Biochemistry

Volume 25, Number 19 September 23, 1986

Accelerated Publications

Complementary DNA Derived Structure of the Amino-Terminal Domain of Human Apolipoprotein B and Size of Its Messenger RNA Transcript†

Christos Cladaras,*,‡ Margarita Hadzopoulou-Cladaras,‡ Rafael Avila,§ Alexander L. Nussbaum,∥ Robert Nicolosi,⊥ and Vassilis I. Zannis*,‡

Section of Molecular Genetics, Cardiovascular Institute, Department of Medicine and Biochemistry, Boston University Medical Center, Boston, Massachusetts 02118, Department of Biochemistry, Harvard Medical School, Boston, Massachusetts 02115, Department of Clinical Sciences, University of Lowell, Lowell, Massachusetts 01854, and Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115

Received June 25, 1986; Revised Manuscript Received July 31, 1986

ABSTRACT: In this paper the sequence of a 5.2-kilobase (kb) cDNA covering the amino-terminal domain of human apolipoprotein B-100 (apoB-100) is reported. The cDNA-derived protein sequence provides the primary structure of 1748 amino acids. This segment of apoB-100 is more hydrophilic than hydrophobic and contains short stretches of predicted helical and β structures that are interrupted by β turns. Blotting analysis of RNA isolated from fetal human and adult monkey tissues and various human cell lines showed synthesis of apoB mRNA by liver and intestine and by cells of hepatic (HepG2) and intestinal (Caco-2) origin. The isolation and characterization of overlapping cDNA clones, which provide a nearly full-length copy of human apoB-100, are also reported. From the length of these clones the size of the cytoplasmic apoB mRNA is estimated to be 14.0 kb and codes for a protein of approximately 512 000 daltons.

Apolipoprotein B (apoB) is a glycoprotein of M_r 400 000–550 000 (Jacobs et al., 1985; Siuta-Mangano et al., 1982; Kane et al., 1980) and comprises 25% of the weight of the low-density lipoprotein (LDL) particle (Goldstein & Brown, 1977). Several studies have shown that human apoB and rat plasma apoB exist in two primary forms, designated B-100 and B-48 (Kane et al., 1980; Kane, 1983). Two other apoB forms, designated B-74 and B-26, may represent degradation products of the B-100 form (Kane et al., 1980). ApoB plays a crucial role in the formation of chylomicrons, very low-density lipoprotein (VLDL), and LDL since these particles are absent from plasma of patients with abetalipoproteinemia, who lack both forms of apoB (Herbert et al., 1982). Recently, a patient has been described who lacks plasma VLDL and LDL (but not chylomicrons) and has a selective deficiency of the B-100 form of apoB (Malloy et al., 1981). This observation suggested but did not prove that the B-100 and B-48 forms might be the products of different genes.

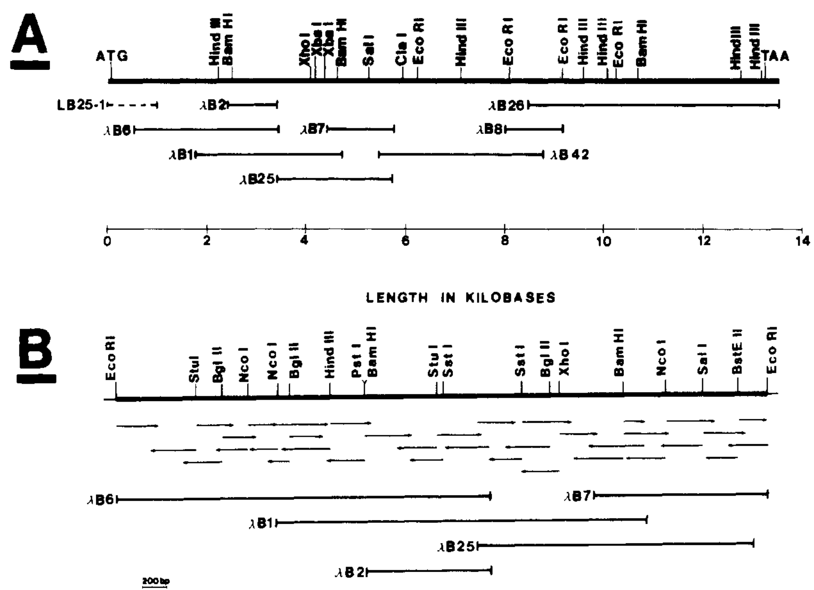
ApoB is the protein determinant for the cellular recognition and catabolism of LDL by the LDL (B/E) receptor (Goldstein & Brown, 1977, 1982). The LDL receptor–apoB interaction and subsequent catabolism mediate the clearance of LDL from plasma and regulate cellular cholesterol biosynthesis (Goldstein & Brown, 1977, 1982). Thus, apoB is thought to play a crucial role in the maintenance of cellular cholesterol homeostasis as well as in the pathogenesis of atherosclerosis (Goldstein & Brown, 1977, 1982). In spite of its importance, studies on the structure and function of apoB have been hampered by its high molecular weight and its unusual physical and chemical properties in the delipidated state (Lee et al., 1981).

Several laboratories have reported recently the isolation and characterization of partial apoB cDNA clones (Deeb et al., 1985; Knott et al., 1985; Huang et al., 1985; Mehrabian et al., 1985; Wei et al., 1985; Carlsson et al., 1985; Shoulders et al., 1985; Law et al., 1985; Protter et al., 1986) covering

† This work was supported by grants from the National Science Foundation (DCB8400173) and the National Institutes of Health (HL26335 and HL36200). V.I.Z. is an Established Investigator of the American Heart Association. R.A. is a fellow of the National Council of Science and Technology of Mexico. This research was performed at Housman Medical Research Center of Boston University Medical Center.

‡ Boston University Medical Center.
§ Harvard School of Public Health.
∥ Harvard Medical School.
⊥ University of Lowell.

0006-2960/86/0425-5351$01.50/0



C

A.A. / Nuc.

TAC ATC CTG AAC ATC AAG AGG GGC ATC ATT TCT GCC CTC CTG GTT CCC CCA GAG ACA GAA GAA GCC AAG CAA GTG TTG TTT CTG GAT ACC 90
Tyr Ile Leu Asn Ile Lys Arg Gly Ile Ile Ser Ala Leu Leu Val Pro Pro Glu Thr Glu Glu Ala Lys Gln Val Leu Phe Leu Asp Thr

GTG TAT GGA AAC TGC TCC ACT CAC TTT ACC GTC AAG ACG AGG AAG GGC AAT GTG GCA ACA GAA ATA TCC ACT GAA AGA GAC CTG GGC CAG 180
31 Val Tyr Gly Asn Cys Ser Thr His Phe Thr Val Lys Thr Arg Lys Gly Asn Val Ala Thr Glu Ile Ser Thr Glu Arg Asp Leu Gly Gln

TGT GAT CGC TTC AAG CCC ATC CGC ACA GGC ATC AGC CCA CTT GCT CTC ATC AAA GGC ATG ACC CGC CCC TTG TCA ACT CTG ATC AGC AGC 270
61 Cys Asp Arg Phe Lys Pro Ile Arg Thr Gly Ile Ser Pro Leu Ala Leu Ile Lys Gly Met Thr Arg Pro Leu Ser Thr Leu Ile Ser Ser

AGC CAG TCC TGT CAG TAC ACA CTG GAC GCT AAG AGG AAG CAT GTG GCA GAA GCC ATC TGC AAG GAG CAA CAC CTC TTC CTG CCT TTC TCC 360
91 Ser Gln Ser Cys Gln Tyr Thr Leu Asp Ala Lys Arg Lys His Val Ala Glu Ala Ile Cys Lys Glu Gln His Leu Phe Leu Pro Phe Ser

TAC AAG AAT AAG TAT GGG ATG GTA GCA CAA GTG ACA CAG ACT TTG AAA CTT GAA GAC ACA CCA AAG ATC AAC AGC CGC TTC TTT GGT GAA 450
121 Tyr Lys Asn Lys Tyr Gly Met Val Ala Gln Val Thr Gln Thr Leu Lys Leu Glu Asp Thr Pro Lys Ile Asn Ser Arg Phe Phe Gly Glu

GGT ACT AAG AAG ATG GGC CTC GCA TTT GAG AGC ACC AAA TCC ACA TCA CCT CCA AAG CAG GCC GAA GCT GTT TTG AAG ACT CTC CAG GAA 540
151 Gly Thr Lys Lys Met Gly Leu Ala Phe Glu Ser Thr Lys Ser Thr Ser Pro Pro Lys Gln Ala Glu Ala Val Leu Lys Thr Leu Gln Glu

CTG AAA AAA CTA ACC ATC TCT GAG CAA AAT ATC CAG AGA GCT AAT CTC TTC AAT AAG CTG GTT ACT GAG CTG AGA GGC CTC AGT GAT GAA 630
181 Leu Lys Lys Leu Thr Ile Ser Glu Gln Asn Ile Gln Arg Ala Asn Leu Phe Asn Lys Leu Val Thr Glu Leu Arg Gly Leu Ser Asp Glu

GCA GTC ACA TCT CTC TTG CCA CAG CTG ATT GAG GTG TCC AGC CCC ATC ACT TTA CAA GCC TTG GTT CAG TGT GGA CAG CCT CAG TGC TCC 720
211 Ala Val Thr Ser Leu Leu Pro Gln Leu Ile Glu Val Ser Ser Pro Ile Thr Leu Gln Ala Leu Val Gln Cys Gly Gln Pro Gln Cys Ser

ACT CAC ATC CTC CAG TGG CTG AAA CGT GTG CAT GCC AAC CCC CTT CTG ATA GAT GTG GTC ACC TAC CTG GTG GCC CTG ATC CCC GAG CCC 810
241 Thr His Ile Leu Gln Trp Leu Lys Arg Val His Ala Asn Pro Leu Leu Ile Asp Val Val Thr Tyr Leu Val Ala Leu Ile Pro Glu Pro

TCA GCA CAG CAG CTG CGA GAG ATC TTC AAC ATG GCG AGG GAT CAG CGC AGC CGA GCC ACC TTG TAT GCG CTG AGC CAC GCG GTC AAC AAC 900
271 Ser Ala Gln Gln Leu Arg Glu Ile Phe Asn Met Ala Arg Asp Gln Arg Ser Arg Ala Thr Leu Tyr Ala Leu Ser His Ala Val Asn Asn

TAT CAT AAG ACA AAC CCT ACA GGG ACC CAG GAG CTG CTG GAC ATT GCT AAT TAC CTG ATG GAA CAG ATT CAA GAT GAC TGC ACT GGG GAT 990
301 Tyr His Lys Thr Asn Pro Thr Gly Thr Gln Glu Leu Leu Asp Ile Ala Asn Tyr Leu Met Glu Gln Ile Gln Asp Asp Cys Thr Gly Asp

GAA GAT TAC 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 1080
331 Glu Asp Tyr 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

AAA TGT GTC CAA AGT ACA AAG CCA TCA CTG ATG ATC CAG AAA GCT GCC ATC CAG GCT CTG CGG AAA ATG GAG CCT AAA GAC AAG GAC CAG 1170
361 Lys Cys Val Gln Ser Thr Lys Pro Ser Leu Met Ile Gln Lys Ala Ala Ile Gln Ala Leu Arg Lys Met Glu Pro Lys Asp Lys Asp Gln

GAG GTT CTT CTT CAG ACT TTC CTT GAT GAT GCT TCT CCG GGA GAT AAG CGA CTG GCT GCC TAT CTT ATG TTG ATG AGG AGT CCT TCA CAG 1260
391 Glu Val Leu Leu Gln Thr Phe Leu Asp Asp Ala Ser Pro Gly Asp Lys Arg Leu Ala Ala Tyr Leu Met Leu Met Arg Ser Pro Ser Gln

GCA GAT ATT AAC AAA ATT GTC CAA ATT CTA CCA TGG GAA CAG AAT GAG CAA GTG AAG AAC TTT GTG GCT TCC CAT ATT GCC AAT ATC TTG 1350
421 Ala Asp Ile Asn Lys Ile Val Gln Ile Leu Pro Trp Glu Gln Asn Glu Gln Val Lys Asn Phe Val Ala Ser His Ile Ala Asn Ile Leu

AAC TCA GAA GAA TTG GAT ATC CAA GAT CTG AAA AAG TTA GTG AAA GAA GTT CTG AAA GAA TCT CAA CTT CCA ACT GTC ATG GAC TTC AGA 1440
451 Asn Ser Glu Glu Leu Asp Ile Gln Asp Leu Lys Lys Leu Val Lys Glu Val Leu Lys Glu Ser Gln Leu Pro Thr Val Met Asp Phe Arg

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 ATA TTT 1530
481 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 Ile Phe

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 1620
511 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

TTG GAA GGA 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 1710
541 Leu Glu Gly 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

GTT AAT GGT CAA GTT CCT GAT GGT GTC TCT AAG GTC TTA GTG GAC CAC TTT GGC TAT ACC AAA GAT GAT AAA CAT GAG CAG GAT ATG GTA 1800
571 Val Asn Gly Gln Val Pro Asp Gly Val Ser Lys Val Leu Val Asp His Phe Gly Tyr Thr Lys Asp Asp Lys His Glu Gln Asp Met Val

AAT GGA ATA ATG CTC AGT GTT GAG AAG CTG ATT AAA GAT TTG AAA TCC AAA GAA GTC CCG GAA GCC AGA GCC TAC CTC CGC ATC TTG GGA 1890
601 Asn Gly Ile Met Leu Ser Val Glu Lys Leu Ile Lys Asp Leu Lys Ser Lys Glu Val Pro Glu Ala Arg Ala Tyr Leu Arg Ile Leu Gly

GAG GAG CTT GGT TTT GCC AGT CTC CAT GAC CTC CAG CTC CTG GGA AAG CTG CTT CTG ATG GGT GCC CGC ACT CTG CAG GGG ATC CCC CAG 1980
631 Glu Glu Leu Gly Phe Ala Ser Leu His Asp Leu Gln Leu Leu Gly Lys Leu Leu Leu Met Gly Ala Arg Thr Leu Gln Gly Ile Pro Gln

ATG ATT GGA GAG GTC ATC AGG AAG GGC TCA AAG AAT GAC TTT TTT CTT CAC TAC ATC TTC ATG GAG AAT GCC TTT GAA CTC CCC ACT GGA 2070
661 Met Ile Gly Glu Val Ile Arg Lys Gly Ser Lys Asn Asp Phe Phe Leu His Tyr Ile Phe Met Glu Asn Ala Phe Glu Leu Pro Thr Gly

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 CAG GCT 2160
691 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 Gln Ala

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 2250
721 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

AAC ACC AAC 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 2340
751 Asn Thr Asn 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

```
    CCA GTC AAG CTG CTC AGT GGA GGC AAC ACA TTA CAT TTG GTC TCT ACC ACC AAA ACG GAG GTG ATC CCA CCT CTC ATT GAG AAC AGG CAG 2430
781 Pro Val Lys Leu Leu Ser Gly Gly Asn Thr Leu His Leu Val Ser Thr Thr Lys Thr Glu Val Ile Pro Pro Leu Ile Glu Asn Arg Gln

    TCC TGG TCA GTT TGC AAG CAA GTC TTT CCT GGC CTG AAT TAC TGC ACC TCA GGC GCT TAC TCC AAC GCC AGC TCC ACA GAC TCC GCC TCC 2520
811 Ser Trp Ser Val Cys Lys Gln Val Phe Pro Gly Leu Asn Tyr Cys Thr Ser Gly Ala Tyr Ser Asn Ala Ser Ser Thr Asp Ser Ala Ser

    TAC TAT CCG CTG ACC GGG GAC ACC AGA TTA GAG CTG GAA CTG AGG CCT ACA GGA GAG ATT GAG CAG TAT TCT GTC AGC GCA ACC TAT GAG 2610
841 Tyr Tyr Pro Leu Thr Gly Asp Thr Arg Leu Glu Leu Glu Leu Arg Pro Thr Gly Glu Ile Glu Gln Tyr Ser Val Ser Ala Thr Tyr Glu

    CTC CAG AGA GAG GAC AGA GCC TTG GTG GAT ACC CTG AAG TTT GTA ACT CAA GCA GAA GGT GCG AAG CAG ACT GAG GCT ACC ATG ACA TTC 2700
871 Leu Gln Arg Glu Asp Arg Ala Leu Val Asp Thr Leu Lys Phe Val Thr Gln Ala Glu Gly Ala Lys Gln Thr Glu Ala Thr Met Thr Phe

    AAA TAT AAT CGG CAG AGT ATG ACC TTG TCC AGT GAA GTC CAA ATT CCG GAT TTT GAT GTT GAC CTC GGA ACA ATC CTC AGA GTT AAT GAT 2790
901 Lys Tyr Asn Arg Gln Ser Met Thr Leu Ser Ser Glu Val Gln Ile Pro Asp Phe Asp Val Asp Leu Gly Thr Ile Leu Arg Val Asn Asp

    GAA TCT ACT GAG GGC AAA ACG TCT TAC AGA CTC ACC CTG GAC ATT CAG AAC AAG AAA ATT ACT GAG GTC GCC CTC ATG GGC GAC CTA AGT 2880
931 Glu Ser Thr Glu Gly Lys Thr Ser Tyr Arg Leu Thr Leu Asp Ile Gln Asn Lys Lys Ile Thr Glu Val Ala Leu Met Gly Asp Leu Ser

    TGT GAC ACA AAG GAA GAA AGA AAA ATC AAG GGT GTT ATT TCC ATA CCC CGT TTG CAA GCA GAA GCC AGA AGT GAG ATC CTC GCC CAC TGG 2970
961 Cys Asp Thr Lys Glu Glu Arg Lys Ile Lys Gly Val Ile Ser Ile Pro Arg Leu Gln Ala Glu Ala Arg Ser Glu Ile Leu Ala His Trp

    TCG CCT GCC AAA CTG CTT CTC CAA ATG GAC TCA TCT GCT ACA GCT TAT GGC TCC ACA GTT TCC AAG AGG GTG GCA TGG CAT TAT GAT GAA 3060
991 Ser Pro Ala Lys Leu Leu Leu Gln Met Asp Ser Ser Ala Thr Ala Tyr Gly Ser Thr Val Ser Lys Arg Val Ala Trp His Tyr Asp Glu

    GAG AAG ATT GAA TTT GAA TGG AAC ACA GGC ACC AAT GTA GAT ACC AAA AAA ATG ACT TCC AAT TTC CCT GTG GAT CTC TCC GAT TAT CCT 3150
1021 Glu Lys Ile Glu Phe Glu Trp Asn Thr Gly Thr Asn Val Asp Thr Lys Lys Met Thr Ser Asn Phe Pro Val Asp Leu Ser Asp Tyr Pro

    AAG AGC TTG CAT ATG TAT GCT AAT AGA CTC CTG GAT CAC AGA GTC CCT CAA ACA GAC ATG ACT TTC CGG CAC GTG GGT TCC AAA TTA ATA 3240
1051 Lys Ser Leu His Met Tyr Ala Asn Arg Leu Leu Asp His Arg Val Pro Gln Thr Asp Met Thr Phe Arg His Val Gly Ser Lys Leu Ile

    GTT GCA ATG AGC TCA TGG CTT CAG AAG GCA TCT GGG AGT CTT CCT TAT ACC CAG ACT TTG CAA GAC CAC CTC AAT AGC CTG AAG GAG TTC 3330
1081 Val Ala Met Ser Ser Trp Leu Gln Lys Ala Ser Gly Ser Leu Pro Tyr Thr Gln Thr Leu Gln Asp His Leu Asn Ser Leu Lys Glu Phe

    AAC CTC CAG AAC ATG GGA TTG CCA GAC TTC CAC ATC CCA GAA AAC CTC TTC TTA AAA AGC GAT GGC CGG GTC AAA TAT ACC TTG AAC AAG 3420
1111 Asn Leu Gln Asn Met Gly Leu Pro Asp Phe His Ile Pro Glu Asn Leu Phe Leu Lys Ser Asp Gly Arg Val Lys Tyr Thr Leu Asn Lys

    AAC AGT TTG AAA ATT GAG ATT CCT TTG CCT TTT GGT GGC AAA TCC TCC AGA GAT CTA AAG ATG TTA GAG ACT GTT AGG ACA CCA GCC CTC 3510
1141 Asn Ser Leu Lys Ile Glu Ile Pro Leu Pro Phe Gly Gly Lys Ser Ser Arg Asp Leu Lys Met Leu Glu Thr Val Arg Thr Pro Ala Leu

    CAC TTC AAG TCT GTG GGA TTC CAT CTG CCA TCT CGA GAG TTC CAA GTC CCT ACT TTT ACC ATT CCC AAG TTG TAT CAA CTG CAA GTG CCT 3600
1171 His Phe Lys Ser Val Gly Phe His Leu Pro Ser Arg Glu Phe Gln Val Pro Thr Phe Thr Ile Pro Lys Leu Tyr Gln Leu Gln Val Pro

    CTC CTG GGT GTT CTA GAC CTC TCC ACG AAT GTC TAC AGC AAC TTG TAC AAC TGG TCC GCC TCC TAC AGT GGT GGC AAC ACC AGC ACA GAC 3690
1201 Leu Leu Gly Val Leu Asp Leu Ser Thr Asn Val Tyr Ser Asn Leu Tyr Asn Trp Ser Ala Ser Tyr Ser Gly Gly Asn Thr Ser Thr Asp

    CAT TTC AGC CTT CGG GCT CGT TAC CAC ATG AAG GCT GAC TCT GTG GTT GAC CTG CTT TCC TAC AAT GTG CAA GGA TCT GGA GAA ACA ACA 3780
1231 His Phe Ser Leu Arg Ala Arg Tyr His Met Lys Ala Asp Ser Val Val Asp Leu Leu Ser Tyr Asn Val Gln Gly Ser Gly Glu Thr Thr

    TAT GAC CAC AAG AAT ACG TCT ACA CTA TCA TGT GAT GGG TCT CTA CGC CAC AAA TTT CTA GAT TCG AAT ATC AAA TTC AGT CAT GTA GAA 3870
1261 Tyr Asp His Lys Asn Thr Phe Thr Leu Ser Cys Asp Gly Ser Leu Arg His Lys Phe Leu Asp Ser Asn Ile Lys Phe Ser His Val Glu

    AAA CTT GGA AAC AAC CCA GTC TCA AAA GGT TTA CTA ATA TTC GAT GCA TCT AGT TCC ATG GGA CCA CAG ATG TCT GCT TCA GTT CAT TTG 3960
1291 Lys Leu Gly Asn Asn Pro Val Ser Lys Gly Leu Leu Ile Phe Asp Ala Ser Ser Ser Met Gly Pro Gln Met Ser Ala Ser Val His Leu

    GAC TCC AAA AAG AAA CAG CAT TTG TTT GTC AAA GAA GTC AAG ATT GAT GGG CAG TTC AGA GTC TCT TCG TTC TAT GCT AAA GGC ACA TAT 4050
1321 Asp Ser Lys Lys Lys Gln His Leu Phe Val Lys Glu Val Lys Ile Asp Gly Gln Phe Arg Val Ser Ser Phe Tyr Ala Lys Gly Thr Tyr

    GGC CTG TCT TGT CAG AGG GAT CCT AAC ACT GGC CGG CTC AAT GGA GAG TCC AAC CTG AGG TTT AAC TCC TCC TAC CTC CAA GGC ACC AAC 4140
1351 Gly Leu Ser Cys Gln Arg Asp Pro Asn Thr Gly Arg Leu Asn Gly Glu Ser Asn Leu Arg Phe Asn Ser Ser Tyr Leu Gln Gly Thr Asn

    CAG ATA ACA GGA AGA TAT GAA GAT GGA ACC CTC TCC CTC ACC TCC ACC TCT GAT CTG CAA AGT GGC ATC ATT AAA AAT ACT GCT TCC CTA 4230
1381 Gln Ile Thr Gly Arg Tyr Glu Asp Gly Thr Leu Ser Leu Thr Ser Thr Ser Asp Leu Gln Ser Gly Ile Ile Lys Asn Thr Ala Ser Leu

    AAG TAT GAG AAC TAC GAG CTG ACT TTA AAA TCT GAC ACC AAT GGG AAG TAT AAG AAC TTT GCC ACT TCT AAC AAG ATG GAT ATG ACC TTC 4320
1411 Lys Tyr Glu Asn Tyr Glu Leu Thr Leu Lys Ser Asp Thr Asn Gly Lys Tyr Lys Asn Phe Ala Thr Ser Asn Lys Met Asp Met Thr Phe

    TCT AAG CAA AAT GCA CTG CTG CGT TCT GAA TAT CAG GCT GAT TAC GAG TCA TTG AGG TTC TTC AGC CTG CTT TCT GGA TCA CTA AAT TCC 4410
1441 Ser Lys Gln Asn Ala Leu Leu Arg Ser Glu Tyr Gln Ala Asp Tyr Glu Ser Leu Arg Phe Phe Ser Leu Leu Ser Gly Ser Leu Asn Ser

    CAT GGT CTT GAG TTA AAT GCT GAC ATC TTA GGC ACT GAC AAA ATT AAT AGT GGT GCT CAC AAG GCG ACA CTA AGG ATT GGC CAA GAT GGA 4500
1471 His Gly Leu Glu Leu Asn Ala Asp Ile Leu Gly Thr Asp Lys Ile Asn Ser Gly Ala His Lys Ala Thr Leu Arg Ile Gly Gln Asp Gly

    ATA TCT ACC AGT GCA ACG ACC AAC TTG AAG TGT AGT CTC CTG GTG CTG GAG AAT GAG CTG AAT GCA GAG CTT GGC CTC TCT GGG GCA TCT 4590
1501 Ile Ser Thr Ser Ala Thr Thr Asn Leu Lys Cys Ser Leu Leu Val Leu Glu Asn Glu Leu Asn Ala Glu Leu Gly Leu Ser Gly Ala Ser

    ATG AAA TTA ACA ACA AAT GGC CGC TTC AGG GAA CAC AAT GCA AAA TTC AGT CTG GAT GGG AAA GCC GCC CTC ACA GAG CTA TCA CTG GGA 4680
1531 Met Lys Leu Thr Thr Asn Gly Arg Phe Arg Glu His Asn Ala Lys Phe Ser Leu Asp Gly Lys Ala Ala Leu Thr Glu Leu Ser Leu Gly

    AGT GCT TAT CAG GCC ATG ATT CTG GGT GTC GAC AGC AAA AAC ATT TTC AAC TTC AAG GTC AGT CAA GAA GGA CTT AAG CTC TCA AAT GAC 4770
1561 Ser Ala Tyr Gln Ala Met Ile Leu Gly Val Asp Ser Lys Asn Ile Phe Asn Phe Lys Val Ser Gln Glu Gly Leu Lys Leu Ser Asn Asp

    ATG ATG GGC TCA TAT GCT GAA ATG AAA TTT GAC CAC ACA AAC AGT CTG AAC ATT GCA GGC TTA TCA CTG GAC TTC TCT TCA AAA CTT GAC 4860
1591 Met Met Gly Ser Tyr Ala Glu Met Lys Phe Asp His Thr Asn Ser Leu Asn Ile Ala Gly Leu Ser Leu Asp Phe Ser Ser Lys Leu Asp

    AAC ATT TAC AGC TCT GAC AAG TTT TAT AAG CAA ACT GTT AAT TTA CAG CTA CAG CCC TAT TCT CTG GTA ACT ACT TTA AAC AGT GAC CTG 4950
1621 Asn Ile Tyr Ser Ser Asp Lys Phe Tyr Lys Gln Thr Val Asn Leu Gln Leu Gln Pro Tyr Ser Leu Val Thr Thr Leu Asn Ser Asp Leu

    AAA TAC AAT GCT CTG GAT CTC ACC AAC AAT GGG AAA CTA CGG CTA GAA CCC CTG AAG CTG CAT GTG GCT GGT AAC CTA AAA GGA GCC TAC 5040
1651 Lys Tyr Asn Ala Leu Asp Leu Thr Asn Asn Gly Lys Leu Arg Leu Glu Pro Leu Lys Leu His Val Ala Gly Asn Leu Lys Gly Ala Tyr

    CAA AAT AAT GAA ATA AAA CAC ATC TAT GCC ATC TCT TCT GCT GCC TTA TCA GCA AGC TAT AAA GCA GAC ACT GTT GCT AAG GTT CAG GGT 5130
1681 Gln Asn Asn Glu Ile Lys His Ile Tyr Ala Ile Ser Ser Ala Ala Leu Ser Ala Ser Tyr Lys Ala Asp Thr Val Ala Lys Val Gln Gly

    GTG GAG TTT AGC CAT CGG CTC AAC ACA GAC ATC GCT GGG CTG GCT TCA GCC ATT GAC ATG AGC ACA AAC TAT AAT TCA GAC TCA CTG CAT 5220
1711 Val Glu Phe Ser His Arg Leu Asn Thr Asp Ile Ala Gly Leu Ala Ser Ala Ile Asp Met Ser Thr Asn Tyr Asn Ser Asp Ser Leu His

    TTC AGC AAT GTC TTC CGT TCT GTA                                                                                    5244
1741 Phe Ser Asn Val Phe Arg Ser Val
```

FIGURE 1: Panel A: Composite restriction map of the inserts of apoB cDNA clones providing the structure of a full-length apoB cDNA. The clone LB25-1 was reported by Protter et al. (1986), and it is indicated by a dashed line. ATG and TAA indicate the initiation and termination codons of apoB-100, respectively. Panel B: Composite restriction map of the inserts of apoB cDNA clones λB_1, λB_2, λB_6, λB_7, and λB_{25} and strategy used for their sequencing. The individual clones are shown at the bottom. Restriction sites used for sequencing are shown (not all *Nco*I and *Bst*EII sites are presented). Panel C: cDNA and the derived amino acid sequences of the five overlapping apoB cDNA clones. The protein sequence corresponding to peptide R3-1 is underlined. (+) indicates that the R3-1 peptide sequence has Lys at this position. The area corresponding to the synthetic oligonucleotide probe is boxed. The extent of overlap of our sequence with the sequences reported by Deeb et al. (1985), Law et al. (1985), and Protter et al. (1986) is enclosed between the symbols *, **, and ***, respectively. (●) indicates differences between this sequence and the previously published sequences as follows: The sequence reported by Deeb et al. (1985) has T, A, and C at residues 1815, 1946, and 1947, respectively, and deletions of nucleotide residues 1921, 1934, and 1974. These deletions change the amino acid sequence from residues 641 to 658. The sequence reported by Law et al. (1985) has C's at residues 1400 and 2670. The former substitution changes Val-467 to Ala. Clone λB_{25} has a G at position 2872, while clone λB_2 has a C at the same position. This G-C transversion results in a substitution of Asp-958 to His. The first amino acid of our sequence corresponds to amino acid 125 of the sequence reported by Protter et al. (1986).

mostly (Knott et al., 1985; Mehrabian et al., 1985; Wei et al., 1985; Carlsson et al., 1985; Shoulders et al., 1985) the carboxy-terminal region of apoB-100. The apoB gene has been mapped in the short arm of chromosome 2 (Knott et al., 1985; Law et al., 1985; Deeb et al., 1986), and the presence of apoB mRNA has been demonstrated in liver and intestine (Deeb et al., 1985, 1986; Knott et al., 1985; Huang et al., 1985; Mehrabian et al., 1985; Carlsson et al., 1985; Shoulders et al., 1985; Law et al., 1985).

We now report the identification and characterization of cDNA clones covering a nearly full-length copy of human apoB-100 (14.0 kb), as well as the complete cDNA sequence of a 5.2-kb amino-terminal domain along with the predicted secondary structure of the deduced protein segment. In addition, we report the distribution of apoB mRNA in fetal human and adult monkey tissues and cell lines of human origin.

Materials and Methods

Library Screening and Characterization of ApoB cDNA Clones. The adult human liver λgt10 cDNA library was a generous gift of Dr. E. F. Fritsch, Genetics Institute, Boston, MA. The double-stranded cDNAs were synthesized by oligo(dT) priming of adult human liver poly(A+) mRNA using reverse transcriptase and the Klenow fragment of DNA polymerase I. The subsequent steps for the preparation of the library were performed as described by Huynh et al. (1985) (E. F. Fritsch, personal communication). The oligonucleotide probe was synthesized by the solid-phase phosphite triester method (Caruthers, 1985) using an automated oligonucleotide synthesizer (Applied Biosystems 380-A) and purified by high-pressure liquid chromatography. The oligonucleotide probe used to screen the cDNA library was 17 nucleotides long and 64-fold degenerate [5′-CCCAT(A/G)TTNGTNAC-(A/G)AA-3′] and corresponded to the apoB amino acid sequence Phe-Val-Thr-Asn-Met-Gly belonging to peptide R3-1 (LeBoeuf et al., 1984).

Growth of recombinant λgt10 clones and phage cDNA purification were as described (Maniatis et al., 1982). For screening the cDNA library with the oligonucleotide probe, hybridizations were carried out at 33 °C in 5 × standard saline citrate (SSC), 5 × Denhardt's reagent, 0.05% sodium pyrophosphate, and 100 μg/mL tRNA with 1×10^6 cpm/filter; washes were at 37 °C in 6 × SSC (Lee et al., 1985). Positive clones were plaque-purified, and DNA isolated from these clones was subcloned into a pUC18 vector. Plasmid DNA was purified by centrifugation to equilibrium in cesium chloride–ethidium bromide density gradients (Maniatis et al., 1982). Sequencing was done by the method of Maxam and Gilbert (1977). Screening of the cDNA library with a nick-translated probe was done in 4 × SET (0.6 M NaCl, 0.08 M Tris-HCl, 4 mM EDTA, pH 7.8), 10 × Denhardt's reagent, 0.2% sodium dodecyl sulfate (SDS), and 50 μg/mL denatured salmon sperm DNA at 65 °C with 1×10^6 cpm/filter; washes were at 65 °C in 2 × SET twice and 1 × SET once (Beltz et al., 1983).

Computer Analysis of ApoB cDNA and Protein Sequences. Hydropathy plots were carried out with the hydropathicity program of Intelegenetics using a three-residue windowing average (Kyte & Doolittle, 1982). The predicted secondary structure was derived by a computer program of Intelegenetics. The program identifies secondary structural domains by using the Chou–Fasman pseudoprobabilities for protein sequence information (Chou & Fasman, 1978). Analysis for internal protein and DNA sequence homology was performed by a program described by Pustell and Kafatos (1986). Analysis for homologies to other proteins was performed by the computer program of Lipman and Pearson (1985) using the National Biomedical Research Foundation data bank (released as of Aug 1985).

Isolation and Blotting Analysis of RNA. Human fetal tissues were obtained from 20–22-week-old human abortuses under a protocol approved by the Research Advisory Committee of the Brigham and Women's Hospital, Boston, MA, as described previously (Zannis et al., 1985). RNA was prepared by the guanidine thiocyanate method and used for Northern and dot blotting analysis (Maniatis et al., 1982; Zannis et al., 1985). The blots were hybridized with the *Sst*I fragments of apoB clone λB_1 labeled by nick translation (Maniatis et al., 1982).

Results

Isolation and Sequence of Human ApoB cDNA Clones. By screening the adult human liver λgt10 cDNA library, we have isolated overlapping cDNA clones covering nearly the entire length copy of the apoB-100 (Figure 1A). Three cDNA clones were initially isolated by screening 50000 recombinant plaques with the oligonucleotide probe described under Materials and Methods. Two of these clones (λB_1 and λB_2) were found to contain a 3.0-kb and a 1.3-kb insert, respectively. The nucleotide sequence of the region that hybridized with the oligonucleotide probe was determined and found to encode an open reading frame that included a stretch of 24 amino acids identical with the sequence of apoB peptide R3-1 (LeBoeuf et al., 1984). Additional apoB cDNA clones extending in the 5′ and 3′ directions of λB_2 were isolated by screening of the λgt10 cDNA library with apoB nick-translated probes. The clones λB_6, λB_1, λB_2, λB_7, and λB_{25}, which cover a 5.2-kb domain of apoB-100, were completely sequenced. The individual clones, a composite restriction map of the five clones, and the sequencing strategy used are shown in Figure 1B. The nucleotide and derived amino acid sequence of these clones are shown in Figure 1C. The 5′ end of this sequence overlaps by 420 base pairs (bp) with the LB25-1 clone (Protter et al., 1986), which contains the 5′ end of the mRNA coding for part of the apoB-26. This overlap of our clone with the LB25-1 identifies the initiator methionine for the apoB-100 and also establishes that the apoB form B-26 represents the amino-terminal region of apoB-100. The clone λB_{26} is 5.6 kb in length and codes for the carboxy-terminal region of apoB-100. The presence of termination codon TAA was confirmed by partial DNA sequence analysis of this clone (data not shown). Published clones have also established the position of the terminator sequence (Knott et al., 1985; Wei et al., 1985).

Analysis of the DNA and Derived Protein Sequence of ApoB cDNA Clones. The hydropathy profile and the predicted secondary structure of the amino-terminal domain of apoB obtained by computer analysis are shown in Figure 2. This analysis showed that this sequence is more hydrophilic than hydrophobic and consists of 40% α-helical structure, 19% β structure, 17% β turns or loops, and 24% random coil. The amino acid composition of this segment of apoB agrees with the reported composition of apoB-100 (Kane et al., 1980). The positions of the six potential N-glycosylation sites and cysteine residues are also shown in Figure 2. The computer analysis did not show any substantial internal amino acid or nucleotide sequence homologies in this apoB segment. Comparison with the data bank protein sequences showed that the apoB segment from amino acid residue 410–570 has a 19% homology with 160 amino acid residues of streptococcal dihydrofolate reductase (Gleisner et al., 1974). Although this comparison gave a z value of 7.5 (probably significant) when compared to 50 randomly permuted dihydrofolate reductase sequences, significant homologies were not found with human and other

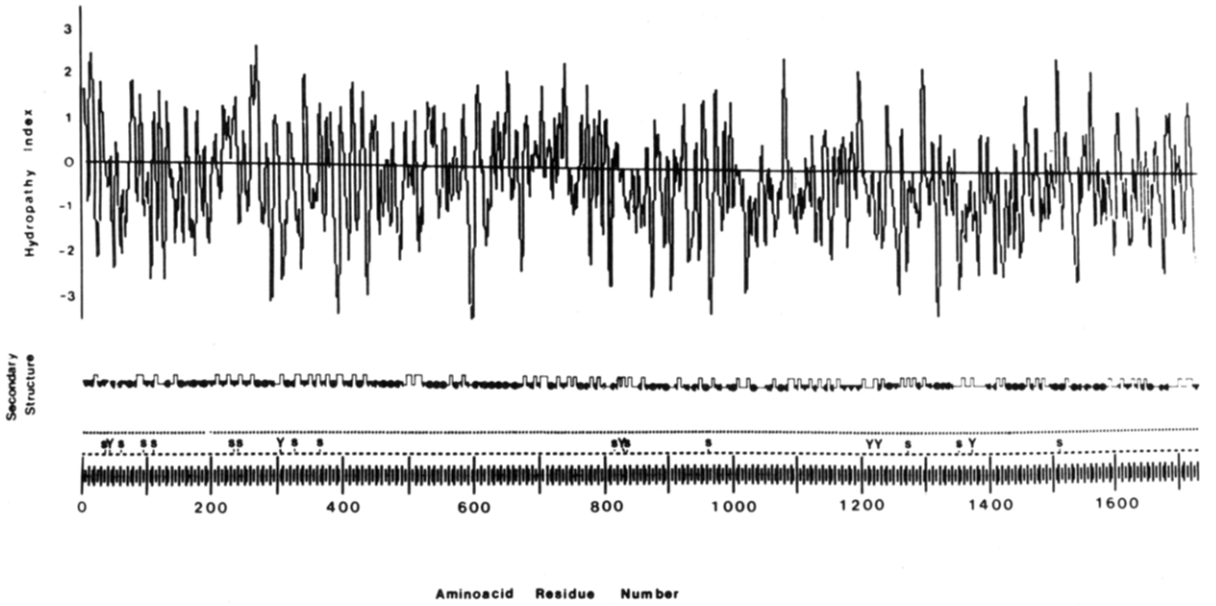


FIGURE 2: Predicted hydropathy profiles and secondary structure of the 1748 amino acid segment of human apoB. Hydropathy profiles and secondary structures were generated as described under Materials and Methods. The secondary structure is represented as follows: (●) 10 residues of α helix; (▼) 10 residues of β structure; (⊓) β-turn or loop structure; and (—) random coil. The amino acid residue numbers are indicated. S and Y designate the position of cysteine residues and the potential N-glycosylation sites, respectively.

mammalian dihydrofolate reductases.

Tissue Distribution of ApoB mRNA. Blotting analysis of RNA isolated from different tissues and cell lines showed the presence of apoB mRNA in fetal human and adult monkey liver and intestine, as well as in HepG2 cells and the human colon carcinoma cell line Caco-2 (Pinto et al., 1983). ApoB mRNA was absent from a variety of fetal human and adult monkey tissues and cells including adrenal gland, brain, gonads, spleen, lung, kidney, heart, stomach, muscle, lymph nodes, thyroid gland, artery, aorta, monocyte macrophages, and skin fibroblasts. Northern blotting of rabbit intestinal RNA gave two discrete species of approximate length 15 and 8 kb, respectively (Figure 3). Reproducible results were obtained in four different RNA preparations.

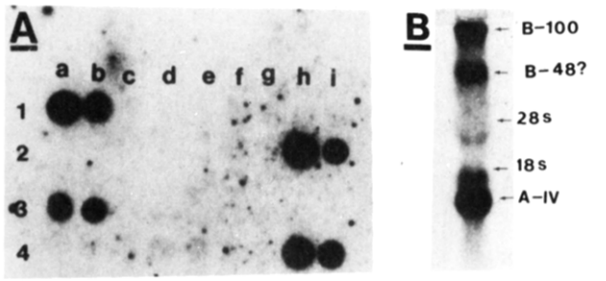


FIGURE 3: Panel A: Dot blot analysis of RNA isolated from various human and monkey tissues as well as of various cell lines of human origin. Rows 1 and 2 contain 10 μg of RNA obtained from fetal human tissues and human cells. Row 1: (a) liver, (b) intestine, (c) adrenal, (d) brain, (e) spleen, (f) gonads, (g) lung, (h) kidney, and (i) heart. Row 2: (a) stomach, (b) thymus, (c) muscle, (d) pancreas, (e) peripheral blood human monocyte macrophage cultures, (f) SV40 transformed human fibroblasts, (g) U937 cells, (h) HepG2 cells, and (i) Caco-2 cells. Rows 3 and 4 contain 10 μg of RNA isolated from monkey tissues. Row 3: same order of tissues as in row 1. Row 4: (a) stomach, (b) muscle, (c) lymph nodes, (d) thyroid, (e) artery, (f) aorta, (g) pancreas, (h) HepG2 cells, and (i) human liver. The monkey liver RNA was obtained from cebus, and all other RNAs were obtained from cynomolgus monkeys. Panel B: Blotting analysis of 30 μg of RNA isolated from rabbit liver. The blot was hybridized sequencially with human apoB and human apoA-IV cDNA probes (Karathanasis et al., 1986). The positions of the apoA-IV and the two apoB mRNA species, as well as of 28S and 18S rRNAs, are indicated.

Discussion

A well-characterized function of the human apoB is the binding to the LDL receptor, which leads to the cellular catabolism of LDL (Goldstein & Brown, 1977, 1982). Domains of this protein may be important for binding to heparin (Mahley et al., 1979) and to the immunoregulatory receptors of lymphocytes (Hui et al., 1980) and for the receptor-independent catabolism of LDL by arterial wall cells (Carew et al., 1984).

In this study we present overlapping cDNA clones that provide a nearly full-length copy of apoB-100. Our findings suggest that the size of the coding region of the cytoplasmic apoB-100 mRNA is 14.0 kb and codes for a protein of approximately 512 000 daltons. In addition, we have obtained the cDNA and derived protein sequence of a 5.2-kb amino-terminal region of apoB-100 as a step toward the elucidation of its function. Computer-aided analysis indicated that this region of apoB contains helical and β structures that are interrupted by β turns. The majority of α helices are in short stretches, and none of them shows the presence of a typical amphipathic α-helical structure. In contrast, the other apolipoproteins contain long amphipathic α-helical structures (Kaiser & Kezdy, 1983; Bogusky et al., 1984; Karathanasis et al., 1986). It is possible that the β structures of apoB may play a more important role for the lipid binding properties of this protein (Knott et al., 1985). ApoB-100 contains both high mannose and complex-type oligosaccharide chains (Siuta-Mangano et al., 1982). Examination of the derived amino acid sequence showed six potential N-linked glycosylation sites at amino acid residues 34, 305, 832, 1217, 1226, and 1372. The 305 and 1372 sites occur in β-turn structures and may represent true glycosylation sites (Beeley, 1977).

In this study analysis of a wide spectrum of fetal human and adult monkey tissues confines the synthesis of apoB-100 to fetal and adult liver and intestine. In addition, the rabbit intestinal mRNA shows the presence of two apoB mRNA species of approximate lengths 15 and 8 kb. The relationship of apoB-100 to the apoB-48 form and their specific expression by hepatic and intestinal cells has been a topic of extensive investigation (Kane, 1983; Edge et al., 1985; Wu & Windmueller, 1981; Lee et al., 1984; Deeb et al., 1986; Protter et

al., 1986). Genetic and biochemical evidence is consistent with the hypothesis that the two forms of apoB are products of the same gene (Herbert et al., 1982; Marcel et al., 1982; Young et al., 1986; Protter et al., 1986). These data include (a) the absence of both apoB forms in patients with abetalipoproteinemia (Herbert et al., 1982), (b) the description of monoclonal antibodies that recognize both apoB-100 and apoB-48 forms (Marcel et al., 1982), (c) the presence of the same genetic polymorphism in both apoB-100 and apoB-48 forms in human subjects, and (d) the recognition of both apoB-100 and apoB-48 by antisera raised against a synthetic peptide corresponding to the amino-terminal region of apoB-100 (Protter et al., 1986). However, the report by Malloy et al. (1981) of a genetic condition characterized by selective deficiency of apoB-100 left open the possibility that apoB-48 and apoB-100 might be the products of different but closely related genes.

Previous studies showed that a high molecular weight (15–22-kb) apoB mRNA form was present in both liver and intestine (Deeb et al., 1985, 1986; Knott et al., 1985; Huang et al., 1985; Mehrabian et al., 1985; Carlsson et al., 1985; Shoulders et al., 1985; Law et al., 1985). Some studies also described the presence of a lower molecular weight (6.5–9-kb) apoB mRNA form in some intestinal RNA preparations (Deeb et al., 1986; Mehrabian et al., 1985). In the latter of the two studies, however, the 9-kb band was not present consistently in different RNA preparations and was interpreted as a degradation product of apoB-100 mRNA (Mehrabian et al., 1985). In contrast, in other studies blotting analysis of human (Knott et al., 1985) and rabbit (Shoulders et al., 1985) intestinal RNA showed only a high molecular weight apoB mRNA form when the blots were hybridized with a carboxy-terminal apoB cDNA probe. Since the hybridization probe utilized in our RNA analysis was the *Sst*I to *Sst*I fragment of the amino-terminal clone λB_1 (Figure 1A,B), our data suggest that the lower molecular weight apoB mRNA form has either extensive homology or sequence identity with the amino-terminal region of apoB-100 mRNA. Taken together with previous findings (Wei et al., 1985; Knott et al., 1985; Shoulders et al., 1985; Protter et al., 1986; Deeb et al., 1986; Young et al., 1986), our data suggest that the apoB-48 mRNA form may be generated by some form of differential splicing of the apoB-100 primary transcript (Schwarzbauer et al., 1983; Periasamy et al., 1984). It is hoped that the structural characterization of the apoB-100 gene will help elucidate the relationship between apoB-100 and apoB-48 forms.

Several studies have indicated that elevated LDL cholesterol or apoB levels are associated with increased risk of coronary heart disease (Kannel et al., 1971; Sniderman et al., 1980). In contrast, decreased LDL cholesterol (hypobetalipoproteinemia) is associated with longevity (Herbert et al., 1982). Some of the variations in LDL cholesterol and apoB observed in humans may be the result of yet unidentified structural apoB gene abnormalities. The availability of cDNA and genomic apoB sequence will allow us to test this hypothesis in the future and help identify human diseases caused by structural or regulatory mutations in the apoB gene.

ACKNOWLEDGMENTS

We thank Gayle Forbes and Elizabeth Walsh for their expert assistance.

REFERENCES

Beeley, J. G. (1977) *Biochem. Biophys. Res. Commun. 76*, 1051–1055.

Beltz, G. A., Jacobs, K. A., Eickbush, T. H., Cherbas, P. T., & Kafatos, F. C. (1983) *Methods Enzymol. 100*, 266–285.

Bogusky, M. S., Elshourbagy, N., Taylor, J. M., & Gordon, J. L. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5021–5025.

Carew, T. E., Pittman, R. C., Marchard, E. R., & Steinberg, D. (1984) *Arteriosclerosis (Dallas) 4*, 214–224.

Carlsson, P., Olofsson, S. O., Bondjers, G., Darnfors, C., Wiklund, O., & Bjursell, G. (1985) *Nucleic Acids Res. 13*, 8813–8826.

Caruthers, M. H. (1985) *Science (Washington, D.C.) 230*, 281–285.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

Deeb, S. S., Motulsky, A. G., & Albers, J. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4983–4986.

Deeb, S. S., Disteche, C., Motulsky, A. G., Lebo, R. B., & Kan, Y. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 419–422.

Edge, S. B., Hoeg, J. M., Schneider, P. D., & Brewer, H. B., Jr. (1985) *Metab., Clin. Exp. 34*, 726–730.

Gleisner, J. M., Peterson, D. L., & Blakley, R. L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 3001–3005.

Goldstein, J. L., & Brown, M. S. (1977) *Annu. Rev. Biochem. 46*, 897–930.

Goldstein, J. L., & Brown, M. D. (1982) in *The Metabolic Basis of Inherited Disease* (Stanbury, J. B., Wyngaarden, J. B., Frederickson, D. S., Goldstein, J. L., & Brown, M. D., Eds.) 5th ed., pp 672–712, McGraw-Hill, New York.

Herbert, P. N., Assmann, G., Gotto, A. M., Jr., & Frederickson, D. S. (1982) in *The Metabolic Basis of Inherited Disease* (Stanbury, J. B., Wyngaarden, J. B., Fredrickson, D. S., Goldstein, J. L., & Brown, M. D., Eds.) 5th ed., pp 589–651, McGraw-Hill, New York.

Huang, L. S., Bock, S. C., Feinstein, S. I., & Breslow, J. L. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 6825–6829.

Hui, D. Y., Harmony, J. A. K., Innerarity, T. L., & Mahley, R. W. (1980) *J. Biol. Chem. 255*, 11775–11781.

Huynh, T. V., Young, R. A., & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach* (Glover, D. M., Ed.) p 49, IRL, Oxford.

Jacobs, J. C., Schumaker, V. N., & Puppione, D. L. (1985) *Biochemistry 24*, 1569–1578.

Kaiser, E. T., & Kezdy, F. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1137–1143.

Kane, J. P. (1983) *Annu. Rev. Physiol. 45*, 637–650.

Kane, J. P., Hardman, D. A., & Paulus, H. E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2465–2469.

Kannel, W. B., Castelli, W. P., Gordon, T., & McNamara, P. M. (1971) *Ann. Intern. Med. 74*, 1–12.

Karathanasis, S. K., Yunis, I., & Zannis, V. I. (1986) *Biochemistry 25*, 3962–3970.

Knott, T. J., Rall, S. C., Jr., 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 (Washington, D.C.) 230*, 37–43.

Kyte, S., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

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. U.S.A. 82*, 8340–8344.

LeBoeuf, R. C., Miller, C., Shiveley, J. E., Schumaker, V. N., Balla, M. A., & Lusis, A. J. (1984) *FEBS Lett. 170*, 105–108.

Lee, D. C., Rose, T. M., Webb, N. R., & Todaro, G. J. (1985) *Nature (London) 313*, 489–491.
Lee, D. M., Valente, A. J., Ku. H., & Maeda, H. (1981) *Biochim. Biophys. Acta 666*, 133–146.
Lee, D. M., Koren, E., Singh, S., & Mok, T. (1984) *Biochem. Biophys. Res. Commun. 123*, 1149–1156.
Lipman, D. J., & Pearson, W. R. (1985) *Science (Washington, D.C.) 227*, 1435–1441.
Mahley, R. W., Weisgraber, K. H., & Innerarity, T. L. (1979) *Biochim. Biophys. Acta 575*, 81.
Malloy, M. J., Kane, J. P., Hardman, D. A., & Hamilton, R. L. (1981) *J. Clin. Invest. 67*, 1441–1450.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Marcel, Y. L., Hogue, M., Theolis, R., Jr., & Milne, R. W. (1982) *J. Biol. Chem. 257*, 13165–13168.
Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.
Mehrabian, M., Schumaker, V. N., Fareed, G. C., West, R., Johnson, D. F., Kirchgessner, T., Lin, H. C., Wang, X., Yuanhong, M., Mendiaz, E., & Lusis, A. J. (1985) *Nucleic Acids Res. 13*, 6937–6953.
Periasamy, M., Strehler, E. E., Garfinkel, L. I., Gubits, R. M., Ruiz-Opazo, N., & Nadal-Ginard, B. (1984) *J. Biol. Chem. 259*, 13595–13604.
Pinto, M., Robine-Leon, S., Appay, M.-D., Kedinger, M., Triadou, N., Dussaulx, E., Lacroix, B., Simon-Assmann, P., Haffen, K., Fogh, J., & Zweibaum, A. (1983) *Biol. Cell (1981) 47*, 323–330.
Protter, A. A., Hardman, D. A., Schilling, J. W., Miller, J., Appleby, V., Chen, G. C., Kirscher, S. W., McEnroe, G., & Kane, J. P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1467–1471.
Pustell, J., & Kafatos, F. C. (1986) *Nucleic Acids Res. 14*, 479–488.
Schwarzbauer, J. E., Tamkum, J. W., Lemischka, I. R., & Hynes, R. O. (1983) *Cell (Cambridge, Mass.) 35*, 421–431.
Shoulders, C. C., Myant, N. B., Sidoli, A., Rodriquez, J. C., Cortese, C., Baralle, F. E., & Cortese, R. (1985) *Atherosclerosis (Dallas) 58*, 277–289.
Siuta-Mangano, P., Howard, S. C., Lennarz, W. J., & Lane, M. D. (1982) *J. Biol. Chem. 257*, 4292–4300.
Sniderman, A., Shapiro, S., Marpole, D., Skinner, B., Teng, B., & Kwiterovich, P. O., Jr. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 601–608.
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. U.S.A. 82*, 7265–7269.
Wu, A., & Windmueller, H. G. (1981) *J. Biol. Chem. 256*, 3615–3618.
Young, S. G., Bertics, S. J., Scott, T. M., Dubois, B. W., Curtiss, L. K., & Witztum, J. L. (1986) *J. Biol. Chem. 261*, 2995–2998.
Zannis, V. I., Cole, S. F., Forbes, G., Karathanasis, S. K., Kurnit, D. M., & Jackson, C. (1985) *Biochemistry 24*, 4450–4455.

Articles

Nuclear Magnetic Resonance Studies of Complex Formation between the Oligonucleotide d(TATC) Covalently Linked to an Acridine Derivative and Its Complementary Sequence d(GATA)

G. Lancelot* and N. T. Thuong
Centre de Biophysique Moléculaire, CNRS, 45071 Orleans Cedex 2, France
Received February 10, 1986; Revised Manuscript Received April 4, 1986

ABSTRACT: The oligodeoxynucleotide d(TATC) was covalently attached to the 9-amino group of 2-methoxy-6-chloro-9-aminoacridine (Acr) through its 3′-phosphate via a pentamethylene linker (m_5). Complex formation between d(TATC)m_5Acr and the complementary strand d(GATA) in aqueous solution was investigated by nuclear magnetic resonance. The COSY and NOESY connectivities allowed us to assign all the proton resonances of the bases, the sugars (except the overlapping 5′/5″ resonances), the acridine, and the pentamethylene chain. Structural informations derived from relative intensities of COSY and NOESY maps revealed that the duplex d(TATC)–d(GATA) adopts a B-type conformation and that the deoxyriboses preferentially adopt a 2′-endo conformation. The NOE connectivities observed between the protons of the bases or of the sugars and the protons of the dye and of the pentamethylene chain led us to propose a model involving an equilibrium between two families of configurations. In the first family, the acridine derivative is intercalated between base pairs C_4-G_4 and T_3-A_3. In the second family, the acridine derivative is sandwiched between two aggregated duplexes. The structure of the intercalated complex as well as that of the aggregated species is discussed.

Gene expression is usually controlled by specific proteins that recognize a base sequence or a nucleic acid local structure. The interactions between functional groups in protein–nucleic acid complexes have been recently reviewed (Hélène & Lancelot, 1982). The binding of an oligonucleotide to its complementary sequence is a highly specific process governed by

0006-2960/86/0425-5357$01.50/0