Evidence for a Multigene Family Involved in Bile Acid 7-Dehydroxylation in *Eubacterium* sp. Strain VPI 12708

W. BRUCE WHITE,[†] CLIFTON V. FRANKLUND, JAMES P. COLEMAN,[‡] AND PHILLIP B. HYLEMON*

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

Received 28 March 1988/Accepted 5 July 1988

Eubacterium sp. strain VPI 12708 is a human intestinal isolate which has an inducible bile acid 7-dehydroxylation activity. At least two cholic acid-induced polypeptides, with molecular masses of 27,000 and 45,000 daltons, respectively, coelute with bile acid 7-dehydroxylation activity. The 45,000-dalton polypeptide appears to be encoded by a cholic acid-induced mRNA species of greater than 6 kilobases, which suggests that the gene coding for this polypeptide is part of a larger operon. A gene has been cloned which flanks the gene encoding the 45,000-dalton polypeptide, in the upstream (5') direction. This gene appears to encode a second 27,000-dalton polypeptide. The gene bears striking homology at both the nucleotide (80%) and deduced amino acid sequence (89%) levels with the gene which encodes the 27,000-dalton polypeptide that has been shown previously to be involved in the bile acid 7-dehydroxylation reaction sequence. The implications of this homology and the possible function(s) of the two homologous genes in bile acid 7-dehydroxylation are discussed. Evidence is presented which suggests that the two homologous genes involved in bile acid 7-dehydroxylation may be part of a larger multigene family in *Eubacterium* sp. strain VPI 12708.

Cholic acid and chenodeoxycholic acid are the major bile acids synthesized from cholesterol in the human liver. During their enterohepatic circulation, these primary bile acids are transformed into a variety of metabolites by the intestinal microflora (9, 10). The metabolism of bile acids by the intestinal microflora is limited by the constraints inherent in the strictly anaerobic environments of the colon. Quantitatively, the most important biotransformation of the primary bile acids by the intestinal microflora is the 7-dehydroxylation, yielding deoxycholic acid and lithocholic acid, which together make up 20 to 25% of the total biliary bile acid pool in humans (26). These secondary bile acids differ markedly from their 7-hydroxylated precursors in physicochemical properties, as well as in physiological effects (2, 4, 13-15, 21, 22, 24).

Studies in our laboratory have focused upon the mechanism of bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. This human intestinal isolate possesses a cholic acid-inducible bile acid 7-dehydroxylation activity (30). The induction of bile acid 7-dehydroxylation activity by sodium cholate in this bacterium coincides with the appearance of at least four new polypeptides with apparent molecular masses of 23,500, 27,000, 45,000, and 77,000 daltons (23, 29). The 27,000- and 45,000-dalton polypeptides coelute with bile acid 7-dehydroxylation activity (23, 31). Both the 27,000-dalton (6) and 45,000-dalton (23, 31) polypeptides have been purified, and the genes which encode these polypeptides have been cloned (6, 31). The nucleotide sequence of the gene which encodes the 27,000-dalton polypeptide and flanking regions has been determined (7).

During preliminary sequence analysis of the gene which encodes the 45,000-dalton polypeptide, we observed that the

flanking upstream sequence bears striking homology with the nucleotide sequence of the gene which encodes the 27,000-dalton polypeptide. In this article we present the nucleotide sequence of a gene which appears to code for an additional 27,000-dalton polypeptide involved in bile acid 7-dehydroxylation.

MATERIALS AND METHODS

Bacterial strains, bacteriophages, and media. *Eubacterium* sp. strain VPI 12708 was grown and bile acid 7-dehydroxylation activity was induced as described previously (30). *Escherichia coli* JM105 (33) and DH5 α (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) were used as host strains for bacteriophage M13mp18 and plasmid pUC8. *E. coli* DH5 α was grown on LB broth (16) or LB broth supplemented with ampicillin at 100 μ g/ml. *E. coli* JM105 was grown on YT broth (18).

Oligonucleotide synthesis and purification. Oligonucleotides were synthesized with a Biosearch Cyclone DNA synthesizer (Biosearch, Inc., San Rafael, Calif.). The oligonucleotides were deprotected and purified as described previously (31).

Nucleic acid isolation. *Eubacterium* sp. strain VPI 12708 DNA was isolated by the method of Marmur (17). Plasmid DNA was isolated by the Ish-Horowitz and Burke modification of the method of Birnboim and Doly (16) and further purified by CsCl gradient centrifugation. Single-stranded M13 DNA was prepared for sequencing by using the procedure described by Bethesda Research Laboratories.

RNA was isolated from exponentially growing cells of *Eubacterium* sp. strain VPI 12708 as described previously (31) and stored as an ethanol precipitate at -20°C until use.

Southern blotting and labeling of oligonucleotides. DNA restriction fragments were transferred from 0.8% agarose gels to nitrocellulose as described by Southern (27). Baked filters were washed, prehybridized, and hybridized as described by Woods (32). DNA-DNA hybridizations were performed at 42°C. Hybridization mixtures contained 10⁶

* Corresponding author.

[†] Present address: Plant Hormone Laboratory, Beltsville Agricultural Research Center, Beltsville, MD 20705.

[‡] Present address: Department of Microbiology and Immunology, East Carolina University School of Medicine, Greenville, NC 27858-4354.

cpm of ^{32}P -labeled synthetic oligonucleotide probe per ml (10^9 cpm/ μg).

Synthetic oligonucleotide probes were labeled with [γ - ^{32}P]ATP ($>3,000$ Ci/mmol) by using T4 polynucleotide kinase as described by Maniatis et al. (16). Unincorporated label was removed with Nensorb 20 cartridges (Du Pont, NEN Research Products, Boston, Mass.) as specified by the manufacturer.

Recombinant DNA methods. Restriction endonuclease digests were carried out as specified by the suppliers. Electrophoresis of DNA, ligation reactions, and other nucleic acid manipulations were performed by the methods of Maniatis et al. (16).

In preparation for DNA sequencing, each strand of the 2.6-kilobase *Eco*RI insert fragment from a pUC8-derived recombinant plasmid shown previously to contain a gene which encodes a 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation (31) was subcloned into M13mp18. *Sau*3AI and *Hpa*II subfragments of the 2.6-kilobase *Eco*RI insert were also subcloned into M13mp18 in preparation for DNA sequencing. For M13 subcloning, *E. coli* DH5 α was transformed with M13mp18-derived recombinant plasmids by a method suggested by the supplier (Bethesda Research Laboratories). M13mp18-derived recombinant phages were then produced with *E. coli* JM105, with *E. coli* DH5 α as a transient host.

DNA sequencing. Dideoxynucleotide chain termination sequence analysis was performed by the methods of Sanger et al. (25). Both strands were completely sequenced. A universal sequencing primer (17-mer) complementary to the M13 sequencing vector (Bethesda Research Laboratories) was used to sequence DNA inserts proximal to the universal primer-binding site. Oligonucleotide primers (17-mers) were chemically synthesized to sequence internal areas of the inserts. Regions of ambiguous DNA sequence were further analyzed by replacing dGTP with 7-deaza-dGTP in the dideoxy sequencing reactions (3, 19); the protocol provided by the supplier (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) was followed. Avian myeloblastosis virus reverse transcriptase (8) was used in place of the DNA polymerase I large fragment (Klenow fragment) in some sequencing reactions. [^{35}S]dATP (600 Ci/mmol; Du Pont, NEN) was used for all sequencing reactions.

Primer extension assays. Primer extension assays were performed on total RNA isolated from cholic acid-induced and uninduced cultures of *Eubacterium* sp. strain VPI 12708 with avian myeloblastosis virus reverse transcriptase by using a method described by Alam et al. (1), except that [^{35}S]dATP (>600 Ci/mmol) was used instead of [α - ^{32}P]dATP. A synthetic oligonucleotide (21-mer) specific for the gene which encodes the 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation (31) was used as the primer in these reactions. The location and orientation of this gene on the cloned 2.6-kilobase *Eco*RI fragment have been reported previously (31). The sequence of the primer used in the RNA primer extension reactions was 5'-CTTAAGCCCTGCAAGAGCTCC-3'. This sequence, located on the coding strand within the region coding for the 45,000-dalton polypeptide, begins 50 bases downstream from the A of the initiator methionine codon.

Sequence data analysis. Analyses of nucleic acid and protein sequence data were performed by using either the Microgenie Sequence Analysis Program (Beckman Instruments, Inc., Fullerton, Calif.) or the IBI/Pustell DNA and Protein Sequence Analysis System (International Biotechnologies, Inc., New Haven, Conn.).

A.

B.

FIG. 1. Partial restriction maps of cloned fragments containing two homologous *Eubacterium* genes associated with bile acid 7-dehydroxylation (27K-1 and 27K-2). Fragment A has been shown previously to contain a gene encoding a 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation (31). The gene designated 27K-2 is the subject of this paper. Fragment B has been shown previously to contain a gene encoding a 27,000-dalton polypeptide involved in bile acid 7-dehydroxylation catalysis (6). Abbreviations: E, *Eco*RI; T, *Taq*I; kb, kilobases.

RESULTS

Physical mapping and DNA sequence analysis. Preliminary analysis indicated that the nucleotide sequence directly upstream from the gene that encodes the 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation (the 45K gene) appeared to be very similar to the nucleotide sequence of a previously described gene (the 27K-1 gene). The 27K-1 gene encodes the 27,000-dalton polypeptide (6), which has been implicated as a major catalytic component in bile acid 7-dehydroxylation (23). Southern hybridization analysis and restriction mapping of *Eubacterium* sp. strain VPI 12708 chromosomal DNA and cloned fragments containing the 45K and 27K-1 genes, respectively, indicated that the 45K and 27K-1 genes were not located on identical fragments (Fig. 1).

Determination of the nucleotide sequence of 1 kilobase of DNA which flanks the 45K gene in the upstream direction indicated that there was an open reading frame directly upstream of the 45K gene, which appeared to encode a polypeptide with an estimated molecular mass of 26,518 daltons. This open reading frame was designated the 27K-2 gene (Fig. 1). The complete DNA sequence is shown in Fig. 2, along with the deduced amino acid sequence of the 27K-2 gene. The start and stop codons for the 27K-2 gene are bounded by residues 221 and 967, respectively. A putative ribosomal-binding site (AAGGCAGG) was present upstream from the initiation codon at nucleotide positions 204 to 211. The stop codon for the 27K-2 gene was located 55 bases upstream from the initiation codon of the 45K gene.

Alignment of the sequence of the 27K-2 gene with the previously reported sequence of the 27K-1 gene (7) indicated that there is 80% homology at the nucleotide sequence level between the two genes (Fig. 3). Significant homology between the two sequences was not observed either upstream of the initiation codon or downstream of the stop codon for the two genes. Comparison of the nucleotide sequences of the two genes indicated that both genes contained the same number of nucleotide residues.

Amino acid sequence analysis. Alignment of the deduced amino acid sequence of the 27K-1 polypeptide (7) with the deduced amino acid sequence of the putative 27K-2 polypeptide indicated that there is 89% homology between the

10	20	30	40	50	60
A AGA TAG ACT GAT CTT TAC	GGA CGG TAA GTA CAA AGA	CGT AGG AAT CAA TGG	CGG CGC GTT		
70	80	90	100	110	120
CTA TAC AGA CAA ATA TGA	GAA GAT AGA CGG CCA	GTG GTA CAT CCT TGA	AAC CGG CTA	TGT	
130	140	150	160	170	180
ACG AAT CTA TGA AGA ACA	TTT CAT GCG TGA TCC	AAA GAT CCA TAT CAC	GAT GAA CAT GCA		
190	200	210	220	230	240
CAA ATA AGA ATA TTG TAA	AAG <u>AAA GGC AGG</u>	AGT AAG AGT <u>ATG</u>	AAT <u>ATG</u> CTT	GTA CAA GAC	AAA
			Met	Asn	Gln
250	260	270	280	290	300
GTT ACG ATC ATC ACA GGC	GGC ACA AGA GGT	ATT GGA TTC GCT	GCT GGC AAA	ATA TTT	ATC
Val Thr Ile Ile Thr Gln	Gly Thr Arg Gly Ile	Gly Phe Ala Ala	Ala Lys Lys	Ile Phe	Ile
310	320	330	340	350	360
GAC AAT GGC GCA AAA GTA	TCC ATC TTC GGA	GAG ACG CAG GAA	GTA GAT ACA	CGC CTT	
Asp Asn Gly Ala Lys Val	Ser Ile Phe Gly Glu	Thr Thr Glu Glu	Val Asp Thr	Ala Leu	
370	380	390	400	410	420
GCA CAG TTA AAA GAA CTT	TAT CCG GAA GAA	GAG GTT GGA TTC	GCG CCG	GAT CTT	ACA
Ala Gln Leu Lys Glu Leu	Gly Thr Glu Val	Phe Thr Glu Lys	Tyr Gly Arg	Leu Asp	
430	440	450	460	470	480
TCC AGA GAC GCA GTT ATG	GCA GCG GTA GGC	CAG GGA GCA CAG	AAA TAT GGC	AGA CTG	GAT
Ser Arg Asp Ala Val Met	Leu Ile Val Thr	Gly Val Ala Gln	Lys Tyr Gly Arg	Leu Asp	
490	500	510	520	530	540
GTC ATG ATC AAC AAT GCA	GGA ATT ACC AGC	AAC AAG GGA TTC	TCC AGA	GTG TCT	GAA
Val Met Ile Asn Asn Ala	Gly Ile Thr Ser Asn	Asn Val Phe Ser	Arg Val	Gly Leu	Glu
550	560	570	580	590	600
GAG TTC AAG CAT ATT ATG	GAC ATC AAC GGA	ACA GGC GTA TTC	GTC TGC	TGG TGC	GCA
Glu Phe Lys His Ile Met	Leu Ile Asn Val	Thr Thr Gly Val	Phe Asn Gly	Ala Trp	Cys
610	620	630	640	650	660
TAC CAG TGC ATG AAG GAT	GCC AAA AAG GGC	GTT ATC ATC AAC	ACG GCA TCC	GTT ACA	GGC
Tyr Gln Cys Met Lys Ser	Met Lys Lys Val	Ile Val Ile Ile Asn	Thr Thr Val	Thr Thr	Gly
670	680	690	700	710	720
ATC TTC GGA TCA CTC TCA	GCG GTA GGA TAT	CCG GCC AGC AAG	GCA AGC	GTG ATC	GGA
Ile Phe Gly Ser Leu Ser	Leu Thr Asp Met	Leu Thr Pro Ala	Ser Lys Ala	Ser Val	Ile
730	740	750	760	770	780
ACC CAT GGA CTT GGA AGA	GAG ATC ATC CCG	AAG AAT ATC CGT	GTA GGA GTG	GCT CCT	
Thr His Gly Leu Gly Arg	Glu Ile Ile Arg Lys	Asn Ile Arg Val	Val Gly Val	Ala Pro	
790	800	810	820	830	840
GGA GTT GTG AAC ACG GAT	ATG ACC AAT GGC	AAT CCT CCG	GAG ATC ATG	GAA GGA	TAT
Gly Val Leu Val Asn Thr	Met Thr Asn Gly	Asn Pro Glu	Ile Met Gly	Tyr Leu	
850	860	870	880	890	900
AAG GCG CTT CCG ATG AAG	AGA ATG CTT GAG	CCG GAA GAG	ATC GCT	AAT GTA	TAC
Lys Ala Leu Pro Met Lys	Arg Met Leu Glu	Pro Glu Glu	Ile Ala Asn	Val Tyr	Leu
910	920	930	940	950	960
CTG GCA TCT GAC TTG	GCA AGC GGC ATT	ACG GCT ACT	ACG GTC AGC	GTA GAC	GCG
Leu Ala Ser Asp Leu Ala	Ser Gly Ile Thr Ala	Thr Thr Val	Ser Val	Asp Gly	Ala
970	980	990	1000	1010	1020
AGA CCA TAA TTT TAA	TTT TTA CTA AGT	AGA ATA TGT	GAT ATA <u>GAG GAG</u>	ATA TAA	AA
Arg Pro					
1030	1040	1050	1060	1070	1080
AC ATG GCT GGA ATA AAA	GAT TTT CCA AAA	TTC GGA GCT	CTT GCA GGG	CTT AAG	ATA
Met Ala Gly Ile Lys	Asp Phe Pro Lys	Phe Phe Gly Ala	Leu Ala Gly Leu	Lys Ile Leu	Asp

FIG. 2. Nucleotide sequence of the noncoding strand of the 27K-2 gene and its flanking regions. The deduced amino acid sequence for the 27K-2 gene is printed below the nucleotide sequence. The initiator methionine codon of the 27K-2 gene is overlined (positions 221 to 223). The first 59 bases of the 3'-flanking gene, shown previously to encode a cholic acid-induced 45,000-dalton polypeptide associated with bile acid 7-dehydroxylation (23, 31), are also shown. The N-terminal amino acid sequence of the 45,000-dalton polypeptide reported previously (31) is shown below the corresponding nucleotides. Putative ribosomal-binding sites are underlined.

two at the amino acid sequence level (Fig. 4). Single amino acid substitutions were found at 22 different positions. There was one run of five amino acids which was different in the two sequences (positions 159 to 164). As was reported previously for the 27K-1 polypeptide (7), a computer-aided protein sequence search indicated a significant similarity between the 27K-2 deduced amino acid sequence and several previously sequenced alcohol-polyol dehydrogenases, specifically the ribitol dehydrogenase from *Klebsiella aerogenes* (20), the alcohol dehydrogenase from *Drosophila melanogaster* (28), and the glucose dehydrogenase from

Bacillus megaterium (11). These enzymes belong to a group of pyridine-linked dehydrogenases referred to as "short" or "non-zinc" dehydrogenases (12). The area of highest sequence similarity between the putative 27K-2 protein, the 27K-1 protein, and the above proteins is shown in Fig. 5. This conserved amino acid sequence has previously been postulated to contain two β -sheet areas responsible for pyridine nucleotide coenzyme binding (12).

Transcriptional mapping. To determine whether the 27K-2 gene and the 45K gene are cotranscribed, we conducted primer extension assays as described in Materials and Methods. Multiple bands were observed in primer extension assays with RNA from cholic acid-induced cells as the template, with two major bands at nucleotide positions 957 and 975 (Fig. 2) and minor bands extending to the top of the sequencing gel, into the coding sequence for the putative 27K-2 polypeptide (Fig. 6). No bands were observed in primer extension assays with RNA from uninduced cells as the template.

Evidence for an additional *Eubacterium* sp. strain VPI 12708 gene related to the 27K-1 and 27K-2 genes. Southern hybridization analysis of *Eco*RI-cut chromosomal DNA probed with an oligonucleotide consisting of a common sequence found within the coding regions of both the 27K-1 and 27K-2 genes (Fig. 2, nucleotides 473 to 495) repeatedly showed three bands with sizes of 2.2, 2.6, and approximately 3.5 kilobases, respectively (Fig. 7). The hybridization of the 2.2- and 2.6-kilobase fragments is consistent with the restriction maps of the 27K-1 and 27K-2 genes and flanking regions (Fig. 1). Southern hybridization analysis of *Taq*I-cut chromosomal DNA showed two bands with sizes of 1.9 and 1.1 kilobases. The hybridization of these two fragments is also consistent with the restriction maps of the 27K-1 and 27K-2 genes and flanking regions.

DISCUSSION

The results of physical mapping studies (Fig. 1) and DNA sequence analysis (Fig. 2 and 3) strongly suggest that the 27K-1 and 27K-2 genes share a common evolutionary origin. The high degree of homology shared by the 27K-1 and 27K-2 genes at both the nucleotide (80%) and deduced amino acid sequence (89%) levels (Fig. 3 and 4) suggests strongly that the 27K-2 gene encodes a 27,000-dalton polypeptide very similar in both structure and function to the polypeptide encoded by the 27K-1 gene. Our results show that the 27K-2 gene is physically linked to the 45K gene.

We have shown previously by Northern (RNA) blot analysis that the 45,000-dalton polypeptide is encoded by a cholic acid-induced mRNA species which appears to be >6 kilobases in size (31). This suggested that the 45K gene was part of a larger operon. Our RNA primer extension results suggest that the 27K-2 gene and the 45K gene are cotranscribed, although the results do not rule out the possibility of an internal promoter within the coding sequence for the putative 27K-2 polypeptide. Additional studies are required to determine whether the multiple mRNA species we routinely detect in primer extension assays are a result of degradation associated with the isolation of the RNA or are the result of some form of RNA processing. We have shown previously that the 27K-1 gene product is encoded by a different mRNA species, which appears to encode only that polypeptide (7).

The discovery of a gene which appears to encode a polypeptide very similar to the 27,000-dalton polypeptide, which has been previously implicated in bile acid 7-dehy-

	1	10	20	30	40	50		610	620	630	640	650
27-1	ACGCGATGCTGGATCTATAGGGAACAAATAGTGATAGTGTGGCAAAC						27-1	TTATCAGTGCATGAAAGATGCAAAGCAGGGCGTTATCATCAATACGGCAT				
27-2	AAGATAGACTGATCTTTACGGACGGTAAGTACAAAGACGTAGGAATCAAT						27-2	ATACCAGTGCATGAAGGATGCCAAAAAGGGCGTTATCATCAACACGGCAT				
	60	70	80	90	100			660	670	680	690	700
27-1	TTTTGTCCATGGACTGCTTATATTGCAATTAAAAAGAACTTTACAA					27-1	CTGTAACCGGAATCTATGGTTCCTTATCAGGAATCGGATATCCATACCAG					
27-2	GGCGGCGGCTTCTATACAGACAAATATGAGAAGATAGACGGCCAGTGGTA					27-2	CCGTTACAGGCATCTTCGGATCACTCTCAGGCGTAGGATATCCGGCCAGC					
	110	120	130	140	150			710	720	730	740	750
27-1	GTGTGAAGATGCCGTGTGATTTTCCAATGTCGCGTCTGTAAAAATGTTAA					27-1	CCAGGCGGCGTAATCGGCCTGACTCATGGTCTTGGAAAGAGAGATTATCCG					
27-2	CATCCTTGAAACCGGCTATGTACGAATCTATGAAGAACATTTTCATGCGTG					27-2	AAGGCAAGCGTGATCGGACTCACCATGGACTTGGAAAGAGAGATCATCCG					
	160	170	180	190	200			760	770	780	790	800
27-1	AGTTGTATCAATCGATACGATCTTTGGCAGATATGATAAGCCAAAGGAA					27-1	TAAGAAATCCGTGTAGTTGGCGTTGCACCTGGCGTTGTAGATACAGATA					
27-2	ATCCAAAGATCCATATCACGATGAACATGCACAAATAAGAATATGTGAAA					27-2	CAAGAATATCCGTGTAGTAGGAGTGGCTCTGGAGTTGTGAACACGGATA					
	210	220	230	240	250			810	820	830	840	850
27-1	AAGAAAGGAAGGAAAAGTTCATGAAACTTGTACAGGACAAATTACAATT					27-1	TGACGAAGGGGCTTCCACCGGAGATCCTGGAGGACTACTTGAAGACACTG					
27-2	AGAAAGGCAGGAGTAAGAGTATGAATCTCGTACAAGACAAAGTTACGATC					27-2	TGACCAATGGCAATCCTCCGGAGATCATGGAAGGATATCTGAAGGCGCTT					
	260	270	280	290	300			860	870	880	890	900
27-1	ATCACAGGCGGAACCCGTGGAATCGGATTCGCAGCAGCAAACTCTTTAT					27-1	CCAATGAAGAGAATGCTTAAGCCGGAAGAGATCGCGAATGTATATCTGTT					
27-2	ATCACAGGCGGCACAAGAGGTATTTGGATTGCGCGCTGCCAAAATATTAT					27-2	CCGATGAAGAGAATGCTTGAGCCGGAAGAGATCGCTAATGTATACCTGTT					
	310	320	330	340	350			910	920	930	940	950
27-1	TGAGAATGGAGCAAAAGTCTCCATATTTGGCGAGACCCAGGAAGAGGTAG					27-1	CCTTGCACTCCGACCTGGCTAGCGGCATCACGGCTACGACGATCAGCGTAG					
27-2	CGACAATGGCGCAAAAGTATCCATCTTCGGAGAGACGCAGGAAGAAAGTAG					27-2	CCTGGCATCTGACTTGGCAAGCGGCATTACGGCTACTACGGTCAGCGTAG					
	360	370	380	390	400			960	970	980	990	1000
27-1	ACACAGCGCTGGCTCAGTTAAAGGAACCTATCCGGAGGAAGAGGTATTA					27-1	ATGGGGCTTACAGGCCATAGAAAAGACATACTGCTATTAAATCCATAGTT					
27-2	ATACAGCGCTTGCACAGTTAAAGAACTTTATCCGGAAGAAAGAGGTTCTG					27-2	ACGGGGCTTACAGACCATAATTTTAAATTTTACTAAGTAGAATATGTGAT					
	410	420	430	440	450			1010	1020	1030	1040	1050
27-1	GGATTGCTCCAGACCTTACATCAAGAGATGCTGTATGGCAGCAGTTGG					27-1	CATACTCCAAGAACAGGCAGGCAAGAGGCATTTCGCTTTTAGCGCGGATG					
27-2	GGATTGCGCCGGATCTTACATCCAGAGACGCAGTTATGGCAGCGTTAGG					27-2	ATAGAAAAGGAGATATAAAAACATGGCTGGAATAAAGATTTTCCAAAAT					
	460	470	480	490	500			1060	1070	1080		
27-1	AACGGTTGCACAGAAGTACGGAAGACTGGATGTCATGATCAACAACGCAG					27-1	CCCGGGCTGCTGCTGATTAAATTCAGCTGGT					
27-2	CCAGGTAGCACAGAAATATGGCAGACTGGATGTCATGATCAACAATGCAG					27-2	TCGGAGCTCTTGCAGGGCTTAAGATACTTG					
	510	520	530	540	550							
27-1	GCATTACAATGAATCTGTATTCTCCAGGGTATCAGAAAGAGGATTTCAA											
27-2	GAATTACCAGCAACAACGTATTCTCCAGAGTGTCTGAAGAAGAGTTCAAG											
	560	570	580	590	600							
27-1	AATATAATGGACATCAATGTTAACGGCGTATTCAATGGCGCATGGTCTGC											
27-2	CATATATGGACATCAACGTAAACAGGCGTATTCAACGGCGCATGGTCTGC											

FIG. 3. Comparison of the nucleotide sequences of the 27K-1 and 27K-2 genes and corresponding flanking regions. Sequences of the noncoding strands are shown. Numbering of the bases is the same as in Fig. 2. The nucleotide sequence of the 27K-1 gene and its flanking regions has been reported previously (7). The initiator methionine codon of the 27K-1 gene is overlined (positions 221 to 223), as is the amber codon at the 3' end of the 27K-1 gene (positions 968 to 970). Asterisks indicate positions at which the bases are identical for the two sequences being compared.

FIG. 4. Comparison of the deduced amino acid sequences for the 27K-1 and 27K-2 genes. The deduced amino acid sequence for the 27K-1 gene has been reported previously (7), as has the N-terminal amino acid sequence of the corresponding 27,000-dalton polypeptide (6). Residue numbering refers to both 27K-1 and 27K-2 deduced amino acid sequences. Boxes surround positions which are identical for both sequences.

droxylation (23), is noteworthy in light of new evidence concerning the mechanism of bile acid 7-dehydroxylation. A novel bile acid-nucleotide conjugation activity has been identified which coelutes with bile acid 7-dehydroxylation activity on gel filtration high-performance liquid chromatography. One product of the bile acid-nucleotide conjugation activity has been identified as 12 α -hydroxy-3-oxo-4-cholenic acid linked at C-24 with an anhydride bond to the β phosphate 5' of ADP 3'-phosphate (5). This bile acid adenosine nucleotide appears to be an intermediate in bile acid 7-dehydroxylation. The structure of the bile acid adenosine nucleotide strongly suggests a new mechanism for bile acid 7-dehydroxylation (5, 7). This proposed multiple-step mechanism implies the involvement of several steroid oxidoreductases in bile acid 7-dehydroxylation.

Immunoinhibition experiments implicated the 27,000-dalton polypeptide encoded by the 27K-1 gene in bile acid

FIG. 6. Primer extension assays. Lanes: U, primer extension with RNA from uninduced cells as the template; I, primer extension with RNA from cholic acid-induced cells as the template; A, C, G, and T, dideoxynucleotide sequencing reactions with single-stranded M13mp18 DNA as the template and the same primer as in lanes U and I. Avian myeloblastosis virus reverse transcriptase was the enzyme used in both primer extension and dideoxynucleotide sequencing reactions. The band at the bottom of the figure in lane T corresponds to the base complementary to the A at nucleotide 992 in Fig. 2.

FIG. 5. Alignment of deduced amino acid sequences for 27K-1 and 27K-2 genes with amino acid sequences of several pyridine nucleotide-linked dehydrogenases. Abbreviations: DADH, *D. melanogaster* alcohol dehydrogenase (28); GDH, *B. megaterium* glucose dehydrogenase (11); RDH, *K. aerogenes* ribitol dehydrogenase (20). The numbers after each protein designation indicate the residue number of the first amino acid shown in that particular line. Boxed areas indicate regions of identity (or containing Ile, Val, and Leu residues). The Leu and Phe residues below DADH position 26 are boxed, since there is an even distribution between these two possibilities.

FIG. 7. Southern hybridization analysis of *Eubacterium* sp. strain VPI 12708 genomic digests probed with 32 P-labeled oligonucleotide (23-mer) consisting of a common sequence found within the coding regions of both 27K-1 and 27K-2 genes. The probe sequence is 5'-AGACTGGATGTCATGATCAACAA-3'. The sequence is that shown at nucleotides 473 to 495 in Fig. 2 and 3. Abbreviation: kb, kilobases.

7-dehydroxylation activity (23). We have suggested that this polypeptide may be involved in at least one of the pyridine nucleotide-dependent steroid oxidation-reduction reactions postulated for the bile acid 7-dehydroxylation pathway (7). The deduced amino acid sequence of this polypeptide contains regions very similar to sequences believed to encompass the pyridine nucleotide-binding domain of several alcohol-polyol dehydrogenases (12). The deduced amino acid sequence for the putative polypeptide encoded by the 27K-2 gene shows the same similarity (Fig. 5). However, the two deduced amino acid sequences are not identical in the region of similarity with the pyridine nucleotide-linked dehydrogenases. We propose that the putative 27,000-dalton polypeptide encoded by the 27K-2 gene is also involved in at least one of the pyridine nucleotide-dependent steroid oxidation-reduction reactions postulated for the bile acid 7-dehydroxylation pathway. Additional studies are required to determine whether both polypeptides function in the same step or in different steps in the bile acid 7-dehydroxylation pathway.

The results of Southern hybridization experiments suggest that the 27K-1 and 27K-2 genes may be members of a larger multigene family in *Eubacterium* sp. strain VPI 12708 (Fig. 7). In addition to the two *Eco*RI fragments, which contain sequences within the 27K-1 and 27K-2 genes, respectively, a strongly hybridizing *Eco*RI fragment of approximately 3.5 kilobases was also repeatedly observed. Investigators in our laboratory have recently cloned the strongly hybridizing 3.5-kilobase *Eco*RI fragment in lambda gt11. Efforts are in progress to map and sequence this *Eco*RI fragment. The Southern hybridization results rule out the possibility that either of the two 27K genes which we have isolated and sequenced is the result of a cloning artifact.

We are presently attempting to clone the rest of the large bile acid-inducible operon, which contains the 27K-2 gene, the 45K gene, and perhaps other genes associated with bile acid 7-dehydroxylation. Additional studies are required to define the specific functions of the products of these genes in bile acid 7-dehydroxylation.

ACKNOWLEDGMENTS

We thank Brian F. Bauer for help with oligonucleotide synthesis, Diane Parlow for typing the manuscript, and W. Michael Holmes for discussion.

This work was supported in part by Program Project Grant PO1 DK38030 from the National Institutes of Health.

LITERATURE CITED

- Alam, J., R. A. Whitaker, D. W. Krogman, and S. E. Curtis. 1986. Isolation and sequence of the gene for ferredoxin I from the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **168**:1265-1271.
- 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.
- Barr, P. J., R. M. Thayer, P. Laybourn, R. C. Najarian, F. Seela, and D. R. Tolan. 1986. 7-Deaza-2'-deoxyguanosine-5'-triphosphate: enhanced resolution of M13 dideoxy sequencing. *Biotechniques* **5**:428-432.
- Carey, M. C. 1985. Physical-chemical properties of bile acids and their salts, p. 345-403. In H. Danielson and J. Sjövall (ed.), *Sterols and bile acids*. New comprehensive biochemistry, vol. 12. Elsevier Science Publishing, Inc., New York.
- 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 7 α -dehydroxylation by an intestinal *Eubacterium* species. *J. Biol. Chem.* **262**:4701-4707.
- 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.
- Coleman, J. P., W. B. White, M. Lijewski, and P. B. Hylemon. 1988. Nucleotide sequence and regulation of a gene involved in bile acid 7-dehydroxylation by *Eubacterium* sp. strain VPI 12708. *J. Bacteriol.* **170**:2070-2077.
- Graham, A., J. Steven, D. McKechnie, and W. J. Harris. 1985. Direct DNA sequencing using avian myeloblastosis virus and Moloney-murine leukemia virus reverse transcriptase. *Focus* **8**:2.
- Hylemon, P. B. 1985. Metabolism of bile acids in intestinal microflora, p. 331-343. In H. Danielson and J. Sjövall (ed.), *Sterols and bile acids*. New comprehensive biochemistry, vol. 12. Elsevier Science Publishing, Inc., New York.
- Hylemon, P. B., and T. L. Glass. 1983. Biotransformation of bile acids and cholesterol by the intestinal microflora, p. 189-213. In D. J. Hentges (ed.), *Human intestinal microflora in health and disease*. Academic Press, Inc., New York.
- Jany, K. D. 1983. Complete amino acid sequence of glucose dehydrogenase from *Bacillus megaterium*. *FEBS Lett.* **165**:6-10.
- Jornvall, 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.
- 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.
- Kulkarni, M. S., P. M. Heidepriem, and K. L. Yielding. 1980. Production by lithocholic acid of DNA strand breaks in L1210 cells. *Cancer Res.* **40**:2666-2669.
- Low-Beer, T. S., and S. Nutter. 1978. Colonic activity, biliary cholesterol saturation, and the pathogenesis of gallstones. *Lancet* **ii**:1063-1065.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Marmur, J. 1961. A procedure for the isolation of deoxyribonucleic acid from microorganisms. *J. Mol. Biol.* **3**:208-218.
- Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**:20-78.
- Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination of DNA sequencing by use of deoxy-7-deazaguanine triphosphate in place of GTP. *Nucleic*

- Acids Res. 14:1319-1324.
20. Moore, C. H., S. S. Taylor, M. J. Smith, and B. S. Hartley. 1978. Ribitol dehydrogenase-*Klebsiella aerogenes*, p. 68. In M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Silver Spring, Md.
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. 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.
23. Paone, D. A. M., and P. B. Hylemon. 1984. HPLC purification and preparation of antibodies to cholic acid-inducible polypeptides from *Eubacterium* sp. V.P.I. 12708. J. Lipid Res. 25:1343-1349.
24. Reddy, B. A., J. H. Weisburger, and E. L. Wynder. 1978. Colon cancer: bile salts as tumor promoters, p. 454-464. In T. J. Slaga, A. Savik, and R. J. Boutwell (ed.) Carcinogenesis, vol 2. Mechanisms of tumor promotion and carcinogenesis. Raven Press, New York.
25. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
26. Sjövall, J. 1960. Bile acids in man under normal and pathological conditions: bile acids and steroids 73. Clin. Chem. Acta 5:33-41.
27. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.
28. Thatcher, D. R. 1980. The complete amino acid sequence of three alcohol dehydrogenase alleloenzymes (Adh^{N-11} , Adh^S , and Adh^{UF}) from the fruitfly *Drosophila melanogaster*. Biochem. J. 187:875-883.
29. White, B. A., A. F. Cacciapuoti, R. J. Fricke, T. R. Whitehead, E. H. Mosbach, and P. B. Hylemon. 1981. Cofactor requirements for 7 α -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.
30. White, B. A., R. L. Lipsky, R. J. Fricke, and P. B. Hylemon. 1980. Bile acid induction specificity of 7 α -dehydroxylase activity in an intestinal *Eubacterium* species. Steroids 35:103-109.
31. 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.
32. Woods, D. 1984. Oligonucleotide screening of cDNA libraries. Focus 6:1-2.
33. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.