. **Kulkarni, M. S., P. M. Heideprien, and K. L. Yielding.** 1980. Production by lithocholic acid of DNA strand breaks in L1210 cells. Cancer Res. **40**:2666–2669.
14. **Lee, D. C., and R. G. Roeder.** 1981. Transcription of adenovirus type 2 genes in a cell-free system: apparent heterogeneity of initiation at some promoters. Mol. Cell. Biol. **1**:635–651.
15. **Liao, H. H., and J. C. Rabinowitz.** 1980. Clostridial apoferredoxin messenger ribonucleic acid assay and partial purification. Biochim. Biophys. Acta **608**:301–314.
16. **Lipsky, R. H., and P. B. Hylemon.** 1980. Characterization of a NADH:flavin oxidoreductase induced by cholic acid in a 7α-dehydroxylating intestinal *Eubacterium* species. Biochim. Biophys. Acta **612**:328–336.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Moore, C. H., S. S. Taylor, M. J. Smith, and B. S. Hartley.** 1978. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, Suppl. 3, p. 68. National Biomedical Research Foundation, Silver Spring, Md.
19. **Moran, C. P., Jr., N. Lang, S. F. J. LeGrice, G. Lee, M. Stephens, A. L. Sonenshein, J. Pero, and R. Losick.** 1982. Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. Mol. Gen. Genet. **186**:339–346.
20. **Murray, M. D.** 1986. Use of sodium trichloroacetate and mung bean nuclease to increase sensitivity and precision during transcript mapping. Anal. Biochem. **158**:165–170.
21. **Narisawa, T., N. E. Magadia, J. H. Weisburger, and E. L. Wynder.** 1974. Promoting effect of bile acids on colon carcinogenesis after intrarectal instillation of *N*-methyl-*N'*-nitrosoguanidine in rats. J. Natl. Cancer Inst. **53**:1093–1097.
22. **Newbury, S. F., N. H. Smith, E. C. Robinson, I. D. Hiles, and C. F. Higgins.** 1987. Stabilization of translationally active mRNA by prokaryotic REP sequences. Cell **48**:297–310.
23. **Owen, R. W., and R. F. Bilton.** 1983. The degradation of cholic acid by *Pseudomonas* sp. NCIB 10590 under anaerobic conditions. Biochem. J. **216**:641–654.
24. **Palmer, R. H.** 1976. Toxic effects of lithocholate on the liver and biliary tree, p. 227–240. *In* W. Taylor (ed.), The hepatobiliary system. Fundamental and pathological mechanisms. Plenum Publishing Corp., New York.
25. **Paone, D. A. M., and P. B. Hylemon.** 1984. HPLC purification and preparation of antibodies to cholic acid-inducible polypeptide from *Eubacterium* sp. VPI 12708. J. Lipid Res. **25**:1343–1349.
26. **Platt, T., and D. A. Bear.** 1983. Role of RNA polymerase, rho factor, and ribosomes in transcription termination, p. 123–161. *In* J. Beckwith, J. Davies, and J. Gallant (ed.), Gene function in procaryotes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Reddy, B. A., J. H. Weisburger, and E. L. Wynder.** 1978. Colon cancer: bile salts as tumor promoters, p. 453–464. *In* T. J. Slaga, A. Sivak, and R. K. Boutwell (ed.), Carcinogenesis, vol. 2. Mechanisms of tumor promotion and cocarcinogenesis. Raven Press, New York.
28. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
29. **Samuelsson, B.** 1960. Bile acids and steroids 96. On the mechanism of the biological formation of deoxycholic acid from cholic acid. J. Biol. Chem. **235**:361–366.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Sjövall, J.** 1960. Bile acids in man under normal and pathological conditions: bile acids and steroids 73. Clin. Chim. Acta **5**:33–41.
32. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015–1026.
33. **Thatcher, D. R.** 1980. The complete amino acid sequence of 3-alcohol dehydrogenase alleloenzymes (Adh^{N-11}, Adh^{S}, and Adh^{UF}) from the fruitfly *Drosophila melanogaster*. Biochem. J. **187**:875–883.
34. **Tinoco, I., P. N. Boner, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature

(London) New Biol. **246**:40–41.

35. **Wallace, R. B., M. J. Johnson, S. V. Suggs, K. Miyoshi, R. Bhatt, and K. Itakura.** 1981. A set of synthetic oligodeoxyribonucleotide primers for DNA sequencing in the plasmid vector pBR322. Gene **16**:21–26.
36. **White, B. A., A. F. Cacciapuoti, R. J. Fricke, T. R. Whitehead, E. H. Mosbach, and P. B. Hylemon.** 1981. Cofactor requirements for 7α-dehydroxylation of cholic and chenodeoxycholic acid in cell extracts of the intestinal anaerobic bacterium, *Eubacterium* species VPI 12708. J. Lipid Res. **22**:891–898.
37. **White, B. A., R. H. Lipsky, R. J. Fricke, and P. B. Hylemon.** 1980. Bile acid induction specificity of 7α-dehydroxylase activity in an intestinal *Eubacterium* species. Steroids **35**:103–109.
38. **White, B. A., D. A. M. Paone, A. F. Cacciapuoti, R. J. Fricke, E. H. Mosbach, and P. B. Hylemon.** 1983. Regulation of bile acid 7-dehydroxylase activity by NAD^+ and NADH in cell extracts of *Eubacterium* species VPI 12708. J. Lipid Res. **24**:20–27.
39. **Woese, C. R., R. Gutell, R. Gupta, and H. F. Noller.** 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol. Rev. **47**:621–669.
40. **Woese, C. R., E. Stackebrandt, T. J. Macke, and G. E. Fox.** 1985. A phylogenetic definition of the major eubacterial taxa. Syst. Appl. Microbiol. **6**:143–151.
41. **Wong, K. C., and S. Chang.** 1986. Identification of a positive retroregulator that stabilizes mRNAs in bacteria. Proc. Natl. Acad. Sci. USA **83**:3233–3237.