Biol. Chem. Hoppe-Seyler
Vol. 368, pp. 419–425, April 1987

Studies on the Organization of the Human Apolipoprotein B 100 Gene

Raimund WAGENER, Reinhard PFITZNER and Wilhelm STOFFEL

Institut für Physiologische Chemie der Universität zu Köln

(Received 2 February 1987)

Summary: The organization of five exons of the 3′ terminal end of the human apolipoprotein B 100 (apo B 100) gene 1906, 184, 115, 7572 and 374 bp long have been determined from two overlapping EMBL3 human genomic clones extending over 18 kb. They encode more than 70% of the apo B 100 amino-acid sequence. The introns between these five exons were sequenced revealing the common intron/exon splice junction sequences. The 7572 bp exon is the longest exon so far reported for mammalian genes with the proposed sequence coding for the LDL receptor binding site. Its possible relationship to apolipoprotein B 48 is discussed.

Untersuchungen über die Organisation des menschlichen Apolipoprotein-B-100-Gens

Zusammenfassung: Fünf Exons des menschlichen Apolipoprotein-B100- (Apo B100-)Gens wurden aus zwei überlappenden menschlichen genomischen EMBL3-Klonen bestimmt, die sich über 18 kb erstrecken. Sie sind 1906, 184, 115, 7572 sowie 374 bp lang und kodieren mehr als 70% der Apo B 100-Aminosäure-Sequenz. Die die fünf Exons verbindenden vier Introns wurden sequenziert. Sie weisen Intron-Exon-Übergänge mit den bekannten Konsensus-Sequenzen auf. Das 7572 bp lange Exon ist das bisher längste beschriebene Säugerexon, das die angenommene Sequenz für die LDL-Rezeptor-Bindungsstelle enthält. Seine mögliche Beziehung zum Apolipoprotein B48 wird diskutiert.

Key words: Apolipoprotein B 100.

Low-density lipoproteins (LDL) represent the main lipoprotein class involved in cholesterol homeostasis. These cholesterylester-rich particles are the end product of the intravascular degradation of the very low density lipoprotein particles (VLDL) synthesized as triacylglycerol-rich particles in the liver. LDL integrity is preserved by one large apoprotein, apo B 100, of 514 kDa which, in addition, is the specific ligand for the LDL receptor in peripheral cells and the hepatic apo B/E receptor[1–3].

Enzymes:
Restriction endonucleases: *Bam*HI (EC 3.1.23.6), *Cla*I (EC 3.1.23.74), *Eco*RI (EC 3.1.23.13), *Pst*I (EC 3.1.23.31), *Sal*I (EC 3.1.23.37).

Abbreviations:
cDNA, complementary desoxyribonucleic acid; apo B 100, apolipoprotein B 100, the apolipoprotein of LDL; apo B 48, apolipoprotein B 48, apolipoprotein of chylomicrones; LDL, low density lipoprotein; VLDL, very low density lipoprotein; bp, base pairs; kb, kilo base pairs; kDa, kilo Dalton; TBE, Tris-borate EDTA buffer, SSPE and SSC see Materials and Methods; $NaDodSO_4$, sodium dodecyl sulphate; RFLP, restriction fragment length polymorphism.

Perturbation of the interaction between ligand (LDL) and receptor inevitably leads to hypercholesterolemia caused by elevated LDL plasma concentrations; a main risk factor for the premature development of coronary heart disease[1].

Hypercholesterolemia in patients with normal numbers and activity of LDL receptor may be related to a mutated amino-acid sequence of its ligand apo B 100. It would be of particular importance to establish the region of the LDL receptor-binding domain and probe for mutations of the responsible sequence which might result in altered LDL and lipid metabolism.

Several laboratories[4–14] including our own[15] have been challenged by the unusual structural and functional properties of apo B because of its role in cellular cholesterol homeostasis and its role in the pathogenesis of atherosclerosis. The method of choice to approach this large protein is to obtain cDNA sequence data by screening human liver cDNA expression libraries either with apo B-specific antibodies or oligonucleotides as probes which are derived from partial amino-acid sequence data.

In six recent reports the complete cDNA-derived protein sequence of apolipoprotein B 100 has been communicated[16–21]. It contains 4563 amino-acid residues including a hydrophobic signal peptide of 27 amino acids. Thus the molecular mass of apo B 100 is approximately 514 kDa. The accumulated cDNA data rendered the isolation of the human apo B 100 gene feasible and desirable for studies on apo B gene expression and its regulation in cholesterol homeostasis. A most important domain of the apo B protein from the biochemical point of view is the LDL receptor binding site. Amino-acid sequences very similar to the apo E receptor-binding site can be recognized in the C-terminal part of the apo B-specific cDNA[5]. Knowledge of the exact location of this binding site on the human genome would enable a distinct genetic examination of patients suffering from atherosclerosis. Furthermore the gene structure might give information about the expression of apo B48 synthesized only in the enterocyte.

Using the previously analysed human apolipoprotein B 100 cDNA clone isolated by immunoscreening from a hepatic cDNA library two human genomic apo B 100 clones (λB 1 and λB 2) were isolated from a human genomic library in the EMBL 3 vector[22]. The clones were overlapping and encompassed five exons of 1906, 184, 115, 7572 and 374 bp (exons A_e-E_e) with four introns 508, 403, 108 and 934 bp length (introns A_i-D_i). The five exons encode more than 70% of the apo B 100 amino-acid sequence. The exon/intron organization of this part of the human apolipoprotein B gene is in full agreement with the recently reported complete structure of the apo B gene[23] except for three minor differences in the intron lengths of introns A_i, B_i and D_i. The 7572 bp long exon is the longest exon among mammalian genes which has been recorded so far.

Materials and Methods

The λ bacteriophage EMBL 3 library was kindly provided by *A. M. Frischauf* (EMBL, D-6900 Heidelberg). *Escherichia coli* strain NM539 was used as the host cell[22].

Hybridization probes

Apo B 100 cDNA clone λgt-B1[15] contains a 1321-bp insert (nucleotides 12781–14101 of the complete cDNA) in the *Eco*RI site of phage λgt11[24]. Apo B 100 genomic clone EMBL3-1 (λB1) contains a 959 bp fragment *Sal*I/*Eco*RI (nucleotides 5547–6506 of the full length cDNA). These fragments were nick-labeled[25] and used for hybridization.

Isolation of genomic apo B clones

2.1×10^5 Plaques of the λ phage EMBL3 library were screened with the ^{32}P-labeled cDNA fragment and the genomic *Eco*RI/*Sal*I fragment. Nitrocellulose filters (BA 85, Schleicher & Schüll) were prehybridized for 6 h at 42 °C in 50% (vol/vol) formamide/5 × Denhardt's solution (0.02% polyvinylpyrolidone, 0.02% Ficoll, 0.02% bovine serum albumin)/5 × SSPE (1 × SSPE = 0.15M NaCl, 10mM NaH_2PO_4, 1mM EDTA), 100 μg of sonified salmon sperm DNA per m*l* and subsequently hybridized with the probe at 10^5 cpm/m*l*. Filters were washed in 2 × SSC (1 × SSC = 0.15M NaCl, 0.015M sodium citrate, pH 7.0)/0.1% $NaDodSO_4$ at 65 °C for 30 min and autoradiographed at – 70 °C overnight with intensifying screens.

Mapping of the apo B gene

λ phage DNA was cleaved by single and double digestion with the corresponding enzymes and the fragments size-fractionated by gel electrophoresis (0.7% agarose and/or 5% (w/vol) polyacrylamide in TBE buffer (0.089M Tris-borate/0.089 M boric acid/2mM EDTA).

DNA sequence determination

Restriction fragments of phage DNA were size-fractionated by agarose gel electrophoresis (0.7% in TBE buffer), transferred onto a DEAE filter membrane (NA 45, Schleicher & Schüll) by electrophoresis, and eluted with 1M NaCl in the presence of 1% arginine. The fragments were subcloned into pUC plasmids[26] and M13 vectors[27]. Sequencing by the dideoxy chain termination method[28] was performed using the respective synthetic oligonucleotides as primers. Primers were synthesized by the solid-phase phosphoramidite method on an automated Applied Biosystems synthesizer (Model 380A).

Results

Isolation and mapping of restriction sites and exon-intron sequencing of the partial human apolipoprotein B gene.

The EMBL3 human genomic library was screened with the labeled *Eco*RI fragment from a human apolipoprotein B 100-specific cDNA clone spanning a total of 1321 bp, 1040 bp of the C-terminal coding region and 281 bp of the noncoding 3′ region up to the polyadenylation signal. This clone has been described previously[15]. It yielded the genomic clone λB1. In addition a 959 bp restriction fragment of this clone, λB1, was used as hybridization probe. A second genomic clone, λB2, was found which overlaps with the λB1 sequence by 6 kb.

Fig. 1 summarizes the restriction sites of the two genomic clones encompassing five exons A_e-E_e encoding Ser^{1254} to Leu^{4536} which correspond to 72.5% of the amino-acid sequence. About 6900 bp of the analysed part of our genomic clones were sequenced, 1900 bp in both directions. Exons B_e, C_e and E_e were sequenced completely. About 25% of exon A_e and 40% of exon D_e have been determined. Length of restriction fragments of these exons and restriction fragments of subcloned fragments digested with additional restriction enzymes could be evaluated with an accuracy of ten bases in comparison with the cDNA data[16]. As intron lengths of ten bases are rather unlikely, we postulate the continuity of the reading frames of these exons. This postulation is in agreement with the finding of Blackhart et al.[23], who have determined the complete exon sequences. Thus we have sequenced the exon-intron junctions of these two large exons and therefore know the first and the last coding nucleotides. Considering the cDNA data[16] we were able to determine the length of the exons to be 1906 and 7572 bp.

The four introns A_i-D_i are 508, 403, 108 and 934 bp long. The complete nucleotide sequences of the four introns were determined. They are summarized together with 50 nucleotides of the 5′- and 3′-flanking coding sequences in Fig. 2. The length of the intron D_i = 508 bp differs from that reported recently[23] by one nucleotide from the corresponding intron 25, intron B_i = 108 bp by 1 nucleotide from intron 27 and intron A_i = 934 bp by 8 nucleotides from intron 28. The reported differences between cDNA sequence and gene sequence as given by Blackhart et al.[23] are in agreement with the exon sequences so far examined by us.

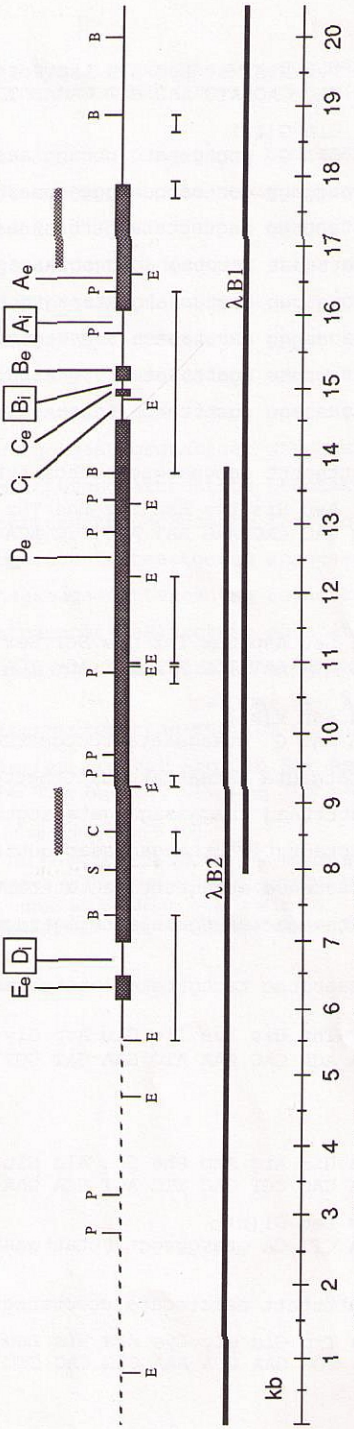

Fig. 1. Restriction map of the partial human apolipoprotein B 100 gene.
The dark, shaded areas correspond to the five exons encoding Ser^{1254} to the polyadenylation signal. Thin continuous lines represent fragments which were sequenced by single or double strand dideoxy chain termination method. The dashed line shows the part of the gene which has not further been analysed. Localization of the hybridization probes are indicated by small bars above the restriction map. The two solid bottom lines correspond to the overlapping clones λB1 and λB2. Introns that differ from Blackhart et al.[23] are within boxes. Abbreviations: B = *Bam*HI, C = *Cla*I, E = *Eco*RI, P = *Pst*I, S = *Sal*I.

Intron D_i

```
1361  Arg Tyr His Met Lys Ala Asp Ser Val Val Asp Leu Leu Ser Tyr Asn
4289  CGT TAC CAC ATG AAG GCT GAC TCT GTG GTT GAC CTG CTT TCC TAC AAT

1377  Val Gln G(ly)
4338  GTG CAA G  gtgagctatg ctcaggtaaa gggtgcaccg ggctagttca tggcaggctc    50
      taagaggaga gcctcctcca gggaggaaag gactttggct ttctagcaga taatcttcct   110
      tgctacttgg aagtctttta ttttattcaa caaatagaaa tatttattaa acatatcacg   170
      tgtattaaat attctagtag gcagtaacag aaagtagaca gataagccag caattataat   230
      tcagtgtgag aggtgctatg ataaagtgta gtatataagt ataaggtaga gtggaagcac   290
      tcaacaaggg aacctaaaca aagcctgtgg tggtcaggca aggcttcctg gaggaatgcc   350
      ttttgctatc agattttatc tttgcattac agatggagga gtctattgca caattggccc   410
      agaaaaatgg ggctttatta ttgaaagact ttcaacatag agattgctct ggaaatgtac   470

1379                                          (G)ly Ser Gly Glu Thr Thr
4345  tgcttaattt aaccaatgtc ttttcatttt tatgttag GA TCT GGA GAA ACA ACA     508

1385  Tyr Asp His Lys Asn Thr Phe Thr Leu Ser Cys Asp Gly Ser Leu Arg
4362  TAT GAC CAC AAG AAT ACG TTC ACA CTA TCA TGT GAT GGG TCT CTA CGC
```

Intron C_i

```
3885  Val Leu Asp Ser Thr Cys Ser Ser Thr Val Gln Phe Leu Glu Tyr Glu
11862 GTC CTG GAT TCC ACA TGC AGC TCA ACC GTA CAG TTC CTA GAA TAT GAA

3901  Leu Asn V(al)
11910 CTA AAT G  gtaagaaata tcctgcctcc tctcctagat actgtatatt ttcaatgaga    50
      gttatgagta aataattatg tatttagttg tgagtagatg tacaattact caatgtcaca   110
      aaatttaag taagaaaaga gatacatgta taccctacac gtaaaaacca aactgtagaa   170
      aatctagtgt cattcaagac aaacagcttt aaagaaaatg gatttttctg taattatttt   230
      aggactaaca atgtctttta actatttatt ttaaaataag tgtgagctgt acattgcata   290
      ttttaaacac aagtgaaata tctggttagg atagaattct cccagtttta caatgaaaac   350

3903                                                          (V)al Leu
11917 atcaacgtcc tactgttatg aatctaataa aatacaaaat ctctcctata cag TT TTG   403

3905  Gly Thr His Lys Ile Glu Asp Gly Thr Leu Ala Ser Lys Thr Lys Gly
11922 GGA ACA CAC AAA ATC GAA GAT GGT ACG TTA GCC TCT AAG ACT AAA GGA
```

Intron B_i

```
3923  Ala His Arg Asp Phe Ser Ala Glu Tyr Glu Glu Asp Gly Lys Tyr Glu
11976 GCA CAC CGT GAC TTC AGT GCA GAA TAT GAA GAA GAT GGC AAA TAT GAA

3939  Gly Leu Gl(n)
12024 GGA CTT CA gtatggagct tttattgaat tgaaacctta tacctttga aaactcattg    50

3941                                                              (Gl)n
12032 tgattttctt catctccata ccccttccgt gatagctcat ctgtttttct gctttcag G   108

3942  Glu Trp Glu Gly Lys Ala His Leu Asn Ile Lys Ser Pro Ala Phe Thr
12033 GAA TGG GAA GGA AAA GCG CAC CTC AAT ATC AAA AGC CCA GCG TTC ACC
```

Intron A_i

```
 3985  Asp Met Asp Glu Asp Asp Asp Phe Ser Lys Trp Asn Phe Tyr Tyr Ser
12162  GAT ATG GAT GAA GAT GAC GAC TTT TCT AAA TGG AAC TTC TAC TAC AGC

 4001  Pro Gln
12210  CCT CAG    gtaaatacca cctaatgagt gacacgcccc caagagcgag tggagaattg    50
       gggcagatac atttaattca ggaccaaata ttcagagatt ccccaaacta ggtgaaagac   110
       aggcggtaag caacttcttc tctgaggaaa tattctctag aaagtattac aatgagtcct   170
       tgattgattt taatgtttag atgcacacat gacatcccat cagcactatt atttattaat   230
       tctgggcaaa tccaggaaga tgagggttat acctcatcat ctaaatcata ggcaagctca   290
       gccataggca gggtatattt ttcagagagg actggtttct gtagtattta aaactttcaa   350
       cattcttccc cacaatagaa ttgctagatg agatacatca aattcctctc atgtcattta   410
       caagctctgc cagggccaaa tcaagggtga cattaccaga ggagaagacc aaacatggtt   470
       ctatgactgt tactaaaagt ttgtcatggg cttggagaat gcgtactgat gttgggattc   530
       tgggtctctg cagggtgggc tccaacttgc ctttttgct atttcttctt ttcctatctg   590
       tcatttcctg actcttcttc tctctcctct tctttctctt ccccccactc ctcttccagt   650
       tttcagtcct aggaaggctt taattttaag tgtcacaatg taaatgacaa acagcaagcg   710
       tttttgttaa atcctttctg gggcatgtga taaagagaaa ttaacaacag tagacttatt   770
       taaccataaa acaaacacat gaactgacat atgaaagata aatccctttc agtatatgaa   830
       agattctctg atctttattt ttaactgcta atgaagtttt agtgtactat attgtgtaat   890

 4003                                                  Ser Ser Pro Asp
12216  tggagtaatt gaaaacatgt tatttttttt ttctctctgt ttag TCC TCT CCA GAT    934

 4007  Lys Lys Leu Thr Ile Phe Lys Thr Glu Leu Arg Val Arg Glu Ser Asp
12228  AAA AAA CTC ACC ATA TTC AAA ACT GAG TTG AGG GTC CGG GAA TCT GAT
```

Fig. 2. Nucleotide sequences of the four 3′-terminal introns of apolipoprotein B 100.

About 50 nucleotides at the 5′- and 3′-adjacent sequences of the exons are included. Exon nucleotide numbers corresponding to the cDNA sequence and amino-acid numbers are given on the left side, intron nucleotide numbers on the right side. Exon nucleotides are written in upper case letters, intron nucleotides in lower case letters.
These nucleotide sequences have been submitted to the EMBL Data Bank with the accession number X04866.

Discussion

This paper reports the exon-intron organization of the 3′ region of the human apo B 100 gene which encompasses more than 70% of the coding sequences. The structure of the five 3′ exons and their four introns is extended over approximately 12 kb. The most striking observation is the presence of two long open reading frames of 1906 and 7572 bp. Our data are in full agreement with the most recent report[23] on the structure of the 43 kb human apo B 100 gene except for the length of three introns.

The 5′ splice donor and 3′ acceptor splice sites of the four introns follow the GT-AG rule[29]. We included the introns in the nucleotide sequence determination to contribute to the accumulation of extensive data about the apo B 100 gene. Furthermore these intron sequences may contain recognition sites for restriction enzymes which could determine new restriction length polymorphisms (RFLP). As will be discussed later apo B48 possibly arises from the B 100 gene by differential splicing[30]. The search for possible hidden splicing sites requires complete sequence data for the gene including intron sequences. Recent RFLP studies[31–34] try to correlate genetic variants of apo B 100 with dyslipoproteinemias.

The 7572 bp long exon, the largest exon so far detected in vertebrates, is of particular interest for future studies. It contains a presumed LDL receptor binding site assigned to a peptide domain enriched in positively charged amino acid residues analogous to that deduced for apo E binding to the LDL receptor[2]. Synthetic peptides corresponding to this region (3345–3381) can substitute for the LDL receptor binding site in reconstituted lipoproteins[19]. In addition three flanking peptide regions with positively charged hydrophobic amino-acid residues between residues 3174 and 3681 may also be involved.

We also wish to discuss the relationship between apo B 100 and apo B48. The liver only expresses apo B 100 whereas intestinal mucosa cells synthesize B48[35]. Genetic, biochemical and immunological evidence[12,36–38] indicates that the two forms are the products of the same gene. B48 is less than half the size of apo B 100 and does not recognize the LDL receptor. Thus far the primary structure of the apo B48 has not been resolved. There is evidence from immunological studies that it bears N-terminal parts[12] as well as C-terminal parts[38] of the apo B 100. Differential splicing could be one reason for this observation. We suggest that the loss of the large 7572 bp exon could lead to a mRNA of about 8 kb, such as has been detected by several groups[7,21,39]. This would reduce the molecular mass from 514 kDa to 228 kDa, which is in the range of the estimated molecular mass of apo B48[3].

The unusually large size of the 7572 bp exon can also be explained by our theory. For it is less difficult to block only one splice site for a large exon than many splice sites for many small exons. The coding of probable sequences for the LDL receptor-binding site in this large exon lends further support to our suggestion. Our working hypothesis is being studied in current experiments.

The support of the *Deutsche Forschungsgemeinschaft*, SFB 74, and of the *Fritz-Thyssen-Stiftung* is gratefully acknowledged. We thank Mrs. *C. Geyer*, Mrs. *J. Hülsmann* and Mr. *M. Körner* for their excellent technical assistance.

Literature

1 Goldstein, J.L. & Brown, M.S. (1983) in *The Metabolic Basis of Inherited Diseases* (Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S., eds.) pp. 672–712, McGraw Hill, New York.
2 Mahley, R.W. & Innerarity, T.L. (1983) *Biochim. Biophys. Acta* **737**, 197–222.
3 Kane, J.P. (1983) *Annu. Rev. Physiol.* **45**, 637–650.
4 Deeb, S.S., Motulsky, A.G. & Albers, J.J. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4983–4986.
5 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* **230**, 37–43.
6 Huang, L.-S., Bock, S.C., Feinstein, S.I. & Breslow, J.L. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6825–6829.
7 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.
8 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.
9 Carlsson, P., Olofsson, S.O., Bondjers, G., Darnfors, C., Wiklund, O. & Bjursell, G. (1985) *Nucleic Acids Res.* **13**, 8813–8826.
10 Shoulders, C.C., Myant, N.B., Sidoli, A., Rodriguez, J.C., Cortese, C., Baralle, F.E. & Cortese, R. (1985) *Atherosclerosis* **58**, 277–289.
11 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.
12 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. U.S.A.* **83**, 1467–1471.

13 Protter, A.A., Hardman, D.A., Sato, K.Y., Schilling, J.W., Yamanaka, M., Hort, Y.J., Hjerrild, K.A., Chen, G.C. & Kane, J.P. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5678–5682.
14 Cladaras, C., Hadzopoulou-Cladaras, M., Avila, R., Nussbaum, A.L., Nicolosi, R. & Zannis, V.I. (1986) *Biochemistry* **25**, 5351–5357.
15 Pfitzner, R., Wagener, R. & Stoffel, W. (1986) *Biol. Chem. Hoppe-Seyler* **367**, 1077–1083.
16 Knott, T.J., Wallis, S.C., Powell, L.M., Pease, R.J., Lusis, A.J., Blackhart, B., McCarthy, B.J., Mahley, R.W., Levy-Wilson, B. & Scott, J. (1986) *Nucleic Acids Research* **14**, 7501–7503.
17 Chen, S.-H., Chao, Y.-Y., Chen, P.-F., Setzer, D., Tanimura, M., Li, W.-H., Gotto, A.M. & Chan, L. (1986) *J. Biol. Chem.* **261**, 12918–12921.
18 Knott, T.J., Pease, R.J., Powell, L.M., Wallis, S.C., Rall, S.C., Innerarity, T.L., Blackhart, B., Taylor, W.H., Marcel, Y., Milne, R., Johnson, D., Fuller, M., Lusis, A.J., McCarthy, B.J., Mahley, R.W., Levy-Wilson, B. & Scott, J. (1986) *Nature (London)* **323**, 734–738.
19 Yang, C.-Y., Chen, S.-H., Gianturco, S.H., Bradley, W.A., Sparrow, J.T., Tanimura, M., Li, W.-H., Sparrow, D.A., DeLoof, H., Rosseau, M., Lee, F.-S., Gu, Z.-W., Gotto, A.M. & Chan, L. (1986) *Nature (London)* **323**, 738–742.
20 Law, S.W., Grant, S.M., Higuchi, K., Hospattankar, A., Lackner, K., Lee, N., Brewer, H.B. Jr. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8142–8146.
21 Cladaras, C., Hadzopoulou-Cladaras, M., Nolte, R.T., Atkinson, D. & Zannis, V.I. (1986) *EMBO J.* **5**, 3495–3507.
22 Frischauf, A.M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170**, 827–842.
23 Blackhart, B.D., Ludwig, E.M., Pierotti, V.R., Caiati, L., Onasch, M.A., Wallis, S.C., Powell, L., Pease, R., Knott, T.J., Chu, M.L., Mahley, R.W., Scott, J., McCarthy, B.J. & Levy-Wilson, B. (1986) *J. Biol. Chem.* **261**, 15364–15367.
24 Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1194–1198.
25 Rigby, P.W.J., Gieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
26 Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
27 Messing, J., Crea, K. & Seeburg, P.H. (1981) *Nucleic Acids Research* **9**, 309–321.
28 Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
29 Breathnach, R. & Chambon, R. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
30 Schwarzbauer, J.E., Tamkum, J.W., Lemischka, I.R. & Hynes, R.O. (1983) *Cell* **35**, 421–431.
31 Law, A., Wallis, S.C., Powell, L.M., Pease, R.J., Brunt, H., Priestley, L.M., Knott, T.J., Scott, J., Altman, D.G., Miller, G.J., Rajput, J. & Miller, N.E. (1986) *Lancet* **8493**, 1301–1303.
32 Talmud, P. & Humphries, S. (1986) *Lancet* **8498**, 104.
33 Berg, K., Powell, L.M., Wallis, S.C., Pease, R., Knott, T.J. & Scott, J. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7367–7370.
34 Hegele, R.A., Huang, L.-S., Herbert, P.N., Blum, C.B., Buring, J.E., Hennekens, C.H. & Breslow, J.L. (1986) *N. Engl. J. Med.* **315**, 488–500.
35 Kane, J.P., Hardman, D.A. & Paulus, H.E. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2465–2469.
36 Herbert, P.N., Assmann, G., Gotto, A.M. & Fredrickson, D.S. (1982) in *The Metabolic Basis of Inherited Diseases* (Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. & Brown, M.S., eds.) pp. 589–651, MaGraw Hill, New York.
37 Marcel, Y.L., Hogue, M., Theolis, R. Jr. & Milne, R.W. (1982) *J. Biol. Chem.* **257**, 13165–13168.
38 Hospattankar, A.V., Fairwell, T., Meng, M., Ronan, R. & Brewer, H.B., Jr. (1986) *J. Biol. Chem.* **261**, 9102–9104.
39 Deeb, S.S., Disteche, C., Motulsky, A.G., Lebo, R.V. & Kan, Y.W. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 419–422.

Note added in proof (12.3.1987):
While this paper was prepared an additional article concerning the apo B gene was published [Carlson, P., Darnfors, C., Olofson, S.-O. & Bjursell, G. (1986) *Gene* **49**, 29–51]. There are some differences to our sequence data. The exon corresponding to our exon D_e lacks a 60 bp *Eco* RI fragment (b 8287-b 8346). Intron C_i consists of 407 bp instead of 403 bp, intron B_i is 109 bp long instead of 108 bp and intron A_i is 950 bp instead of 934 bp in our gene structure.

Raimund Wagener, Reinhard Pfitzner and Wilhelm Stoffel*, Institut für Physiologische Chemie der Universität zu Köln,
Joseph-Stelzmann-Str. 52, D-5000 Köln 41.

* To whom correspondence should be addressed.