Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 8142–8146, November 1986
Biochemistry

Human liver apolipoprotein B-100 cDNA: Complete nucleic acid and derived amino acid sequence

(low density lipoproteins/low density lipoprotein receptor/mRNA/oligonucleotide probe)

S. W. LAW, S. M. GRANT, K. HIGUCHI, A. HOSPATTANKAR, K. LACKNER, N. LEE, AND H. B. BREWER, JR.

Molecular Disease Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by Donald S. Fredrickson, August 1, 1986

ABSTRACT **Human apolipoprotein B-100 (apoB-100), the ligand on low density lipoproteins that interacts with the low density lipoprotein receptor and initiates receptor-mediated endocytosis and low density lipoprotein catabolism, has been cloned, and the complete nucleic acid and derived amino acid sequences have been determined. ApoB-100 cDNAs were isolated from normal human liver cDNA libraries utilizing immunoscreening as well as filter hybridization with radiolabeled apoB-100 oligodeoxynucleotides. The apoB-100 mRNA is 14.1 kilobases long encoding a mature apoB-100 protein of 4536 amino acids with a calculated amino acid molecular weight of 512,723. ApoB-100 contains 20 potential glycosylation sites, and 12 of a total of 25 cysteine residues are located in the amino-terminal region of the apolipoprotein providing a potential globular structure of the amino terminus of the protein. ApoB-100 contains relatively few regions of amphipathic helices, but compared to other human apolipoproteins it is enriched in β-structure. The delineation of the entire human apoB-100 sequence will now permit a detailed analysis of the conformation of the protein, the low density lipoprotein receptor binding domain(s), and the structural relationship between apoB-100 and apoB-48 and will provide the basis for the study of genetic defects in apoB-100 in patients with dyslipoproteinemias.**

Human apolipoprotein B (apoB) is the principal apolipoprotein on chylomicrons, very low density lipoproteins, intermediate density lipoproteins, and low density lipoproteins (LDLs) (1–4). In human plasma, apoB exists in two forms, designated apoB-48 and apoB-100 that are separable by $NaDodSO_4$/PAGE (4). ApoB-48 and apoB-100 play several important roles in lipoprotein biosynthesis and catabolism. B-48 and B-100 apolipoproteins are required for lipoprotein particle assembly in the liver and intestine. A defect in apoB-48 and apoB-100 secretion results in abetalipoproteinemia, a disease characterized by a deficiency of the B apolipoproteins as well as all apoB-containing lipoproteins including chylomicrons, very low density lipoprotein, and LDL (4). ApoB-100 also plays a pivotal role in lipoprotein metabolism as the ligand on LDL that interacts with the LDL receptor and initiates receptor-mediated endocytosis and LDL catabolism (5). Defects in the LDL receptor are associated with severe plasma elevations of apoB-100 and LDL (5).

The structure and physicochemical properties of apoB-100 have been extensively studied for nearly a decade. The analysis of apoB-100 has been difficult because delipidated apoB-100 is insoluble in aqueous solution, and generally aggregates in buffers containing $NaDodSO_4$, urea, or guanidine hydrochloride (1–4). The molecular size of apoB-100 has been controversial, and values ranging from 8 to 400 kDa have been reported (1–4, 6). The heterogeneity in apoB-100 molecular size has been attributed to the propensity of delipidated apoB-100 to aggregate and to the reported sensitivity of apoB-100 to protease cleavage (3, 4, 7). We (8, 9) and others (10–16) have reported the partial sequence of human and rat (17) liver apoB-100 cDNA. We now report the complete nucleotide sequence of apoB-100 mRNA and its derived amino acid sequence.

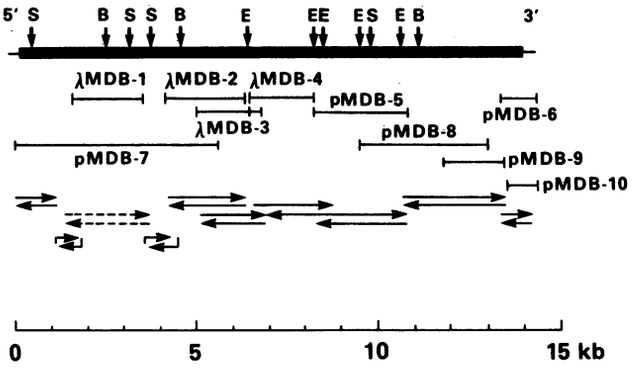

FIG. 1. Restriction endonuclease cleavage map and sequencing strategy for human liver apoB-100 cDNA. The thick solid line represents the coding region of apoB-100 mRNA. The 5'-noncoding and 3'-noncoding sequences are represented by thin lines flanking the coding region. Restriction enzyme sites are represented by a single letter code. (E, *Eco*RI; B, *Bam*HI; S, *Sac* I). The regions of the mRNA corresponding to each of the 10 partial cDNA clones are indicated below the map. Arrows below the clones represent direction and length of sequences obtained by the dideoxy procedure on M-13 subclones (solid arrows →) or supercoil plasmid DNAs (bent arrows ↱). Broken arrows (- - →) represent regions sequenced by the Maxam–Gilbert procedure. The scale indicates the size of the apoB-100 mRNA in kb.

MATERIALS AND METHODS

Extraction of Human Liver mRNA. RNA was isolated from adult human liver obtained from a 34-year-old female automobile accident victim. Frozen liver tissue was pulverized in dry ice by a stainless steel tissue grinder. Liver powder was then homogenized by a polytron in 4 M guanidine thiocyanate (Fluka AG) containing 0.5% sarcosyl, 0.1 M 2-mercaptoethanol, and 0.1% antifoam A followed by centrifugation through a 5.7 M CsCl cushion as described (18, 19). RNA pellets were dissolved in sterile 10 mM Tris·HCl, pH 7.4, containing 0.5% $NaDodSO_4$ and 1 mM EDTA and then extracted with phenol/chloroform/isoamylalcohol (24:24:1; vol/vol). The RNA in the aqueous phase was precipitated by

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 lipoprotein; apoB-100, apolipoprotein B-100; kb, kilobase(s).

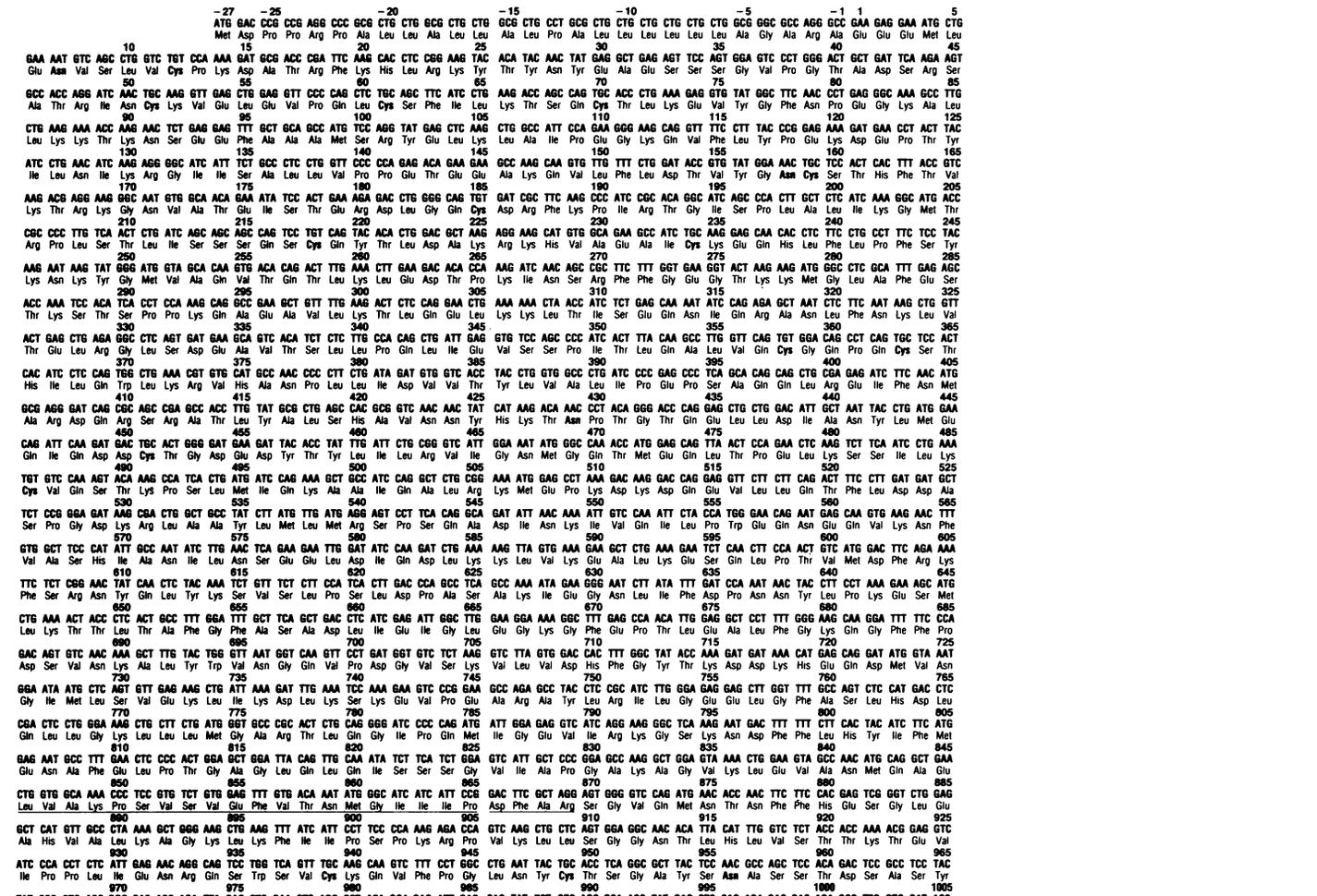

FIG. 2. *(Figure continues on the following page.)*

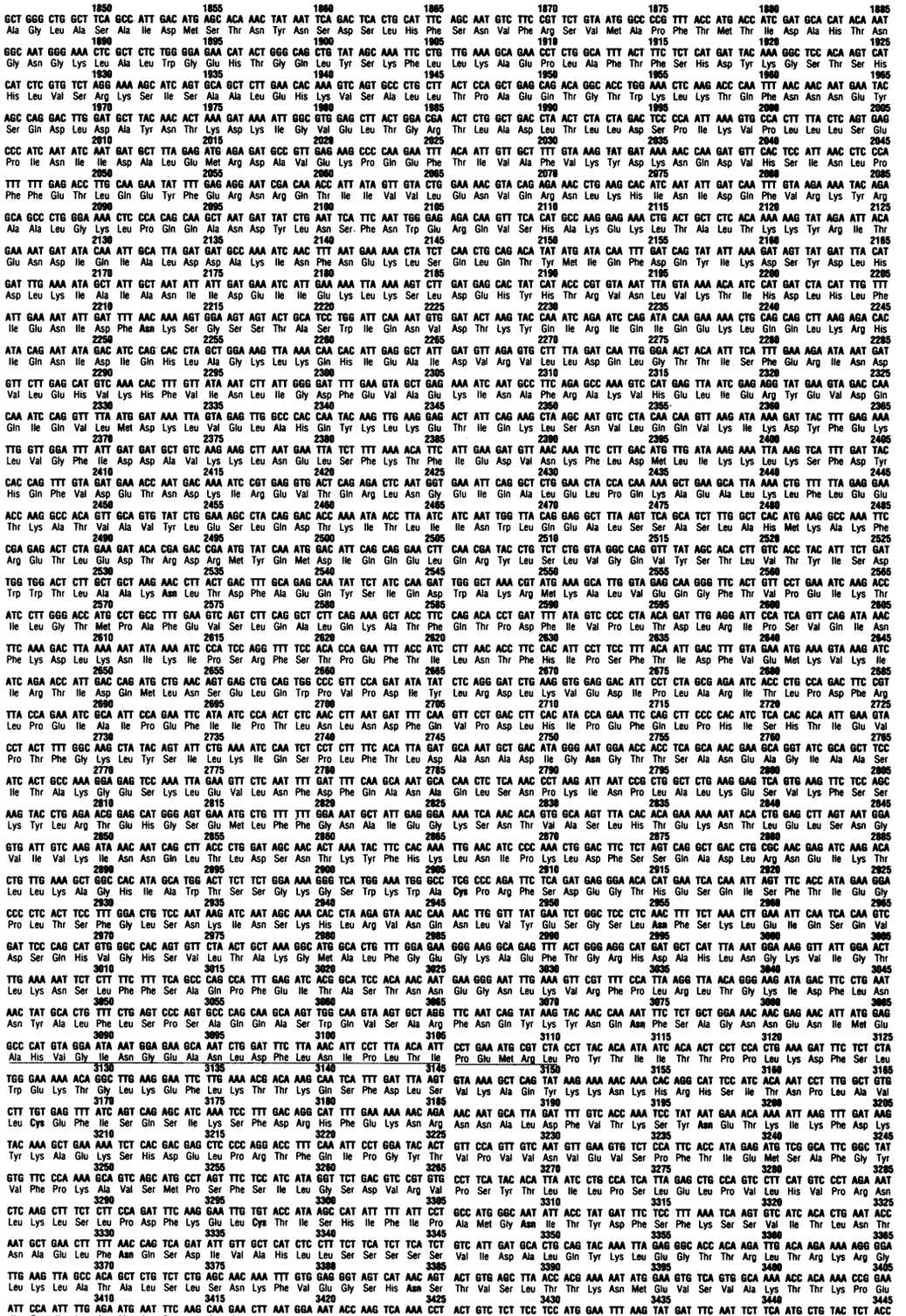

FIG. 2. *(Figure continues on the following page.)*

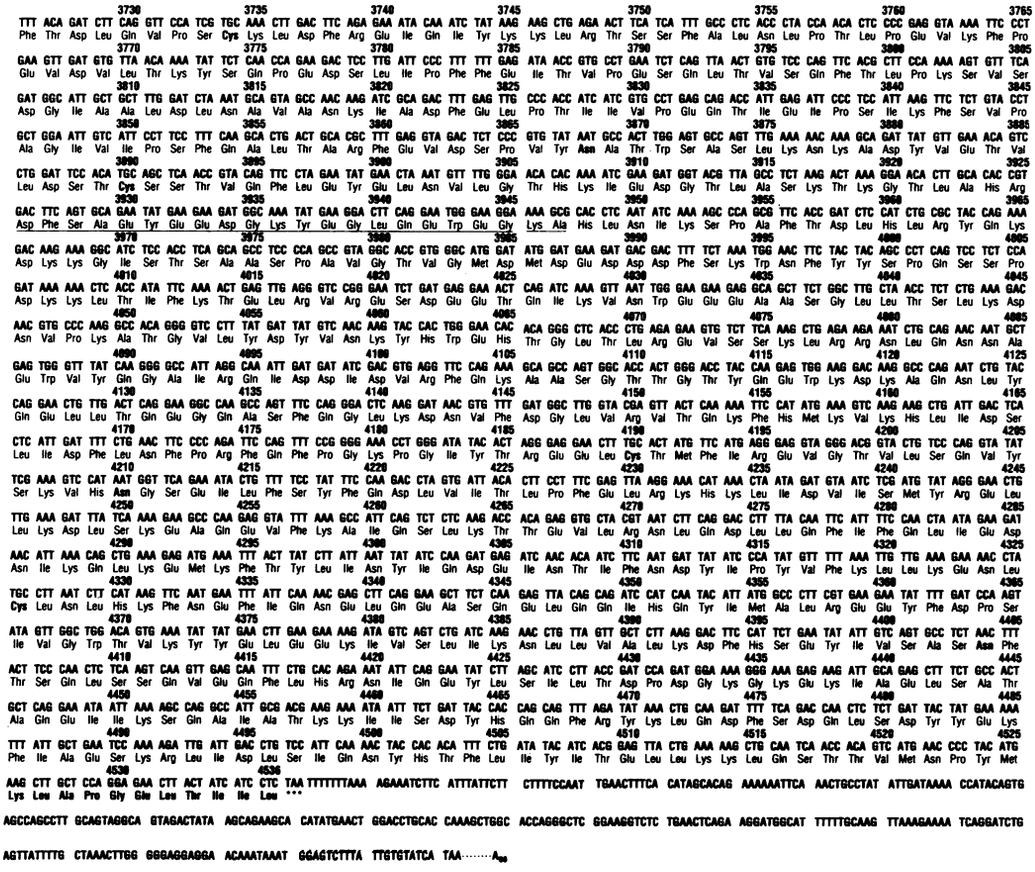

FIG. 2. Nucleotide sequence and derived amino acid sequence of human liver apoB-100 cDNA. Amino acids are numbered below the sequence. Residue 1 is the NH_2 terminus of the mature protein. Negative numbers refer to the prepeptide sequence. Cysteine residues and potential glycosylation sites are printed in bold type. The sequences of apoB peptides that have been published are underlined (22, 23).

ethanol. Poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose affinity chromatography.

Preparation of Human Liver cDNA Libraries. Several human liver cDNA libraries were utilized in this study. A cDNA library established in pBR322 has been described (20). A cDNA library established in λgt11 was kindly provided by G. Ricca. A cDNA library established in pAT153 PvuII/8 was a gift from R. D. Campbell (21).

Screening the Human Liver cDNA Libraries for ApoB-100 cDNA Clones. The 5′-end-labeled synthetic oligonucleotides and nick-translated apoB-100 cDNAs were used as hybridization probes to screen the plasmid and the λgt10 cDNA libraries. Monospecific, polyclonal antibodies against apoB-100 as well as monospecific, polyclonal antibodies raised against synthetic apoB-100 peptides (18, 22, 23) were used to screen the λgt11 expression library as described (18).

DNA Sequence Determination. The entire apoB-100 mRNA sequence was included in 10 overlapping cDNA clones. Both the Maxam and Gilbert chemical modification procedure (24) and the Sanger's dideoxynucleotides chain-termination procedure (25) were used. In the latter procedure, single-stranded DNA templates were generated by either subcloning into M13mp18 and M13mp19 or by denaturation of supercoil plasmid DNA by NaOH (26). Universal sequencing primers were purchased from Bethesda Research Laboratories. Synthetic oligonucleotide primers were purchased from OCS Labs (Denton, TX).

RESULTS

Identification of Human Liver ApoB-100 cDNA Clones. The λgt11 human liver cDNA library was screened with a monospecific antibody to human apoB-100 (GR-22) and with synthetic oligonucleotides SN-R3-1 (22) and SN-MDB-18 (23) that were based on the corresponding peptides present in apoB-100. Four overlapping apoB-100 cDNA clones were isolated. Clone λMDB-1 is 1.9 kilobases (kb) long and located 1.5 kb from the NH_2-terminal end of apoB-100 (8, 9). The other three overlapping clones (λMDB-2, λMDB-3, λMDB-4) are located 0.7 kb away from the 3′ end of clone MDB-1 and have a combined length of 4.2 kb (Fig. 1). To obtain the entire sequence of apoB-100 mRNA, 30-base synthetic oligonucleotides homologous to the 5′ and 3′ end of our apoB cDNA inserts were employed as hybridization probes to screen several other cDNA libraries. One of the clones, pMDB-7 (Fig. 1) was isolated from a human liver cDNA library, which used the *Pvu* II site of pAT153/Pvu II/8 (21). This clone hybridized to oligonucleotide probes at the 5′ end of λMDB-1 and λMDB-2 and thus contained a minimum of 4 kb of the 5′ end of the apoB-100 mRNA. Restriction enzyme analysis revealed an insert size of >6.0 kb indicating this clone included sequences bridging the gap between clones λMDB-1 and λMDB-2. Clones containing the 3′ half of the apoB-100 mRNA were obtained by screening the plasmid cDNA library (18, 20) with a nick-translated cDNA probe of λMDB-4 and synthetic oligonucleotides based on the sequence of the apoB peptide MDB-18. Several clones were isolated, clone pMDB-5 (Fig. 1) contained a 3.0-kb insert overlapping the 3′ end of clone λMDB-4. Clone pMDB-6 contained the 3′-untranslated region of apoB-100 mRNA and 750 base pairs of the coding region. Several clones containing the rest of the apoB-100 mRNA sequence were obtained from a plasmid cDNA library using the 30-base synthetic oligonucleotides homologous to clones pMDB-5 and pMDB-6.

Each of the putative apoB-100 cDNAs was verified by both RNA gel and Southern hybridization analysis. Oligonucleotides of each clone were also synthesized and used as hybridization probes to human liver poly(A)$^+$ RNA by RNA gel blot analysis. All cDNA and synthetic probes hybridized to the 14.1-kb apoB-100 mRNA band. In addition, monospecific antibodies prepared against synthetic peptides based on the derived amino acid sequence of the cDNA clones immunoblotted to apoB-100 separated by $NaDodSO_4$/PAGE (23).

Restriction Endonuclease Cleavage Map and Sequence Determination of ApoB-100 mRNA. The restriction endonuclease cleavage map of apoB-100 mRNA is illustrated in Fig. 1. cDNA inserts were isolated and subcloned into M13mp18 and M13mp19. In some instances, a modification of the dideoxy sequencing protocol using end-labeling was employed (27). Deletion subcloning (28) was used when sequencing large cDNA inserts as in the case of clone pMDB-7. Both the plus and minus strands were sequenced, and sequence data were edited with the aid of a computer.

Structural Analysis of ApoB-100 mRNA and the ApoB-100 Protein. The results of our sequence studies established that the apoB-100 mRNA is 14.1 kb long encoding a plasma B-100 apolipoprotein of 4536 amino acid residues with an amino acid molecular weight of 512,723. The completed sequence of apoB-100 mRNA is in good agreement with the published partial sequences of apoB-100 (8–16), except for the 3′-untranslated region that is different from that reported by Knott *et al.* (10) in many places. Computer analysis of the nucleic acid and derived amino acid sequence revealed several interesting features of apoB-100. Twelve out of a total of 25 cysteine residues are located within the first 500 amino acids at the amino terminus of apoB-100 (Fig. 2). There are 20 potential glycosylation sites, the majority of which are located in the middle of the protein.

DISCUSSION

In the present report we have described the complete nucleic acid and derived amino acid sequence of human liver mature apoB-100. At 4536 amino acids, to our knowledge, apoB-100 is the largest protein cloned and sequenced. The structure and physicochemical properties of apoB-100 have been extensively studied for nearly a decade. However, the analysis of this protein has been difficult because delipidated apoB-100 is insoluble in aqueous solution and forms aggregates even in buffers containing $NaDodSO_4$, urea, or guanidine hydrochloride. Our cloning of the entire apoB-100 cDNA permitted the definitive determination of the molecular weight of apoB-100 protein as 512,723. The first 500 residues of apoB-100 contains 12 cysteines that may form intramolecular disulfide bridges resulting in the NH_2-terminal region of apoB-100 being folded into a globular structure. In this respect, this region of apoB-100 is similar to that of human serum albumin (29, 30) that has 34 cysteine residues forming 17 pairs of intramolecular disulfide bonds. It is known that the introns of the serum albumin gene are located in a characteristic manner at the 5′ end and reflect the domain structure of the protein. Based on the similarity in structure of the two proteins, we can anticipate that there are multiple introns in the first 500 amino acid region of apoB-100.

The secondary structure of apoB-100 based on computer analysis revealed 40% α-helix, 25% β-structure, and 35% random coil but very little amphipathic helical structure. Thus, unlike other apolipoproteins, amphipathic helices are not a characteristic feature of apoB-100. ApoB-100, however, has a greater percent of β-structure when compared to other apolipoproteins, and segments of the β-structure contain distinct hydrophobic and hydrophilic faces.

The elucidation of the complete amino acid sequence of apoB-100 will permit a detailed analysis of the conformation of the protein, the LDL receptor binding domain(s), the structural relationship between apoB-100 and apoB-48, and the ability to study the genetic defects in apoB structure in patients with dyslipoproteinemias.

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. Goldstein, J. L. & Brown, M. S. (1977) *Annu. Rev. Biochem.* **46**, 897–930.
6. Elovson, J., Jacobs, J. C., Schumaker, V. N. & Puppione, D. L. (1985) *Biochemistry* **24**, 1567–1578.
7. Cardin, A. D., Witt, K. R., Chao, J., Margolis, H. S., Donaldson, V. H. & Jackson, R. L. (1984) *J. Biol. Chem.* **259**, 8522–8528.
8. Law, S. W., Lee, N., Monge, J. C., Sakaguchi, A. Y., Naylor, S. L. & Brewer, H. B., Jr. (1985) *Biochem. Biophys. Res. Commun.* **131**, 1003–1012.
9. Law, S. W., Lackner, K. J., Hospattankar, A. V., Anchors, J. M., Sakaguchi, A. Y., Naylor, S. L. & Brewer, H. B., Jr. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8340–8344.
10. Knott, T. J., Rall, S. C., Innerarity, T. L., Jacobson, S. F., Urdea, M. S., Levy-Wilson, B., Powell, L. M., Pease, R. J., Eddy, R., Nakai, H., Byers, M., Priestley, L. M., Robertson, E., Rall, L. B., Betsholtz, C., Shows, T. B., Mahley, R. W. & Scott, J. (1985) *Science* **230**, 37–43.
11. Protter, A. A., Hardman, D. A., Schilling, J. W., Miller, J., Appleby, V., Chen, G. C., Kirsher, S. W., McEnroe, G. & Kane, J. P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1467–1471.
12. Wei, C.-F., Chen, S.-H., Yang, C.-Y., Marcel, Y. L., Milne, R. W., Li, W.-H., Sparrow, J. T., Gotto, A. M., Jr. & Chan, L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7265–7269.
13. Mehrabian, M., Schumaker, V. N., Fareed, G. C., West, R., Johnson, D. F., Kirchgessner, T., Lin, H. C., Wang, X., Ma, Y., Mendiaz, E. & Lusis, A. J. (1985) *Nucleic Acids Res.* **13**, 6937–6953.
14. Carlsson, P., Olofsson, S. O., Bondjers, G., Darnfors, C., Wiklund, O. & Bjursell, G. (1985) *Nucleic Acids Res.* **13**, 8813–8826.
15. Deeb, S. S., Motulsky, A. G. & Alberts, J. J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4983–4986.
16. Shoulders, C. C., Myant, N. B., Sidoli, A., Rodriquez, J. C., Cortese, C., Baralle, F. E. & Cortese, R. (1985) *Atherosclerosis* **58**, 277–289.
17. Lusis, A. J., West, R., Mehrabian, M., Reuben, M. A., LeBoeuf, R. C., Kaptein, J. S., Johnson, D. F., Schumaker, V. N., Yuhasz, M. P., Schotz, M. C. & Elovson, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4597–4601.
18. Law, S. W. & Brewer, H. B., Jr. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 66–70.
19. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **24**, 5294–5299.
20. Law, S. W., Gray, G. & Brewer, H. B., Jr. (1983) *Biochem. Biophys. Res. Commun.* **112**, 257–264.
21. Tertia Belt, K., Carroll, M. C. & Porter, R. R. (1984) *Cell* **36**, 907–914.
22. LeBoeuf, R. C., Miller, C., Shively, J. E., Schumaker, V. N., Balla, M. A. & Lusis, A. J. (1984) *FEBS Lett.* **170**, 105–108.
23. Hospattankar, A. V., Fairwell, T., Meng, M., Ronan, R. & Brewer, H. B., Jr. (1986) *J. Biol. Chem.* **20**, 9102–9104.
24. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
25. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
26. Yanish-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
27. Brunner, A. M., Schimenti, J. C. & Duncan, C. H. (1986) *Biochemistry*, in press.
28. Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
29. Law, S. W. & Dugaiczyk, A. (1981) *Nature (London)* **291**, 201–205.
30. Minghetti, P. P., Ruffner, D. E., Kuang, W. J., Dennison, D. E., Hawkins, J. W., Beattie, W. G. & Dugaiczyk, A. (1986) *J. Biol. Chem.* **261**, 6747–6757.