Characterisation of mRNAs encoding the precursor for human apolipoprotein C1

T.J.Knott, M.E.Robertson, L.M.Priestley, M.Urdea¹, S.Wallis² and J.Scott

Division of Molecular Medicine, MRC Clinical Research Centre, Harrow, Middlesex HA1 3UJ, UK, ¹Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA, and ²Department of Biochemistry, St. Mary's Hospital Medical School, London W2 1PG, UK

Received 2 April 1984; Accepted 13 April 1984

ABSTRACT

cDNA clones encoding human apolipoprotein C1 have been isolated from an adult liver cDNA library. Apo C1 mRNA was shown to have two species of approximately 580 and 560 bases by RNA blot hybridisation. The intracellular precursor of apo C1 was inferred from the cDNA sequence to be an 83 amino acid polypeptide consisting of the 57 residue mature protein and an additional 26 residue amino terminal signal peptide. The 5' untranslated regions of the messages are 63 and 40 bases as determined by primer extension and the 3' untranslated region 111 bases. A polyadenylation signal is situated 10 bases 3' of the poly(A) tail. The mRNA level of apo C1 in human liver was significantly greater than that of apo AII and apo E.

INTRODUCTION

Apolipoprotein C1 (apo C1) is a major component (10%) of human very low density lipoprotein (VLDL) and a more minor component (2%) of high density lipoprotein (HDL) (1). Synthesis is believed to occur only in the liver and intestine. In man the protein is a single polypeptide chain of 57 amino acid residues, the sequence of which is known (2,3).

Apo C1 binds phospholipid and can activate lecithin cholesterol acyl transferase (LCAT) *in vitro* (4,5). The capacity to activate LCAT is less than that of apo AI, when unsaturated acyl donors like plasma phosphatidyl choline are used. The relative importance of apo AI and apo C1 in the physiological activation of LCAT remains to be resolved. Apo AI, AII, C1 and CIII also share amino acid sequence homologies that suggest evolution by gene duplication from a common precursor (6). A prominent internal periodicity of 11 amino acid residues is evident in apo AI. This is reflected at the nucleotide level (7). Apo AII, C1 and CIII have a more or less similar amino acid periodicity (6). The linkage of apo AI and CIII on the short arm of chromosome 11 further supports the close relationship of the apo A and C proteins (8).

This report describes the nucleotide sequence of cDNAs corresponding

to two mRNAs coding for apo CII. These results will allow structural comparison between the mRNAs coding for the other apo A, C and E peptides.

MATERIALS AND METHODS

Preparation of Oligonucleotides.

The synthetic oligonucleotide probes used to screen the cDNA library were constructed using the apo CII heptapeptide Lys-Met-Arg-Glu-Trp-Phe-Ser (residues 37 - 43), chosen for minimal degeneracy of coding sequence. Two pools of 64 fold degenerate 20 mers (the third base of Ser being excluded) were synthesized by solid-phase phosphoramidite methodology (9) and purified from 20% sequencing gels. The sequences in each pool were:

Pool I 3'-TTCTACG^NCNCTCTACCAAGTC

Pool II 3'-TTCTACG^NCNCTCTACCAAGAG

('N' indicates a mixture of all four bases).

The oligonucleotide 3'-GCGAGGGCCAGGACCACCAC was used for primer extension.

Screening of adult human liver cDNA library.

The cDNA library was provided by D. Woods and has been described elsewhere (10). The library was screened with the ³²P-labelled oligonucleotide as previously described (11).

Preparation of plasmid DNA and restriction analysis

Plasmid DNA was isolated by the cleared lysate method (12) and further purified on CsCl equilibrium density gradients. Restriction enzymes were bought from BRL or New England Biolabs and used as recommended by the vendor.

DNA sequence analysis.

Dideoxy chain termination sequencing was performed essentially as described by Sanger *et al* (13). Restriction fragments were isolated from agarose gels by electrophoresis on to DE81 paper (14), eluted, ethanol precipitated and ligated into M13mp8 or M13mp9 vectors (15). After transfection into the *E. coli* host JM101, recombinant phage were selected and single stranded DNA templates prepared (16). Sequencing reactions were primed with a universal 17mer obtained from Collaborative Research Inc.

cDNAs were cloned also in the plasmid pUC8 for sequencing by the chemical cleavage method as described by Maxam and Gilbert (17).

Primer Extension.

Primed synthesis on adult human liver poly(A) RNA was performed

using the ^{32}P -labelled oligonucleotide as previously described (11).

RNA preparation and Northern blot analysis.

Total RNA was extracted from adult human liver by the guanidinium thiocyanate method and purified by lithium chloride precipitation and phenol extraction (18). Poly(A) RNA was prepared by two cycles of oligo(dT) chromatography (19). Glyoxalated RNA was separated on 2% agarose gels, transferred to nitrocellulose (20) and hybridised to nick-translated cDNA probes. The characterization of apo AII and apo E cDNA probes is described elsewhere (21,22).

RESULTS.

8,900 clones were screened with each of the pools of oligonucleotide probes; seven gave strong positive signals with pool two. Plasmid DNA was prepared and the restriction patterns of the inserts were similar. The cDNAs ranged between 400 and 480 base pairs (bp) in length. The largest cDNA was selected for DNA sequence analysis. The same clone was labelled and hybridised to a Northern blot of human liver poly(A) RNA. Two RNA species of approximately 580 and 560 nucleotides were detected (Fig. 1).

The longest cDNA clone included 362 bases complementary to apo CII mRNA and a 35 bp poly(A) tract in addition to poly(dC) tails (Fig. 2). The remainder of the 5' sequence not represented in the cDNA library was determined by primed synthesis on human liver mRNA using an oligonucleo-

Figure 1.

Northern blot of adult human liver poly(A)⁺ RNA hybridised with nick-translated apo CII cDNA. Two distinct mRNA species are visible.

Structure of human apolipoprotein C1 mRNA. The positions of restriction enzyme sites within the longest cDNA clone are illustrated. Additional sites for *Sau3A*I and *A1*uI occur in the 5' untranslated region. Regions sequenced by the M13 dideoxy (dotted arrows) and chemical cleavage (solid arrows) methods are indicated.

CCCGCAGCTCAGCCACGGCACAGATCAGCACCACGACCCCTCCCTCGGGGCTC													53
-26													
	Met	Arg	Leu	Phe	Leu	Ser	Leu	Pro	Val	Leu	Val	Val	
GCC	ATG	AGG	CTC	TTC	CTG	TCG	CTC	CCG	GTC	CTG	GTG	GTG	92
Val	Leu	Ser	Ile	Val	Leu	Glu	Gly	Pro	Ala	Pro	Ala	Gln	
GTT	CTG	TCG	ATC	GTC	TTG	GAA	GGC	CCA	GCC	CCA	GCC	CAG	131
- 1 + 1													
Gly	Thr	Pro	Asp	Val	Ser	Ser	Ala	Leu	Asp	10	Lys	Leu	Lys
GGG	ACC	CCA	GAC	GTC	TCC	AGT	GCC	TTG	GAT	AAG	CTG	AAG	170
20													
Glu	Phe	Gly	Asn	Thr	Leu	Glu	Asp	Lys	Ala	Arg	Glu	Leu	
GAG	TTT	GGA	AAC	ACA	CTG	GAG	GAC	AAG	GCT	CGG	GAA	CTC	209
30													
Ile	Ser	Arg	Ile	Lys	Gln	Ser	Glu	Leu	Ser	Ala	Lys	Met	
ATC	AGC	CGC	ATC	AAA	CAG	AGT	GAA	CTT	TCT	GCC	AAG	ATG	248
40													
Arg	Glu	Trp	Phe	Ser	Glu	Thr	Phe	Gln	Lys	Val	50	Lys	Glu
CGG	GAG	TGG	TTT	TCA	GAG	ACA	TTT	CAG	AAA	GTG	AAG	AAG	287
57													
Lys	Leu	Lys	Ile	Asp	Ser	***							
AAA	CTC	AAG	ATT	GAC	TCA	TGA	GGACCTGAAGGGTGACATCCAGGA						
GGGGCCTCTGAAATTTCCACACCCACGCGCTGTGCTGAGGACTCCCGCCAT													
GTGGCCCCAGGTGCCACCAATAAAATCTACCG(A) _n													

The nucleotide sequence of human apolipoprotein C1 cDNA. Translation begins at the first ATG triplet. Amino acids are numbered relative to the N-terminal threonine residue (+ 1) of mature apo C1. The polyadenylation signal AATAAA is underlined.

Figure 4.

Northern blot of adult human liver RNA. Apo AII, apo CI and apo E cDNAs were ^{32}P -labelled to similar specific activities (2×10^8 cpm/ μg DNA). An equal number (2×10^6 cpm/ml) of counts of each probe were hybridised to lanes containing 15 μg of total RNA. Lane a) apo AII, b) apo CI, c) apo E.

tide complementary to nucleotides 73 - 92 (Fig. 3). The cDNA product of reverse transcription appeared as two bands of similar intensity which correspond to species of 98 bases and 75 bases when sized on DNA sequence gels. The DNA sequence of the extended primer could be read to within 10 bases of the 5' end. The mRNAs have 5' untranslated regions of approximately 63 and 40 bases. The first in frame methionine codon is followed by an open reading frame coding for an 83 amino acid polypeptide. The translation termination codon TGA is followed by 111 untranslated bases. The sequence AAUAAA, characteristically found near the 3' terminal of most eukaryote mRNAs, is located 10 nucleotides upstream of the poly(A) tail.

Northern blots of adult human liver RNA have been compared for the level of mRNA for apo AII, apo CI and apo E (Fig. 4).

DISCUSSION

The abundance of apo CI cDNA clones in the library is less than 0.1%. Apo AI, apo AII and apo E cDNAs have a two fold greater abundance (21-24). Comparison of the mRNA levels for apo AII, apo CI and apo E, however, indicates that the mRNA for apo CI is most abundant, being several times that of apo AII and apo E. Apo CI appears to be under-represented in the cDNA library, with the caveat that human tissue is taken under unphysiological conditions. This may be due to loss of cDNAs corresponding to small mRNAs during preparation of the library.

The mRNA has a single open reading frame of 249 bases. When

translated, the 57 residues corresponding to the known sequence of apo C1 are preceded by an additional 26 amino acids which are absent from the mature form of the protein. There are no additional carboxy terminal residues. The amino terminal extension has the characteristics of a signal peptide (25), namely a basic residue next to the initiating methionine, a hydrophobic core region and a C-terminal glycine.

In common with rat apo AIV, rat apo E and avian VLDL apoprotein, apo C1 has no propeptide joining the signal peptide to the mature protein (26). In contrast, human and rat apo AI, the other LCAT activator, has a unique peptide with Gln-Gln proceeding the carboxy-terminus of this mature protein and appears to be processed extracellularly by a specific enzyme (27,28). Apo AII has a propeptide with the usual dibasic residues at its cleavage site (21).

In the cDNA coding for the mature protein, the weak internal homology seen in the amino acid sequence is not found at a nucleotide level (7).

The cDNA probes described in this study will facilitate the structural analysis of the apo C1 gene, chromosomal localization and the determination of expression in other tissues. Genetic variants that may predispose to atherosclerotic vascular disease can also be sought and phylogenetic relationships with other apolipoproteins analysed.

ACKNOWLEDGEMENTS

We are grateful to S. Humphries and D. Woods for the cDNA library, O. Fernando for his valuable contributions and Mrs. S. Schwarz and Miss. C. Boxall for preparation of this manuscript.

REFERENCES

1. Herbert, P.N., Assman, G., Gotto, A.M. Jr. and Fredrickson, D.S. (1983) In: *The Metabolic Basis of Inherited Disease*, Stanbury, J.B., Wyngaarden, J.B., Fredrickson, D.S., Goldstein, J.L. and Brown, M.S. Eds. McGraw-Hill, New York, pp. 589-621.
2. Jackson, R.L., Sparrow, J.T., Baker, H.N., Morrisett, J., Taunton, O.D. and Gotto, A.M. Jr. (1974) *J. Biol. Chem.* 249, 5308-5313.
3. Shulman, R.S., Herbert, P.N., Wehrly, K. and Fredrickson, D.S. (1975) *J. Biol. Chem.* 250, 182-190.
4. Soutar, A.K., Garner, C.W., Baker, H.N., Sparrow, J.T., Jackson, R.L., Gotto, A.M. Jr. and Smith, L.C. (1975) *Biochem.* 14, 3057-3064.
5. Albers, J.J., Lin, J. and Roberts, G.P. (1979) *Artery* 5, 61-75.
6. Barker, W.C. and Dayhoff, M.O. (1977) *Comp. Biochem. Physiol.* B57, 309-315.
7. Karathanasis, S.K., Zannis, V.I. and Breslow, J.L. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 6147-6151.

8. Gerald, P.S. and Grzeschik, K.H. (1984) *Cytogenet. Cell Genet.* 7, 104-107.
9. Beaucage, S.L. and Caruthers, M.H. (1981) *Tetrahedron Lett.* 22, 1859-1862.
10. Woods, D.E., Markham, A.F., Ricker, A.T., Goldberger, G. and Colten, H.R. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5661-5665.
11. Scott, J., Selby, M., Urdea, M., Quinoga, M., Bell, G.I. and Rutter, W.J. (1983) *Nature*, 302, 538-540.
12. Clewell, D.B. (1972) *J. Bacteriol.* 110, 667-676.
13. Sanger, F., Coulson, A., Barrell, B., Smith, A. and Roe, B. (1980) *J. Mol. Biol.* 143, 161-178.
14. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) *Analyt. Biochem.* 112, 295-298.
15. Messing, J. and Vieira, J. (1982) *Gene* 19, 269-276.
16. Winter, G. and Fields, S. (1980) *Nucl. Acids Res.* 8, 1965-1974.
17. Maxam, A.M. and Gilbert, W. (1980) *Methods in Enzymol.* 65, 499-560.
18. Auffray, C. and Rougeon, F. (1980) *Europ. J. Biochem.* 107, 310-314.
19. Bantle, J.A., Maxwell, I.H. and Hahn, W.E. (1976) *Analyt. Biochem.* 72, 413-427.
20. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* 77, 5201-5205.
21. Knott, T.J., Priestly, L.M., Urdea, M. and Scott, J. (1984) (in press).
22. Wallis, S.C., Rogne, S., Gill, L., Markham, A., Edge, M., Woods, D., Williamson, R. and Humphries, S. (1983) *EMBO. J.* 2, 2369-2373.
23. Breslow, J.L., McPherson, J., Nassbaum, A.L., Williams, H.W., Lofquist-Kahl, F., Karathanasis, K. and Zannis, V.I. (1982) *J. Biol. Chem.*, 257, 14639-14641
24. Breslow, J.L., Ross, D., McPherson, J., Williams, H.W., Kurnit, D., Nassbaum, A.L., Karathanasis, S.K. and Zannis, V.I. (1982) *Proc. Natl. Acad. Sci. (U.S.A.)*, 79, 5661-5665.
25. Sabatini, D.D., Kreibach, G., Morimoto, T. and Adesnik, M. (1982) *J. Cell Biol.* 92, 1-22.
26. Shoulders, C.C., Kornblihtt, A.R., Munro, B.S. and Baralle, F.E. (1983) *Nucl. Acids Res.* 11, 2827-2837.
27. Gordon, J.I., Smith, D.P., Andy, R., Alpers, D.H., Schonfeld, G. and Strauss, A.W. (1982) *J. Biol. Chem.*, 257, 971-978.
28. Edelstein, C., Gordon, J.I., Toscas, K., Sims, H.F., Strauss, A.W. and Scanu, A.M. (1983) *J. Biol. Chem.*, 258, 11430-11433.