JOURNAL OF BACTERIOLOGY, Dec. 1990, p. 7011–7019
0021-9193/90/127011-09$02.00/0

Vol. 172, No. 12

Cloning and Sequencing of a Bile Acid-Inducible Operon from *Eubacterium* sp. Strain VPI 12708

DARRELL H. MALLONEE, W. BRUCE WHITE,† AND PHILLIP B. HYLEMON*

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

Received 31 May 1990/Accepted 23 September 1990

Two bile acid-inducible polypeptides from *Eubacterium* sp. strain VPI 12708 with molecular weights of 27,000 and approximately 45,000 have previously been shown to be encoded by genes residing on a 2.9-kb *Eco*RI fragment. We now report the cloning and sequencing of three additional overlapping DNA fragments upstream from this *Eco*RI fragment. Together, these four fragments contain a large segment of a bile acid-inducible operon which encodes the 27,000- and 45,000-M_r (now shown to be 47,500-M_r) polypeptides and open reading frames potentially coding for four additional polypeptides with molecular weights of 59,500, 58,000, 19,500, and 9,000 to 11,500. A bile acid-inducible polypeptide with an apparent M_r of 23,500, as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, was purified to homogeneity, and the N-terminal amino acid sequence that was obtained matched the sequence deduced from the open reading frame coding for the 19,500-M_r polypeptide. A short DNA segment containing the 3′ downstream end of the gene coding for the 47,500-M_r polypeptide was not successfully cloned but was directly sequenced from DNA fragments synthesized by polymerase chain reaction. The mRNA initiation site for the bile acid-inducible operon was shown by primer extension to be immediately upstream from the gene encoding the 58,000-M_r polypeptide. A potential promoter region upstream from the mRNA initiation site displayed significant homology with the promoter regions of previously identified bile acid-inducible genes from *Eubacterium* sp. strain VPI 12708. We hypothesize that this bile acid-inducible operon codes for most of the enzymes involved in the bile acid 7α-dehydroxylation pathway in this bacterium.

During enterohepatic circulation, bile acids can undergo numerous biotransformations carried out by anaerobic intestinal bacteria (10, 11). One of the quantitatively most important bile acid biotransformations is 7-dehydroxylation. Cholic and chenodeoxycholic acids undergoing 7-dehydroxylation yield deoxycholic and lithocholic acids, respectively. Deoxycholic acid makes up approximately 20 to 25% of the total bile acid pool in humans (19).

Eubacterium sp. strain VPI 12708 is an anaerobic intestinal bacterium that possesses a bile acid 7-dehydroxylation activity which is induced by culturing in the presence of unconjugated C-24 bile acids that possess a 7α-hydroxyl group (22). This activity is thought to proceed by a multistep pathway in which the bile acid is first linked to an adenosine nucleotide upon entering the cell. The bile acid undergoes a pair of oxidation reactions followed by loss of the 7α-hydroxyl group in a dehydration step and then undergoes a series of reduction steps, yielding the final dehydroxylated product (5).

Following cholic acid induction, the *Eubacterium* strain synthesizes at least six new polypeptides with approximate molecular weights of 77,00, 56,000 (two polypeptides), 45,000, 27,000, and 23,500, as determined by one- and two-dimensional sodium dodecyl sulfate (SDS-polyacrylamide gel electrophoresis (PAGE) (17, 21). The genes coding for two identical copies of the 27,000-M_r polypeptide (*baiA1* and *baiA3*) have been cloned and sequenced from separate chromosomal DNA fragments (4, 5, 8). The gene coding for most of the 45,000-M_r polypeptide has also been cloned on a separate 2.9-kb *Eco*RI fragment (24). The gene coding for a third copy of the 27,000-M_r polypeptide (*baiA2*) is also located on the 2.9-kb *Eco*RI fragment, immediately upstream from the 45,000-M_r polypeptide (25). The *baiA2* gene shares 81% nucleotide sequence identity with the *baiA1* and *baiA3* genes, while the polypeptide encoded by the *baiA2* gene shares 92% amino acid sequence identity with the other two 27,000-M_r polypeptides (8, 25).

Northern (RNA) blot analysis of transcripts from the *baiA1* and *baiA3* genes has suggested that they code for small (~950-bases) monocistronic mRNAs (5). However, Northern blot analysis of the transcript containing the *baiA2* gene and the gene encoding the 45,000-M_r polypeptide indicated the presence of an mRNA species greater than 6 kb in length (24). This paper reports the cloning and sequencing of chromosomal DNA fragments adjacent to the *baiA2* gene and the gene encoding the 45,000-M_r polypeptide.

MATERIALS AND METHODS

Bacterial strains and culture conditions. *Eubacterium* sp. strain VPI 12708 stock cultures were maintained in chopped meat medium as described by Holdeman and Moore (9). The *Eubacterium* strain was grown anaerobically as previously described (22) for protein and DNA isolation. Strains of *Escherichia coli* used in the various cloning experiments are identified in the appropriate sections.

Polypeptide purification. For isolation of the 23,500-M_r polypeptide, *Eubacterium* sp. strain VPI 12708 was grown in 4 liters of medium and induced by addition of sodium cholate as previously described (17, 22). The cells were harvested by

* Corresponding author.
† Present address: Anaerobe Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

centrifugation, resuspended in a minimum volume of 50 mM sodium phosphate buffer (pH 6.8) containing 100 μg of DNase I, and lysed by two passages through a French pressure cell at 12,000 lb/in^2. The cell lysate was centrifuged at 105,000 × *g* for 2 h, and the supernatant was collected and dialyzed (4°C) against 50 mM sodium phosphate buffer (pH 6.8). The sediment from a 35 to 75% ammonium sulfate precipitation of the soluble cell extract was then dialyzed and passed through a high-performance liquid chromatography (HPLC) system composed of two Altex TSK 3000 SW gel filtration columns attached in tandem and equilibrated with 50 mM sodium phosphate buffer (pH 6.8)–100 mM NaCl. Fractions displaying the 23,500-M_r polypeptide bands, as determined by SDS-PAGE, were pooled, dialyzed, and concentrated with a PF10 filter attached to a positive-pressure stirred-cell filtration devise (Amicon Corp., Danvers, Mass.). Concentrated samples were applied to an analytical Beckman DEAE 3SW HPLC column equilibrated with 20 mM sodium phosphate buffer (pH 6.0). Fractions containing the 23,500-M_r polypeptide from a 0 to 500 mM NaCl gradient were subjected to a second gel filtration purification step and finally to a second DEAE-HPLC (pH 5.0) purification step. Final fractions containing the 23,500-M_r polypeptide were extensively dialyzed against HPLC-grade water and concentrated in Amicon Centricon 10 filtration units. Protein concentrations were determined by the Bio-Rad protein assay system (Bio-Rad Laboratories, Richmond, Calif.).

Polypeptide sequencing. A 25-μg aliquot from a final pool of 120 μg of the purified 23,500-M_r polypeptide was used for N-terminal amino acid sequence determination on an Applied Biosystems (Foster City, Calif.) 477A protein sequencer at the University of Illinois Biotechnology Center.

Western immunoblot procedures. For Western blot analysis, crude cell extracts and fractions collected from HPLC fractionations were separated by SDS-PAGE and electroblotted to nitrocellulose filters. The filters were then exposed to purified immunoglobulin G prepared against HPLC fractions containing 7-dehydroxylation activity from *Eubacterium* sp. strain VPI 12708 (17). Reaction to antibody was detected by use of a goat anti-rabbit secondary antibody with a colorimetric horseradish peroxidase detection system as instructed by the manufacturer (Bio-Rad).

Recombinant DNA methods. *Eubacterium* sp. strain VPI 12708 chromosomal DNA was isolated by the method of Marmur (15). For preparation of the λgt11 libraries, *Eubacterium* chromosomal DNA was digested to completion with *Eco*RI and size fractionated on 0.7% agarose gels. Fragments of the appropriate size were cut out of the gel and purified with GeneClean (Bio 101 Inc., La Jolla, Calif.) or by electroelution in dialysis tubing, followed by phenol-chloroform extraction and ethanol precipitation. Purified fragments were ligated to *Eco*RI-digested λgt11 arms (Bethesda Research Laboratories, Gaithersburg, Md. [BRL]), packaged, and used to infect *E. coli* Y1090 (27). λgt11 DNA was isolated as described by Davis et al. (6).

For construction of the pUC19 libraries, *Eubacterium* DNA was digested to completion by the appropriate restriction endonuclease and size fractionated on appropriate agarose gels. Fragments of the appropriate size were cut out of the gels and purified with GeneClean or by electroelution, followed by phenol-chloroform extraction and ethanol precipitation. The fragments were ligated to appropriately digested pUC19 DNA that had been treated with bacterial alkaline phosphatase (BRL). The ligated mixture was then used to transform HB101, DH5α, or DH5αMCR competent cells as described by the supplier of the cells (BRL). Plasmid DNA was isolated by the method of Birnboim and Doly (2).

Fragments used for subcloning into M13mp18 and M13mp19 vectors were obtained by digesting DNA from pUC19 or λgt11 clones with appropriate restriction endonucleases and separating the fragments on agarose or polyacrylamide gels. Fragments were cut out of the gels and purified as described above. Purified fragments were ligated to M13 DNA that had been digested with the appropriate restriction endonucleases and treated with bacterial alkaline phosphatase. Ligated mixtures were used to transform *E. coli* DH5αF′ or DH5αF′IQ competent cells (BRL). For sequencing, M13 clones were grown in *E. coli* JM101 (26) and single-stranded DNA was prepared according to the procedure described by Davis et al. (6).

DNA hybridization procedures. Oligonucleotides were synthesized on either a Cyclone DNA synthesizer (Millipore Corp., Burlington, Mass.) or an Applied Biosystems model 380A DNA synthesizer and purified as previously described (4). Several of the oligonucleotides used were synthesized at the Medical College of Virginia/Virginia Commonwealth University Nucleic Acids Core Facility. Purified oligonucleotides were endlabeled for Southern blots and screening of clone banks with T4 DNA kinase and [γ-^{32}P]ATP (3,000 Ci/mmol). Unincorporated label was removed with Nensorb 20 cartridges (Dupont, NEN Research Products, Boston, Mass.).

For Southern blot analyses, DNA restriction fragments were transferred from agarose gels to Dupont GeneScreen membranes as described in the package insert. Prehybridization, hybridization to end-labeled DNA probes, and washing of the membranes were performed as described by Davis et al. (6).

The procedure of Davis et al. (6) was used for λgt11 plaque lifts to nitrocellulose filters and for hybridization of the filters with end-labeled synthetic oligonucleotide probes. Colonies of *E. coli* transformed with plasmid DNA were lifted from agar plates to nitrocellulose filters and placed sequentially in puddles on plastic wrap containing 0.5 M NaOH for 5 min, 1.5 M NaCl and 0.5 M Tris (pH 7.4) for 5 min, and 1.5 M NaCl and 2× SSC (0.3 M NaCl, 0.03 M sodium citrate) for 5 min. The filters were air dried for 1 h at room temperature and then baked for 2 h in an 80°C vacuum oven. Baked filters were prewashed for 1 to 2 h at 37°C in a solution of 1 M NaCl–1 mM EDTA–0.1% SDS–50 mM Tris hydrochloride, pH 8. Prehybridization, hybridization, and washing of the filters were performed as described by Davis et al. (6). Filters from the Southern blots and from the plaque and colony lifts were placed in cassettes with X-Omat RP or X-Omat AR film (Eastman Kodak Co., Rochester, N.Y.) and a DuPont Cronex Lightning-Plus intensifying screen for 4 to 48 h at −70°C before the film was developed.

DNA sequencing. DNA sequences were obtained by the dideoxy-chain termination method (18), using the Sequenase procedure of United States Biochemical Corp. (Cleveland, Ohio) with either double-stranded plasmid DNA or single-stranded M13 DNA as the template. Commercially available universal sequencing primers (17-mers) and other synthetic oligonucleotides were used in the sequencing procedures. Regions of ambiguous DNA sequence were further analyzed by use of dITP in the sequencing reactions. The DNA was labeled with [α-^{35}S]dATP. All sequence information was obtained from both strands of overlapping fragments.

Primer extension. The primer extension procedure was performed as described by Ausubel et al. (1).

PCR procedures. Inverse polymerase chain reaction (PCR)

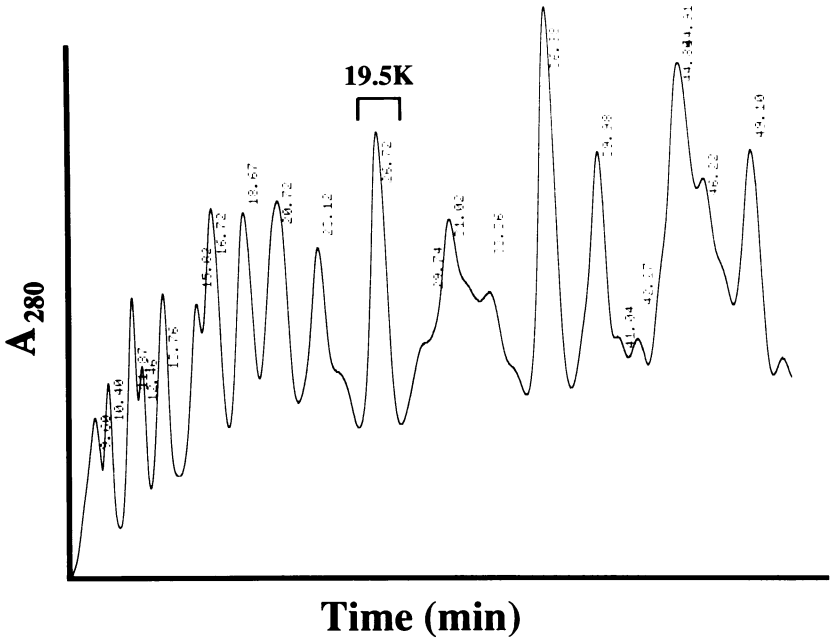


FIG. 1. Fractionation of the 23,500-M_r (19.5K) polypeptide on a DEAE-HPLC column (pH 6.0). The column was injected with approximately 25 mg of protein from fractions containing the 23,500-M_r polypeptide from an initial gel filtration HPLC fractionation step.

was used to amplify downstream (3′) segments of the operon. *Eubacterium* sp. strain VPI 12708 chromosomal DNA was first digested with restriction endonuclease *Hae*III and size fractionated on a 0.7% agarose gel. Gel slices containing the appropriate-size fragments, as determined by Southern blot analysis, were cut out of the gels and purified by GeneClean. Purified fragments were ligated with T4 DNA ligase. Approximately 10 to 20 ng of the ligated mixture was subjected to PCR amplification, using a pair of 23-mer oligonucleotide primers located within the sequenced region of the fragment and hybridizing to opposite strands. The Perkin Elmer (Norwalk, Conn.) PCR kit was used according to insert instructions, with cycling temperatures of 94°C (1 min), 55°C (2 min), and 72°C (1 min) on a Perkin Elmer thermal cycler for 30 cycles.

Asymmetric PCR was used to obtain DNA for sequencing and was performed both on the starting ligated mixtures and on fragments obtained from a first round of PCR. A 50:1 ratio of primers was used with the ligated mixtures, and a single primer was used with the PCR material. The asymmetric PCRs were run for 25 to 30 cycles as described above.

Analysis of sequence data. Analysis of nucleic acid and protein sequence data was performed with the IBI/Pustell DNA analysis program (International Biotechnologies, Inc., New Haven Conn.) and the Genetics Computer Group program (University of Wisconsin Biotechnology Center, Madison, Wis.).

Nucleotide sequence accession number. The GenBank accession number for the sequences reported in this article is M36292.

RESULTS

Sequencing of the 2.9-kb *Eco*RI fragment. Sequencing was completed on a 2.9-kb *Eco*RI fragment that had been previously cloned in λgt11 and shown to contain the gene coding for a 27,000-M_r bile acid-inducible polypeptide (*baiA2*) and the major portion of a gene coding for a bile acid-inducible polypeptide with an apparent M_r of 45,000 (as determined by SDS-PAGE of cholic acid-induced cell extracts; 24, 25). An open reading frame potentially coding for a polypeptide containing 166 amino acids and having a calculated M_r of 19,514 was found immediately upstream (5′) from the *baiA2* gene. Immediately upstream from the open reading frame coding for the hypothetical 19,500-M_r polypeptide was what appeared to be the 3′ end of a fourth open reading frame.

Isolation of the 19,500-M_r polypeptide. To determine whether the hypothetical 19,500-M_r polypeptide was identical to the 23,500-M_r cholic acid-induced polypeptide observed on SDS-PAGE, the 23,500-M_r polypeptide was purified to homogeneity as described in Materials and Methods. Separation of this polypeptide on a DEAE-HPLC column (pH 6) is shown in Fig. 1, and analysis of the purification steps by SDS-PAGE is shown in Fig. 2. The purified polypeptide was shown by Western blot analysis to react with antibodies prepared to HPLC fractions containing 7-dehydroxylation activity from extracts of cholic acid-induced *Eubacterium* sp. strain VPI 12708 (Fig. 3). No detectable 23,500-M_r polypeptide was observed in cell extracts prepared from uninduced cultures. The purified polypeptide was subjected to N-terminal amino acid sequence analysis, and 31 residues of amino acid sequence were obtained. Comparison of this sequence with the deduced amino acid sequence of the proposed 19,500-M_r polypeptide from the open reading frame on the 2.9-kb *Eco*RI fragment revealed a 100% amino acid sequence identity with the exception of the second amino acid, where a questionable residue was suggested to be a phenylalanine instead of a threonine (Fig. 4). Southern blots using an oligonucleotide probe made from this N-terminal sequence showed a single hybridizing band corresponding to a 2.9-kb *Eco*RI fragment. The purified polypeptide and the hypothetical polypeptide

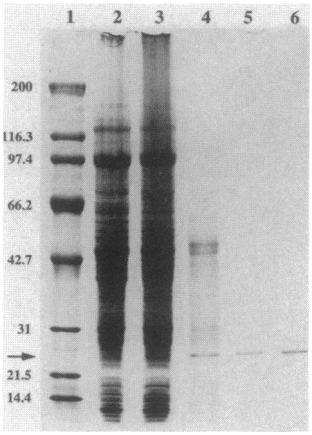


FIG. 2. Purification steps for the 23,500 (19,500)-M_r polypeptide (arrow) demonstrated by SDS-PAGE with a 7 to 20% polyacrylamide gradient. Lanes: 1, molecular weight markers; 2, uninduced cell extract from *Eubacterium* sp. strain VPI 12708; 3, cholic acid-induced cell extract; 4, pooled fractions from the initial gel-filtration HPLC step; 5, pooled fractions from the first DEAE-HPLC step; 6, pooled fractions from the second gel filtration HPLC step.

coded for by the open reading frame were thus judged to be identical, migrating as a 23,500-M_r polypeptide on SDS-PAGE, and will be referred to hereafter as the 19,500-M_r polypeptide.

Cloning of a 300-bp *Nrul* fragment. Since it was apparent that the entire bile acid-induced operon was not contained on the 2.9-kb *Eco*RI fragment, attempts were made to clone larger surrounding DNA fragments. Attempts were made to clone a 6-kb *Bam*HI fragment that contained the entire

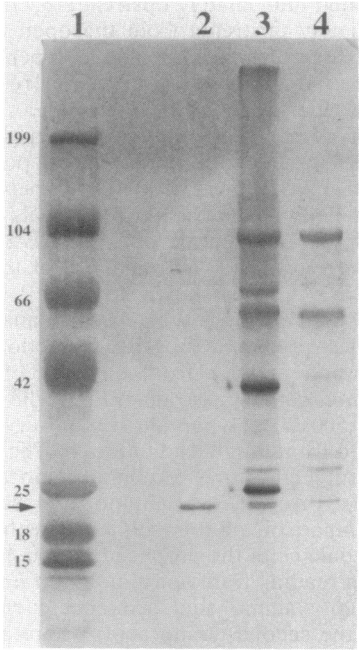


FIG. 3. Western immunoblot using antibodies made to HPLC fractions from cholic acid-induced extracts of *Eubacterium* sp. strain VPI 12708. Lanes: 1, prestained molecular weight markers; 2, purified 23,500 (19,500)-M_r polypeptide (arrow); 3, cholic acid-induced *Eubacterium* cell extract; 4, uninduced cell extract.

	1	2	3	4	5	6	7	8	9	10	11
N-term:	Met	***Phe***	Leu	Glu	Glu	Arg	Val	Glu	Ala	Leu	Glu
Operon:	Met	Thr	Leu	Glu	Glu	Arg	Val	Glu	Ala	Leu	Glu

	12	13	14	15	16	17	18	19	20	21	22
N-term:	Lys	Glu	Leu	Gln	Glu	Met	Lys	Asp	Ile	Glu	Ala
Operon:	Lys	Glu	Leu	Gln	Glu	Met	Lys	Asp	Ile	Glu	Ala

	23	24	25	26	27	28	29	30	31
N-term:	***Ile***	Lys	***Glu***	Leu	Lys	Gly	Lys	Tyr	Phe
Operon:	Ile	Lys	Glu	Leu	Lys	Gly	Lys	Tyr	Phe

FIG. 4. Comparison of amino acid sequences obtained from N-terminal amino acid sequencing of purified 23,500 (19,500)-M_r polypeptide (N-term) and the deduced amino acid sequence from an open reading frame on the operon (Operon). Highlighted amino acids at positions 2, 23, and 25 represent residues that could not be accurately determined by N-terminal sequence analysis. Unmatched amino acids at position 2 are boxed.

2.9-kb *Eco*RI fragment and to clone a 4-kb *Bgl*II fragment that overlapped the *Eco*RI fragment and contained additional 5′ sequence. Attempts were also made to clone a large (15- to 20-kb) fragment from a partial *Sau*3A digest cloned into λDASH (Stratagene, La Jolla, Calif.; 8). None of these clones were isolated. In fact, of 23 λDASH clones obtained that hybridized to a probe recognizing all three copies of the *baiA* gene, none contained the *baiA2* gene residing on the 2.9-kb *Eco*RI fragment.

We therefore decided to try to clone a smaller fragment overlapping the upstream (5′) *Eco*RI site in order obtain a DNA probe for possible cloning of another *Eco*RI fragment in λgt11. For this purpose, an attempt was made to obtain a 300-bp *Nru*I fragment overlapping the upstream *Eco*RI site. This clone was obtained by inserting the *Nru*I fragment into the *Sma*I site of a pUC19 vector and using *E. coli* DH5α as the host strain. This fragment was sequenced and shown to contain the N-terminal coding region for a possible polypeptide containing from 87 to 105 amino acids, depending on the correct initiation codon, and having a calculated M_r of 9,099 to 11,447. The fragment also contained what appeared to be the 3′ end of a fifth open reading frame.

Cloning of a 3-kb *Kpn*I fragment. The *Eco*RI fragment overlapping the *Nru*I fragment on the upstream side was only 270 bp in length. Although this fragment was obtained in a λgt11 clone, it was not used for sequencing. Instead, a 3-kb *Kpn*I fragment that had its 3′ restriction site within the *Nru*I fragment was obtained (Fig. 5). This *Kpn*I fragment was cloned into the *Kpn*I site of pUC19, using *E. coli* DH5α as the host strain. Sequencing of this fragment revealed the presence of an open reading frame potentially coding for a fifth polypeptide containing 540 amino acids and having a calculated M_r of 59,513. The 3′ end of this open reading frame is contained in the *Nru*I fragment. Once again there also appeared to be the 3′ end of an additional open reading frame within the *Kpn*I fragment.

Cloning of a 2.4-kb *Eco*RI fragment. A fourth overlapping DNA segment was obtained by cloning a 2.4-kb *Eco*RI fragment into λgt11. This fragment contained an approximately 600-bp segment of DNA upstream from the 5′ *Kpn*I site (Fig. 5). Sequencing of this fragment revealed the 5′ end of an open reading frame extending into the *Kpn*I fragment and potentially coding for a sixth polypeptide containing 521 amino acids and having a calculated M_r of 58,272. Upstream from the sixth open reading frame was a possible promoter structure.

Sequencing of downstream PCR fragments. To determine the sequence of the 3′ end of the gene coding for the 45,000-M_r polypeptide, attempts were made to clone a

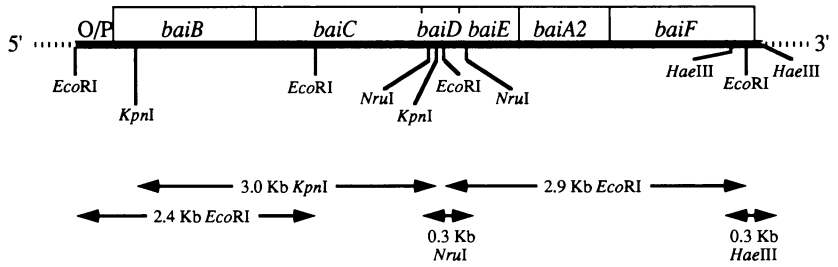


FIG. 5. Partial restriction map and locations of open reading frames for the proposed bile acid-inducible operon from *Eubacterium* sp. strain VPI 12708. O/P, Operator-promoter region.

300-bp overlapping *Hae*III fragment and a 600-bp overlapping *Bcl*I fragment. Both of these fragments overlapped the 2.9-kb *Eco*RI fragment by about 150 bases. When these cloning attempts were unsuccessful, amplification of the 300-bp *Hae*III fragment by inverse PCR was attempted. Amplified fragments were then subjected to a second round of asymmetric PCR amplification of each strand before sequencing of the fragments was attempted. Sequencing results revealed the 3′ end of the gene coding for the 45,000-M_r polypeptide and resulted in an open reading frame coding for 426 amino acids and a polypeptide with a calculated M_r of 47,448. Insufficient sequence information was obtained to determine whether an additional open reading frame existed in the 3′ sequence.

We propose to label the open reading frames coding for the individual polypeptides as bile acid-inducible (*bai*) genes. The genes coding for the hypothetical 58,000-, 59,500-, and 9,000-M_r polypeptides will be called *baiB*, *baiC*, and *baiD*, respectively. The genes coding for the 19,500- and 47,500-M_r polypeptides will be called *baiE* and *baiF*, respectively. The gene within the operon coding for the 27,000-M_r polypeptide has previously been called the *baiA2* gene (8). Figure 5 shows the locations of open reading frames on this operon and the locations of the overlapping clones. Figure 6 shows the entire nucleotide and amino acid sequences for the six open reading frames.

Primer extension. The 5′ end of the mRNA species containing the bile acid-inducible operon was determined by primer extension to be 68 bases upstream from the proposed initiation codon for the 58,000-M_r polypeptide (Fig. 7). The DNA sequence containing the first 2 bases of the proposed mRNA species and extending 127 additional bases upstream from this point displayed 69.8% sequence identity with a stretch of DNA containing the proposed promoter regions for the *baiA1* and *baiA3* bile acid-inducible genes from *Eubacterium* sp. strain VPI 12708 (5, 8; Fig. 8).

Data bank searches for sequence similarities. Data bank searches revealed no amino acid sequences significantly similar to the 59,500-, 47,500-, 19,500-, or 9,000- to 11,500-M_r polypeptides. As previously reported, the amino acid sequence for the 27,000-M_r polypeptide showed significant homology to sequences for several alcohol/polyol dehydrogenases (5, 25). Significant homology was also obtained between the amino acid sequences for the 58,000-M_r polypeptide and for several polypeptides involved in reacting cyclic carboxylated compounds with ATP to form active adenylated compounds. These homologous polypeptides include a 4-coumarate:coenzyme A ligase from *Petroselinum crispum* (parsley; 14), tyrocidine synthetase 1 (*tycA* gene product) from *Bacillus brevis* (20), luciferase from *Photinus pyralis* (North American firefly; 7), and a polypeptide (*entE* gene product) involved in the activation of 2,3-dihydroxybenzoate during enterobactin (enterochelin) synthesis in *E. coli* (13). The sequence homologies for these polypeptides with the 58,000-M_r polypeptide ranged from 19 to 31% amino acid sequence identity over a span of 118 to 521 amino acids with optimum alignment, including gaps.

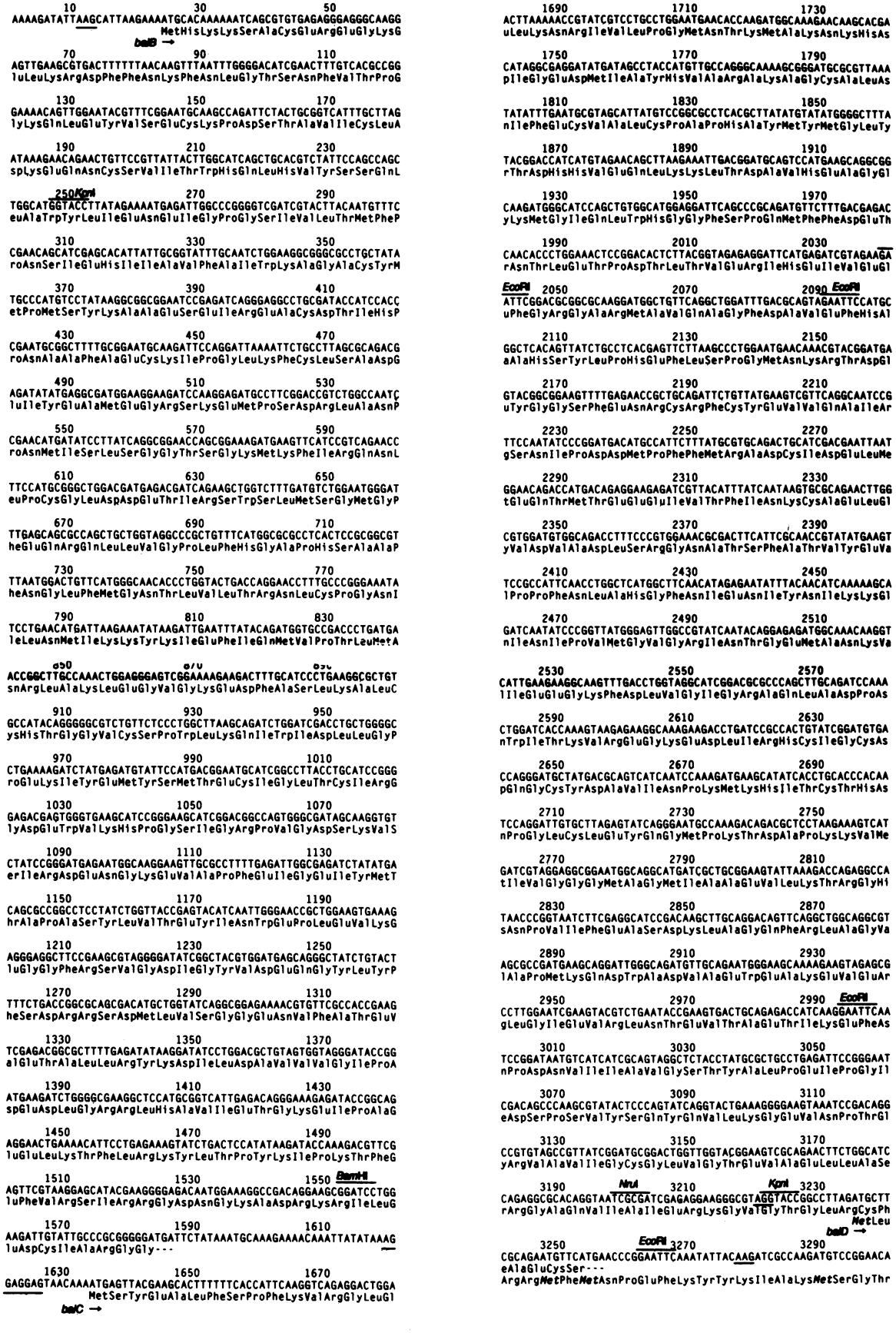
DISCUSSION

We report here the cloning and sequencing of the major portion of a large bile acid-inducible operon from *Eubacterium* sp. strain VPI 12708. At least six open reading frames are included in this operon, potentially coding for polypeptides with molecular weights of (5′ to 3′) of 58,000, 59,500, 9,000 to 11,500, 19,500, 27,000, and 47,500.

Three of these polypeptides have been conclusively shown to be bile acid induced. The 47,500-M_r polypeptide has been previously isolated from cholic acid-induced extracts of the *Eubacterium* strain, and the N-terminal amino acid sequence obtained matches the sequence deduced from the cloned gene (24). It has also been previously shown that there are three copies of the *baiA* gene, which codes for the bile acid-inducible 27,000-M_r polypeptides. One of the copies (*baiA2*) resides on the large operon, while the other two (*baiA1* and *baiA3*) reside on separate chromosomal fragments producing monocistronic messages (5, 8). The 19,500-M_r polypeptide was shown in this report to be the same polypeptide that had been previously described as a 23,500-M_r bile acid-induced polypeptide (21). The reason for the discrepancy between the apparent M_r and the calculated M_r is unclear.

Characterizing the other three hypothetical polypeptides (58,000, 59,500, and 9,000 M_r) as bile acid-induced polypeptides is based on several lines of reasoning. First, it was previously reported that there appeared to be two bile acid-induced polypeptides of about 56,000 M_r, as determined by two-dimensional SDS-PAGE of induced versus uninduced cell extracts (21). It is possible that those two polypeptides are the 58,000- and 59,500-M_r polypeptides reported in this study. In regard to the hypothetical 9,000-M_r polypeptide, it was reported (23) that a substantial portion of 7-dehydroxylation activity lost in HPLC-purified protein fractions could be recovered by adding material from low-molecular-weight (8,000 to 14,000) eluting fractions.

The open reading frame coding for the proposed 9,000- to 11,500-M_r polypeptide (*baiD*) appears to be the only open reading frame on the operon with a questionable initiation codon. There are 4 methionine residues in the first 19 residues of this open reading frame (Fig. 6). However, the last of these methionines appears to be the correct initiation codon when one looks for possible ribosome-binding sites


10 30 50
AAAAGATATTAAGCATTAAGAAAATGCACAAAAAATCAGCGTGTGAGAGGGAGGGCAAGG
MetHisLysLysSerAlaCysGluArgGluGlyLysG
baiB →
70 90 110
AGTTGAAGCGTGACTTTTTTAACAAGTTTAATTTGGGGACATCGAACTTTGTCACGCCGG
luLeuLysArgAspPheAsnLysPheAsnLeuGlyThrSerAsnPheValThrProG
130 150 170
GAAAACAGTTGGAATACGTTTCGGAATGCAAGCCAGATTCTACTGCGGTCATTTGCTTAG
lyLysGlnLeuGluTyrValSerGluCysLysProAspSerThrAlaValIleCysLeuA
190 210 230
ATAAAGAACAGAACTGTTCCGTTATTACTTGGCATCAGCTGCACGTCTATTCCAGCCAGC
spLysGluGlnAsnCysSerValIleThrTrpHisGlnLeuHisValTyrSerSerGlnL
250KpnI 270 290
TGGCATGGTACCTTATAGAAAATGAGATTGGCCCGGGGTCGATCGTACTTACAATGTTTC
euAlaTrpTyrLeuIleGluAsnGluIleGlyProGlySerIleValLeuThrMetPheP
310 330 350
CGAACAGCATCGAGCACATTATTGCGGTATTTGCAATCTGGAAGGCGGGCGCCTGCTATA
roAsnSerIleGluHisIleIleAlaValPheAlaIleTrpLysAlaGlyAlaCysTyrM
370 390 410
TGCCCATGTCCTATAAGGCGGCGGAATCCGAGATCAGGGAGGCCTGCGATACCATCCACC
etProMetSerTyrLysAlaAlaGluSerGluIleArgGluAlaCysAspThrIleHisP
430 450 470
CGAATGCGGCTTTTGCGGAATGCAAGATTCCAGGATTAAAATTCTGCCTTAGCGCAGACG
roAsnAlaAlaPheAlaGluCysLysIleProGlyLeuLysPheCysLeuSerAlaAspG
490 510 530
AGATATATGAGGCGATGGAAGGAAGATCCAAGGAGATGCCTTCGGACCGTCTGGCCAATC
luIleTyrGluAlaMetGluGlyArgSerLysGluMetProSerAspArgLeuAlaAsnP
550 570 590
CGAACATGATATCCTTATCAGGCGGAACCAGCGGAAAGATGAAGTTCATCCGTCAGAACC
roAsnMetIleSerLeuSerGlyGlyThrSerGlyLysMetLysPheIleArgGlnAsnL
610 630 650
TTCCATGCGGGCTGGACGATGAGACGATCAGAAGCTGGTCTTTGATGTCTGGAATGGGAT
euProCysGlyLeuAspAspGluThrIleArgSerTrpSerLeuMetSerGlyMetGlyP
670 690 710
TTGAGCAGCGCCAGCTGCTGGTAGGCCCGCTGTTTCATGGCGCGCCTCACTCCGCGGCGT
heGluGlnArgGlnLeuLeuValGlyProLeuPheHisGlyAlaProHisSerAlaAlaP
730 750 770
TTAATGGACTGTTCATGGGCAACACCCTGGTACTGACCAGGAACCTTTGCCCGGGAAATA
heAsnGlyLeuPheMetGlyAsnThrLeuValLeuThrArgAsnLeuCysProGlyAsnI
790 810 830
TCCTGAACATGATTAAGAAATATAAGATTGAATTTATACAGATGGTGCCGACCCTGATGA
leLeuAsnMetIleLysLysTyrLysIleGluPheIleGlnMetValProThrLeuMetA
ACCGGCTTGCCAAACTGGAGGGAGTCGGAAAAGAAGACTTTGCATCCCTGAAGGCGCTGT
snArgLeuAlaLysLeuGluGlyValGlyLysGluAspPheAlaSerLeuLysAlaLeuC
910 930 950
GCCATACAGGGGGCGTCTGTTCTCCCTGGCTTAAGCAGATCTGGATCGACCTGCTGGGGC
ysHisThrGlyGlyValCysSerProTrpLeuLysGlnIleTrpIleAspLeuLeuGlyP
970 990 1010
CTGAAAAGATCTATGAGATGTATTCCATGACGGAATGCATCGGCCTTACCTGCATCCGGG
roGluLysIleTyrGluMetTyrSerMetThrGluCysIleGlyLeuThrCysIleArgG
1030 1050 1070
GAGACGAGTGGGTGAAGCATCCGGGAAGCATCGGACGGCCAGTGGGCGATAGCAAGGTGT
lyAspGluTrpValLysHisProGlySerIleGlyArgProValGlyAspSerLysValS
1090 1110 1130
CTATCCGGGATGAGAATGGCAAGGAAGTTGCGCCTTTTGAGATTGGCGAGATCTATATGA
erIleArgAspGluAsnGlyLysGluValAlaProPheGluIleGlyGluIleTyrMetT
1150 1170 1190
CAGCGCCGGCCTCCTATCTGGTTACCGAGTACATCAATTGGGAACCGCTGGAAGTGAAAG
hrAlaProAlaSerTyrLeuValThrGluTyrIleAsnTrpGluProLeuGluValLysG
1210 1230 1250
AGGGAGGCTTCCGAAGCGTAGGGGATATCGGCTACGTGGATGAGCAGGGCTATCTGTACT
luGlyGlyPheArgSerValGlyAspIleGlyTyrValAspGluGlnGlyTyrLeuTyrP
1270 1290 1310
TTTCTGACCGGCGCAGCGACATGCTGGTATCAGGCGGAGAAAACGTGTTCGCCACCGAAG
heSerAspArgArgSerAspMetLeuValSerGlyGlyGluAsnValPheAlaThrGluV
1330 1350 1370
TCGAGACGGCGCTTTTGAGATATAAGGATATCCTGGACGCTGTAGTGGTAGGGATACCGG
alGluThrAlaLeuLeuArgTyrLysAspIleLeuAspAlaValValValGlyIleProA
1390 1410 1430
ATGAAGATCTGGGGCGAAGGCTCCATGCGGTCATTGAGACAGGGAAAGAGATACCGGCAG
spGluAspLeuGlyArgArgLeuHisAlaValIleGluThrGlyLysGluIleProAlaG
1450 1470 1490
AGGAACTGAAAACATTCCTGAGAAAGTATCTGACTCCATATAAGATACCAAAGACGTTCG
luGluLeuLysThrPheLeuArgLysTyrLeuThrProTyrLysIleProLysThrPheG
1510 1530 1550 BamHI
AGTTCGTAAGGAGCATACGAAGGGGAGACAATGGAAAGGCCGACAGGAAGCGGATCCTGG
luPheValArgSerIleArgArgGlyAspAsnGlyLysAlaAspArgLysArgIleLeuG
1570 1590 1610
AAGATTGTATTGCCCGCGGGGGATGATTCTATAAATGCAAAGAAAACAAATTATATAAAG
luAspCysIleAlaArgGlyGly---
1630 1650 1670
GAGGAGTAACAAAATGAGTTACGAAGCACTTTTTTCACCATTCAAGGTCAGAGGACTGGA
MetSerTyrGluAlaLeuPheSerProPheLysValArgGlyLeuGl
baiC →
1690 1710 1730
ACTTAAAAACCGTATCGTCCTGCCTGGAATGAACACCAAGATGGCAAAGAACAAGCACGA
uLeuLysAsnArgIleValLeuProGlyMetAsnThrLysMetAlaLysAsnLysHisAs
1750 1770 1790
CATAGGCGAGGATATGATAGCCTACCATGTTGCCAGGGCAAAAGCGGGATGCGCGTTAAA
pIleGlyGluAspMetIleAlaTyrHisValAlaArgAlaLysAlaGlyCysAlaLeuAs
1810 1830 1850
TATATTTGAATGCGTAGCATTATGTCCGGCGCCTCACGCTTATATGTATATGGGGCTTTA
nIlePheGluCysValAlaLeuCysProAlaProHisAlaTyrMetTyrMetGlyLeuTy
1870 1890 1910
TACGGACCATCATGTAGAACAGCTTAAGAAATTGACGGATGCAGTCCATGAAGCAGGCGG
rThrAspHisHisValGluGlnLeuLysLysLeuThrAspAlaValHisGluAlaGlyGl
1930 1950 1970
CAAGATGGGCATCCAGCTGTGGCATGGAGGATTCAGCCCGCAGATGTTCTTTGACGAGAC
yLysMetGlyIleGlnLeuTrpHisGlyGlyPheSerProGlnMetPhePheAspGluTh
1990 2010 2030
CAACACCCTGGAAACTCCGGACACTCTTACGGTAGAGAGGATTCATGAGATCGTAGAAGA
rAsnThrLeuGluThrProAspThrLeuThrValGluArgIleHisGluIleValGluGl
EcoRI 2050 2070 2090 EcoRI
ATTCGGACGCGGCGCAAGGATGGCTGTTCAGGCTGGATTTGACGCAGTAGAATTCCATGC
uPheGlyArgGlyAlaArgMetAlaValGlnAlaGlyPheAspAlaValGluPheHisAl
2110 2130 2150
GGCTCACAGTTATCTGCCTCACGAGTTCTTAAGCCCTGGAATGAACAAACGTACGGATGA
aAlaHisSerTyrLeuProHisGluPheLeuSerProGlyMetAsnLysArgThrAspGl
2170 2190 2210
GTACGGCGGAAGTTTTGAGAACCGCTGCAGATTCTGTTATGAAGTCGTTCAGGCAATCCG
uTyrGlyGlySerPheGluAsnArgCysArgPheCysTyrGluValValGlnAlaIleAr
2230 2250 2270
TTCCAATATCCCGGATGACATGCCATTCTTTATGCGTGCAGACTGCATCGACGAATTAAT
gSerAsnIleProAspAspMetProPhePheMetArgAlaAspCysIleAspGluLeuMe
2290 2310 2330
GGAACAGACCATGACAGAGGAAGAGATCGTTACATTTATCAATAAGTGCGCAGAACTTGG
tGluGlnThrMetThrGluGluGluIleValThrPheIleAsnLysCysAlaGluLeuGl
2350 2370 2390
CGTGGATGTGGCAGACCTTTCCCGTGGAAACGCGACTTCATTCGCAACCGTATATGAAGT
yValAspValAlaAspLeuSerArgGlyAsnAlaThrSerPheAlaThrValTyrGluVa
2410 2430 2450
TCCGCCATTCAACCTGGCTCATGGCTTCAACATAGAGAATATTTACAACATCAAAAAGCA
lProProPheAsnLeuAlaHisGlyPheAsnIleGluAsnIleTyrAsnIleLysLysGl
2470 2490 2510
GATCAATATCCCGGTTATGGGAGTTGGCCGTATCAATACAGGAGAGATGGCAAACAAGGT
nIleAsnIleProValMetGlyValGlyArgIleAsnThrGlyGluMetAlaAsnLysVa
2530 2550 2570
CATTGAAGAAGGCAAGTTTGACCTGGTAGGCATCGGACGCGCCCAGCTTGCAGATCCAAA
lIleGluGluGlyLysPheAspLeuValGlyIleGlyArgAlaGlnLeuAlaAspProAs
2590 2610 2630
CTGGATCACCAAAGTAAGAGAAGGCAAAGAAGACCTGATCCGCCACTGTATCGGATGTGA
nTrpIleThrLysValArgGluGlyLysGluAspLeuIleArgHisCysIleGlyCysAs
2650 2670 2690
CCAGGGATGCTATGACGCAGTCATCAATCCAAAGATGAAGCATATCACCTGCACCCACAA
pGlnGlyCysTyrAspAlaValIleAsnProLysMetLysHisIleThrCysThrHisAs
2710 2730 2750
TCCAGGATTGTGCTTAGAGTATCAGGGAATGCCAAAGACAGACGCTCCTAAGAAAGTCAT
nProGlyLeuCysLeuGluTyrGlnGlyMetProLysThrAspAlaProLysLysValMe
2770 2790 2810
GATCGTAGGAGGCGGAATGGCAGGCATGATCGCTGCGGAAGTATTAAAGACCAGAGGCCA
tIleValGlyGlyGlyMetAlaGlyMetIleAlaAlaGluValLeuLysThrArgGlyHi
2830 2850 2870
TAACCCGGTAATCTTCGAGGCATCCGACAAGCTTGCAGGACAGTTCAGGCTGGCAGGCGT
sAsnProValIlePheGluAlaSerAspLysLeuAlaGlyGlnPheArgLeuAlaGlyVa
2890 2910 2930
AGCGCCGATGAAGCAGGATTGGGCAGATGTTGCAGAATGGGAAGCAAAAGAAGTAGAGCG
lAlaProMetLysGlnAspTrpAlaAspValAlaGluTrpGluAlaLysGluValGluAr
2950 2970 2990 EcoRI
CCTTGGAATCGAAGTACGTCTGAATACCGAAGTGACTGCAGAGACCATCAAGGAATTCAA
gLeuGlyIleGluValArgLeuAsnThrGluValThrAlaGluThrIleLysGluPheAs
3010 3030 3050
TCCGGATAATGTCATCATCGCAGTAGGCTCTACCTATGCGCTGCCTGAGATTCCGGGAAT
nProAspAsnValIleIleAlaValGlySerThrTyrAlaLeuProGluIleProGlyIl
3070 3090 3110
CGACAGCCCAAGCGTATACTCCCAGTATCAGGTACTGAAAGGGGAAGTAAATCCGACAGG
eAspSerProSerValTyrSerGlnTyrGlnValLeuLysGlyGluValAsnProThrGl
3130 3150 3170
CCGTGTAGCCGTTATCGGATGCGGACTGGTTGGTACGGAAGTCGCAGAACTTCTGGCATC
yArgValAlaValIleGlyCysGlyLeuValGlyThrGluValAlaGluLeuLeuAlaSe
3190 NruI 3210 KpnI 3230
CAGAGGCGCACAGGTAATCGCGATCGAGAGGAAGGGCGTAGGTACCGGCCTTAGATGCTT
rArgGlyAlaGlnValIleAlaIleGluArgLysGlyValTGlyThrGlyLeuArgCysPh
MetLeu
baiD →
3250 EcoRI 3270 3290
CGCAGAATGTTCATGAACCCGGAATTCAAATATTACAAGATCGCCAAGATGTCCGGAACA
eAlaGluCysSer---
ArgArgMetPheMetAsnProGluPheLysTyrTyrLysIleAlaLysMetSerGlyThr

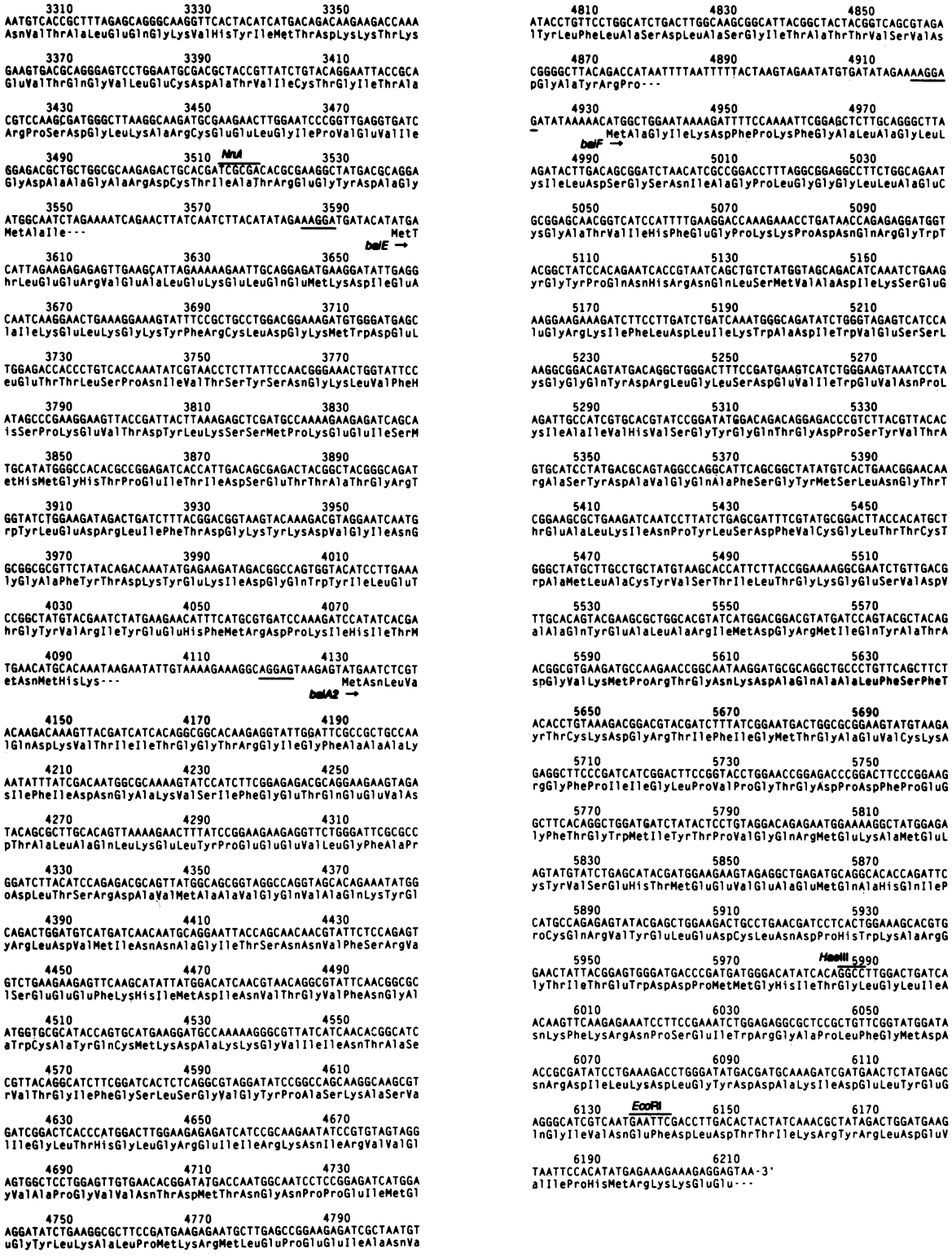


FIG. 6. Complete nucleotide and amino acid sequences of the proposed bile acid-inducible operon from *Eubacterium* sp. strain VPI 12708. Putative ribosome-binding sites are underlined. Restriction sites are overlined and appropriately labeled. Four methionine residues within the first nineteen amino acid residues of open reading frame *baiD* are presented in bold italic type. The correct initiation codon for this proposed *baiD* gene is unknown.

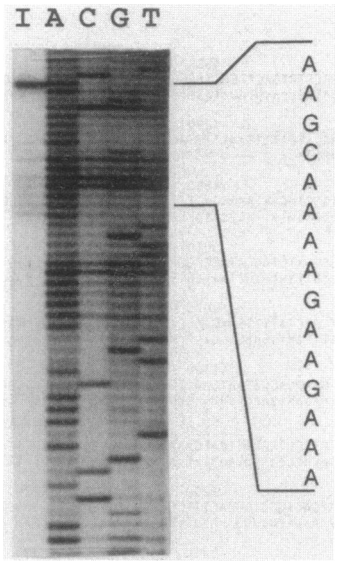


FIG. 7. Primer extension analysis for the 5′ mRNA initiation site. Lanes: I, primer extension using mRNA isolated from cholic acid-induced *Eubacterium* sp. strain VPI 12708 and a ^{32}P-labeled primer; A, C, G, and T, sequencing reactions using ^{35}S-labeled DNA.

and takes intergenic spacing into consideration. The intergenic spacing between the proposed *baiC* and *baiD* genes would then be similar to the spacing between the other genes on the operon (Fig. 7). The possibility of multiple initiation sites cannot be discounted for this proposed polypeptide.

The question of whether all six of the open reading frames reside on the same operon is partly answered by noting that a potential promoter region is situated immediately upstream from these open reading frames and that primer extension analysis reveals that mRNA synthesis initiates in this region. This potential promoter region shares considerable sequence homology with the promoter regions of previously reported bile acid-inducible genes from *Eubacterium* sp. strain VPI 12708 (5, 8). Also, the previously reported mRNA length for this operon was greater than 6 kb (24). These data together strongly suggest that all six open reading frames are transcribed as a polycistronic message from a single bile acid-induced operon. Determination of whether there are additional open reading frames downstream from the *baiF* gene will require further work, as it has been difficult to clone DNA restriction fragments from this region of the chromosome. Determination of the regulatory functions of the promoter-operator regions for the bile acid-inducible operon and the other bile acid-inducible genes will also require further study. However, the extensive homology exhibited in this regulatory region (Fig. 8) suggests the possibility of binding sites for regulatory proteins.

It is hypothesized that most or all of the polypeptides encoded by this bile acid-inducible operon are involved in the multistep 7-dehydroxylation pathway. Antibodies prepared against the 27,000-M_r polypeptides have been shown to inhibit 7-dehydroxylation activity (17). Antibodies prepared against HPLC fractions containing 7-dehydroxylation activity (17) have also been shown to react against the 19,500-, 27,000-, and 47,500-M_r polypeptides and possibly against the 58,000-, and 59,500-M_r polypeptides (4, 24; Fig. 3).

Data bank searches for amino acid sequences similar to those of the *bai* genes have previously revealed that the 27,000-M_r polypeptides exhibit extensive homology with several alcohol/polyol dehydrogenases (5, 12, 25). Therefore, the 27,000-M_r polypeptides may catalyze the oxidation-reduction of the 3α-hydroxy group of bile acid substrates in the 7α-dehydroxylation pathway. We report here that the 58,000-M_r polypeptide has sequence homology with several polypeptides from wide-ranging species. These polypeptides share the common activity of adenylation of compounds containing cyclic ring structures. Therefore, the 58,000-M_r polypeptide may be involved in the formation of the bile acid adenosine nucleotide described by Coleman et al. (3). Two of the homologous polypeptides, the firefly luciferase and the 4-coumarate:coenzyme A ligase from parsley, have similar M_rs of 61,000 and 60,000, respectively (7, 14). The tyrocidine synthetase 1 from *B. brevis* has a reported M_r of approxi-

```
27K-1&3:    CTATAGGGAAACAAAATAGTGATAGTGTTTGCAAACTTTTTGTCCA
5'           *  *   *  *    * **    ******* **      ** ********
OPERON:     ATGTTAGAAGTTATAAATCTGATAGTATTATAGCACATTTTGTCCC

27K-1&3:    TGGACTGCTTATATTTTGCAATTAAAAAAG-AA-CTTTACAAGTTG
            *  *    *** ** ** * *   ****** ** ****** * ***
OPERON:     TTAATCAATTAGATATTACCAAC-AAAAAGTAATCTTTACTA-TTG

                                    -35
27K-1&3:    TAAGATGCCGTGTGATTTTCCAATGTCGCGTCCTGT-AAAATGTTA
            *********  *  ** *** ** * *   *  *  ********
OPERON:     TAAGATGCCACGAAATTATCCGATATTGTACCAGGCAAAAATGTTA

            -10          mRNA
27K-1&3:    AAGTTGTATCAATCGATACGATACTTTGGCAGATATGATAAGCCAA
            ***** ****** * * *  * *   **    * *     *  *
OPERON:     AAGTTATATCAAGCAAAAGAAGAAAATGTTGAAAAAATGTACAAAT

27K-1&3     AGGAAAAGAAAGGAAGGAAAAGTTC ATG
            ** *  *  **   *   ***                    3'
OPERON      AGCACGAAGAAAAGATATTAAGCATTAAGAAA ATG
```

FIG. 8. Comparison of promoter regions for the cholic acid-inducible operon and the *baiA1* and *baiA3* bile acid-inducible genes (27K-1&3). The proposed −10 and −35 regions are boxed, and the mRNA initiation sites are indicated. A region containing significant homology is shaded. The ATG sequences on the 3′ ends represent the proposed initiation codons for the *baiA1* and *baiA3* genes and for the *baiB* gene (OPERON).

mately 120,000 (16), while the M_r of the polypeptide from *E. coli* involved in the activation of 2,3-dihydroxybenzoate has not been determined and the gene coding for this polypeptide has only been partially sequenced (13). The task of assigning specific catalytic activities to these polypeptides will require further study.

ACKNOWLEDGMENTS

We thank Janet Adams, Pamela Melone, and Sarah Jacobs for technical assistance in this study and Bryan A. White for assistance with the protein sequencing.

This work was supported by Public Health Service research grants DK40986 and P01DK38030 from the National Institutes of Health. D.H.M. was supported by Public Health Service postdoctoral training grant T32-CA-0956404 from the National Institutes of Health.

LITERATURE CITED

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **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.
4. **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.
5. **Coleman, J. P., W. B. White, M. Lijewski, and P. B. Hylemon.** 1988. Nucleotide sequence and regulation of a gene involved in bile acid 7α-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **170**:2070–2077.
6. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. Basic methods in molecular biology. Elsevier Science Publishing, Inc., New York.
7. **de Wet, J. R., K. V. Wood, M. Deluca, D. R. Helinski, and S. Subramani.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. Mol. Cell. Biol. **7**:725–737.
8. **Gopal-Srivastava, R., D. H. Mallonee, W. B. White, and P. B. Hylemon.** 1990. Multiple copies of a bile acid-inducible gene in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **172**:4420–4426.
9. **Holdeman, L. V., and W. E. C. Moore (ed.).** 1977. Anaerobic laboratory manual, 4th edition. Virginia Polytechnic Institute and State University, Blacksburg.
10. **Hylemon, P. B.** 1985. Metabolism of bile acids in intestinal microflora, p. 331–343. *In* H. Danielson and J. Sjovall (ed.), Sterols and bile acids. New comprehensive biochemistry, vol. 12. Elsevier Science Publishing, Inc., New York.
11. **Hylemon, P. B., and T. L. Glass.** 1983. Biotransformation of bile acids and cholesterol by the intestinal microflora, p. 189–213. *In* D. J. Henteg (ed.), Human intestinal microflora in health and disease. Academic Press, Inc., New York.
12. **Jörnvall, H., H. vonBahr-Lindström, K. D. Jany, W. Ulmer, and M. Fröschle.** 1984. Extended super family of short alcohol-polyol-sugar-dehydrogenase: structural similarities between glucose and ribitol dehydrogenases. FEBS Lett. **165**:190–196.
13. **Liu, J., K. Duncan, and C. T. Walsh.** 1989. Nucleotide sequence of a cluster of *Escherichia coli* enterobactin biosynthesis genes: identification and purification of its product 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase. J. Bacteriol. **171**:791–798.
14. **Lozoya, E., H. Hoffmann, C. Douglas, W. Schulz, D. Scheel, and K. Hahlbrock.** 1988. Primary structures and catalytic properties of isonenzymes encoded by the two 4-coumarate:CoA ligase genes in parsley. Eur. J. Biochem. **176**:661–667.
15. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol. **3**:208–218.
16. **Mittenhuber, G., R. Weckermann, and M. A. Marahiel.** 1989. Gene cluster containing the genes for tyrocidine synthetases 1 and 2 from *Bacillus brevis*: evidence for an operon. J. Bacteriol. **171**:4881–4887.
17. **Paone, D. A. M., and P. B. Hylemon.** 1984. HPLC purification and preparation of antibodies to cholic acid-inducible polypeptides from *Eubacterium* sp. VPI 12708. J. Lipid Res. **25**:1343–1349.
18. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
19. **Sjövall, J.** 1960. Bile acids in man under normal and pathological conditions: bile acids and steroids 73. Clin. Chim. Acta **5**:33–41.
20. **Weckermann, R., R. Fürbass, and M. A. Marahiel.** 1988. Complete nucleotide sequence of the *tycA* gene coding the tyrocidine synthetase 1 from *Bacillus brevis*. Nucleic Acids Res. **16**:11841.
21. **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 V.P.I. 12708. J. Lipid Res. **22**:891–898.
22. **White, B. A., R. L. 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.
23. **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 V.P.I. 12708. J. Lipid Res. **24**:20–27.
24. **White, W. B., J. P. Coleman, and P. B. Hylemon.** 1988. Molecular cloning of a gene encoding a 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **170**:611–616.
25. **White, W. B., C. V. Franklund, J. P. Coleman, and P. B. Hylemon.** 1988. Evidence for a multigene family involved in bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J. Bacteriol. **170**:4555–4561.
26. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
27. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probe. Proc. Natl. Acad. Sci. USA **80**:1194–1198.