Volume 14 Number 23 1986 Nucleic Acids Research

Human lecithin-cholesterol acyltransferase gene: complete gene sequence and sites of expression

John McLean, Karen Wion, Dennis Drayna, Christopher Fielding* and Richard Lawn

Department of Molecular Biology, Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080 and *Cardiovascular Research Institute and Department of Physiology, University of California, San Francisco, CA 94143, USA

Received 14 August 1986; Revised and Accepted 31 October 1986

ABSTRACT

The human lecithin-cholesterol acyltransferase (LCAT) gene has been sequenced to completion. The gene is divided into six exons spanning ~4,200 bp. Exon five codes for amino acids homologous to the interfacial active site of several lipases, and also codes for an amphipathic α-helix resembling the carboxy terminus of apolipoprotein E. Blot hybridization data suggest that there is only one LCAT gene in humans. The 1550 base LCAT mRNA can be detected in liver and HepG2 (hepatocyte) cells, but not in small intestine, spleen, pancreas, placenta or adrenal tissue.

INTRODUCTION

The plasma protein lecithin-cholesterol acyltransferase (LCAT) is a key component in the process of transport of cholesterol from peripheral tissues to the liver for catabolism (1). LCAT catalyzes the transfer of acyl groups from lecithin to the 3-OH of free cholesterol. This results in the conversion of diffusible cholesterol to insoluble cholesteryl ester molecules which reside in the core of circulating lipoprotein particles. This conversion maintains a gradient of cholesterol concentration between cell membranes and plasma. When LCAT activity is genetically defective, cholesterol accumulates in tissues, causing a number of clinical symptoms (2).

The complete human LCAT cDNA has been cloned and sequenced, and used to drive the expression of active LCAT by transfected tissue culture cells (3). Translation of the cDNA sequence predicts a mature protein of 416 amino acids with a hydrophobic leader sequence of 24 residues. The glycoprotein migrates on SDS gels with an apparent M_r ~63,000. The human LCAT sequences share a region of homology with the "interfacial" active site of porcine pancreatic lipase and rat lingual lipase (3). To provide a basis for comparison of the gene structure of LCAT with genes of related lipid interacting enzymes and with the LCAT genes of patients deficient in plasma LCAT activity, we used our cDNA clone to isolate the LCAT gene from a phage

λ/human genomic library. Here we report the complete DNA sequence of 6.9 kb comprising the LCAT gene and flanking DNA, and discuss the tissues of origin of LCAT mRNA.

METHODS

LCAT clones were recovered from a human genomic DNA library (4) utilizing an LCAT cDNA clone (3) as probe by methods previously described (3). DNA sequencing was accomplished by dideoxy chain termination methods (5). Approximately 80 percent of the genomic DNA was sequenced on both strands. All other regions were sequenced at least twice. Details concerning RNAse mapping as in (6) and Northern blot hybridization as in (3) are given in the figure legends. Primer extension analysis was performed by adapting our normal cDNA synthesis procedure (7): 0.4 μg of ^{32}P-end labeled 17 base oligonucleotide was annealed to 20 μg of poly(A^+) RNA and incubated for 3 hours at 42°C with 85 units of AMV reverse transcriptase.

RESULTS AND DISCUSSION

Isolation and Sequence of the LCAT Gene

Utilizing the LCAT cDNA clone pL12 (3) as probe, we isolated the LCAT gene from a library of human DNA in the bacteriophage λ Charon 30 (4). Five hybridizing clones all appeared to contain the entire mRNA coding sequence. Subsequent analysis showed that all of the sequences contained in the full-length cDNA clone lie in a single, 18 kb genomic BamHI fragment. Nearly 7,000 nucleotides of this portion of the genome were sequenced by subcloning restriction fragments into M13 vectors, some of which were shortened by controlled exonuclease digestion (8), and performing dideoxy chain termination reactions (5) with primer oligomers hybridizing to either M13 sequences or to LCAT genomic sequences already determined. A restriction map and sequencing strategy is shown in figure 1 and the DNA sequence in figure 2. In all exons the genomic sequence precisely matches the cDNA sequence we reported previously (3).

The LCAT gene is divided into six exons, so that the ~1,400 bp of mRNA coding sequence are spread out over ~4,200 bp of the genome. The signal prepeptide and the first 27 amino acids of the mature protein are encoded in the first exon. Exon five contains two regions of homology with other proteins that interact with lipids. One of these is the hexapeptide (residues 178-183) that is identical to the amino acid sequence at the interfacial active site of porcine pancreatic lipase, as we previously

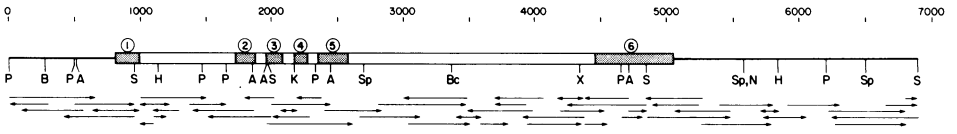

Figure 1. The human LCAT gene: restriction map and sequencing strategy. The human LCAT gene is depicted below a size scale in base pairs. From left to right are the 5' flanking region, exons (filled bars with numbers above), introns (open bars), and the 3' flanking region. All restriction sites for AccI (A), BamHI (B), BclI (B), HindIII (H), KpnI (K), NsiI (N), SstI (S), SphI (Sp), and PstI (P) are shown below their location in the DNA. Arrows indicate the extent and direction of DNA sequencing reactions.

reported (3). In addition, the first 59 nucleotides of exon five share a striking 66% DNA sequence homology with the 3' terminal coding region of the apolipoprotein E gene (9). The portion of LCAT encoded by this region, plus several surrounding nucleotides, could form an amphipathic α-helix, a structure implicated in lipid binding (10). The amphipathic helix could follow the proline at residue 150 (encoded at the end of exon 4) and extend to the proline at residue 174. Hence exon five may encode a protein domain which interacts with the lipid interface containing the substrates of the LCAT reaction. Further protein characterization will be necessary before detailed comparisons of protein domains and exon boundaries can be made.

The last exon contains nearly half of the protein and the extremely short 3'-untranslated region. As noted earlier, the 3'-untranslated region of LCAT is so short (23 nucleotides) that the polyadenylation signal AATAAA (11) is partially contained in the codon for the C-terminal glutamate (GAA) and the translation stop codon TAA (3). The indicated site of polyadenylation is the principal (if not exclusive) poly(A) site of LCAT mRNA, based on: analysis of four independent cDNA clones; the size and homogeneity of hybridizing RNA; the occurrence of a TATGTTTC "consensus" sequence 3' of the splice site (12), and negative results of Northern blots probed with sequences 3' to this site. It can now also be seen that no other AATAAA-like polyadenylation signal sequences occur for at least 726 bp downstream of the stop codon, where the sequence AATTAAA is found.

All exon-intron borders conform to the GT-AG rule and to reasonable variants of the surrounding splice site consensus sequences (13). The largest intron in this gene, between exons 5 and 6, contains three Alu type repetitive elements (14) in the same orientation (Fig. 2). The only other Alu elements in the genomic region we have sequenced are four repeat


exon 1
exon 2
exon 3
exon 4
exon 5
exon 6
Alu sequence
Alu sequence
Alu sequence
Alu sequence
Alu sequence
Alu sequence

elements downstream of the 3' end of message synthesis. All four lie in the orientation opposite that of the Alu repeats in intron 5.

The location of the 5' terminus of the mRNA has been determined by primer extension and RNAse mapping (Fig. 3). Based on the sequence of cDNA clones, and on Northern blot determination of the size of hybridizing mRNA in liver, we estimated that the 5'-untranslated region of the LCAT message was less than 100 nucleotides (3). The longest cDNA clones we reported contained 11 bp upstream of the initiation codon ATG. We have subsequently analyzed a cDNA clone which contains 24 bp of upstream sequence, matching sequence from genomic clones.

A 17-base oligonucleotide was synthesized to correspond to the complement of the coding sequence of residues 13-18 of the mature protein (nucleotides 946-962, ending in a SstI site; fig. 2). This oligonucleotide was ^{32}P-labeled and extended into cDNA with reverse transcriptase in the presence of liver poly(A^+) RNA. The pattern of extension is shown in figure 3 and indicates a major product 154 bases long, as well as a second product of ~440 bases. The results of RNAse mapping are consistent with a 5' message start site indicated by the smaller primer extension product. A 461 bp PstI-SstI cloned genomic fragment, which contains 340 bp of sequence 5' to the initiator ATG and commences at codon 18, was inserted into pSP65 for synthesis of a uniformly labeled RNA probe (15). This probe was annealed to poly(A^+) liver RNA and digested with single strand specific RNAses A and T1. A series of protected RNA fragments ranging from ~144-156 bases are seen in figure 3. RNAse dilution experiments (not shown) suggest that the largest protected fragment at 156 bases represents the true end of the RNA:RNA duplex. The series of smaller fragments may be due to overdigestion by the enzymes and may also reflect the specificity of the enzymes, which only hydrolyze ribo G, U and C. No RNAse protected fragment can be seen which corresponds to the larger 440 base primer extension

Figure 2. The human LCAT gene sequence. The DNA sequence of the LCAT gene and flanking regions are shown, with nucleotides numbered to the left of each row. The derived amino acids in exons are given above the nucleotides, and are numbered on the left; phe (double underlined) at +1 is the first residue of the mature protein (3). Arrows indicate the approximate 5' message start site and the polyadenylation site. 5' and 3' untranslated regions are underlined and the AATAAA polyadenylation signal sequence is double underlined. Proposed CAAT and TATA box sequences are enclosed. Alu repeat sequences are underlined and labeled. The three Alu repeats in intron 5 are oriented 3' to 5', while the four Alu repeats flanking the gene are oriented 5' to 3'.

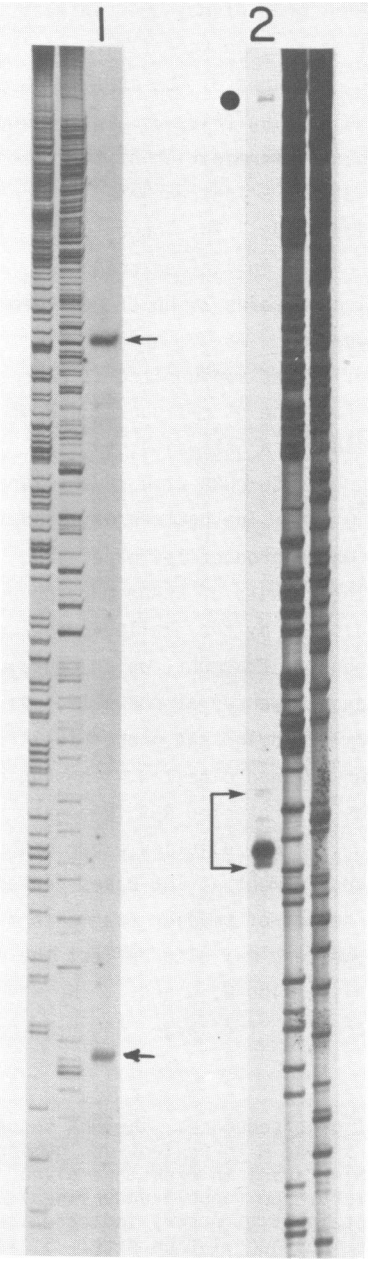

Figure 3. Determination of the mRNA start site. At the left are shown two of four lanes of a known DNA sequence ladder employed as a size standard. Lane 1 shows cDNA primed with a ^{32}P-end labeled 17-base oligonucleotide. Arrows indicate the primer extension products of ~440 and 154 nucleotides.

Lane 2 is the product of an RNAse mapping reaction. A uniformly labeled, 486 base RNA probe (containing 5' LCAT genomic sequences and flanking DNA of the vector plasmid pSP65) was annealed to human liver RNA and digested with single strand specific RNAses. The series of protected fragments ranging from 144–156 nucleotides are indicated. These fragments are not seen in yeast tRNA control lanes (not shown). Remaining undigested probe of 486 bases, marked by a dot, can be seen near the top of the lane. Two lanes of a sequencing ladder are at the right (the photograph enlargement scale for lanes 1 and 2 are slightly different). For cDNA synthesis, 0.4 µg of end-labeled oligonucleotide was incubated with 20 µg of human liver poly(A^+) RNA, 1 mM deoxynucleotide triphosphates, 85 units of AMV reverse transcriptase at 42°C for 3 hr under conditions described in (7). After phenol extraction and ethanol precipitation, one-half of the sample was loaded onto a 6 percent polyacrylamide, 7 M urea gel. For RNAse mapping, a genomic fragment was cloned into the SstI and PstI sites of pSP65 for synthesis of radiolabeled RNA, followed by annealing to 10 µg of liver poly(A^+) RNA for 15 hours at 37°C and digestion with 0.4 µg/ml of RNAse A and 0.2 µg/ml of RNAse T1 in 10 mM Tris (pH 7.5), 5 mM EDTA, 200 mM NaCl, 100 mM LiCl for 30 minutes at 24°C as described in (15). The end point of both the oligonucleotide primer and genomic fragment probes is the SstI site in exon 1 of the LCAT gene.

product. We tentatively assign the mRNA start site to nucleotide 809±2 in figure 2. Both cDNA extension and RNAse mapping results predict an mRNA terminus at that position. The fortuitous location of an intron at nucleotide 809 could also result in an RNAse mapping fragment of the observed size, but no consensus splice acceptor site occurs there. In addition, the measured size of human LCAT mRNA (see figure 4b) predicts a short 5' untranslated region consistent with this message start site assignment and none of the approximately 50 LCAT cDNA clones identified extended further upstream. The nature of the longer cDNA extension product remains unknown. Although it is not likely to indicate the true 5' terminus of LCAT mRNA for the above reasons, this possibility has not been ruled out.

The closest resemblance to a TATA box consensus sequence in the region of the proposed start site is the sequence GATAA commencing 23 bp upstream. At 100 bp 5' to the start site, the sequence GGCAATCT resembles the consensus CAAT box sequence GGPyCAATCT (16). These two consensus sequences comprise elements of many eukaryotic gene promoters (16), although a recognizable TATA box cannot be found in all genes. For example, the human apolipoprotein A-I gene, which is a major protein component of high density lipoproteins and an activator of LCAT, contains no TATA-like sequence in the proximity of its transcription start site (17,18).

Southern Blot Analysis

Genomic Southern blot analysis was performed in order to compare the restriction map derived from mapping and sequencing cloned genomic DNA with

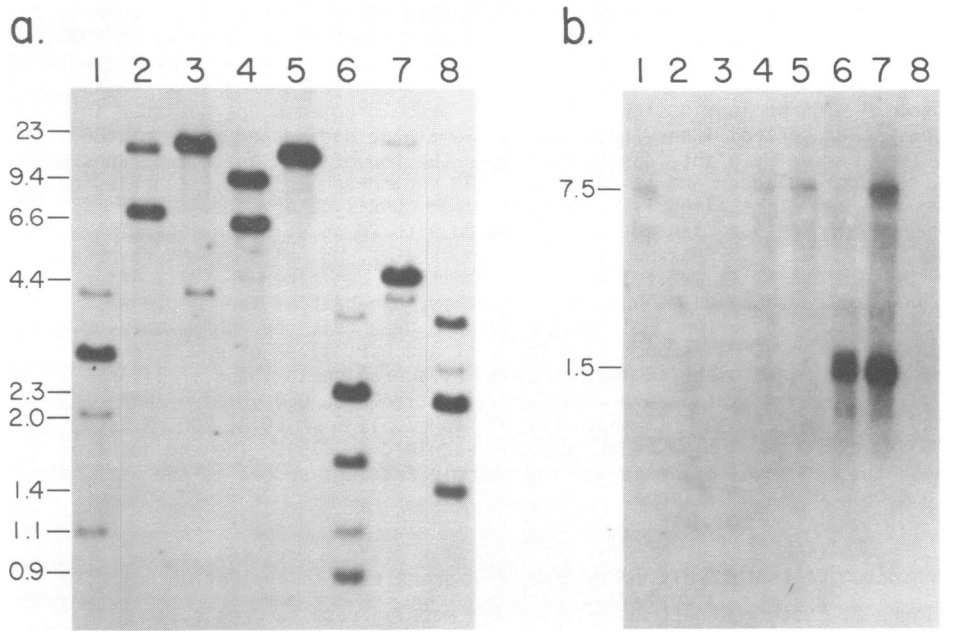

Figure 4. Blot hybridization analysis of LCAT DNA and RNA. A. Southern blots of human genomic DNA cleaved with various restriction enzymes were probed with ^{32}P-labeled cDNA insert of LCAT clone pL12 (3) at 42°C in 5xSSC, 50 mM sodium phosphate (pH 6.8), 5x Denhardt's solution, 100 µg/ml sonicated salmon sperm DNA, 50 percent formamide and 10 percent dextran sulfate and washed at 65°C in 0.2x SSC, 0.1 percent SDS. Enzymes used in each lane are: 1, SstI; 2, KpnI, 3, BamHI; 4, BclI; 5, BglII; 6, PstI; 7, HindIII; 8, AccI. Size standards in bp, indicated on the left, derive from HindIII-digested λ and HaeIII-digested ϕX DNA. All fragments above the detection limit of 500 bp that are predicted by DNA sequencing are seen in genomic DNA. With some enzymes, a single, faint, unpredicted band appears, which may be due to an additional region of homology with a portion of the probe.

B. Northern blots of poly(A^+) liver from human tissues were hybridized to LCAT cDNA under the same conditions as in A. Each lane contains 5 µg of RNA from: 1, U937 (monocyte-like) cell line; 2, white blood cells; 3, pancreas, 4, placenta; 5, adrenal; 6, liver; 7, hepG2 (hepatocyte) cell line; 8, small intestine (terminal ileum). As indicated, LCAT mRNA is ~1550 nucleotides; a larger hybridizing RNA of ~7500 b is seen in some samples. Ribosomal RNAs and HindIII-cut λ DNA were used as size standards.

uncloned DNA, and to estimate the number of LCAT genes. Figure 4a shows the results of digestion of human DNA with several different restriction enzymes and probing with the isolated insert fragment of the human LCAT cDNA clone pL12 (2). (This probe was chosen because it contains nearly all of the

coding region, but no poly(A) tail.) In all cases, hybridizing fragments are seen to be consistent in size and intensity with predictions based upon DNA sequencing. There is no evidence of multiple LCAT genes. However, digestion with several enzymes yields a single, faint, anomalous hybridizing fragment which might indicate that homology exists elsewhere to a limited part of the cDNA sequence. No more than one, faint unpredicted band has been seen with any enzyme tested. Thus we conclude that the human LCAT cDNA we have characterized (3) is the product of a single gene whose DNA sequence is reported here.

Tissue Distribution of LCAT RNA

The LCAT glycoprotein circulates in the bloodstream associated with lipoprotein particles whose various components are synthesized in the liver and, in some cases, other tissues as well. We previously cloned LCAT cDNA from human liver RNA. In order to further delineate the sites of synthesis, we extracted poly(A^+) RNA from a number of human cell types for Northern blot hybridization with ^{32}P-labeled LCAT cDNA clone pL12 (3). As shown in figure 4b, a band corresponding to the ~1550 base LCAT message derives only from liver and the hepatocyte-derived cell line HepG2. No hybridizing RNA of that size occurs in samples from spleen, pancreas, placenta, adrenal gland, small intestine (terminal ileum), white blood cells or the monocyte-like cell line U937. The measured size of 1550 bases is consistent with data derived from our analysis of LCAT clones, i.e. an open reading frame of 1320 bases, a 5'-untranslated region of ~28 bases, a 3'-untranslated region of 23 bases, and a poly(A) tail which is commonly estimated to contain ~200 bases. We do not know the cause of the larger, ~7500 base, hybridizing band seen in some samples.

In summary, the human LCAT gene spans 4200 bp and is divided into six exons. The gene is known to reside in the q21-22 region of human chromosome 16 (19). Blot hybridization data indicate that there is only one LCAT gene in humans, although more distantly related genes may well exist. The gene product, which is chiefly associated with the HDL fraction of plasma lipoprotein particles, is synthesized in liver (2), and no extrahepatic sites of synthesis have been identified in this study. The characterization of the human LCAT gene reported here can now serve as a basis of identification and detailed characterization of the mutations which underlie congenital LCAT deficiency and thus contribute to the understanding of the activity and substrate interactions of this enzyme.

REFERENCES

1. Fielding, C.J. and Fielding, P.E. (1981) Proc. Natl. Acad. Sci. USA 78, 3911-3914.
2. Glomset, J.A., Norum, K.R. and Gjone, E. (1983) in The Metabolic Basis of Inherited Disease, 5th Ed. (Stanbury, J., Wyngaarden, J., Fredrickson, D., Goldstein, J. and Brown, M., eds.) pp. 643-654, McGraw-Hill, New York.
3. McLean, J., Fielding, C., Drayna, D., Dieplinger, H., Baer, B., Kohr, W., Henzel, W. and Lawn, R. (1986) Proc. Natl. Acad. Sci. USA 83, 2335-2339.
4. Wood, W., Capon, D., Simonsen, C., Eaton, D., Gitschier, J., Keyt, B., Seeburg, P., Smith, D., Hollingshead, P., Wion, K., Delwart, E., Tuddenham, E., Vehar, G. and Lawn, R. (1984) Nature 312, 330-337.
5. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
6. Wion, K., Kelly, D., Summerfield, J., Tuddenham, E. and Lawn, R. (1985) Nature 317, 726-729.
7. Lawn, R., Adelman, J., Bock, S., Franke, A., Houck, C., Najarian, R., Seeburg, P., and Wion, K. (1981) Nucleic Acids Res. 9, 6103-6114.
8. Dale, R.M.K., McClure, B.A. and Houchins, J.P. (1985) Plasmid 13, 31-40.
9. McLean, J., Elshourbagy, N., Chang, D., Mahley, R. and Taylor, J. (1984) J. Biol. Chem. 259, 6498-6504.
10. Seqrest, J., Jackson, R., Morrisett, J. and Gotto, A. (1974) FEBS Lett., 38, 247-253.
11. Proudfoot, N. and Brownlee, G. (1981) Nature 252, 359-362.
12. McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) Nucl. Acids Res. 13, 1347-1368.
13. Mount, S.M. (1982) Nucl. Acids Res. 10, 459-472.
14. Schmidt, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.
15. Melton, D.A., Krieg, P.A., Regaliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
16. Breatnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
17. Karanthanasis, S.K., Zannis, V.I., and Breslow, J.L. (1983) Proc. Natl. Acad. Sci. USA 80, 6147-6151.
18. Shoulders, C.C., Kornblihtt, A.R., Munro, B.S., and Baralle, F.E. (1983) Nucleic Acids Res. 11, 2827-2837.
19. Teisberg, P., Gjone, E., and Olaisen, B. (1975) Ann. Hum. Genet. 38, 327-331.