Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 8340–8344, December 1985
Biochemistry

Human apolipoprotein B-100: Cloning, analysis of liver mRNA, and assignment of the gene to chromosome 2

(low density lipoproteins/expression vector/low density lipoprotein receptor)

SIMON W. LAW*, KARL J. LACKNER*, ASHOK V. HOSPATTANKAR*, J. MICHAEL ANCHORS*, ALAN Y. SAKAGUCHI†, SUSAN L. NAYLOR†, AND H. BRYAN BREWER, JR.*

*Molecular Disease Branch, Building 10, Room 7N117, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20205; and †Department of Cellular and Structural Biology, University of Texas Health Science Center, San Antonio, TX 78284

Communicated by Martha Vaughan, August 8, 1985

ABSTRACT **Human apolipoprotein B-100 (apo B-100) is the major apolipoprotein of low density lipoproteins and the principal ligand for interaction with the low density lipoprotein receptor. The human apo B-100 gene has been inserted into a λgt-11 expression vector, and the apo B-100 cDNA clones have been identified by screening with a monospecific apo B-100 antiserum, by screening with synthetic oligonucleotides based on the amino acid sequence of peptides isolated from apo B-100, and by immunoblot analysis of the expressed protein with a monoclonal antibody to apo B-100. The complete nucleotide and derived-amino acid sequence of a 1.7-kilobase cDNA clone of apo B-100 was determined. The 560-amino acid residues of apo B-100 contain no unique linear or repeating sequences of amino acids. The computer-predicted conformation of the apo B-100 protein contains segments of helical structure; however, a large portion of the protein is organized into β-structure. The β-structure may be important in lipid–apo B-100 interactions in low density lipoprotein and may contribute to the insolubility of delipidated apo B-100 in aqueous buffers. RNA blot hybridization analysis of liver mRNA utilizing a *Nco* I/*Hin*dIII apo B-100 cDNA probe revealed that the apo B-100 mRNA is 15–18 kilobases long, which is of sufficient size to code for a 250–387 kDa apolipoprotein, the proposed molecular size of delipidated plasma apo B-100. The gene for human apo B-100 has been localized to chromosome 2 by filter hybridization of human–mouse somatic cell hybrids utilizing a 400-base-pair *Nco* I/*Hin*dIII apo B-100 cDNA probe. This location is in contrast to the low density lipoprotein receptor that has been localized to chromosome 19. The cloning of human apo B-100 has provided new insights into the structure and physicochemical properties of apo B-100 and will facilitate studies on the factors modulating apo B-100 biosynthesis and the expression of the apo B-100 gene in patients with dyslipoproteinemias.**

Apolipoprotein B is the principal apolipoprotein on chylomicrons, very low density lipoproteins, intermediate density lipoproteins, and low density lipoproteins (LDL) (1–4). In human plasma, apolipoprotein B is heterogenous and exists in two major forms, apolipoprotein B-48 and B-100 (4). Apolipoprotein B-48 is synthesized exclusively by the intestine, and secreted on triglyceride-rich chylomicrons that undergo hydrolysis by lipoprotein lipase resulting in the formation of chylomicron remnants that are ultimately removed by the liver through an apolipoprotein E-mediated receptor process (4–6). Apolipoprotein B-100 (apo B-100) is synthesized by the liver (4, 7) and is secreted on triglyceride-rich very low density lipoprotein (1–4, 6). Liver very low density lipoproteins also undergo hydrolysis by lipoprotein lipase and are converted to intermediate density lipoproteins and, finally, to LDL (1–4, 6). Apo B-100, the major apolipoprotein on LDL, is the protein determinant that interacts with the high-affinity LDL receptor and initiates the process of receptor mediated endocytosis that culminates in LDL catabolism (8).

The structure and physicochemical properties of apo B-100 have been extensively studied for nearly a decade. The analysis of apo B-100 has been difficult because delipidated apo B-100 is insoluble in aqueous solution, and generally aggregates in buffers containing $NaDodSO_4$, urea, and guanidine hydrochloride (1–4). Structural studies of apo B-100 have met with limited success due to the insolubility of the native protein and its peptide fragments. The molecular size of apo B-100 remains controversial, and values ranging from 8 to 400 kDa have been reported (1–4, 9). The heterogeneity in apo B-100 molecular size has been attributed to the propensity of delipidated apo B-100 to aggregate and to the reported sensitivity of apo B-100 to protease cleavage (3, 4, 10). The structure of delipidated apo B-100 and LDL–apo B-100 has been explored with monoclonal apo B-100 antibodies. Monoclonal antibodies have been characterized that will block the interaction of LDL with the LDL receptor (11, 12), and studies with a Fab fragment have been interpreted as indicating that there is a single apo B-100 per LDL particle (11). Analysis of the epitopes present on apo B-100 and apolipoprotein B-48 has suggested that apo B-100 contains a structural domain which is not present in apolipoprotein B-48 (11).

To determine the primary amino acid sequence and to elucidate the biosynthesis and processing of apo B-100 we have isolated and characterized the cDNA of human liver apo B-100 from a λgt-11 cDNA expression library.

MATERIALS AND METHODS

Apo B-100 Antisera. Apo B-100 for use as an immunogen in rabbits was isolated from delipidated LDL by preparative 3.5% polyacrylamide gel electrophoresis. The apo B-100 antiserum (GR-22) reacted with apo B-100 and apolipoprotein B-48 by immunoblot analysis but did not bind to apolipoproteins A-I, A-II, A-IV, C-I, C-II, C-III, D, E, and H. A monoclonal antibody (AB-B-5) to apo B-100 was purchased from Radioimmunoassay (Scarborough, ON, Canada). Antibodies present in the apo B-100 antisera that were directed against *Escherichia coli* or phage antigens were adsorbed with lysate from a lysogen *E. coli* BNN97 by incubating for 48 hr at 25°C (13).

Synthetic Oligonucleotides. Synthetic oligonucleotides SN-R3-1 (3′-TTRTACCCNTA (R,T)TA-5′; 1:24) and SN-MDB-18 (3′-CTYATRCTYCTYCTYCC-5′; 1:32) based on the sequences of apo B-100 peptides R3-1 and MDB-18 were synthesized by the modified phosphite triester method (OCS Laboratory, Denton, TX). The 24-amino acid peptide R3-1 was reported by LeBoeuf *et al.* (14), and the second peptide

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LDL, low density lipoproteins; apo B-100, apolipoprotein B-100; kb, kilobase(s); bp, base pair(s).

MDB-18 (Asp-Phe-Ser-Ala-Glu-Tyr-Glu-Glu-Asp-Gly-Lys-Tyr-Glu-Gly-Leu-Gln-Glu-Trp) has been isolated and sequenced in our laboratory.

Screening of λgt-11 Human Liver cDNA Expression Library. The human λgt11-cDNA expression library was the gift of W. Droham and G. Ricca, Meloy Laboratories (Springfield, VA). *E. coli* Y1090 cultures (0.1 ml, 3×10^8 bacteria) grown in LB medium and supplemented with ampicillin (100 μg/ml), 0.4% maltose, and 10 mM $MgSO_4$ were infected with 2×10^4 plaque-forming units of the λgt11-cDNA phages (13) at 37°C for 20 min. After incubation, the cultures were mixed with 7 ml of 0.7% agarose (FMC, Rockland, ME) in LB medium containing 10 mM $MgSO_4$, were plated onto 150-mm diameter LB agar plates, and were incubated at 42°C for 5 hr. Following incubation, a 140-mm diameter nitrocellulose filter (Schleicher & Schuell or Millipore) saturated with 10 mM isopropylthio β-D-galactoside (Bethesda Research Laboratories) was placed over each plate and was incubated at 37°C for an additional 16 hr. Plaque filters were then incubated for two 20-min periods with TBS [0.02 M Tris·HCl (pH 7.4), 0.5 M NaCl] containing 3% (wt/vol) gelatin and 0.01% NaN_3 at 25°C. The filters were transferred to a TBS solution containing a 1:250 dilution of the apo B-100 antiserum (GR-22) in TBS and 1.0% (wt/vol) gelatin. After a 48-hr incubation at 25°C, filters were washed briefly in distilled water and for two 20-min periods in TBS at 25°C. Filters were then transferred individually to a TBS solution containing a 1:2500 dilution of horseradish-peroxidase-conjugated goat anti-rabbit IgG (Bio-Rad) and were incubated at 25°C for 2 hr with shaking. Following this incubation, filters were washed as described above for the first antibody incubation, and the immunocomplex was detected by utilizing the chromogenic substrate 4-chloro-1-naphthol (Kirkegaard and Perry Laboratory, Gaithersburg, MD). Clones were also screened by hybridization with radiolabeled synthetic oligonucleotides SN-R3-1 and SN-MDB-18. Filter hybridizations with radiolabeled synthetic oligonucleotides were performed as reported (15).

Isolation of apo B-100 cDNA Clones. λgt-11 recombinant clones which were positive with the apo B-100 antibodies were plaque purified to homogeneity. Plate lysate stocks were prepared by infection of *E. coli* Y1088 (13). Recombinant phage DNA was isolated by polyethylene glycol precipitation followed by cesium chloride density gradient centrifugation (16). Approximately 1 mg of DNA was isolated from a 500-ml culture.

Restriction Enzyme Mapping and Sequence Analysis of Apo B-100 cDNA. DNA fragments were generated by digestion with restriction endonucleases (Bethesda Research Laboratories, Boehringer Mannheim, or New England Biolabs) using the conditions reported (15). DNA sequences were determined by the chemical cleavage procedure of Maxam and Gilbert (17).

Analysis of Lysogen for Fusion Protein Containing Sequences of Apo B-100. *E. coli* Y1089 were grown in LB medium containing 10 mM $MgSO_4$, 0.4% maltose, and ampicillin (100 μg/ml) for 18 hr at 32°C. Cell cultures were diluted 1:5 with 0.01 M Tris·HCl (pH 7.5)/0.1 M NaCl/0.01 M $MgSO_4$/0.05% gelatin and infected with λgt-11 apo B-100 clones at 37°C for 20 min. Aliquots of infected cultures were spread on agar plates and recombinant lysogens were identified by their ability to grow at 32°C but not at 42°C (13).

For identification of the fusion protein, recombinant lysogens were grown in LB medium at 32°C for 18 hr. Cell cultures were diluted 1:100 with fresh LB medium and incubated at 32°C until the absorbency at 550 nm reached 0.5. The lysogens were then induced to synthesize the hybrid protein by addition of 0.2 mM isopropylthio β-D-galactoside and incubation for 3 hr at 37°C. The fusion protein was analyzed by $NaDodSO_4$/PAGE (7.5% acrylamide, 0.1% $NaDodSO_4$) of the cell lysate, and the gels were either stained with Coomassie blue or immunoblotted with an apo B-100 monoclonal antibody as reported (18).

RNA Blot Hybridization. Total, poly(A)$^+$, and poly(A)$^-$ RNA were separated on a 1% agarose gel in the presence of 6% (vol/vol) formaldehyde at 3.5 volts per cm for 6 hr and transferred to nitrocellulose filters (16, 19). The filters were hybridized with a nick-translated 400-bp apo B-100 *Nco* I/*Hin*dIII cDNA probe (10^7 cpm) in 10 ml of hybridization solution containing 5× standard saline citrate (SCC; 1× = 0.15 M NaCl/0.015 M sodium citrate, pH 7), 5× Denhardt's solution (1× = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), 0.1% $NaDodSO_4$, 20 mM Tris·HCl (pH 7.4), denatured salmon sperm DNA (100 μg/ml), and 50% (vol/vol) formamide at 42°C for 20 hr. After hybridization, filters were washed in 1× SSC containing 0.1% $NaDodSO_4$ for five 5-min periods at 25°C, for two 30-min periods at 65°C, and finally for one 30-min period at 65°C in 0.5× SSC followed by autoradiography.

Human–Mouse Somatic Cell Hybrids and Filter Hybridization. Human–mouse somatic cell hybrids were constructed by fusing mouse LM/TK (thymidine kinase deficient) or RAG (hypoxanthine phosphoribosyltransferase deficient) fibroblasts with human fibroblasts or leukocytes (20–22). Proliferating cell hybrids were selected in hypoxanthine/aminopterin/thymidine medium, and the cell hybrids characterized for chromosomal content by trypsin–Giemsa banding (21). Isolated DNAs (10 μg) were digested with 30 units of *Eco*RI (Bethesda Research Laboratories) for 3 hr at 37°C. The restriction fragments were separated by 0.7% agarose gel electrophoresis, transferred to nitrocellulose sheets (Schleicher & Schuell), and hybridized as reported (22) utilizing a 400-bp *Nco* I/*Hin*dIII apo B-100 cDNA probe.

RESULTS

Identification of Apo B-100 cDNA Clones. The λgt11 human liver cDNA library was screened with a monospecific antibody to human apo B-100 (GR-22) and with synthetic

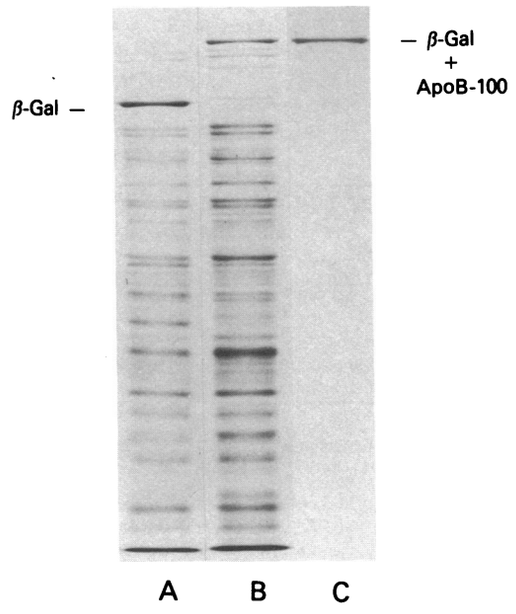

FIG. 1. Electrophoretogram and immunoblot of cell lysates separated by $NaDodSO_4$/PAGE. Lane A, BNN97 lysate stained with Coomassie blue; lanes B and C, lysates of a recombinant lysogen containing the fusion protein stained with Coomassie blue and an immunoblot utilizing an apo B-100 monoclonal antibody, respectively. The electrophoretic positions of β-galactosidase (118 kDa) and the fusion protein (185 kDa) containing β-galactosidase and apo B-100 are indicated.

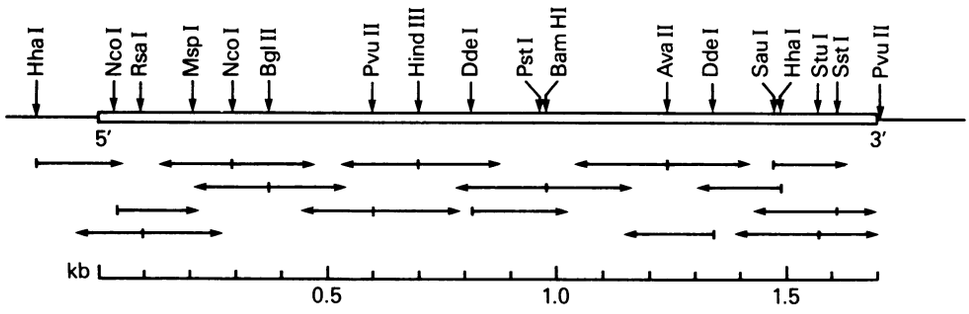

FIG. 2. Restriction map of 15λgtB-100. The open bar represents the cloned cDNA which is orientated 5' to 3' from left to right. Note that the left phage arm is shown on the right. The length and direction of the sequenced restriction fragments are illustrated by the arrows.

oligonucleotides SN-R3-1 and SN-MDB-18 that were based on the corresponding peptides present in apo B-100. A total of 70,000 plaques were screened, and 5 positive clones containing 1.0- to 2.3-kilobase (kb) inserts were identified by antibody screening. Two different clones were positive to each of the synthetic oligonucleotides.

To confirm the specificity of the screening of the library utilizing monospecific apo B-100 antiserum and synthetic

```
ACC TAT TTG ATT CTG CGG GTC ATT GGA AAT ATG GGC CAA ACC ATG GAG CAG TTA ACT CCA GAA CTC AAG TCT TCA ATC CTG AAA TGT GTC CAA AGT ACA AAG CCA  105
Thr Tyr Leu Ile Leu Arg Val Ile Gly Asn Met Gly Gln Thr Met Glu Gln Leu Thr Pro Glu Leu Lys Ser Ser Ile Leu Lys Cys Val Gln Ser Thr Lys Pro

TCA CTG ATG ATC CAG AAA GCT GCC ATC CAG GCT CTG CGG AAA ATG GAG CCT AAA GAC AAG GAC CAG GAG GTT CTT CTT CAG ACT TTC CTT GAT GAT GCT TCT CCG  210
Ser Leu Met Ile Gln Lys Ala Ala Ile Gln Ala Leu Arg Lys Met Glu Pro Lys Asp Lys Asp Gln Glu Val Leu Leu Gln Thr Phe Leu Asp Asp Ala Ser Pro

GGA GAT AAG CGA CTG GCT GCC TAT CTT ATG TTG ATG AGG AGT CCT TCA CAG GCA GAT ATT AAC AAA ATT GTC CAA ATT CTA CCA TGG GAA CAG AAT GAG CAA GTG  315
Gly Asp Lys Arg Leu Ala Ala Tyr Leu Met Leu Met Arg Ser Pro Ser Gln Ala Asp Ile Asn Lys Ile Val Gln Ile Leu Pro Trp Glu Gln Asn Glu Gln Val

AAG AAC TTT GTG GCT TCC CAT ATT GCC AAT ATC TTG AAC TCA GAA GAA TTG GAT ATC CAA GAT CTG AAA AAG TTA GTG AAA GAA GCT CTG AAA GAA TCT CAA CTT  420
Lys Asn Phe Val Ala Ser His Ile Ala Asn Ile Leu Asn Ser Glu Glu Leu Asp Ile Gln Asp Leu Lys Lys Leu Val Lys Glu Ala Leu Lys Glu Ser Gln Leu

CCA ACT GTC ATG GAC TTC AGA AAA TTC TCT CGG AAC TAT CAA CTC TAC AAA TCT GTT TCT CTT CCA TCA CTT GAC CCA GCC TCA GCC AAA ATA GAA GGG AAT CTT  525
Pro Thr Val Met Asp Phe Arg Lys Phe Ser Arg Asn Tyr Gln Leu Tyr Lys Ser Val Ser Leu Pro Ser Leu Asp Pro Ala Ser Ala Lys Ile Glu Gly Asn Leu

ATA TTT GAT CCA AAT AAC TAC CTT CCT AAA GAA AGC ATG CTG AAA ACT ACC CTC ACT GCC TTT GGA TTT GCT TCA GCT GAC CTC ATC GAG ATT GGC TTG GAA GGA  630
Ile Phe Asp Pro Asn Asn Tyr Leu Pro Lys Glu Ser Met Leu Lys Thr Thr Leu Thr Ala Phe Gly Phe Ala Ser Ala Asp Leu Ile Glu Ile Gly Leu Glu Gly

AAA GGC TTT GAG CCA ACA TTG GAA GCT CTT TTT GGG AAG CAA GGA TTT TTC CCA GAC AGT GTC AAC AAA GCT TTG TAC TGG GTT AAT GGT CAA GTT CCT GAT GGT  735
Lys Gly Phe Glu Pro Thr Leu Glu Ala Leu Phe Gly Lys Gln Gly Phe Phe Pro Asp Ser Val Asn Lys Ala Leu Tyr Trp Val Asn Gly Gln Val Pro Asp Gly

GTC TCT AAG GTC TTA GTG GAC CAC TTT GGC TAT ACC AAA GAT GAT AAA CAT GAG CAG GAT ATG GTA AAT GGA ATA ATG CTC AGT GTT GAG AAG CTG ATT AAA GAT  840
Val Ser Lys Val Leu Val Asp His Phe Gly Tyr Thr Lys Asp Asp Lys His Glu Gln Asp Met Val Asn Gly Ile Met Leu Ser Val Glu Lys Leu Ile Lys Asp

TTG AAA TCC AAA GAA GTC CCG GAA GCC AGA GCC TAC CTC CGC ATC TTG GGA GAG GAG CTT GGT TTT GCC AGT CTC CAT GAC CTC CAG CTC CTG GGA AAG CTG CTT  945
Leu Lys Ser Lys Glu Val Pro Glu Ala Arg Ala Tyr Leu Arg Ile Leu Gly Glu Glu Leu Gly Phe Ala Ser Leu His Asp Leu Gln Leu Leu Gly Lys Leu Leu

CTG ATG GGT GCC CGC ACT CTG CAG GGG ATC CCC CAG ATG ATT GGA GAG GTC ATC AGG AAG GGC TCA AAG AAT GAC TTT TTT CTT CAC TAC ATC TTC ATG GAG AAT 1050
Leu Met Gly Ala Arg Thr Leu Gln Gly Ile Pro Gln Met Ile Gly Glu Val Ile Arg Lys Gly Ser Lys Asn Asp Phe Phe Leu His Tyr Ile Phe Met Glu Asn

GCC TTT GAA CTC CCC ACT GGA GCT GGA TTA CAG TTG CAA ATA TCT TCA TCT GGA GTC ATT GCT CCC GGA GCC AAG GCT GGA GTA AAA CTG GAA GTA GCC AAC ATG 1155
Ala Phe Glu Leu Pro Thr Gly Ala Gly Leu Gln Leu Gln Ile Ser Ser Ser Gly Val Ile Ala Pro Gly Ala Lys Ala Gly Val Lys Leu Glu Val Ala Asn Met

CAG GCT GAA CTG GTG GCA AAA CCC TCC GTG TCT GTG GAG TTT GTG ACA AAT ATG GGC ATC ATC ATT CCG GAC TTC GCT AGG AGT GGG GTC CAG ATG AAC ACC AAC 1260
Gln Ala Glu Leu Val Ala Lys Pro Ser Val Ser Val Glu Phe Val Thr Asn Met Gly Ile Ile Ile Pro Asp Phe Ala Arg Ser Gly Val Gln Met Asn Thr Asn

TTC TTC CAC GAG TCG GGT CTG GAG GCT CAT GTT GCC CTA AAA GCT GGG AAG CTG AAG TTT ATC ATT CCT TCC CCA AAG AGA CCA GTC AAG CTG CTC AGT GGA GGC 1365
Phe Phe His Glu Ser Gly Leu Glu Ala His Val Ala Leu Lys Ala Gly Lys Leu Lys Phe Ile Ile Pro Ser Pro Lys Arg Pro Val Lys Leu Leu Ser Gly Gly

AAC ACA TTA CAT TTG GTC TCT ACC ACC AAA ACG GAG GTG ATC CCA CCT CTC ATT GAG AAC AGG CAG TCC TGG TCA GTT TGC AAG CAA GTC TTT CCT GGC CTG AAT 1470
Asn Thr Leu His Leu Val Ser Thr Thr Lys Thr Glu Val Ile Pro Pro Leu Ile Glu Asn Arg Gln Ser Trp Ser Val Cys Lys Gln Val Phe Pro Gly Leu Asn

TAC TGC ACC TCA GGC GCT TAC TCC AAC GCC AGC TCC ACA GAC TCC GCC TCC TAC TAT CCG CTG ACC GGG GAC ACC AGA TTA GAG CTG GAA CTG AGG CCT ACA GGA 1575
Tyr Cys Thr Ser Gly Ala Tyr Ser Asn Ala Ser Ser Thr Asp Ser Ala Ser Tyr Tyr Pro Leu Thr Gly Asp Thr Arg Leu Glu Leu Glu Leu Arg Pro Thr Gly

GAG ATT GAG CAG TAT TCT GTC AGC GCA ACC TAT GAG CTC CAG AGA GAG GAC AGA GCC TTG GTG GAT ACC CTG AAG TTT GTA ACT CAA GCA GAA GGC GCG AAG CAG 1680
Glu Ile Glu Gln Tyr Ser Val Ser Ala Thr Tyr Glu Leu Gln Arg Glu Asp Arg Ala Leu Val Asp Thr Leu Lys Phe Val Thr Gln Ala Glu Gly Ala Lys Gln
```

FIG. 3. Nucleic acid sequence and derived amino acid sequence of 15λgtB-100. The nucleic acids are numbered beginning at the first triplet of the cloned apo B-100 sequence. The first triplet of the open reading frame of apo B-100 is an ACC codon for threonine. The sequence of peptide R3-1 is underlined.

oligonucleotides, a clone designated 24λgtB-100 with a 1.5-kb insert was induced with isopropylthio β-D-galactoside to produce fusion protein for analysis. The lysate of the recombinant lysogen was analyzed by $NaDodSO_4$/PAGE and immunoblotted with the monospecific apo B-100 antiserum and a monoclonal antibody to apo B-100. The fusion protein with an apparent molecular size of 185 kDa reacted with the monoclonal antibody to apo B-100 (Fig. 1). The molecular size of this protein is consistent with the size of the cDNA insert. Thus, the screening procedure that utilized monospecific apo B-100 antiserum and synthetic oligonucleotides effectively identified apo B-100 cDNA clones. A clone designated 15λgtB-100, which was also positive when screened by the SN-R3-1 oligonucleotide probe, was selected for restriction enzyme mapping and sequence analysis.

Restriction Endonuclease Map and Sequence Analysis of 15λgtB-100 Clone. A restriction endonuclease cleavage map of the 15λgtB-100 clone is illustrated in Fig. 2. Based on this map, restriction enzyme cleavage sites were selected for the isolation of DNA fragments for use in sequence analysis. A total of 24 fragments were used to determine the sequence of the 1.7-kb apo B-100 insert. The restriction endonuclease map and the sequence strategy to determine the nucleotide sequence of 15λgtB-100 are shown in Fig. 2.

The complete nucleotide sequence of the apo B-100 insert in the 15λgtB-100 clone was determined by the Maxam and Gilbert procedure (17) (Fig. 3). The correct reading frame of the nucleotide sequence of 15λgtB-100 was ascertained by the identification of the nucleotide and derived-amino acid sequence of the R3-1 peptide present in apo B-100 and known to be present in the clone by SN-R3-1 hybridization analysis. This reading frame was also confirmed by sequence analysis of a second clone, 24λgtB-100 (Fig. 1), which was in phase with the *lacZ* gene reading frame. An identical reading frame of over 600 nucleotides was determined in 24λgtB-100 confirming that the correct reading frame has been selected for 15λgtB-100.

The nucleic acid and derived-amino acid sequence of 15λgtB-100 is illustrated in Fig. 3. The amino acid sequence of peptide R3-1 derived from the cloned cDNA sequence of 15λgtB-100 is identical to the reported sequence of R3-1 (14) except at residue 21 where the cDNA predicts an aspartic acid residue instead of a lysine.

RNA Blot Hybridization Analysis of Human Apo B-100 mRNA. Apo B-100 mRNA in human liver was evaluated by RNA blot hybridization analysis employing a 400-bp *Nco* I/*Hin*dIII cDNA probe. cDNA probes for apolipoprotein A-I and human serum albumin containing 866 bp and 700 bp, respectively, were prepared as reported (15, 23). All probes were radiolabeled by nick-translation (15) and total, poly(A)$^+$, and poly(A)$^-$ human liver RNAs were analyzed (Fig. 4). The probes for apolipoprotein A-I, human serum albumin, and apo B-100 hybridized to total and poly(A)$^+$ RNA but not to poly(A)$^-$ RNA. An apo B-100 mRNA of identical size to the liver apo B-100 mRNA was present in human intestine. The apo B-100 mRNA was 15–18 kb that is considerably larger than the 1.0 kb and 2.1 kb of apolipoprotein A-I and human serum albumin mRNA, respectively.

Chromosomal Localization of the Apo B-100 Gene. High molecular weight DNAs were extracted from a panel of 11 human–mouse somatic cell hybrids and analyzed by the Southern blotting technique. The *Nco* I/*Hin*dIII apo B-100 cDNA probe hybridized to a single 1.9-kb *Eco*RI fragment of human DNA. Under the same hybridization conditions, a 6.0-kb *Eco*RI fragment of mouse DNA was also positive. No difficulty in chromosome assignment was encountered since the *Eco*RI fragments of human and mouse DNA were significantly different in size. The results of the complete analysis of the cell hybrids are shown in Table 1. The gene for apo B-100 segregated with chromosome 2.

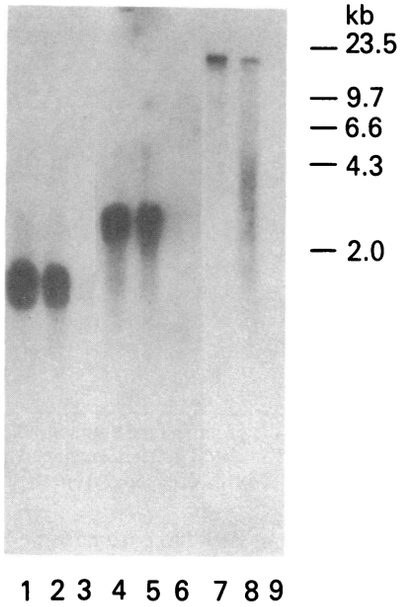

FIG. 4. Autoradiogram of the RNA blot hybridization analysis of human liver RNA. Lanes 1–3, 4–6, and 7–9 were hybridized to cDNA probes to apolipoprotein A-I, human serum albumin, and apo B-100, respectively. Lanes 1, 4, and 7 contain poly(A)$^+$ RNA (1 μg); lanes 2, 5, and 8 contain total RNA (10 μg); and lanes 3, 6, and 9 contain poly(A)$^-$ RNA (1 μg). Molecular size markers are λ*Hin*dIII fragments.

DISCUSSION

In this report we have described the cloning of human apo B-100 utilizing a λgt11 expression vector. Positive clones were identified by screening with monospecific apo B-100 antiserum, synthetic oligonucleotides, and by immunoblot analysis of the expressed protein with a monoclonal antibody to apo B-100. The complete nucleic acid and derived amino acid sequence of a 1.7-kb apo B-100 clone was determined. An amino acid sequence identical to peptide R3-1 (14) except for a substitution of a lysine for an aspartic acid residue was present in the protein sequence encoded by the cloned apo B-100. Inspection of the 560-amino acid sequence of the apo B-100 protein revealed no long stretches of hydrophobic residues or unusual sequences. Computer analysis of the nucleic acid and derived-amino acid sequence of apo B-100 did not reveal any internal repeats or sequence homology to any other known human plasma apolipoprotein. Evaluation of the predicted secondary structure of the apo B-100 protein revealed the presence of helices; however, a significant portion of the protein was β-structure. The large degree of β-structure in apo B-100 is unusual and not observed in other human apolipoproteins (4, 25, 26). The β-structure may play an important role in the lipid–protein interaction of apo B-100 with LDL because, in contrast to other apolipoproteins, apo B-100 does not dissociate from the lipoprotein particle (3), and delipidated apo B-100 is insoluble in aqueous buffers (3, 4, 10).

RNA blot hybridization analysis revealed that the liver apo B-100 mRNA is 15–18 kb. The mRNA of apo B-100 is of sufficient size to code for a protein of 250–386 kDa. These results are consistent with the 250 and 386 kDa molecular size of purified apo B-100 determined by hydrodynamic techniques by Steele and Reynolds (27) and Elovson *et al.* (9), respectively. In addition, the coding of apo B-100 as a single large apolipoprotein rather than several cross-linked small proteins is consistent with the biosynthesis of a single-chain high molecular weight B apolipoprotein in cell culture studies reported by Siuta-Mangano *et al.* (28).

Table 1. Segregation of the apo B-100 gene with human chromosomes in human–mouse somatic cell hybrids

Hybrid	Apo B-100	Presence of chromosome																						Translocation
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	
ICL-15	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	−	−	+	+	−	−	
TSL-2	+	−	+	−	−	+	+	−	−	−	+	−	+	−	−	−	−	−	+	−	+	+	−	+	17/3
ATR-13	+	+	+	+	+	−	+	+	+	−	+	−	+	−	+	+	−	+	+	−	−	−	−	−	5/X
WIL-5	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	+	−	+	
WIL-6	+	−	+	−	+	+	+	+	+	−	+	+	−	−	+	−	−	+	−	+	+	+	−	+	
WIL-2	−	−	−	−	−	−	−	−	+	−	−	−	+	−	−	−	−	+	−	−	−	+	−	+	
WIL-8X	−	−	−	+	+	+	+	+	−	−	+	+	+	−	+	−	−	+	+	+	+	+	−	+	
WIL-7	+	−	+	+	−	+	+	−	+	−	+	+	−	+	+	−	−	+	+	−	−	+	−	+	
NSL-9	−	−	−	−	−	+	−	−	+	−	+	−	+	+	+	+	+	+	−	−	+	+	+	−	17/9
WIL-14	−	−	−	+	−	−	−	−	+	−	+	−	−	−	+	−	−	+	−	−	−	−	−	+	
XER-11	−	+	−	+	+	−	+	+	+	−	+	−	+	−	−	+	+	+	+	+	+	+	+	−	11/X X/11

A + designation under the chromosome number indicates that the chromosome is present in 10% of the metaphases examined. The presence (+) or absence (−) of human apo B-100 in a given cell hybrid is indicated in the column designated apo B-100. The translocation chromosomes located in the designated hybrids originated from the human parental cells used for cell fusions (24).

The localization of apo B-100 to chromosome 2 is the first apolipoprotein to be identified on this chromosome and is not in synteny with the LDL receptor that is located on chromosome 19 (29). The other major ligand for the LDL receptor, apolipoprotein E, is, however, present on chromosome 19 (30).

The cloning of human apo B-100, the determination of the size of human liver apo B-100 mRNA, and the elucidation of the nucleic acid sequence of a portion of the gene for apo B-100 have provided new insights into the structure and biochemical properties of apo B-100. The availability of cDNA clones of apo B-100 has also permitted the evaluation of the apolipoprotein B-48 mRNA in the intestine. An apo B mRNA of 7–9 kb in addition to the 15- to 18-kb apo B-100 mRNA was identified in the human intestine (data not shown) consistent with the synthesis of apolipoprotein B-48 by the intestine (4). These studies, therefore, indicate that both apo B-100 and B-48 are synthesized by the human intestine. The cloning of the apo B-100 gene will now facilitate the study of the coordinate control of the expression of apo B-100 and the LDL receptor and the characterization of the gene for apo B-100 in abetalipoproteinemia and other dyslipoproteinemias.

The authors thank Mr. Thomas Manes, Ms. Nancy Lee, Ms. Martha Meng, and Ms. Cheri Bishop for their expert technical assistance; Ms. Nadine Mitchell for the preparation of the manuscript; Dr. L. S. Yeh of the National Biomedical Research Foundation and Dr. M. Kanehisa of the National Institutes of Health for computer analysis of the nucleic acid and peptide sequences; Dr. Y. Yamada of the Laboratory of Development Biology and Anomalies for advice on preparation of the recombinant lysogens. This work was supported in part by National Cancer Institute Grant CA39186 to A.Y.S.

1. Jackson, R. L., Morrisett, J. D. & Gotto, A. M, Jr. (1976) *Physiol. Rev.* **56,** 259–316.
2. Scanu, A. M. & Teng, T. (1979) in *The Biochemistry of Atherosclerosis*, ed. Scanu, A. M. (Dekker, New York), pp. 107–122.
3. Osborne, J. C., Jr., & Brewer, H. B., Jr. (1977) *Adv. Protein Chem.* **31,** 253–337.
4. Kane, J. P. (1983) *Annu. Rev. Physiol.* **45,** 637–650.
5. Mahley, R. W. & Innerarity, T. L. (1983) *Biochim. Biophys. Acta* **737,** 197–222.
6. Brewer, H. B., Jr., Zech, L. A., Gregg, R. E., Schwartz, D. & Schaefer, E. J. (1983) *Ann. Intern. Med.* **98,** 623–640.
7. Edge, S. B., Hoeg, J. M., Schneider, P. D. & Brewer, H. B., Jr. (1983) *Arteriosclerosis* **3,** 499–500.
8. Goldstein, J. L. & Brown, M. S. (1977) *Annu. Rev. Biochem.* **46,** 897–930.
9. Elovson, J., Jacobs, J. C., Schumaker, V. N. & Puppione, D. L. (1985) *Biochemistry* **24,** 1569–1578.
10. Cardin, A. D., Witt, K. R., Chao, J., Margolis, H. S., Donaldson, V. H. & Jackson, R. L. (1984) *J. Biol. Chem.* **259,** 8522–8528.
11. Marcel, Y. L., Hogue, M., Theolis, R., Jr. & Milne, R. W. (1982) *J. Biol. Chem.* **257,** 13165–13168.
12. Tikkanen, M. J., Dargar, R., Pfleger, B., Gonen, B., Davie, J. M. & Schonfeld, G. (1982) *J. Lipid Res.* **23,** 1032–1038.
13. Young, R. A. & Davis, R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1194–1198.
14. LeBoeuf, R. C., Miller, C., Shivety, J. E., Schumaker, V. N., Balla, M. A. & Lusis, A. J. (1984) *FEBS Lett.* **170,** 105–108.
15. Law, S. W. & Brewer, H. B., Jr. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 66–70.
16. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
17. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
18. Sprecher, D. L., Taam, L. & Brewer, H. B., Jr. (1984) *Clin. Chem.* **30,** 2084–2092.
19. Lehrach, H., Diamond, D., Wozney, J. M., and Boedtker, H. (1977) *Biochemistry* **16,** 4743–4750.
20. Szybalski, W. S., Szybalski, E. H. & Ragni, G. (1962) Monograph 7 (Natl. Cancer Inst., Bethesda, MD), pp. 75–89.
21. Seabright, M. (1971) *Lancet* **ii,** 971.
22. Law, S. W., Gray, G., Brewer, H. B., Jr., Sakaguchi, A. Y. & Naylor, S. L. (1984) *Biochem. Biophys. Res. Commun.* **118,** 934–942.
23. Dugaiczyk, A., Law, S. W. & Dennison, O. E. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 71–75.
24. Sakaguchi, A. Y., Naylor, S. R., Shows, T. B., Toole, J. J., McCoy, M. & Weinberg, R. A. (1983) *Science* **219,** 1081–1082.
25. Assmann, G. & Brewer, H. B., Jr. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 989–993.
26. Jackson, R. L., Morrisett, J. D., Gotto, A. M., Jr., & Segrest, J. D. (1975) *Mol. Cell. Biochem.* **6,** 43–50.
27. Steele, J. C. H., Jr., & Reynolds, J. A. (1979) *J. Biol. Chem.* **254,** 1639–1643.
28. Siuta-Mangano, P., Howard, S. C., Lennarz, W. J. & Lane, M. D. (1982) *J. Biol. Chem.* **257,** 4292–4300.
29. Francke, U., Brown, M. S. & Goldstein, J. L. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2826–2830.
30. Olaisen, B., Teisberg, P. & Gedde-Dahl, T. J. (1982) *Hum. Genet.* **62,** 233–236.