Human apolipoprotein B: identification of cDNA clones and characterization of mRNA

Margarete Mehrabian, Verne N.Schumaker, George C.Fareed¹, Ruth West, David F.Johnson, Todd Kirchgessner, Hun-Chi Lin¹, Xingbo Wang, Yuanhong Ma, Elizabeth Mendiaz¹ and Aldons J.Lusis*

Departments of Medicine, Microbiology and Molecular Biology Institute, UCLA, Los Angeles, CA 90024, and ¹International Genetic Engineering, Inc., 1545 17th Street, Santa Monica, CA 90404, USA

Received 15 July 1985; Accepted 9 September 1985

ABSTRACT

Apolipoprotein B (apoB) is a major protein component of low density and very low density lipoproteins. Because of its large size and heterogeneity, molecular studies of apoB have been difficult, and its structure and regulation remain poorly understood. We now report the identification of human apoB cDNA clones by antibody screening of hepatoma libraries in the expression vector λ gt11. Both oligo(dT) primed and random primed libraries were constructed and screened with polyclonal antibodies to intact apoB, as well as with antibodies raised against a synthetic peptide based on the limited amino acid sequence available for apoB. The identity of the clones was unambiguously established by comparisons of the cloned cDNA sequences with apoB amino acid sequences. The clones hybridize to an exceptionally large 20 kb mRNA that is present in liver and intestine but not other tissues examined, consistent with the distribution expected from protein biosynthetic studies. The properties of the mRNA have implications for the biogenesis of the multiple apoB molecular weight forms secreted by liver and intestine.

INTRODUCTION

Lipids are transported through the circulation as lipoprotein complexes consisting of a hydrophobic core of cholesterol ester and triglyceride surrounded by phospholipids, free cholesterol and various apolipoproteins. Several classes of lipoproteins have been distinguished on the basis of density. Among these are chylomicrons and very low density lipoproteins (VLDL), which are secreted by intestine and liver, respectively, low density lipoproteins (LDL), which are derived in the circulation by lipolysis of VLDL, and high density lipoproteins (HDL), which are derived from precursor particles secreted by liver (1,2). Apolipoprotein B (apoB) is the major protein component of chylomicrons, VLDL and LDL, and it serves at least two crucial functions in lipoprotein metabolism. First, it is essential in the assembly and secretion of both VLDL and chylomicrons, since individuals with a rare genetic defect in the production of apoB, termed abetalipoproteinemia, are unable to produce these lipoproteins (3). Second, it serves as a recogni-

tion marker for the removal of LDL from the circulation by receptor-mediated uptake into a variety of cells (4). It differs from other apolipoproteins in that it is an integral component of lipoproteins; while other apolipoproteins can exchange between lipoprotein particles and are soluble in the absence of detergents, apoB is extremely hydrophobic and remains tightly associated with lipoprotein particles. Also, whereas other apolipoproteins range in size from about 6 kDa-45 kDa, apoB is among the largest peptides known with molecular weight estimates ranging from 250-550 kDa (3,5-9). The hydrophobicity and size of apoB have made its study difficult, and only recently have we succeeded in obtaining limited amino acid sequence information (10). Furthermore, the protein exhibits size heterogeneity, with at least two major forms being present in the circulation. In humans, the liver synthesizes and incorporates into VLDL a species which we estimate has a molecular weight of about 400 kDa, while intestine incorporates into chylomicrons a smaller 210 kDa species. These 400 and 210 kDa species have been referred to as B100 and B48, respectively, and are synonymous with the PI and PIII species in Elovson's nomenclature (3,7,8). The B100 but not the B48 species contains the recognition marker for uptake by the LDL receptor which has been extensively studied by Brown, Goldstein, and collaborators (3). The mechanisms leading to the production of the two species are unclear. Although most forms of abetalipoproteinemia are characterized by failure in the production of both apoB100 and apoB48, others are characterized by continued expression of apoB48 but not apoB100, raising the possibility that the two species are encoded by separate genes (3). On the other hand, peptide fingerprinting and immunological studies are consistent with their being derived from a common gene (8, 11-14).

Because of the difficulties in studying apoB at the protein level, we have sought to clone sequences encoding apoB to study questions concerning its structure and biogenesis. Also, a variety of studies have strongly implicated LDL, the major cholesterol carrier in human plasma, in the development of atherosclerosis, and accumulating evidence suggests that genetic variations affecting the structure and regulation of apoB may contribute to hypercholesterolemia and to atherosclerosis (15-19). The availability of probes for the apoB gene should allow this possibility to be examined by providing markers to trace the inheritance of the gene in family studies. Efforts to clone apoB have been hampered by inability to translate apoB mRNA either in vitro or by microinjection into frog oocytes. Therefore, we adopted the approach of antibody screening of cDNA expression libraries

using the λ gt11 vector developed by Young and Davis (20-22). We previously reported the isolation of rat apoB cDNA clones by antibody screening of a liver cDNA library (23). The clones were found to hybridize to a 20 kb mRNA present in liver. Unfortunately, since there is no amino acid sequence information for rat apoB, the identity of the clones could be established only by immunological characterization of the expressed fusion proteins.

We now report the identification and characterization of human apoB cDNA clones by antibody screening of hepatoma cDNA expression libraries. In order to unambiguously establish the identity of the clones by sequence comparisons, we specifically isolated clones corresponding to the very limited known apoB amino acid sequence (10) by screening the libraries with antibody prepared against a synthetic peptide constructed on the basis of that apoB amino acid sequence. The cloned probes identify a 20 kb mRNA in both liver and intestine, although the liver produces an apoB peptide roughly twice the size of that produced in intestine.

MATERIALS AND METHODS

Materials

Radioisotopes and enzymes were from Amersham and Bethesda Research Laboratories, respectively. Human jejunum tissue was obtained from gastric bypass surgery and was immediately placed in cold saline. Within 45 min., a 2 cm piece was cut longitudinally, washed in cold saline with shaking and frozen on dry ice. Rhesus monkey tissues were generous gifts of the Oregon Primate Center and of Dr. William Conner, University of Oregon Medical School. All tissues were stored under liquid nitrogen until use.

Cell cultures

The Hep G2 cell line (24, 25) was grown in flasks with Eagles minimum essential medium supplemented with 10% fetal calf serum. Rates of apoB synthesis were estimated by culturing cells in methionine-free Eagles medium containing 35 S-methionine for periods of 0.5-2 hrs. Homogenates were prepared in 1 mM phenylmethylsulfonyl fluoride, 7 mM EDTA (pH7), 0.1% sarcosyl, 0.2% Triton X-100. Radioactivity incorporated into apoB was determined following immunoprecipitation and SDS polyacrylamide gel electrophoresis.

Antibodies

Rabbit antiserum to intact apoB was raised in rabbits against highly purified human LDL. It was shown to be monospecific for apoB by immunoblotting with human plasma. Affinity purified antibody was prepared by passing antiserum over a column of immobilized LDL followed by elution with

0.1 M glycine, pH3.

We previously isolated two proteolytic fragments of human apoB resulting from digestion with S. aureus protease in the presence of SDS and determined their amino terminal amino acid sequences (10). A peptide corresponding to a portion of the sequence of one of these (R2-5) was synthesized by the solid phase method (26) and purchased from Peninsula Laboratories, California. The sequence of the synthetic peptide was Leu-Asp-Phe-Leu-Asn-Ile-Pro-Leu-Arg-Ile-Pro-Pro-Met-Arg. Antiserum specific for this peptide was prepared by immunizing rabbits with 200 µg keyhole limpet hemocyanin (KLH) coupled to the synthetic peptide in three injections as described by Green, et al. (27). The antipeptide serum generated in this way was able to specifically react with µg quantities of denatured apoB. This antiserum had a titer of 1:5000 when assayed for binding to 0.5 µg apoB immobilized on nitrocellulose filters (28).

RNA and DNA isolation and filter hybridization

RNA was prepared from tissues by homogenization in guanidine thiocyanate followed by centrifugation over a cesium chloride cushion (29). Poly(A) containing RNA was enriched by oligo(dT) cellulose chromatography. Genomic DNA was purified from isolated nuclei by incubation in 10 mM EDTA, pH8, 0.2% SDS, 600 µg/ml proteinase K (Sigma) at 37°C for 24 hr followed by phenol extraction and ethanol precipitation (30). For Northern analysis RNA was electrophoresed on 0.75% agarose gels in the presence of formaldehyde and hybridized to cDNA probes as previously described (32). Insert cDNA was isolated by preparative electrophoresis and labeled by oligonucleotide priming (33) with ³²P-dCTP to a specific activity of about 10⁹ cpm/µg.

Subcloning and DNA sequencing

Inserts were isolated by EcoRI restriction of lambda phage DNA and subcloned into M13 (31). Sequencing was carried out by the dideoxy termination method of Sanger et al. (34). Computer analysis of DNA sequences was performed using the programs and sequence files of Staden (35), Queen and Korn (36) and Orcutt and Dayhoff (37).

Construction and screening of human hepatoma cDNA libraries in λgt11

Two separate Hep G2 cell line cDNA libraries were constructed using either random oligonucleotide priming or oligo(dT) priming. Calf thymus DNA oligonucleotides were prepared according to Taylor et al. (38). Double stranded cDNA was prepared from poly(A⁺) RNA essentially according to the method of Gubler and Hoffman (39) in which second strand synthesis is mediated by RNase H and polymerase I. For random primed first strand

synthesis, an optimal ratio of 1:50 (wt/wt) for poly(A⁺) RNA:oligonucleotide primer was chosen on the basis of the yield and size of cDNA transcripts. Reverse transcriptase was from Life Sciences, Inc. The double stranded cDNA was methylated at EcoRI restriction sites, blunt ended with T4 DNA polymerase and ligated to EcoRI linkers (Pharmacia P-L Biochemicals) according to Young and Davis (20, 21). Excess linkers were removed from cDNA by two cycles of ethanol precipitation in 0.2M ammonium acetate followed by Sephadex G75 gel filtration. To decrease the background of phage lacking inserts, the EcoRI digested λ gt11 DNA was treated with calf intestinal alkaline phosphatase (Boehringer Mannheim Biochemicals) prior to ligation to cDNA. Ligation was performed using about 50 ng cDNA per μ g λ gt11 DNA. The ligated DNA was packaged using an extract purchased from Amersham and amplified on E. coli Y1088 (20,21). Phage were plated on E. coli strain Y1090 at a density of 1.5×10^4 phage per 150 mm plate and isopropyl thiogalactoside (IPTG) induced plaques were transferred to nitrocellulose filters (Schleicher and Schuell) (20,21,23). Filters were reacted for 90 min at room temperature with rabbit antibodies diluted 1:1000 followed by treatment with ¹²⁵I-protein A and autoradiography as described (23). Positive phage were taken through two or more rounds of plaque purification.

Characterization of fusion proteins.

Lysogens were prepared by infection of E. coli Y1089 cells (20, 21). Cell lysates, prepared as previously described (23) and suspended in 8M urea were subjected to electrophoresis on 8% SDS polyacrylamide gels. Western blots were prepared by electrophoretic transfer to nitrocellulose. Filters were incubated with the appropriate antibody overnight at room temperature, washed, and developed with ¹²⁵I-protein A followed by autoradiography, as described (23).

RESULTS

Cloning strategy

The cloning and identification of apoB sequences present a special problem because in vitro translation of apoB mRNA has been unsuccessful and because apoB is among the largest known peptides, with a molecular weight of about 400 kDa, which would correspond to a coding sequence of about 11 kb. Moreover, our previous studies of rat apoB suggest that the entire mRNA is about 20 kb in size, more than twice the size of any previously reported mRNA species. With the objective of utilizing oligonucleotide screening of libraries, we obtained partial amino acid sequences for two apoB peptides

isolated after partial proteolysis of human apoB (10). The locations of these sequences along the apoB polypeptide are unknown. However, the use of corresponding short mixed oligonucleotide probes as well as longer unique probes based on codon usage failed to identify apoB sequences in ours and other laboratories. The reasons for this may involve the relatively ambiguous nature of the nucleic acid sequences predicted from the apoB amino acid sequence and the likelihood that the mRNA region encoding the peptide sequence is distant from the poly(A) tail used for priming cDNA synthesis. Therefore, we adopted the following alternative strategy: A synthetic peptide derived from the amino terminal sequence of apoB peptide R2-5 was synthesized, coupled to keyhole limpet hemocyanin, and used to immunize rabbits. The antiserum generated in this way was able to specifically react with μg quantities of denatured apoB and also with the expected proteolytic fragment when tested by the Western immunoblot procedure (data not shown). This antiserum was then used to screen human hepatoma cDNA libraries in $\lambda\text{gt}11$. The $\lambda\text{gt}11$ vector can accommodate up to 8.3 kb of cDNA cloned into a unique EcoRI restriction site near the C-terminal region of the β -galactosidase gene. Thus, when in the correct reading frame and orientation, the inserted cDNA is expressed as an IPTG inducible fusion protein containing the sequence encoded by the insert fused to an enzymatically inactive β -galactosidase protein, allowing clones to be screened with specific antibodies (20). On the basis of the very large size of rat apoB mRNA, we did not expect to synthesize and package full-length apoB cDNA. Therefore, the cDNA synthesis was primed by random oligonucleotides (38), which should yield sequences along the entire length of the mRNA, as well as the conventional method of oligo(dT) priming, which should enrich for sequences at the 3' end of the mRNA.

Isolation of apoB cDNA clones

As a source of mRNA for library construction, we chose the human hepatoma cell line Hep G2. The cells produce and secrete apoB100 (40), and pulse labeling studies indicated that the relative rate of apoB synthesis was 0.01% or greater (data not shown). Double stranded cDNA was synthesized from poly(A⁺) selected Hep G2 RNA according to the method of Gubler (39) using either random oligonucleotides or oligo(dT) to prime synthesis. In vitro packaging of 1 μg of dephosphorylated $\lambda\text{gt}11$ DNA that had been ligated to the Hep G2 cDNA via EcoRI linkers yielded 3×10^6 and 6×10^6 plaque forming units (pfu) for the random primed and oligo(dT) primed libraries, respectively. Greater than 95% of the phage contained inserts as judged

Figure 1. Antibody competition screening of apoB clones identified with antibody to the synthetic apoB peptide. The 12 clones identified with antibody to the amino terminal sequence of apoB peptide R2-5 were examined for the ability of excess peptide to compete for antibody binding to fusion proteins as described in the text. Two clones, RP2 and A18, shown above, exhibited the strongest competition. Section A was treated with antibody to intact apoB (diluted 1:1000). Sections B and C were treated with antibody to the R2-5 peptide sequence (diluted 1:1000) in the absence or presence of 15 $\mu\text{g/ml}$ of synthetic peptide, respectively.

by the inactivation of β -galactosidase, and examination of randomly selected clones indicated that the size of the inserts ranged up to several kilobases. Approximately 2.5×10^5 phage from each of the two hepatoma cDNA expression libraries were screened separately with a monospecific polyclonal antibody to intact human apoB and with a polyclonal antibody to a synthetic apoB peptide. Of a large number of putative clones observed in the primary screens, secondary and tertiary plaque purification yielded 10 positive clones detected by antibody to intact apoB and 12 positive clones detected by antibody to the synthetic apoB peptide.

The clones identified with antibody to the synthetic apoB peptide were further screened by antibody competition studies. Under the same conditions used for the initial plaque screens, the 12 individual clones were plated at a density of about 100 pfu per 85 mm plate and lifts were taken onto IPTG soaked filters. The filters were cut into sections and treated with either: a) antibody to intact apoB; (b) antibody to the R2-5 peptide; (c) antibody to the R2-5 peptide plus a large excess of peptide to compete with the fusion

				15					30					45					60
CTG	AAA	GT•	CGT	TTT	CCA	TTA	AGG	TTA	ACA	GGG	AAG	ATA	GAC	TTC	CTG	AAT	AAC	TAT	GCA
Leu	Lys	Val	Arg	Phe	Pro	Leu	Arg	Leu	Thr	Gly	Lys	Ile	Asp	Phe	Leu	Asn	Asn	Tyr	Ala
				75					90					105					120
CTG	TTT	CTG	AGT	CCC	AGT	GCC	CAG	CAA	GCA	AGT	TGG	CAA	GTA	AGT	GCT	AGG	TTC	AAT	CAG
Leu	Phe	Leu	Ser	Pro	Ser	Ala	Gln	Gln	Ala	Ser	Trp	Gln	Val	Ser	Ala	Arg	Phe	Asn	Gln
				135					150					165					180
TAT	AAG	TAC	AAC	CAA	AAT	TTC	TCT	GCT	GGA	AAC	AAC	GAG	AAC	ATT	ATG	GAG	GCC	CAT	GTA
Tyr	Lys	Tyr	Asn	Gln	Asn	Phe	Ser	Ala	Gly	Asn	Asn	Glu	Asn	Ile	Met	Glu	<u>Ala</u>	<u>His</u>	<u>Val</u>
				195					210					225					240
GGA	ATA	AAT	GGA	GAA	GCA	AAT	CTG	GAT	TTC	TTA	AAC	ATT	CCT	TTA	ACA	ATT	CCT	GAA	ATG
Gly	Ile	Asn	Gly	Glu	Ala	Asn	Leu	Asp	Phe	Leu	Asn	Ile	Pro	Leu	<u>Thr</u>	<u>Ile</u>	<u>Pro</u>	<u>Glu</u>	<u>Met</u>
				255					270					285					300
CGT	CTA	CCT	TAC	ACA	ATA	ATC	ACA	ACT	CCT	CCA	CTG	AAA	GAT	TTC	TCT	CTA	TGG	GAA	AAA
<u>Arg</u>	<u>Leu</u>	<u>Pro</u>	<u>Tyr</u>	<u>Thr</u>	<u>Ile</u>	<u>Ile</u>	<u>Thr</u>	<u>Thr</u>	<u>Pro</u>	<u>Pro</u>	<u>Leu</u>	<u>Lys</u>	<u>Asp</u>	<u>Phe</u>	<u>Ser</u>	<u>Leu</u>	<u>Trp</u>	<u>Glu</u>	<u>Lys</u>
ACA	GGC	TTG	AAG																
Thr	Gly	Leu	Lys																

Peptide sequence:

Ala Leu Val Gly Ile Asn Gly Glu Ala Asn Leu Asp
Phe Leu Asn Ile Pro Leu Arg Ile Pro Pro Met Arg

Figure 2. Nucleic acid sequence and corresponding amino acid sequence of apoB cDNA clone RP2. Clone RP2 was sequenced in both directions by the Sanger dideoxy chain termination method. Shown are the nucleic acid sequence (not including the EcoRI linker sequences) and the amino acid sequence predicted from a single open reading frame continuous with lac Z. The region of the cDNA sequence corresponding to the amino terminal sequence of an apoB proteolytic fragment is underlined. The sequence of the proteolytic fragment is shown below for comparison, and the amino acids which differ from those derived from the cDNA sequence are indicated by underlining.

protein epitope for antibody. Of the 12 clones examined, the signals of two (RP2 and A18) were strongly competed by the presence of excess peptide (Fig. 1) and these were selected for further characterization as discussed below.

Characterization of apoB cDNA and fusion proteins.

In addition to the two clones identified with antibody to the synthetic peptide (RP2 and A18), two clones identified with antibody to intact apoB (A7h and A6c) were selected for further characterization on the basis of intensity of signal obtained with antibody screening and insert size. Three of these, A7h, A6c and A18, containing cDNA inserts of about 600 bp, 3500 bp and 550 bp respectively, were selected from the oligo(dT) primed library, while RP2, with an insert size of about 300 bp, was selected from the random primed library.

Clone RP2 was subcloned into M13 phage and sequenced by the dideoxy chain termination method. It contains a single open reading frame, continuous with lacZ, and the predicted amino acid sequence includes the entire

amino acid sequence we previously obtained for apoB peptide R2-5 as well as flanking sequences (Fig. 2). Three differences were observed between the sequence deduced from the cDNA and the amino terminal sequence determined for apoB peptide R2-5 (Fig. 2). A reexamination of the amino acid sequencing data indicated that these differences are compatible with errors in amino acid sequence determination, although they could also result from allelic variation. The sequence is preceeded by a glutamic acid residue, which is expected since the peptide was derived by partial proteolysis with *S. aureus* V8 protease (10), which cleaves on the carboxyl side of glutamic acid and aspartic acid residues. A search of the Biomedical Research Foundation and Los Alamos data bases (36, 37) showed no extensive homology for the nucleic acid or amino acid sequence of clone RP2. The predicted amino acid sequence did not show the preponderance of amphipathic α -helical structure present in other apolipoproteins (data not shown). As expected, clone RP2 crosshybridized with clone A18, and both clones hybridized to the same mRNA species in liver and intestine (see below). Although clone A18 was identified in the oligo(dT) primed library, it does not appear to represent the 3' region of apoB mRNA. Preliminary sequencing of A18 indicates that it begins with a tract of 14 A residues followed by an open reading frame lacking a polyadenylation sequence (data not shown). Thus, it presumably was derived by priming from an internal A-rich region of the mRNA.

The clones identified with antibody to intact apoB, A6c and A7h, failed to crosshybridize with clones A18 or RP2, but they crosshybridized with each other and also hybridized to common restriction fragments of human genomic DNA (data not shown), indicating that they overlap. Their identity was further examined by immunological characterization of the resulting fusion proteins. Lysogens of each clone were made in Y1089 cells, and freeze-thaw lysates of heat induced lysogens were prepared. These, along with a non-recombinant λ gt11 lysate, were analysed by immunoblotting after sodium dodecyl sulfate gel electrophoresis (Fig. 3). When probed with anti-galactosidase, the nonrecombinant λ gt11 gave the expected 116 kDa β -galactosidase band while the recombinant clones consistently showed multiple bands (Fig. 3). Of these, only the largest species, corresponding to 135 and 120 kDa, also reacted with affinity purified antibody to apoB (Fig. 3B). Thus, the lower molecular weight species presumably represent degraded fusion protein fragments which have lost their apoB portions through proteolysis, a common observation with fusion proteins produced in *E. coli*.

Figure 3. Immunoblotting of fusion proteins produced by apoB clones A7h and A6c. Panel A was reacted with antibody to *E. coli* β -galactosidase (Cappell, diluted 1:2000) and panel B was reacted with affinity purified antibody to intact apoB (diluted 1:1000). The lanes contained 50 μ g protein of induced lysate obtained from lysogens of a λ gt11 control, clone A6c, or clone A7h. The molecular weights of the reacting bands are shown in kDa. Panel B also contained 1 μ g of human LDL, and the position of the apoB100 band is indicated.

Synthesis of both fusion proteins and β -galactosidase was dependent upon induction with IPTG (data not shown). The apparent molecular weight of the A7h fusion protein, 135 kDa, was about that expected for the maximal 20 kDa apoB peptide which could be encoded by a 600 bp insert. The fusion protein produced by A6c, however, was much smaller than that predicted for a coding region of 3500 bp. This could result from the presence of 3' nontranslated sequences in A6c, which would be consistent with the presence of considerable noncoding RNA in the message (see below), or to rapid degradation of a larger fusion protein. The identity of the cDNA clones was further confirmed by showing that they hybridize with the same uncommonly large mRNA identified by clones RP2 and A18 (see below).

Genomic apoB sequences

As judged by Southern blotting analysis of human genomic DNA, the gene

Figure 4. Expression of apoB mRNA in Hep G2 cells, monkey liver and intestine, and rat liver. Poly(A)⁺ RNA (10 µg/lane) was denatured and electrophoresed through 0.75% agarose formaldehyde gels, blotted onto nitrocellulose, hybridized with a ³²P-labeled cDNA insert from clone A7h and autoradiographed overnight at -70°C with an intensifying screen. One lane containing rat liver RNA (shown at the right) was washed under conditions of reduced stringency (0.5 x SSC at 48°C) to increase the strength of the signal. Previously identified rat apoB cDNA clones hybridize to a mRNA species of identical size (about 20 kb) present in rat liver (data not shown). Comparable results were obtained when cDNA isolated from apoB clones A6c, RP2 and A18 were hybridized to RNA from monkey liver. The positions of the major 20 kb apoB band, 18S and 28S ribosomal RNA, and Hind III digested lambda DNA size markers are indicated.

for apoB appears to be present as a single copy in the genome. For example, when human DNA was cut with Hind III, RP2 cDNA hybridized to a single band 2.8 kb in size (data not shown).

Expression of apoB mRNA.

The four cDNA clones characterized above were used as probes to examine the size, abundance and tissue distribution of mRNA for apoB. On Northern blots of Hep G2 or primate liver RNA all clones hybridized to a single major mRNA species corresponding to about 20 kb (Fig. 4). Diffuse hybridization in lower molecular weight regions was observed, but since this

varied between mRNA preparations, it probably represents partial degradation of this very large mRNA species. Clone A7h also crosshybridized with a mRNA species of identical size present in rat liver (Fig. 4). Assuming that apoB100 has a molecular weight of about 400 kDa, which would correspond to a coding region of about 11 kb, this mRNA is nearly twice the size necessary to encode human apoB100.

All four cDNA clones also hybridized to a mRNA species of about 20 kb present in monkey intestine (Fig. 4). Densitometric scanning of mRNA preparations showing minimal degradation (Fig. 4) indicated that the ratio of liver to intestinal hybridization signal was about 1.4. Human jejunum total RNA or poly(A) containing RNA also contained a single major hybridizing band about 20 kb in size (data not shown). The size of the intestinal RNA was not discernibly different from that present in liver as judged by mobility on agarose gels.

In addition to the major 20 kb band, some preparations of hepatic and intestinal mRNA contained discreet lower molecular weight hybridizing bands. Most striking was a band approximately 9 kb in size that was present in two separate preparations of monkey intestinal RNA and one preparation of human intestinal RNA (Fig. 4). Since the primary apoB form secreted by intestine, apoB48, has a molecular weight of about 210 kDa, corresponding to a coding region of about 6 kb, this smaller 9 kb RNA could conceivably encode apoB48. The relative amounts of the lower molecular weight species, however, varied widely between different RNA preparations; for example, the ratio of the 20 kb signal to the 9 kb signal was approximately 4:1 in one preparation of monkey intestinal RNA (Fig. 4B) whereas the 9 kb species predominated in a separate preparation (data not shown). Thus, we suspect that the lower molecular weight RNA species represent degradation products of the 20 kb species rather than, for example, products resulting from alternative splicing. The apoB mRNA appears to be extremely susceptible to degradation, as all of our RNA preparations exhibited some hybridization in lower molecular weight regions. This probably results from the extremely large size of the mRNA, since filter hybridization of the same RNA preparations with probes for smaller mRNA species, such as apoA-I and apoE, failed to reveal significant degradation (data not shown).

No specific hybridization to the apoB probes was observed with RNA isolated from monkey tissues which do not secrete apoB, including spleen, muscle, kidney and adrenal gland (data not shown).

On Northern blots containing monkey liver mRNA, hybridization with

human apoB probe A7h yielded a signal about 25% as intense as that observed after hybridization with a human apoE cDNA probe of similar size and specific radioactivity (data not shown). Since apoE constitutes roughly 0.1% of the mRNA in primate liver, these results indicate that apoB is a very abundant mRNA species in liver and intestine.

DISCUSSION

Several cDNA clones for human apoB have been isolated by antibody screening of human hepatoma cDNA libraries in the expression vector λ gt11. In order to unambiguously establish the identity of the clones, we utilized an antibody specific for a short region of the 400 kDa protein for which sequence information was available. Sequencing of the cDNA thus identified showed that it encoded a 24 amino acid sequence identical in all but three amino acids with that previously derived for a proteolytic fragment of apoB (10). Other, nonoverlapping cDNA clones were isolated by screening the libraries with antibody to intact apoB. Their identity was established by immunological characterization of the resulting fusion proteins and by the fact that they hybridize to the same uncommonly large mRNA species as the clones containing the known apoB sequence. Moreover, the size of primate mRNA identified with these probes was identical to that which we previously observed for rat liver apoB mRNA. The cDNA clones were partially characterized and used as probes to examine apoB expression at the level of mRNA. The tissue distribution of apoB mRNA, with abundant expression in liver and intestine but no detectable mRNA in other tissues examined, is consistent with that expected on the basis of protein production and secretion. However, the very large size of the mRNA and the presence of mRNA species of identical size in liver and intestine, raise important questions concerning the biogenesis of apoB.

Eight major apolipoproteins are associated with the various circulating lipoproteins in mammals. Seven of these apolipoproteins (A-I, A-II, A-IV, C-I, C-II, C-III and E) have previously been cloned and sequenced, and they appear to be members of a dispersed gene family derived from a common ancestral sequence. These apolipoproteins consist largely of amphipathic α -helices which are thought to be involved in lipid binding, and they are internally repetitive in terms of both amino acid and nucleic acid sequences (1, 41-44). In contrast to these seven previously characterized apolipoproteins, the limited sequence information we have thus far derived for rat (23) and human apoB does not show extensive homology with other apolipopro-

teins, nor does the corresponding amino acid sequence exhibit a preponderance of amphipathic α -helical structure. We have used the apoB probes to map the apoB gene to the pter-q323 region of human chromosome 2 (45); thus, the gene for apoB is unlinked to the gene clusters on chromosomes 1, 11 and 19 that encode the other apolipoproteins. No other significant homologies at the amino acid or nucleic acid level were observed with other proteins or genes. However, since the sequenced region represents only a few percent of the apoB coding region, questions of internal and external homologies will remain until larger regions of the sequence are determined. Clearly, this will be a large undertaking, but it should now be feasible by techniques such as primer extension from the presently derived clones.

Due to the intractability of the apoB peptides, their molecular properties remain an enigma. We have estimated the sizes of apoB100 and apoB48 as 400 and 210 kDa, respectively, on the basis of their hydrodynamic properties in guanidine hydrochloride (9), but other estimates for apoB100 range from about 250 kDa to 550 kDa (3). Also, a variety of groups have reported lower molecular weight peptides associated with apoB and have suggested that the high molecular weight forms represent aggregates of these (46-52). Our present finding of a 20 kb mRNA, more than sufficient to encode a protein of 400 kDa, is compatible with synthesis of full-length (400 kDa) apoB peptides, as are the studies of Lane et al. (53) on the elongation kinetics of nascent apoB peptides in cultured chicken hepatocytes. Although these results do not definitively rule out the possibility of a small apoB translation product, we suspect that the smaller peptides observed may represent proteolytic products, since apoB is known to be very susceptible to proteases (54).

Our finding of a major apoB mRNA species of about 20 kb in both liver and intestine was unexpected since the liver incorporates into VLDL a peptide of about 400 kDa (apoB100) while intestine incorporates into chylomicrons a peptide of about 210 kDa (apoB48) (3, 7-9, 55-58). Further studies will be required to determine the explanation for this observation. One possibility is that the liver and intestinal mRNA species, although not discernibly different in size, are not identical as a result of alternative splicing and give rise to separate translation products. Such alternative splicing involving internal exons has been observed in other systems (59-61). Our finding that apoB cDNA clone RP2 hybridizes to a single restriction fragment in Hind III digests of genomic DNA, yet reacts with both liver and intestinal mRNA, strongly supports the concept of a single apoB structural

gene which encodes both apoB100 and apoB48.

Since apoB is required for the assembly and secretion of VLDL and chylomicrons as well as serving as a recognition marker for cellular uptake of LDL (3-5), structural and regulatory variations of apoB could alter the structure and metabolism of these lipoproteins. Indeed, studies of patients with familial combined hyperlipidemia and familial hypertriglyceridemia have revealed altered production and catabolism of apoB (15-18). Also Rapasz and coworkers have carried out alloantigenic studies of apoB in swine which suggest strongly that certain structural alleles of apoB are associated with hyperlipidemias and aortic intimal lipidosis (19). The identification of probes for apoB now makes feasible the identification of restriction fragment length polymorphisms as well as apoB alleles to answer questions concerning the role of apoB in hyperlipidemias and atherosclerosis.

ACKNOWLEDGEMENTS

This work was supported by a grant from the U.S. Public Health Service (HL27481), and was done in part during the tenure of an Established Investigatorship (to A. Lusis) from the American Heart Association with funds contributed in part by the American Heart Association, Los Angeles Affiliate. M. Mehrabian and T. Kirchgeßner were supported in part by Atherosclerosis Research Training Grant HL07-386. We are also grateful to Chad Ben-Avram, Dr. Renee LeBoeuf, Dr. John Shively, and Dr. John Elovson for advice during the course of this work.

*To whom correspondence should be addressed

REFERENCES

1. Mahley, R.W., Innerarity, T.L., Rall, S.C., Jr. and Weisgraber, K.H. (1984) *J. Lipid Res.* 25, 1277-1294.
2. Havel, R.J., Goldstein, J.L. and Brown, M.S. (1980) *Metabolic Control and Disease*, Bandy, P.K. and Rosenberg, L.E. Eds. 8th ed. pp. 393-494. W.B. Saunders, Philadelphia.
3. Kane, J.P. *Annu. Rev. Physiol.* (1983) 45, 637-650.
4. Brown, M.S., Kovanen, P.T. and Goldstein, J.L. (1981) *Science*, 212, 628-635.
5. Smith, R., Dawson, J.R. and Tanford, C. (1977) *J. Biol. Chem.* 247, 3376-3381.
6. Steele, J.C.H., Jr. and Reynolds, J.A. (1979) *J. Biol. Chem.* 254, 1639-1643.
7. Kane, J.P., Hardman, D.A. and Paulus, H.E. (1980) *Proc. Nat. Acad. Sci. USA* 77, 2465-2469.
8. Elovson, J., Huang, Y.O., Baker, N., and Kannan, R. (1981) *Proc. Nat. Acad. Sci. USA* 77, 157-161.
9. Elovson, J., Jacobs, J.C., Schumaker, V.N. and Puppione, D.L. (1985)

- Biochemistry 24, 1569-1578.
10. LeBoeuf, R.C., Miller, C., Shively, J.E., Schumaker, V.N., Balla, M.A. and Lusis, A.J. (1984) FEBS Lett 170, 105-108.
11. Tsuo, B.P., Curtiss, L.K. and Edgington, T.S. (1982) J. Biol. Chem. 257, 15222-15228.
12. Tikkanen, M.J., Parger, R., Pflieger, B., Gonen, B., Davie, J.M. and Schonfeld, G. (1982) J. Lipid Res. 23, 1032-1038.
13. Marcel, Y.L., Hogue, M., Theolis, R. and Milne, R.M. (1982) J. Biol. Chem. 257, 13165-13168.
14. Maloy, M.J., Kane, J.P., Hardman, D.A., Hamilton, R.L. and Dalal, K. (1981) J. Clin. Invest. 67, 1441-1450.
15. Teng, B., Thompson, G.R., Sniderman, A.D., Forte, T.M., Krauss, I.M. and Kwiterovich, P.A., Jr. (1983) Proc. Nat. Acad. Sci. USA 80, 6662-6666.
16. Brunzell, J.D., Sniderman, A.D., Albers, J.J. and Kwiterovich, P.O., Jr. (1984) Arteriosclerosis 4, 79-83.
17. Vega, G.L., Beltz, W.F., Grundy, S.M. (1985) J. Lipid Res. 26, 115-126.
18. Fisher, W.R. (1983) Metabolism 32, 282-291.
19. Rapasz, J., and Hasler-Rapasz, J. (1984) Atherosclerosis and Cardio-vascular Disease, Lenzi, F. and Descovich, G.C. Eds., pp. 99-108. Editrice Compositori, Bologna.
20. Young, R.A. and Davis, R.W. (1983) Proc. Nat. Acad. Sci. USA 80, 1194-1198.
21. Young, R.A. and Davis, R.W. (1983) Science 222, 778-782.
22. Huynh, T.V., Young, R.A. and Davis, R.W. (1985) DNA Cloning Techniques: A Practical Approach, Glover, D. Ed., Vol. I, pp. 49-78. IRL Press, Oxford.
23. Lusis, A.J., West, R.W., Mehrabian, M., Reuben, M.A., LeBoeuf, R.C., Kaptein, J.S., Johnson, D.F., Schumaker, V.N., Yuhasz, M.P., Schotz, M.C. and Elovson, J. (1985) Proc. Nat. Acad. Sci. USA 82, 4597-4601.
24. Knowles, B.B., Howe, C.C. and Aden, D.P. (1980) Science 209, 497-499.
25. Zannis, V.I., Breslow, J.L., San Giacomo, T.R., Aden, D.P. and Knowles, B.B. (1981) Biochemistry 20, 7089-7096.
26. Merrifield, J.M. (1962) J. Amer. Chem. Soc. 85, 2149.
27. Green, N., Alexander, H., Olsen, A., Alexander, S., Shinnick, T., Sutcliffe, J.G. and Lerner, R.A. (1982) Cell 28, 477-487.
28. John, R., Schiebler, W. and Greengard, R. (1984) Proc. Nat. Acad. Sci. USA 81, 1684-1687.
29. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
30. Foder, E.J.B. and Doty, P. (1979) Biochem. Biophys. Res. Commun. 77, 1478-1485.
31. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.
32. Goldberg, D.A. (1980) Proc. Nat. Acad. Sci USA 77, 5794-5798.
33. Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
34. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA. 74, 5463-5467.
35. Staden, R. (1982) Nucl. Acids Res. 10, 2951-2961.
36. Lipman, D.J. and Wilbur, W.J. (1982) RBFA-Programs for Biosequence Similarity Analysis (Mathematical Research Branch, National Institutes of Arthritis, Diabetes, and Digestive and Kidney Diseases, NIH, Bethesda, MD).
37. Orcutt, B.C. and Dayhoff, M.D. (1982) PIR Reports NUC-0484 and PRO-0484 (National Biomedical Research Foundation, Washington DC).
38. Taylor, J.M., Iilmensee, R. and Summers, J. (1976) Biochem. Biophys. Acta 422, 324-330.
39. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263-269.

40. Rash, J.M., Rothblat, G.H. and Sparkes, C.E. (1981) *Biochem. Biophys. Acta* 666, 294-298.
41. Barker, W.C. and Dayhoff, M.O. (1977) *Comp. Biochem. Physiol.* B57, 309-315.
42. Karathanasis, S.K., Zannis, V.I. and Breslow, J.L. (1983) *Proc. Nat. Acad. Sci USA* 81, 5021-5025.
43. Boguski, M.S., Elshourbagy, N., Taylor, J.M. and Gordon, J.I. (1984) *Proc. Nat. Acad. Sci USA* 81, 5021-5025.
44. Segrest, J.P., Jackson, R.L., Morrisett, J.D. and Gotto, A.M. (1974) *FEBS Lett.* 38, 247-253.
45. Mehrabian, M., Sparkes, R.S., Mohandas, T., Schumaker, V.N., Heinzmann, C., Zollman, S. and Lysis, A.J. 8th International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.* (in press).
46. Chen, C.H. and Aladjem, F. (1978) *Biochem. Med.* 19, 178-187.
47. Huang, S.S. and Lee, D.M. (1979) *Biochem. Biophys. Acta* 77, 424-441.
48. Soccoro, L. and Camejo, G. (1979) *J. Lipid Res.* 20, 631-638.
49. Soccoro, L., Lopez, F., Lopez, A., Camejo, G. (1982) *J. Lipid Res.* 23, 1283-1291.
50. Olofsson, S.O., Bostrom, K., Svanberg, U. and Bondjers, G. (1980) *Biochemistry* 19, 1059-1064.
51. Olofsson, S.O., Elias, P., Bostrom, K., Lundholm, K., Norfeldt, P.I., Wiklund, O., Fager, G. and Bondjers, G. (1983) *FEBS Lett.* 156, 63-66.
52. Bostrom, K., Wettesten, M., Wiklund, O., Bondjers, G., Lundholm, K., Elias, P., Norfeldt, P.I. and Olofsson, S.O. (1984) *Eur. J. Biochem.* 143, 101-107.
53. Janero, D. and Lane, M.D. (1983) *J. Biol. Chem.* 258, 14496-14504.
54. Steele, J.C.H., Jr. (1979) *Thromb. Res.* 15, 573-579.
55. Krishnaiah, K.V., Walker, L.F., Borensztajn, G., Schonfeld, G. and Getz, G.S. (1980) *Proc. Nat. Acad. Sci. USA* 77, 3806-3810.
56. Sparks, C.E. and Marsh, J.B. (1981) *J. Lipid Res.* 22, 519-527.
57. Wu, A.L. and Windmueller, H.G. (1981) *J. Biol. Chem.* 256, 3615-3618.
58. Bell-Quint, J., Forte, T. and Graham, P. (1981) *Biochem. Biophys. Res. Commun.* 99, 700-706.
59. Broome, S. and Gilbert, W. (1985) *Cell* 40, 537-546.
60. Schwarzbauer, J.E., Tamkun, J.W., Lamischka, I.R. and Hynes, R.O. (1983) *Cell* 35, 421-431.
61. Kornblihtt, A.R., Vibe-Pedersen, K. and Barelle, F.E. (1984) *EMBO J.* 3, 221-226.