RESEARCH ARTICLE

Human Apolipoprotein B: Structure of Carboxyl-Terminal Domains, Sites of Gene Expression, and Chromosomal Localization

Timothy J. Knott, Stanley C. Rall, Jr., Thomas L. Innerarity
Shellie F. Jacobson, Mickey S. Urdea, Beatriz Levy-Wilson
Lyn M. Powell, Richard J. Pease, Roger Eddy, Hiroshi Nakai
Mary Byers, Linda M. Priestley, Elaine Robertson
Leslie B. Rall, Christer Betsholtz, Thomas B. Shows
Robert W. Mahley, James Scott

Abstract. *Apolipoprotein (apo-) B is the ligand responsible for the receptor-mediated catabolism of low density lipoproteins, the principal cholesterol-transporting lipoproteins in plasma. The primary structure of the carboxyl-terminal 30 percent (1455 amino acids) of human apo-B (apo-B100) has been deduced from the nucleotide sequence of complementary DNA. Portions of the protein structure that may relate to its receptor binding function and lipid binding properties have been identified. The apo-B100 messenger RNA is about 19 kilobases in length. The apo-B100 gene is expressed primarily in liver and, to a lesser extent, in small intestine, but in no other tissues. The gene for apo-B100 is located in the p24 region (near the tip of the short arm) of chromosome 2.*

Lipoproteins are the primary carriers of plasma cholesterol and triglycerides and have a specific set of protein constituents, including apolipoproteins A-I, A-II, A-IV, B, C-I, C-II, C-III, and E. The largest of these proteins, apolipoprotein (apo-) B, is an obligatory constituent of chylomicrons, very low density lipoproteins (VLDL), and low density lipoproteins (LDL). Two antigenically related forms of apo-B circulate in plasma as constituents of these lipoproteins. The apo-B100 is synthesized in the liver and is essential for the packaging of triglyceride and cholesteryl ester into VLDL within the cisternae of the endoplasmic reticulum. The apo-B48 is produced by the intestine and has a similar structural role in the formation of chylomicrons (*1*, *2*).

Apolipoprotein B100 is secreted into the plasma by the liver in the form of VLDL. The triglycerides of these lipoproteins are hydrolyzed by the action of lipoprotein lipase, converting the VLDL to cholesterol-enriched LDL. The LDL, possessing apo-B100 as their exclusive apolipoprotein constituent, are the principal cholesterol transporting lipoproteins in human plasma (>60 percent of the plasma cholesterol is present in LDL). Epidemiological studies have established a positive correlation between the amount of LDL cholesterol in the plasma and accelerated coronary artery heart disease (atherosclerosis). Plasma levels of LDL are controlled by the rate of synthesis of VLDL and by catabolism of these lipoproteins by the apo-B,E(LDL) receptor in the liver (*1*).

Apolipoprotein B100 is the protein component responsible for the cellular recognition and catabolism of LDL via the LDL receptor pathway. The apo-B,E(LDL) receptor recognizes two apolipoprotein ligands, apo-B100 and apo-E, but not apo-B48 (*2*). Studies on selective chemical modification of specific amino acid residues of these ligands have given further insight into the nature of the chemical interaction of apo-B100 and apo-E with the lipoprotein receptors. When either lysine or arginine residues are chemically modified, the binding of both apo-B100– and apo-E–containing lipoproteins to the apo-B,E(LDL) receptor is abolished (*3*). These studies and previous work that has helped to define the region of apo-E responsible for receptor binding (*2*) suggest that the apo-B receptor binding domain should be a region enriched in lysine and arginine residues.

Despite the abundance of apo-B in plasma and the importance of the apo-B,E(LDL) receptor pathway in determining the concentration of plasma cholesterol, few details are known about the structure or biosynthesis of apo-B. The insolubility of apo-B, its tendency to aggregate, and its susceptibility to proteases have impeded progress in determining its structure. When the lipids are removed, apo-B100 is insoluble in aqueous buffers, except in the presence of high concentrations of detergents; even if it remains soluble, it is probably aggregated. The apo-B in LDL or delipidated apo-B is very sensitive to cleavage by proteases and to cation-catalyzed oxidation. Thus, the molecular size of apo-B100 has not been clearly established. A number of reports have indicated that apo-B100 is composed of multiple subunits of less than 100 kD (*4*). However, most investigators now agree that apo-B is most likely a single, large polypeptide between 350 kD and 550 kD (*1*, *2*). Because of the apparent large size and extreme intractability of apo-B, molecular biology techniques are more promising than protein chemistry methods for solving the problem of the structure of this protein. We now report the nucleotide sequence of complementary DNA (cDNA) clones encoding the carboxyl-terminal 30 percent of human liver apo-B (apo-B100).

Isolation and nucleotide sequence of apo-B100 cDNA clones. Because of the apparent large size of the apo-B100 polypeptide (>350 kD), the poly(A)-primed cDNA libraries that we used in our study would not be expected to contain 5′ messenger RNA (mRNA) sequences coding for the amino-terminus of the protein. Therefore, we sought to obtain protein sequence data from fragments of

T. J. Knott, L. M. Powell, R. J. Pease, L. M. Priestley, E. Robertson, and J. Scott are in Molecular Medicine at the Clinical Research Centre, Watford Road, Harrow, Middlesex, United Kingdom; S. C. Rall, Jr., T. L. Innerarity, S. F. Jacobson, B. Levy-Wilson, and R. W. Mahley are at the Gladstone Foundation Laboratories, Cardiovascular Research Institute, Departments of Medicine and Pathology, University of California, San Francisco 94140; M. S. Urdea and L. B. Rall are at Chiron Corporation, 4560 Horton Street, Emeryville, California 94608; R. Eddy, H. Nakai, M. Byers, and T. B. Shows are at the Roswell Park Memorial Institute, Department of Human Genetics, Buffalo, New York 14263; C. Betsholtz is in the Department of Pathology, University Hospital, Uppsala, Sweden.

apo-B100 near the COOH-terminus of the protein so that oligonucleotide probes could be constructed and used for screening cDNA libraries. Apolipoprotein B100 of human plasma LDL was digested with thrombin, and the major fragments were isolated. Previously, Cardin *et al.* (*5*) demonstrated that thrombin digestion yields four fragments of apo-B100. They suggested the order of the thrombolytic fragments along the length of the apo-B molecule (Fig. 1), and these fragments have been designated T1, T2, T3, and T4. We have confirmed the order of these peptides by determining the NH_2-terminal sequence of apo-B100 and of each of the four fragments.

The most COOH-terminal thrombolytic peptide, designated T2, represents approximately 30 percent of the entire apo-B structure (Fig. 1). The NH_2-terminus of the T2 peptide yielded a single sequence as follows: Ala-Val-Ser-Met-Pro-Ser-Phe-Ser-Ile-Leu-Gly-Ser-Asp-Val-X-Val-Pro-Ser-Tyr-Thr-Leu-Ile-Leu-Pro-Ser-Leu-Glu-Leu-Pro. Residue X was tentatively identified as Arg, but could not be confirmed in subsequent analyses. Therefore, the oligonucleotide probe synthesized for use in screening the cDNA libraries was based only on the first 14 amino acid residues of T2 (*6*).

Six of 5000 clones from an adult human liver cDNA library (*7*) hybridized with the ^{32}P-labeled oligonucleotide probe. Each of the six clones was purified, ^{32}P-labeled, and hybridized to dot blots of total RNA prepared from tissues of an adult female rhesus monkey (*8*). One clone, p61D7, hybridized to an mRNA that was abundant in liver and, to a lesser extent, in small intestine (Fig. 2, left). No hybridization to mRNA of colon, spleen, pancreas, kidney, lung, pituitary, heart, or brain could be detected. Clone p61D7, containing an insert of approximately 1500 base pairs, was also used to screen a second adult human liver cDNA library prepared from mRNA fractionated by density gradient centrifugation (>28*S*) (*9*). A series of overlapping clones was isolated and used to determine the size of the putative apo-B100 mRNA. As demonstrated by Northern blotting (Fig. 2, right), the clones hybridized to a 19-kilobase (kb) mRNA in the human hepatoblastoma cell line, designated HepG2, and in the human proximal small intestine. In addition, we examined an autopsy specimen of adult human liver, and it also contained a hybridizable 19-kb mRNA. The size of the mRNA was consistent with its coding for an extremely large polypeptide. Furthermore, the occurrence of the apo-B mRNA principally within the liver is consistent with the liver being the main site of synthesis of apo-B100.

Overlapping cDNA clones, corresponding to more than 5 kb of 3′ cDNA sequence, were used to determine the double-stranded nucleotide sequence of the 3′ end of the putative apo-B mRNA (Fig. 3). The combined nucleotide sequence of these clones revealed one open reading frame of 4365 nucleotides, encoding 1455 amino acids of the COOH-terminus, and a 3′ untranslated region of more than 260 nucleotides following the TAA termination codon. The nucleotide sequence and derived amino acid sequence of the coding portion are presented in Fig. 4.

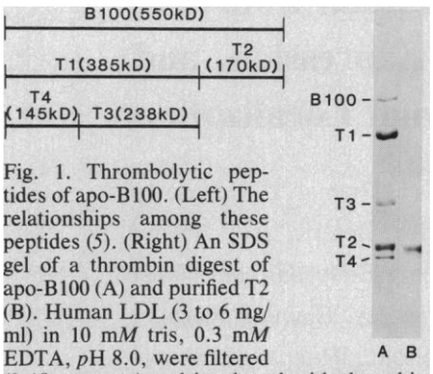


Fig. 1. Thrombolytic peptides of apo-B100. (Left) The relationships among these peptides (*5*). (Right) An SDS gel of a thrombin digest of apo-B100 (A) and purified T2 (B). Human LDL (3 to 6 mg/ml) in 10 m*M* tris, 0.3 m*M* EDTA, *p*H 8.0, were filtered (0.45-μm pore) and incubated with thrombin (enzyme:substrate = 1:100, by weight) for 18 hours at 23°C. Thrombin was then removed by gel filtration chromatography (Sephadex G-50, 2.2 by 90 cm) in 0.15*M* NaCl, 0.3 m*M* EDTA, *p*H 8.2, at 23°C. The digested LDL (ThLDL) from the G-50 column were concentrated by Amicon filtration to 20 ml and dialyzed for 18 hours at 4°C against 50 m*M* Na_2CO_3, 50 m*M* NaCl, 0.3 m*M* EDTA, *p*H 8.6, and filtered (0.45-μm pore). After the addition of glutathione (final concentration 0.05 percent by weight), solid sodium deoxycholate was added very slowly over a 2- to 3-hour period to the ThLDL (final weight ratio 55:1). After each small addition, the sodium deoxycholate was gently dissolved. The delipidated ThLDL were subjected to gel filtration chromatography at 23°C on a Sephadex G-100 column (2.2 by 60 cm) equilibrated with 10 m*M* sodium deoxycholate, 50 m*M* NaCl, 50 m*M* Na_2CO_3, 0.3 m*M* EDTA, *p*H 8.6, to separate the protein from the lipids. The protein fraction was concentrated to about 10 ml on Amicon YM30 membranes. A portion representing 5 to 10 percent of the delipidated ThLDL was dansylated by adding 15 μl of a 10 percent dansyl chloride solution (made fresh in acetone). Excess reagent was immediately removed by passing the treated protein through a small column of Sephadex G-25 equilibrated with 2 percent SDS, 0.01*M* sodium phosphate, *p*H 7.0. The effluent containing the dansylated protein was detected with ultraviolet light. The dansylated ThLDL were added to the remainder of the ThLDL and dialyzed against 0.1 percent SDS, 2.5 m*M* tris-glycine, *p*H 8.3. The thrombolytic fragments of apo-B100 were separated on SDS-polyacrylamide gradient (3 to 7 percent) slab gels (3.0 mm thick). No more than 5 mg of protein could be placed on each gel. When the electrophoresis was completed, ultraviolet light (long wavelength) was used to visualize the dansylated protein bands, which were sliced from the gel. The gel slices were then subjected to electroelution for 18 hours at 400 volts. The elution buffer was 0.1 percent SDS, 2.5 m*M* tris, 19 m*M* glycine. The eluted thrombolytic fragments were extensively dialyzed against 5 m*M* NH_4HCO_3 and 0.1 percent SDS at 4°C and lyophilized. From four preparations of LDL, an average of 82, 56, and 6 percent of the total LDL protein was recovered in ThLDL, delipidated ThLDL, and thrombolytic fragment T2, respectively.

Amino acid sequence of the COOH-terminus of apo-B100. Proof that this derived amino acid sequence represents the COOH-terminus of apo-B100 was obtained by comparing the sequence to protein sequence data derived from various fragments of apo-B100. Although the nucleotide sequence matched with only 32 of the 42 bases of the oligonucleotide probe used to screen the cDNA library (*6*), the derived amino acid sequence in this region corresponded to the 29 NH_2-terminal residues of thrombolytic fragment T2 and revealed that the unidentified residue X was indeed Arg (Fig. 4, residues 169 to 197). The NH_2-terminal sequences of four T2-derived CNBr peptides (*10*) and three apo-B100-derived tryptic peptides (*11*) corresponded to predicted amino acid sequences at seven points along the sequence (Fig. 4). Furthermore, the COOH-terminal sequence of apo-B100 as determined by mass spectrometry of a tryptic peptide (*12*) agreed with the derived COOH-terminus (residues 1446 to 1455). The cDNA sequence also predicted a peptide sequence (residues 5 to 29) that matched at 21 of 25 positions with the peptide R2-5 of apo-B previously reported (*13*). These ten peptide sequences occur within the same open reading frame, thus confirming that these cDNA clones represent apo-B100.

The derived composition of T2 (1287 amino acids) is as follows: Asp_{66} Asn_{60} Thr_{86} Ser_{125} Glu_{91} Gln_{65} Pro_{46} Gly_{49} Ala_{66} Cys_5 Val_{79} Met_{17} Ile_{84} Leu_{147} Tyr_{48} Phe_{69} Trp_{13} Lys_{100} His_{32} Arg_{39}. The calculated mass of 146 kD is somewhat lower than the 170 kD estimated by Cardin *et al.* (*5*), but peptide T2 is also very likely glycosylated (Fig. 4), and therefore the discrepancy may be more apparent than real.

Dot matrix analysis (*14*) of the amino acid sequence did not reveal any significant internally repeated elements. Likewise, analysis of the nucleotide sequence revealed no internal repeats. A comparison of the sequence of apo-B100 with other published protein sequences (*14*) indicated that there were no particularly striking homologies between the COOH-

terminus of apo-B and other proteins. Of particular note is that no apparent homologies to the other apolipoproteins were observed.

Chromosomal localization. Thirty-eight human-mouse somatic cell hybrids were used to map the apo-B gene to chromosome 2 (Table 1 and Fig. 5b). Furthermore, in situ hybridization demonstrated that apo-B could be localized to the p24 region of the short arm of chromosome 2 (Fig. 5a). No other genes relevant to lipid metabolism have been localized in this region. Other apolipoprotein genes (*15*) have been mapped to chromosome 1 (apo-A-II), 11 (apo-A-I, A-IV, and C-III) and 19 (apo-C-I, C-II, and E). The apo-B,E(LDL) receptor gene has also been localized to chromosome 19 (*16*). The localization of apo-B to the p24 region of chromosome 2 therefore gives no clue as to whether there is an evolutionary relationship between the apo-B gene and the other apolipoprotein genes.

Implications of apo-B mRNA distribution. Apolipoprotein B100 is a large, structural protein important in the biosynthesis of VLDL by the liver. Our study demonstrates that apo-B100 is very large, with a 19-kb mRNA. Not only was apo-B100 mRNA detected in human liver, in the HepG2 hepatoma cell line, and in rhesus monkey liver, but it was also detected in the human and monkey small intestine. The most likely explanation for this observation is that the intestine synthesizes apo-B100 in addition to apo-B48. The presence of small quantities of apo-B100 mRNA in the small intestine is consistent with previous observations obtained from studies of the perfused rat intestine and from the analysis of mesenteric lymph lipoproteins, which suggest that a high molecular mass form of apo-B (resembling apo-B100) may be produced by the intestine (*1*, *17*). Nevertheless, it is well established that the principal form of apo-B synthesized and secreted by the small intestine is the smaller apo-B48 (apo-B with a molecular mass that is approximately one-half that of apo-B100) (*1*). However, our data do not exclude the possibility that apo-B100 and apo-B48 may be products of different proteolytic processing of the same translation product synthesized from the same mRNA. Alternatively, the two forms of apo-B could be products of the same gene that arise from different mRNA's produced by differential splicing. This remains to be elucidated.

Role of apo-B100 as a lipid binding protein. We analyzed the amino acid sequence of the COOH-terminus of apo-B100, taking into consideration the functional roles of apo-B100 as a structural protein in the assembly and secretion of hepatic lipoproteins and as the ligand responsible for mediating the binding of LDL to the apo-B,E(LDL) receptor. Predictions of the secondary structure with the Chou-Fasman algorithm (*18*) revealed that the COOH-terminus of apo-B is highly ordered, with approximately 30 α-helical segments encompassing approximately 33 percent of the residues and with a similar number of β structures representing about 25 percent of the residues. Few of these α or β structures extend over long distances because this sequence appears to have numerous β turns. Analysis of the sequence with the Kyte-Doolittle hydrophobicity plot (*19*) indicates that the COOH-terminus of apo-B has, in general, more hydrophilic than hydrophobic character (Fig. 6). The hydrophilic regions possess more α than β structure, while the hydrophobic regions have more β structure. There are frequent predicted crossovers from hydrophobic to hydrophilic character; and thus the apo-B100 polypeptide may be woven into and out of the lipid environment of the lipoprotein particle at irregular intervals and may have numerous lipid anchoring sites rather than a single large lipid binding domain that is characteristic of the smaller apolipoproteins. If the remainder of the apo-B polypeptide has similar features, it would appear that apo-B100 has more characteristics common to an integral membrane protein than to a typical apolipoprotein. Thus, apo-B serves as a nonexchangeable structural component of lipoproteins, while other apolipoproteins readily exchange among the various lipoproteins.

Lipid binding of other apolipoproteins has been ascribed to α-helical segments having amphipathic character (*20*). In some cases (apo-A-I, apo-A-IV, and apo-E) these amphipathic segments occur in tandem repetitions of 22 amino acids, resulting from tandem repetitions of 66 nucleotides in their genes (*21*). In the 1455 amino acids of apo-B thus far sequenced, no such repetitions (or any sort of repeated elements) occur. Furthermore, no typical amphipathic α helices are apparent in the apo-B100 sequence thus far analyzed. However, there are two potentially interesting hydrophobic regions (residues 135 to 270 and residues 600 to 770) that are predicted to have primarily β structures and that may be important in lipid binding. These same regions are highly enriched in proline residues, with more than half of all the prolines of this COOH-terminal segment occurring in these two hydrophobic regions. A frequent characteristic of these two regions is that the prolines are preceded by alternating hydrophilic and hydrophobic amino acid residues such that the hydrophobic side chains are on one face of the β sheet and the

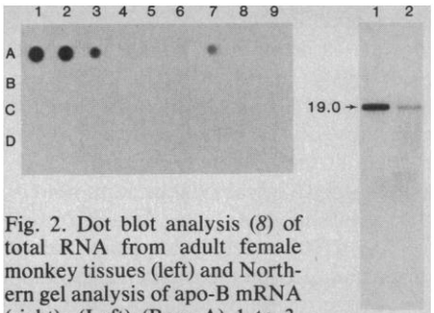


Fig. 2. Dot blot analysis (*8*) of total RNA from adult female monkey tissues (left) and Northern gel analysis of apo-B mRNA (right). (Left) (Row A) 1 to 3, liver; 4, pancreas; 5 and 6, spleen; 7, small intestine; 8 and 9, colon. (Row B) Samples as in row A, but treated with NaOH. (Row C) 1, lung; 2, heart; 3, kidney; 4 and 5, pituitary; 6, frontal cortex; 7, occipital cortex; 8, cerebellum; 9, brainstem. (Row D) Samples as in row C, but treated with NaOH. Each sample was 10 μg of total RNA. (Right) (Lane 1) 5 μg of total RNA from HepG2; lane 2, 20 μg of total RNA from human intestine. The size of the 19-kb mRNA was estimated by comparison to λ Hind III fragments.

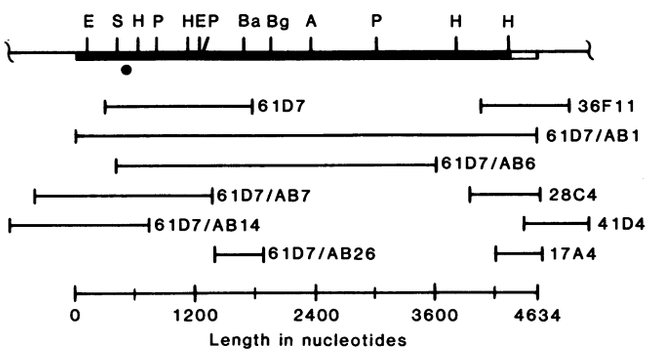


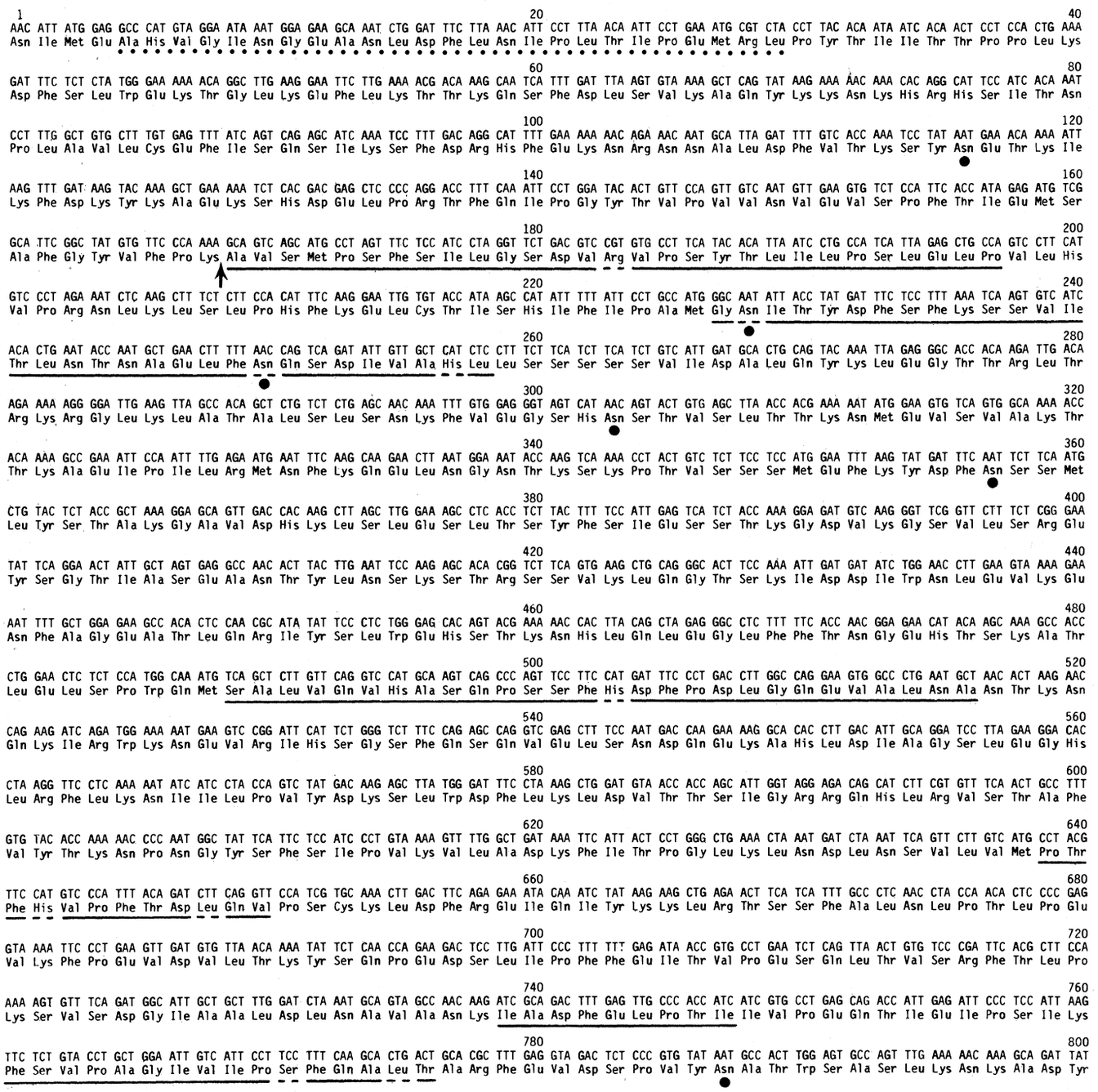
Fig. 3. Restriction map of cDNA clones representing the 3′ end of the apo-B mRNA. E, Eco RI; S, Sst I; H, Hind III; P, Pst I; Ba, Bam HI; Bg, Bgl II; A, Acc I. The filled circle denotes the position of hybridization of the oligonucleotide probe. The thickened line represents the sequenced regions: coding (solid bar) and noncoding (open bar). The DNA sequence was determined by dideoxy chain termination on both strands and across all restriction sites used (*28*). Subcloning of restriction fragments into M13 vectors and preparation of single-stranded DNA templates was performed as described (*29*).

hydrophilic side chains are on the other. These regions could represent amphipathic β sheets analogous to the amphipathic α helices seen in other apolipoproteins. Such a structure has been postulated to be of potential importance in the structure of LDL (*22*).

Previously, apo-B has been shown to be a glycoprotein (with ~8 to 10 percent of its mass as carbohydrate) with oligosaccharides of both the high mannose and complex type (*23*). The role of apo-B glycosylation in the structure and function of this protein remains to be determined. However, the COOH-terminal region of apo-B100 has eight potential *N*-glycosylation sites (Asn-X-Ser/Thr), as indicated in Fig. 4. Interestingly, most of these sites occur near points where the protein is predicted to cross over from hydrophobic to hydrophilic character and, thus, could be at the surface of the particle at the lipid–water interface. Although only *N*-glycosylation, and not *O*-glycosylation, has been demonstrated for apo-B100, the hydroxy amino acids serine and threonine are especially abundant in the region of residues 230 to 430 (representing ~30 percent of all amino acids in this region).

Role of apo-B100 as a receptor binding protein. In addition to the structural role of apo-B100, this apolipoprotein serves as the ligand mediating the binding of LDL to the apo-B,E(LDL) receptor. Because of the similarities between the interaction of apo-B100 and apo-E with the apo-B,E(LDL) receptor (*2*), it is reasonable to consider that the receptor binding domains of the two ligands might be similar. Studies of apo-E (*2*) have clearly established that the region responsible for its direct interaction with the receptor is in the vicinity of residues 140 to 150 of the apo-E molecule, a region enriched in basic amino acids (Fig. 7).

Inspection of the amino acid sequence of the COOH-terminal region of apo-B100 reveals a potentially important sequence that has a structure similar to



that proposed for the receptor binding domain of apo-E. This region (residues 276 to 286) of possible importance is compared to the apo-E binding domain in Fig. 7. In comparing this structure with that of the apo-E binding domain, we find it noteworthy that normal binding is disrupted by genetic variants of apo-E in which neutral amino acid substitutions for arginine or lysine occur at residues 142, 145, and 146 (*2*). These residues in apo-E correspond to amino acids 278, 281, and 282 in the comparable region of apo-B100. The structure of the apo-B,E(LDL) receptor has now been determined, and a postulated ligand binding domain has been shown to be enriched in acidic amino acids (*24*). The

Table 1. Distribution of apo-B with human chromosomes in human-mouse cell hybrids.

Chromosome number		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
Concordant hybrids	(+/+)*	12	18	16	12	14	14	15	14	6	17	15	14	12	15	15	8	16	16	11	11	15	10	15
(No.)	(−/−)	18	20	8	15	13	17	11	10	18	10	12	11	12	6	12	14	5	11	15	10	5	11	9
Discordant hybrids	(+/−)	6	0	2	6	4	4	3	4	12	1	3	4	6	3	3	10	2	2	7	7	3	8	3
(No.)	(−/+)	2	0	12	5	7	3	9	10	2	10	8	9	8	14	8	6	15	9	5	10	15	9	11
Discordancy (%)		21	0	37	29	29	18	32	37	37	29	29	34	37	45	29	42	45	29	32	45	47	45	37

*The first symbol within the parentheses indicates those hybrids that were either positive (+) or negative (−) for apo-B; the second symbol within the parentheses indicates those hybrids that either contained (+) or lacked (−) the particular chromosome.

```
                                                                           820                                                                         840
GTT GAA ACA GTC CTG GAT TCC ACA TGC AGC TCA ACC GTA CAG TTC CTA GAA TAT GAA CTA AAT GTT TTG GGA ACA CAC AAA ATC GAA GAT GGT ACG TTA GCC TCT AAG ACT AAA GGA ACA
Val Glu Thr Val Leu Asp Ser Thr Cys Ser Ser Thr Val Gln Phe Leu Glu Tyr Glu Leu Asn Val Leu Gly Thr His Lys Ile Glu Asp Gly Thr Leu Ala Ser Lys Thr Lys Gly Thr

                                                                           860                                                                         880
CTT GCA CAC CGT GAC TTC AGT GCA GAA TAT GAA GAA GAT GGC AAA TTT GAA GGA CTT CAG GAA TGG GAA GGA AAA GCG CAC CTC AAT ATC AAA AGC CCA GCG TTC ACC GAT CTC CAT CTG
Leu Ala His Arg Asp Phe Ser Ala Glu Tyr Glu Glu Asp Gly Lys Phe Glu Gly Leu Gln Glu Trp Glu Gly Lys Ala His Leu Asn Ile Lys Ser Pro Ala Phe Thr Asp Leu His Leu

                                                                           900                                                                         920
CGC TAC CAG AAA GAC AAG AAA GGC ATC TCC ACC TCA GCA GCC TCC CCA GCC GTA GGC ACC GTG GGC ATG GAT ATG GAT GAA GAT GAC GAC TTT TCT AAA TGG AAC TTC TAC TAC AGC CCT
Arg Tyr Gln Lys Asp Lys Lys Gly Ile Ser Thr Ser Ala Ala Ser Pro Ala Val Gly Thr Val Gly Met Asp Met Asp Glu Asp Asp Asp Phe Ser Lys Trp Asn Phe Tyr Tyr Ser Pro

                                                                           940                                                                         960
CAG TCC TCT CCA GAT AAA AAA CTC ACC ATA TTC AAA ACT GAG TTG AGG GTC CGG GAA TCT GAT GAG GAA ACT CAG ATC AAA GTT AAT TGG GAA GAA GAG GCA GCT TCT GGC TTG CTA ACC
Gln Ser Ser Pro Asp Lys Lys Leu Thr Ile Phe Lys Thr Glu Leu Arg Val Arg Glu Ser Asp Glu Glu Thr Gln Ile Lys Val Asn Trp Glu Glu Glu Ala Ala Ser Gly Leu Leu Thr

                                                                           980                                                                        1000
TCT CTG AAA GAC AAC GTG CCC AAG GCC ACA GGG GTC CTT TAT GAT TAT GTC AAC AAG TAC CAC TGG GAA CAC ACA GGG CTC ACC CTG AGA GAA GTG TCT TCA AAG CTG AGA AGA AAT CTG
Ser Leu Lys Asp Asn Val Pro Lys Ala Thr Gly Val Leu Tyr Asp Tyr Val Asn Lys Tyr His Trp Glu His Thr Gly Leu Thr Leu Arg Glu Val Ser Ser Lys Leu Arg Arg Asn Leu

                                                                          1020                                                                        1040
CAG AAC AAT GCT GAG TGG GTT TAT CAA GGG GCC ATT AGG CAA ATT GAT GAT ATC GAC GTG AGG TTC CAG AAA GCA GCC AGT GGC ACC ACT GGG ACC TAC CAA GAG TGG AAG GAC AAG GCC
Gln Asn Asn Ala Glu Trp Val Tyr Gln Gly Ala Ile Arg Gln Ile Asp Asp Ile Asp Val Arg Phe Gln Lys Ala Ala Ser Gly Thr Thr Gly Thr Tyr Gln Glu Trp Lys Asp Lys Ala

                                                                          1060                                                                        1080
CAG AAT CTG TAC CAG GAA CTG TTG ACT CAG GAA GGC CAA GCC AGT TTC CAG GGA CTC AAG GAT AAC GTG TTT GAT GGC TTG GTA CGA GTT ACT CAA AAA TTC CAT ATG AAA GTC AAG CAT
Gln Asn Leu Tyr Gln Glu Leu Leu Thr Gln Glu Gly Gln Ala Ser Phe Gln Gly Leu Lys Asp Asn Val Phe Asp Gly Leu Val Arg Val Thr Gln Lys Phe His Met Lys Val Lys His

                                                                          1100                                                                        1120
CTG ATT GAC TCA CTC ATT GAT TTT CTG AAC TTC CCC AGA TTC CAG TTT CCG GGG AAA CCT GGG ATA TAC ACT AGG GAG GAA CTT TGC ACT ATG TTC ATA AGG GAG GTA GGG ACG GTA CTG
Leu Ile Asp Ser Leu Ile Asp Phe Leu Asn Phe Pro Arg Phe Gln Phe Pro Gly Lys Pro Gly Ile Tyr Thr Arg Glu Glu Leu Cys Thr Met Phe Ile Arg Glu Val Gly Thr Val Leu

                                                                          1140                                                                        1160
TCC CAG GTA TAT TCG AAA GTC CAT AAT GGT TCA GAA ATA CTG TTT TCC TAT TTC CAA GAC CTA GTG ATT ACA CTT CCT TTC GAG TTA AGG AAA CAT AAA CTA ATA GAT GTA ATC TCG ATG
Ser Gln Val Tyr Ser Lys Val His Asn Gly Ser Glu Ile Leu Phe Ser Tyr Phe Gln Asp Leu Val Ile Thr Leu Pro Phe Glu Leu Arg Lys His Lys Leu Ile Asp Val Ile Ser Met
                                 ●
                                                                          1180                                                                        1200
TAT AGG GAA CTG TTG AAA GAT TTA TCA AAA GAA GCC CAA GAG GTA TTT AAA GCC ATT CAG TCT CTC AAG ACC ACA GAG GTG CTA CGT AAT CTT CAG GAC CTT TTA CAA TTC ATT TTC CAA
Tyr Arg Glu Leu Leu Lys Asp Leu Ser Lys Glu Ala Gln Glu Val Phe Lys Ala Ile Gln Ser Leu Lys Thr Thr Glu Val Leu Arg Asn Leu Gln Asp Leu Leu Gln Phe Ile Phe Gln

                                                                          1220                                                                        1240
CTA ATA GAA GAT AAC ATT AAA CAG CTG AAA GAG ATG AAA TTT ACT TAT CTT ATT AAT TAT ATC CAA GAT GAG ATC AAC ACA ATC TTC AAT GAT TAT ATC CCA TAT GTT TTT AAA TTG TTG
Leu Ile Glu Asp Asn Ile Lys Gln Leu Lys Glu Met Lys Phe Thr Tyr Leu Ile Asn Tyr Ile Gln Asp Glu Ile Asn Thr Ile Phe Asn Asp Tyr Ile Pro Tyr Val Phe Lys Leu Leu

                                                                          1260                                                                        1280
AAA GAA AAC CTA TGC CTT AAT CTT CAT AAG TTC AAT GAA TTT ATT CAA AAC GAG CTT CAG GAA GCT TCT CAA GAG TTA CAG CAG ATC CAT CAA TAC ATT ATG GCC CTT CGT GAA GAA TAT
Lys Glu Asn Leu Cys Leu Asn Leu His Lys Phe Asn Glu Phe Ile Gln Asn Glu Leu Gln Glu Ala Ser Gln Glu Leu Gln Gln Ile His Gln Tyr Ile Met Ala Leu Arg Glu Glu Tyr

                                                                          1300                                                                        1320
TTT GAT CCA AGT ATA GTT GGC TGG ACA GTG AAA TAT TAT GAA CTT GAA GAA AAG ATA GTC AGT CTG ATC AAG AAC CTG TTA GTT GCT CTT AAG GAC TTC CAT TCT GAA TAT ATT GTC AGT
Phe Asp Pro Ser Ile Val Gly Trp Thr Val Lys Tyr Tyr Glu Leu Glu Glu Lys Ile Val Ser Leu Ile Lys Asn Leu Leu Val Ala Leu Lys Asp Phe His Ser Glu Tyr Ile Val Ser

                                                                          1340                                                                        1360
GCC TCT AAC TTT ACT TCC CAA CTC TCA AGT CAA GTT GAG CAA TTT CTG CAC AGA AAT ATT CAG GAA TAT CTT AGC ATC CTT ACC GAT CCA GAT GGA AAA GGG AAA GAG AAG ATT GCA GAG
Ala Ser Asn Phe Thr Ser Gln Leu Ser Ser Gln Val Glu Gln Phe Leu His Arg Asn Ile Gln Glu Tyr Leu Ser Ile Leu Thr Asp Pro Asp Gly Lys Gly Lys Glu Lys Ile Ala Glu
         ●
                                                                          1380                                                                        1400
CTT TCT GCC ACT GCT CAG GAA ATA ATT AAA AGC CAG GCC ATT GCG ACG AAG AAA ATA ATT TCT GAT TAC CAC CAG CAG TTT AGA TAT AAA CTG CAA GAT TTT TCA GAC CAA CTC TCT GAT
Leu Ser Ala Thr Ala Gln Glu Ile Ile Lys Ser Gln Ala Ile Ala Thr Lys Lys Ile Ile Ser Asp Tyr His Gln Gln Phe Arg Tyr Lys Leu Gln Asp Phe Ser Asp Gln Leu Ser Asp

                                                                          1420                                                                        1440
TAC TAT GAA AAA TTT ATT GCT GAA TCC AAA AGA TTG ATT GAC CTG TCC ATT CAA AAC TAC CAC ACA TTT CTG ATA TAC ATC ACG GAG TTA CTG AAA AAG CTG CAA TCA ACC ACA GTC ATG
Tyr Tyr Glu Lys Phe Ile Ala Glu Ser Lys Arg Leu Ile Asp Leu Ser Ile Gln Asn Tyr His Thr Phe Leu Ile Tyr Ile Thr Glu Leu Leu Lys Lys Leu Gln Ser Thr Thr Val Met

                                                  1455
AAC CCC TAC ATG AAG CTT GCT CCA GGA GAA CTT ACT ATC ATC CTC TAA   TTTTTTAAAA GAAATCTCAT TATCTCTTTC CAATGAACTT CACATAGCAC AGAAAAAATC AAACTGCCTA TATTGATAAA
Asn Pro Tyr Met Lys Leu Ala Pro Gly Glu Leu Thr Ile Ile Leu ***

ACCATACAGT GAGCCAGCCT TGCAGTAGGC AGTAGACTAT AAGCAGAAGC ACATATGAAC TGGACCTGCA CCAAAGCTGG CACCAGGGCT CGGAAGGTCT CTGAACTCAG AAGGATGGCA TTTTTTGCAA GTTAAAGAAA

ATCAGGATCT GAGTTATTTT GCTAAACTTG GGGGAGGAGG AACAAA
```

Fig. 4. Nucleotide sequence of cDNA clones representing the 3′ end of apo-B100 mRNA and the derived amino acid sequence of the COOH-terminal 30 percent of apo-B100. The arrow indicates the thrombin cleavage site that gives rise to fragment T2. The lines under the derived amino acid sequence indicate those regions found from peptide sequencing. The broken lines indicate those amino acid residues that could not be assigned with certainty in the peptide sequence analyses. The dotted line indicates the peptide R2-5 of LeBoeuf *et al.* (*13*). Circles indicate potential *N*-glycosylation sites. Asterisks (***) indicate the termination codon. Numbering of the sequence is arbitrary and begins at the most NH_2-terminal amino acid. The sequence of the first 266 nucleotides of the 3′-untranslated region is given, but there is a still-unknown number of nucleotides preceding the poly(A) tail.

sequence of the proposed ligand binding domain is also shown in Fig. 7.

In addition to a role in receptor binding, regions enriched in basic amino acids may also be involved in mediating apo-B100 interaction with heparin. In both apo-E and apo-B, modification of lysine or arginine residues results in loss of heparin binding ability as well as receptor binding ability (*3*). Within the COOH-terminus of apo-B100, there are other sites of potential importance for either receptor or heparin binding. These include residues 40 to 76 and residues 94 to 136.

The importance of these domains in either receptor or heparin binding remains to be determined. However, with respect to receptor binding, it is reasonable to focus attention on a single site within apo-B100 and to focus attention on the COOH-terminus of the molecule. Previously, Milne and Marcel and their co-workers (*25*) demonstrated that the binding of a single monoclonal antibody to the LDL particle blocks receptor binding activity, which suggests that there may be only a single receptor binding domain expressed on the surface of an LDL particle. Furthermore, they have localized the epitopes of monoclonal antibodies that inhibit receptor binding activity to the apo-B74 fragment of apo-B100. This fragment encompasses thrombolytic fragments T3 and T2 (Fig. 1). We have identified a monoclonal antibody (18C4) that also inhibits the binding of LDL to the apo-B,E(LDL) receptor. On immunoblots, it recognizes only fragment T2. The basic region encompassing residues 276 to 286, which most closely resembles the apo-E binding domain, is located in thrombolytic fragment T2 within 110 residues of the thrombin cleavage site. Antibodies to synthetic peptides from the region 276 to 286 and other basic regions are being prepared to determine experimentally whether any of these sites is the receptor binding domain.

In summary, the availability of cDNA clones for apo-B will allow complete structural characterization of the apo-B mRNA, the entire protein, and ultimately the gene. This will eventually allow the unraveling of the steps that control the biosynthesis of VLDL and LDL, thus providing insights into how to control LDL levels and to regulate LDL cholesterol metabolism.

Note added in proof: After this research article was submitted, Deeb *et al.* (*26*) reported a partial cDNA sequence for human apo-B. Their sequence does not overlap any portion of ours. Lusis *et al.* (*27*) reported a partial cDNA sequence for the rat large apo-B. The beginning of their insert shows a 70 percent homology (both nucleotides and amino acids) to our sequence beginning at amino acid residue 929 in Fig. 4.

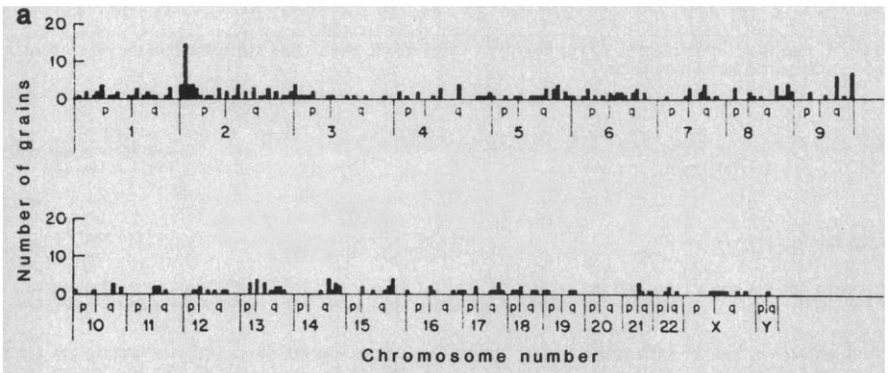


Fig. 5. (a) In situ hybridization of apo-B cDNA in 163 human metaphases. The horizontal axis represents the chromosomes in their relative size proportion; the vertical axis gives the silver grain distribution. The in situ hybridization (*30*) shows significant labeling at the 2p24 site in comparison to the nonspecific background labeling (19.4 percent of all grains are on chromosome 2; 4.7 percent of all grains are on 2p24). (b) Southern hybridization of apo-B cDNA probe (clone p61D7) to human, mouse, and somatic cell hybrid DNA. The DNA was digested with Eco RI (or Hind III, not shown) and hybridized with ^{32}P-labeled cDNA insert. The size of the 1.6-kb fragment was estimated with λ Hind III- and φX174 Hae III–digested DNA markers. Human DNA (H) was prepared from Wi38 cells and mouse DNA (M) from LM/TK$^-$ cells. In the examples of somatic cell hybrids shown, positive hybridization (+) and negative hybridization (−) are indicated. The construction and characterization of the human-mouse somatic cell hybrids, the preparation of DNA, and hybridization conditions were as described (*31*).

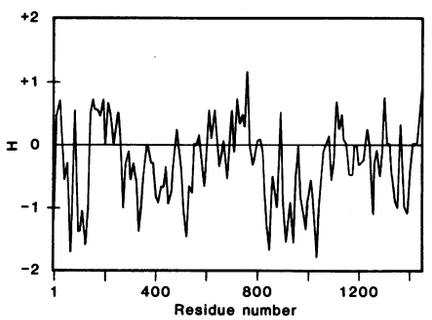


Fig. 6. Hydrophobicity plot of the COOH-terminal region of apo-B100. The hydrophobicity index (H) of Kyte and Doolittle (*19*) was used and is plotted against the sequence using a windowing average of 18 residues. Residues are numbered as in Fig. 4.

Fig. 7. Structural similarities between the receptor binding domain of apo-E and a region of apo-B. The homologous positively charged residues lysine and arginine are boxed in the sequence of apo-E and apo-B. The consensus sequence of the postulated ligand binding site of the apo-B,E(LDL) receptor, which is enriched in the negatively charged residues aspartic and glutamic acids (underlined), is shown for comparison (*24*).

Apo-E (Residues 140-150)	-His-Leu-Arg-Lys-Leu-Arg-Lys-Arg-Leu-Leu-Arg-
Apo-B (Residues 276-286)	-Thr-Thr-Arg-Leu-Thr-Arg-Lys-Arg-Gly-Leu-Lys-
Apo-B,E(LDL) Receptor (Consensus Sequence)	-Cys-Asp-X-X-X-Asp-Cys-X-Asp-Gly-Ser-Asp-Glu-

References and Notes

1. R. J. Havel, J. L. Goldstein, M. S. Brown, in *Metabolic Control and Disease*, P. K. Bondy and L. E. Rosenberg, Eds. (Saunders, Philadelphia, 1980), pp. 393–494; J. P. Kane, *Annu. Rev. Physiol.* **45**, 637 (1983); J. L. Goldstein and M. S. Brown, in *The Metabolic Basis of Inherited Disease*, J. B. Stanbury *et al.*, Eds. (McGraw-Hill, New York, 1983), pp. 672–712.
2. R. W. Mahley and T. L. Innerarity, *Biochim. Biophys. Acta* **737**, 197 (1983); R. W. Mahley, T. L. Innerarity, S. C. Rall, Jr., K. H. Weisgraber, *J. Lipid Res.* **25**, 1277 (1984).
3. R. W. Mahley *et al.*, *J. Biol. Chem.* **252**, 7279 (1977); K. H. Weisgraber, T. L. Innerarity, R. W. Mahley, *ibid.* **253**, 9053 (1978); R. W. Mahley, K. H. Weisgraber, T. L. Innerarity, *Biochim. Biophys. Acta* **575**, 81 (1979).
4. C.-H. Chen and F. Aladjem, *Biochem. Biophys. Res. Commun.* **60**, 549 (1974); W. A. Bradley, M. F. Rohde, A. M. Gotto, Jr., R. L. Jackson, *ibid.* **81**, 928 (1978); L. Socorro and G. Camejo, *J. Lipid Res.* **20**, 631 (1979); L. Socorro, F. López, A. López, G. Camejo, *ibid.* **23**, 1283 (1982).
5. A. D. Cardin *et al.*, *J. Biol. Chem.* **259**, 8522 (1984).

6. A 42-base synthetic oligodeoxynucleotide, based on the first 14 amino acids of thrombolytic fragment T2, was synthesized as described [B. D. Warner *et al.*, *DNA* **3**, 401 (1984)]. Codon choices were made to reduce the degeneracy of the probe. The choices were based on data gathered on codon preferences for other apolipoprotein and membrane proteins. The oligonucleotide synthesized was the complement of the coding sequence for the NH_2-terminus of T2: 3′-CGGCACA$^{G}_{C}$GTACGGGA$^{G}_{C}$GAAGAGG-TA$^{T}_{G}$GACCCGA$^{G}_{C}$GCTGCAC-5′.
7. The cDNA library was constructed by D. E. Woods, A. F. Markham, A. T. Ricker, G. Goldberger, and H. R. Colten [*Proc. Natl. Acad. Sci. U.S.A.* **79**, 5661 (1982)] and screened as described by J. Scott *et al.* [*Nature (London)* **302**, 538 (1983)] and G. I. Bell *et al.* [*ibid.* **310**, 775 (1984)].
8. Plasmid DNA was prepared by a modified alkaline lysis method [T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982)] and ^{32}P-labeled by random priming as described [A. P. Feinberg and B. Vogelstein, *Anal. Biochem* **132**, 6 (1983); *ibid.* **137**, 266 (1984)]. RNA dot blot analysis was performed as described by L. B. Rall *et al.* [*Nature (London)* **313**, 228 (1985)] and L. B. Rall, J. Scott, and G. I. Bell (*Methods Enzymol.*, in press.)
9. K. T. Belt *et al.*, *Cell* **36**, 907 (1984).
10. Lyophilized T2 was dissolved in water at 2 mg/ml (the SDS concentration was about 0.6 to 0.8 percent). The protein solution was added dropwise to four times its volume of 90 percent HCOOH with stirring. Any precipitate that formed was allowed to clear before the next drop was added. Solid CNBr (125:1, by weight) was added, and the mixture was incubated for 8 hours at room temperature. The same weight of CNBr was then added and allowed to react for an additional 16 hours. The digest was then dialyzed (3.5-kD cutoff tubing) against 100 ml of 70 percent HCOOH with two changes over a 2- to 3-hour period. The bag was then opened, and the concentration of SDS was increased to 1 percent. The solution was dialyzed against 1 percent SDS, 5 m*M* NH_4HCO_3 for 4 to 5 hours and then against 0.1 percent SDS, 5 m*M* NH_4HCO_3 for 24 hours. The solution was frozen as a plug and lyophilized to reduce the volume (but not to dryness). The remaining liquid was adjusted to *p*H 7.0 to 7.5 with 1*N* NaOH. A portion of the peptide mixture was dansylated as described in Fig. 1, and the fragments were separated on 15 to 20 percent polyacrylamide gradient gels (1.5 mm thick). The peptides were transferred by electroelution into dialysis bags (3.5-kD cutoff) and dialyzed against 5 m*M* NH_4HCO_3, 0.1 percent SDS, for 48 hours and then lyophilized. Peptides were sequenced in a Beckman 890C Sequencer with a standard 0.1*M* Quadrol program in the presence of 3 mg of Polybrene and 0.5 percent SDS. Amino-terminal sequences were obtained on four peptides. CB1 (25 percent yield) was a 19.8-kD peptide whose NH_2-terminal sequence corresponds to residues 227 to 258 in Fig. 4. The absence of identifiable amino acids at positions 228 and 250 almost certainly indicates that these Asn residues are glycosylated. Furthermore, the molecular size of CB1 is too large for that predicted from the positions of the methionines in Fig. 4; therefore, this peptide is probably behaving anomalously on gels because of glycosylation. Peptide CB2 (30 percent yield) was a 17.2-kD peptide that represents residues 489 to 638 (calculated, 16.95 kD). Peptide CB3 (12 percent yield) was a 30.5-kD peptide that represents residues 639 to 903 (calculated, 29.95 kD). Peptide CB4 (<20 percent yield) was an 18- to 19-kD peptide found as a contaminating component in other fractions. The peptide represents residues 1275 to 1440 (calculated, 18.76 kD).
11. Low density lipoproteins (5 to 7 mg/ml) in 0.15*M* NaCl, 0.01 percent EDTA were filter-sterilized (0.45-μm pore size). Trypsin in 200 m*M* NH_4HCO_3, *p*H 7.6, was added to achieve a final ratio of trypsin to protein of 1:35 (by weight). The solution was then incubated at 37°C in an N_2 atmosphere for 6 hours. The soluble peptides and the trypsin were removed by gel filtration chromatography over a Sephadex G-50 column (2.2 by 90 cm) at 4°C in 0.15*M* NaCl, 1 m*M* EDTA, *p*H 8.2. The void volume fractions were collected and concentrated on an XM50 Amicon membrane to about 10 ml and then filtered through a 1.2-μm filter. The trypsin-treated LDL were dialyzed against 0.3 m*M* EDTA, *p*H 7.4, and lyophilized. A mixture of chloroform and methanol (2:1) was added to the lyophilized lipoprotein and incubated at 23°C for 30 minutes. To reduce the density of the organic solvent, an equal volume of methanol was added, and the protein was sedimented at low speed (3000 rev/min for 10 minutes, Beckman J6). The chloroform-methanol mixture was removed by aspiration, and the peptide mixture was dried under N_2. The procedures for the dansylation, electrophoresis, and elution of the peptides were the same as described for the CNBr fragments of T2, except that the separating gel was a 10 to 15 percent gradient gel. Amino-terminal sequences of three tryptic peptides (of nine that were determined) were located in the COOH-terminal portion of apo-B. These 29-, 23-, and 25-kD peptides, respectively, represent those beginning at residues 739, 761, and 1154, respectively (Fig. 4). Each peptide follows a lysine residue.
12. Peptides were generated by tryptic digestion at 37°C for 6 hours followed by gel filtration on Sephadex G-25. Peptide pools were then subjected to high-performance liquid chromatography using an *N*-propanol/acetic acid system. The sequences of tryptic peptides were determined by interpreting the fragmentation patterns observed by fast atom bombardment (FAB) (VG ZAB mass spectrometer equipped with an M-scan FAB gun; 10 kV and 20 μA xenon ion current [H. R. Morris and M. Panico, *Biochem. Biophys. Res. Commun.* **101**, 623 (1981)]. Where sequences were not complete due to lack of ion fragmentation, the remaining amino acids were assigned by comparing the FAB molecular weight and amino acid analysis. First and last amino acids in each peptide were determined by observing the shift in molecular weight in the mass spectrum after Edman and carboxypeptidase treatment, respectively. Several fractions were chosen for detailed mass spectrometric analysis, although the chromatogram was very complex (a consequence of the very large size of this protein), and none were at a stage of sufficient purity for classical sequencing. A peptide with an $(M + H) = 1039$ was of particular interest because it did not contain a COOH-terminal Lys or Arg (no change in mass after carboxypeptidase B digestion, thus making it a candidate for the true COOH-terminus of the intact apo-B protein). The ratio of Leu to Ile (from amino acid analysis) was 3:2, although these could not be differentiated by the mass data. The composition and partial sequence of this peptide was consistent with the predicted sequence of residues 1446 to 1455 in Fig. 4.
13. R. C. LeBoeuf *et al.*, *FEBS Lett.* **170**, 105 (1984).
14. J. Pustell and F. C. Kafatos, *Nucleic Acids Res.* **12**, 643 (1984); M. O. Dayhoff, W. C. Barker, L. T. Hunt, *Methods Enzymol.* **91**, 524 (1983).
15. B. Olaisen, P. Teisberg, T. Gedde-Dahl, Jr., *Hum. Genet.* **62**, 233 (1982); C. L. Jackson, G. A. P. Bruns, J. L. Breslow, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2945 (1984); T. J. Knott *et al.*, *Biochem. Biophys. Res. Commun.* **125**, 299 (1984); G. A. P. Bruns, S. K. Karathanasis, J. L. Breslow, *Arteriosclerosis* **4**, 97 (1984); O. Schamaun *et al.*, *Hum. Genet.* **68**, 181 (1984).
16. U. Francke, M. S. Brown, J. L. Goldstein, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2826 (1984).
17. A.-L. Wu and H. G. Windmueller, *J. Biol. Chem.* **256**, 3615 (1981); D. M. Lee, E. Koren, S. Singh, T. Mok, *Biochem. Biophys. Res. Commun.* **123**, 1149 (1984).
18. P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.* **47**, 251 (1978).
19. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
20. J. P. Segrest, R. L. Jackson, J. D. Morrisett, A. M. Gotto, Jr., *FEBS Lett.* **38**, 247 (1974).
21. M. S. Boguski, N. A. Elshourbagy, J. M. Taylor, J. I. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 992 (1985).
22. D. Osterman, R. Mora, F. J. Kézdy, E. T. Kaiser, S. C. Meredith, *J. Am. Chem. Soc.* **106**, 6845 (1984).
23. P. Lee and W. C. Breckenridge, *Can. J. Biochem.* **54**, 42 (1976); N. Swaminathan and F. Aladjem, *Biochemistry* **15**, 1516 (1976).
24. T. Yamamoto *et al.*, *Cell* **39**, 27 (1984); T. C. Sudhof, J. L. Goldstein, M. S. Brown, D. W. Russell, *Science* **228**, 815 (1985).
25. R. W. Milne and Y. L. Marcel, *FEBS Lett.* **146**, 97 (1982); Y. L. Marcel, M. Hogue, R. Theolis, Jr., R. W. Milne, *J. Biol. Chem.* **257**, 13165 (1982); R. W. Milne, R. Theolis, Jr., R. B. Verdery, Y. L. Marcel, *Arteriosclerosis* **3**, 23 (1983).
26. S. S. Deeb, A. G. Motulsky, J. J. Albers, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4597 (1985).
27. A. J. Lusis *et al.*, *ibid.*, p. 4597.
28. F. Sanger, A. R. Coulson, B. G. Barrell, A. J. H. Smith, B. A. Roe, *J. Mol. Biol.* **143**, 161 (1980).
29. T. J. Knott *et al.*, *Nucleic Acids Res.* **12**, 3909 (1984).
30. B. U. Zabel, S. L. Naylor, A. Y. Sakaguchi, G. I. Bell, T. B. Shows, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6932 (1983).
31. T. B. Shows *et al.*, *Somat. Cell Genet.* **10**, 315 (1984); T. B. Shows *et al.*, *Cytogenet. Cell Genet.* **21**, 99 (1978); T. B. Shows, A. Y. Sakaguchi, S. L. Naylor, *Adv. Hum. Genet.* **12**, 341 (1982); E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975); S. L. Naylor *et al.*, *J. Exp. Med.* **157**, 1020 (1983).

32. We thank C. Joannon, R. Bartsch, and D. S. Bailey for the intestinal RNA; D. Woods, S. Humphries, K. T. Belt, M. C. Carroll, and R. R. Porter for cDNA libraries; N. Haskins and colleagues of M-Scan, Ltd., Silwood Park, Ascot, U.K., who performed some of the amino acid sequence work and protein chemistry under the direction of Professor Howard R. Morris of the Department of Biochemistry, Imperial College, London, under MRC contract; L. Haley, M. Henry, and C. Young for technical assistance; R. Otter for computer analysis; Barbara Allen for editorial assistance; J. X. Warger and N. J. Gargasz for graphic arts; and K. Humphrey, S. Richmond, and K. Sholly for manuscript preparation. Supported in part by NIH grants GM 20454 and HO 05196 and by American Cancer Society grant CD62 to T.B.S.

23 July 1985; accepted 4 September 1985