Identification and DNA Sequence of a Human Apolipoprotein E cDNA Clone*

(Received for publication, September 22, 1982)

Jan L. Breslow,† Joseph McPherson, Alexander L. Nussbaum, Hilary W. Williams, Frederica Lojquist-Kahi, Sotirios K. Karathanasis, and Vassilis I. Zannis

From the Children's Hospital Medical Center, and the §Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

cDNA clones encoding human apolipoprotein E were identified by screening an adult human liver cDNA library with an oligonucleotide probe. The probe was a mixture of synthetic 14-base long DNA oligomers constructed to correspond to all possible codons for apo-E amino acids 218–222. Plasmids from four of the 20 clones selected by this screening procedure were digested with PstI and all had five internal PstI sites with a total length of the cDNA insert of approximately 900 base pairs. DNA sequence analysis of one of these clones, designated pE-301, revealed that it corresponded to apo-E amino acids 81–299, and contained a standard termination codon, polyadenylation signal, and poly A tail. The DNA sequence examined included the known apo-E polymorphic sites at amino acids 112, 145, and 158, and the mutant apo-E phenotypes can all be explained on the basis of a single base substitution in the first position of each of these codons. This work supports the hypothesis that the apo-E polymorphism is due to mutations in the region of DNA coding for the apo-E structural gene.

Apo-E was first identified in 1973 as a component of normal human VLDL (1). This apoprotein has subsequently been demonstrated in chylomicrons, LDL and HDL, and exists in all mammalian species studied thus far (2–5). Apo-E synthesis occurs in liver and reticuloendothelial cells (6–11). Apo-E is secreted from its sites of synthesis as sialo apo-E and subsequently desialated in plasma (10, 12). Plasma apo-E is involved in receptor-mediated recognition of lipoprotein particles by tissues (13, 14). Extrapleithic and hepatic nonreticuloendothelial cells possess a receptor which recognizes both apo-B and apo-E (13, 14). Hepatic tissues also exhibit a receptor which only recognizes apo-E (14). In humans, studies utilizing two-dimensional polyacrylamide gel electrophoresis have revealed six apo-E phenotypes (15–18). Through family studies it was shown that these apo-E phenotypes are the result of a single structural gene locus for apo-E with three common alleles designated e4, e3, and e2 (16–19). According to a new uniform nomenclature system, these phenotypes and the corresponding genotypes are E4/4 = e4, e4; E3/3 = e3, e3; E2/2 = e2, e2; E4/3 = e4, e3; E3/2 = e3, e2; and E4/2 = e4, e2 (19). Studies of apo-E phenotypes in the general population have shown that the most common allele, e3, occurs with a frequency of 0.74, whereas the other two alleles each occur with frequencies of from 0.11 to 0.15 (18, 20). Further studies have shown that familial type III HLP, a disease characterized by hyperlipidemia, xanthomatosis, and premature atherosclerosis is associated in over 90% of cases with the apo-E phenotype E2/2 (18). In in vitro receptor binding studies, apo-E of the E2/2 phenotype isolated from patients with type III HLP has been shown to bind poorly to apo-B/E receptors (21). Thus, it appears that a mutation in an allele specifying an apo-E structural gene may be the underlying defect that leads to altered lipoprotein metabolism and the clinical symptomatology of this disease. Rall et al. (22–24) have recently studied the protein sequence of apo-E from individuals with different apo-E phenotypes. Through these studies, they have identified three polymorphic amino acid sites. In the most common apo-E polypeptide specified by the e3 allele, these sites are as follows: Cys112, Arg145, and Arg158. The apo-E allele, e4, differs from the e3 allele by having a Cys112 → Arg substitution (22–24). In the case of the apo-E allele, e2, individuals with type III HLP and the apo-E phenotype E2/2 have been found with either an Arg145 → Cys, or an Arg158 → Cys substitution (22–24). Furthermore, Ghiselli, et al. (25) have described an individual with type III HLP who has no detectable apo-E. To better understand the genetic determinants of functional apo-E in humans, as well as the regulation of apo-E synthesis, it will be necessary to study the genomic organization and sequence of the apo-E gene and its transcripts. As a first step towards this goal, we have isolated and characterized bacterial clones containing portions of human apo-E cDNA.

MATERIALS AND METHODS

Preparation of the Oligonucleotide Probe—To probe a human adult liver cDNA library for apo-E cDNA clones, an oligonucleotide mixture was synthesized which corresponds to apo-E amino acids 218 to 222 (Met-Glu-Glu-Met-Gly) (22–24). This region was selected because these amino acids are specified by relatively unambiguous codons. Fig. 1 shows the sequence of mRNA from 5′ to 3′ specifying apo-E amino acids 218 to 222, as well as the cDNA sequence from 3′ to 5′. This cDNA sequence was synthesized by a solid phase phosphite triester method using the reaction conditions and procedures of Matteucci and Caruthers (26, 27) and Beaucage and Caruthers (28). Briefly, a sample of 25 μg of functionalized silica gel (20) charged with 1.3 μmol of DMTrT by attachment of its 3′OH group through ester linkage to the solid phase, was unblocked at the 5′-OH group with a Lewis acid (saturated ZnBr2/aqueous CH3NO2). The sample was condensed with 10 μg of the protected nucleoside phosphoramidite of adenosine; DMTrAP-(MeO)-NEt2, 5′-dimethoxytrityl-2′(methoxy)-diethylaminophosphite; HLP, hyperlipoproteinemia; HPLC, high pressure liquid chromatography; bp, base pairs; SDS, sodium dodecyl sulfate.
Apo-E cDNA Clone

Fig. 1. DNA base sequence of the mixture of oligomers synthesized and used to probe the human adult liver cDNA library.

Apo-E rDNA Probe

A.B. Region 218 219 220 221 222

(M5'-terminus) MET GLU GLU MET GLY

RNA 5' AUG GAG GAG AUG 66 3'

DNA 3' TAC CTC CTC TAC CTC 5'

Fig. 2. Map of clone pE-301. PstI sites are shown as are the positions on the polypeptide chain to which they correspond.

Fig. 3. Complete DNA sequence of clone pE-301.

| Nucleotide | DNA String | Description |
|------------|------------|-------------|
| 1 | GAG | Apo-E cDNA Clone |
| 2 | GAG | Apo-E cDNA Clone |
| 3 | AUG | Apo-E cDNA Clone |
| 4 | AUG | Apo-E cDNA Clone |
| 5 | GAG | Apo-E cDNA Clone |
| 6 | GAG | Apo-E cDNA Clone |
| 7 | AUG | Apo-E cDNA Clone |
| 8 | AUG | Apo-E cDNA Clone |
| 9 | GAG | Apo-E cDNA Clone |
| 10 | GAG | Apo-E cDNA Clone |
| 11 | AUG | Apo-E cDNA Clone |
| 12 | AUG | Apo-E cDNA Clone |
| 13 | GAG | Apo-E cDNA Clone |
| 14 | GAG | Apo-E cDNA Clone |
| 15 | AUG | Apo-E cDNA Clone |
| 16 | AUG | Apo-E cDNA Clone |
| 17 | GAG | Apo-E cDNA Clone |
| 18 | GAG | Apo-E cDNA Clone |
| 19 | AUG | Apo-E cDNA Clone |
| 20 | AUG | Apo-E cDNA Clone |
| 21 | GAG | Apo-E cDNA Clone |
| 22 | GAG | Apo-E cDNA Clone |
| 23 | AUG | Apo-E cDNA Clone |
| 24 | AUG | Apo-E cDNA Clone |
| 25 | GAG | Apo-E cDNA Clone |
| 26 | GAG | Apo-E cDNA Clone |
| 27 | AUG | Apo-E cDNA Clone |
| 28 | AUG | Apo-E cDNA Clone |
| 29 | GAG | Apo-E cDNA Clone |
| 30 | GAG | Apo-E cDNA Clone |

RESULTS

Identification of apo-E cDNA Clones—10,000 cDNA clones were transferred to nitrocellulose filters and hybridized to the oligonucleotide probe as described under "Materials and Methods." Initial washing of the filters at 23°C showed significant nonspecific hybridization, while washings at 30 or 40°C substantially reduced this background, and positive clones could be identified. About 30 positive clones were observed after the 30°C wash, 20 of which maintained the hybridization signal after the 40°C wash. A 50°C wash totally abolished the signal from all clones. Four out of the 20 clones selected after the 40°C wash were used for further analysis. The plasmid preparations from these four clones were digested with PstI and all four plasmids had five internal PstI sites with a total length of the cDNA inserts of approximately 900 bp. DNA sequence analysis of one of these clones was undertaken to verify that it corresponded to apo-E cDNA.

Sequencing of Clone pE-301—Clone pE-301 has been mapped by a combination of PstI digestion and DNA sequencing of the fragments obtained. Fig. 2 shows the location of the PstI sites in pE-301 and their relationship to the amino acid sequence of the apo-E polypeptide as reported by Rall et al. (22, 24) and Weisgraber et al. (23). Fig. 3 shows the complete DNA base sequence of this clone. As can be seen, clone pE-301 corresponds to apo-E amino acids 81-299 and contains a TGA termination codon, a 158-bp long 3' untranslated region and a 44-bp long portion of the poly A tail of the apo-E mRNA. At the polymorphic sites 112, 145, and 158, this cDNA clone specifies cysteine, arginine, and arginine, respectively.

DISCUSSION

In previous studies, we have shown that apo-E is a relatively abundant secretory product of human fetal and adult liver in organ culture and of human hepatoma cells in tissue culture (10, 12). In addition, we have recently shown that approximately 0.2% of the protein synthesized in a cell-free translation system of human liver cytoplasmic poly A containing RNA is apo-E (38). This implies that apo-E mRNA is abundant in a total liver mRNA preparation and, therefore, human liver cDNA libraries should be enriched in clones containing the apo-E cDNA sequence. Based on the amino acid sequence for apo-E, we designed and synthesized a mixture of oligonucleotides and used these as a probe to screen an adult human liver cDNA library in an attempt to isolate these apo-E cDNA clones. The region that spans apo-E amino acids 13-15 bases in length are sufficient for the detection of a unique gene in a genomic yeast DNA library (36).
APo-E POLYMORPHIC SITES

| Position | A (a.m.) | C (g.m.) | Mutation |
|----------|---------|---------|----------|
| 112      | Cys     | Tle     | Arg      |
| 145      | Arg     | Cys     | Tle      |
| 158      | Arg     | Cys     | Tle      |

The single-base mutation in the DNA necessary to accommodate these amino acid changes is indicated.

To establish the hybridization condition for the screening, we used empirical formulae (23) to estimate that the Tm of a perfectly matched 14-base long DNA oligomer with a 50% content of GC should be 51 °C. This indicated that hybridization at 1 M salt at 31-36 °C should satisfy the opposing requirements for sensitivity (low stringency) and specificity (high stringency). As predicted, screening of the cDNA bank with the mixture of labeled 14-base-long DNA oligomers at room temperature (low stringency) showed a high background of nonspecific hybridization. However, hybridizations at 30-40 °C (high stringency) increased the specificity without significantly affecting the sensitivity of the hybridization signal. In this manner, 10,000 clones of the cDNA library were screened using the probe under high stringency hybridization conditions, and 20 clones were identified as strongly hybridizing colonies. Further examination of 4 of these 20 positive clones showed that they contained five internal PstI sites at positions 112, 145, and 158. This particular DNA clone would be of the phenotype E3 and this clone has demonstrated amino acid changes at amino acids 112, 145, and 158. This particular DNA clone would be of the phenotype E3 and this clone has demonstrated amino acid changes at amino acids 112, 145, and 158. This particular DNA clone would be of the phenotype E3 and this clone has demonstrated amino acid changes at amino acids 112, 145, and 158. This particular DNA clone would be of the phenotype E3 and this clone has demonstrated amino acid changes at amino acids 112, 145, and 158. Therefore, the hypothesis that the apo-E polymorphism is due to mutations in the region of DNA coding for the apo-E structural gene is supported by the speculation that the apo-E polymorphism is due to mutations in the region of DNA coding for the apo-E structural gene is, and probably underlie a condition in humans called type III HLP which leads to hyperlipidemia, xanthomatosis, and premature atherosclerosis.

Acknowledgments—We would like to thank Drs. D. Woods, E. Prochownik, and A. Michelson for providing us with the cDNA library, and L. Duda, and M. Hardiman for their expert assistance.

REFERENCES
1. Shore, V. G., and Shore, B. (1973) Biochemistry 12, 502-507.
2. Mahley, R. W. (1978) in Disturbances in Lipid and Lipoprotein Metabolism (Dietschy J., Gotto, A. M., Jr., and Ortko, J. A., eds) pp. 181-197, Clinical Physiology Series, American Physiology Society.
3. Shore, V. G., Shore, B., and Hart, R. G. (1974) Biochemistry 13, 1579-1585.
4. Chapman, M. J., and Mills, G. L. (1977) Biochem. J. 167, 9-21.
5. Zannis, V. I., and Breslow, J. L. (1982) Mol. Cell. Biochem. 42, 3-20.
6. Marsh, J. B. (1976) J. Lipid Res. 17, 85-90.
7. Hamilton, R. L., Williams, M. C., Fielding, C. J., and Havel, R. J. (1978) J. Clin. Invest. 58, 897-909.
8. Felker, T. E., Fainaru, M., Hamilton, R. L., and Havel R. J. (1977) J. Lipid Res. 18, 4654-4673.
9. Wu, A. L., and Windmueller, H. G. (1979) J. Biol. Chem. 254, 7315-7322.
10. Zannis, V. I., Kurnit, D. M., and Breslow, J. L. (1982) J. Biol. Chem. 257, 536-544.
11. Basu, S. K., Ho, Y. K., Brown, M. S., Bilheimer, D. W., Anderson, R. G. W., and Goldstein, J. L. (1982) J. Biol. Chem. 257, 9788-9795.
12. Zannis, V. I., Breslow, J. L., SanGiacomo, T. R., Aden, D. P., and Knowles, B. B. (1981) Biochemistry 20, 7095-7096.
13. Innerarity, T. L., and Mahley, R. W. (1978) Biochemistry 17, 1440-1447.
14. Hui, D. Y., Innerarity, T. L., and Mahley, R. W. (1981) J. Biol. Chem. 256, 5646-5655.
15. Zannis, V. I., and Breslow, J. L. (1980) J. Biol. Chem. 255, 1759-1769.
16. Zannis, V. I., Just, P. W., and Breslow, J. L. (1981) Am. J. Hum. Genet. 33, 11-24.
17. Zannis, V. I., and Breslow, J. L. (1981) Biochemistry 20, 1033-1041.
18. Breslow, J. L., Zannis, V. I., SanGiacomo, T. R., Third, J. L. H., Tracy, T., and Glueck, C. J. (1982) J. Lipid Res. 23, 1224-1235.
19. Zannis, V. I., Breslow, J. L., Uttermann, G., Mahley, R. W., Weisgraber, K. H., Havel, R. J., Goldstein, J. L., Brown, M. S., Schenfeld, G., Hazzard, W. R., and Blum, C. S. (1982) J. Lipid Res. 23, 911-914.
20. Uttermann, G., Langenback, U., Beisiegle, U., and Weber, W. (1980) Am. J. Hum. Genet. 32, 339-347.
21. Schneider, W. J., Kovaren, P. T., Brown, M. S., Goldstein, J. L., Uttermann, G., Weber, W., Havel, R. J., Kotite, L., Eane, J. F., Innerarity, T. L., and Mahley, R. W. (1981) J. Clin. Invest. 68, 1075-1085.
22. Rall, S. C., Weisgraber, K. H., and Mahley, R. W. (1981) J. Biol. Chem. 257, 4171-4178.
23. Weisgraber, K. H., Rall, S. C., and Mahley, R. W. (1981) J. Biol. Chem. 256, 9077-9080.
24. Rall, S. C., Weisgraber, K. H., Innerarity, T. L., and Mahley, R. W. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 4696-4700.
25. Ghiissi, G., Schaefer, E. J., Gascon P., and Brewer, H. B., Jr. (1981) Science 214, 1239-1241.
26. Matteucci, M. D., and Caruthers, M. H. (1981) J. Am. Chem. Soc. 103, 3185-3191.
27. Matteucci, M. D., and Caruthers, M. H. (1981) Tetrahedron Lett. 21, 719-722.
28. Beaucage, S. L., and Caruthers, M. H. (1981) Tetrahedron Lett. 22, 1859-1862.
29. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G., and Colten, H. R. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 5661-5665.
30. Grunstein, M., and Hogness, D. S. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 3961-3965.
31. Birnbaum, H. C., and Doly, J. (1979) Nucleic Acids Res. 5, 1513-1523.
32. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560-564.
33. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467.
34. Messing, J., Crea, R., and Seeburg, P. (1981) Nucleic Acids Res. 9, 309-321.
35. Breslow, J. L., Zannis, V. I., Goldberger, G., Ross, D. G., and Kurnit, D. M. (1982) Circulation, in press.
36. Szostak, J. W., Siles, J. L., Tye, B. K., Chiu, P., Sherman, F., and Wu, R. (1976) Methods Enzymol. 88, 419-427.
37. Davis, R. W., Botstein, D., and Roth, J. R., Jr. (1980) Advanced Bacterial Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.