Volume 213, number 1, 221–226 FEB 04506 March 1987

The human preproapolipoprotein C-II gene

Complete nucleic acid sequence and genomic organization

Silvia S. Fojo, Simon W. Law and H. Bryan Brewer, jr

Molecular Disease Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

Received 8 January 1987

The complete nucleic acid sequence of human preproapolipoprotein (apo) C-II has been determined from 2 apoC-II clones isolated from 2 different human genomic DNA libraries. The cloned fragments were approx. 14 and 18 kb long, and sequence analysis established that the apoC-II gene consists of 3338 nucleotides containing 3 intervening sequences of 2391, 167, and 298 bases. The first intron is located within the 5′-untranslated region of apoC-II and contains 4 Alu type sequences. The second intron interrupts the codon specifying amino acid −11 of the apoC-II signal peptide. The last intron, which contains a 38 bp sequence which is repeated 6 times, interrupts the codon specifying for amino acid +44 of the mature apolipoprotein.

Preproapolipoprotein C-II; Genomic sequence; Oligonucleotide probe; Lipoprotein metabolism

1. INTRODUCTION

Human apolipoprotein C-II plays a central role in lipid metabolism as a cofactor for lipoprotein lipase, the enzyme that hydrolyzes triglycerides in plasma chylomicrons and very low density lipoprotein [1–3]. The amino acid sequence of human apoC-II has been previously reported by our laboratory [4]. The major plasma isoform of apoC-II is a proprotein composed of a single polypeptide chain of 79 amino acids which then undergoes proteolytic cleavage to form the mature 73 amino acid apolipoprotein [5]. Synthetic oligonucleotides based on this amino acid sequence enabled us to isolate an apoC-II clone from the cDNA libraries. Nucleic acid sequence analyses of human apoC-II cDNA clones revealed that the apoC-II mRNA encodes a 101 amino acid long apolipoprotein which includes a 22 amino acid signal peptide and a 79 amino acid residue plasma apolipoprotein [6–9]. By utilizing human-mouse somatic cell hybrids we have localized the apoC-II gene to chromosome 19 [10]. Southern blot analyses of DNA of patients with familial apoC-II deficiency have thus far revealed no major insertions or deletions in the apoC-II gene [11,12]. However, the deficiency of apoC-II in the plasma of these kindreds may be due to either minor mutations in the structural or regulatory regions of the apoC-II gene or to a derangement in post-transcriptional processes.

In order to better understand the biosynthesis and processing of apoC-II in both normal and diseased states, we have cloned the gene for normal human apoC-II. In this report, we present the complete nucleic acid sequence and genomic organization of the preproapoC-II gene as determined from 2 different libraries and compare it with the previously published sequence of Wei et

Correspondence address: S.S. Fojo, Molecular Disease Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

Abbreviations: apo, apolipoprotein; VLDL, very low density lipoproteins; bp, base pair

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

al. [13]. We believe this work will provide the basis for understanding the underlying defects in patients with apoC-II deficiency.

2. MATERIALS AND METHODS

2.1. *Cloning and screening of genomic libraries for apoC-II clones*

A human placental genomic DNA phage library was a generous gift of Dr Philip Leder. A second library was made by ligating the 15–20 kb size DNA obtained by a partial *Mbo*I digest of normal human white blood cells with EMBL3 vector DNA. *E. coli* P2392 was then infected with the EMBL3 recombinant phage. Screening of both libraries was performed as described [14].

2.2. *Large scale preparation of phage DNA and subcloning into pBR322*

Large scale preparations of phage DNA were performed as described by Maniatis et al. [15] with minor modifications [14]. 5 kb *Bam*HI and 3.7 kb *Eco*RI fragments of insert genomic DNA isolated from the phage containing the apoC-II insert were purified by agarose gel electrophoresis and inserted into the *Bam*HI and *Eco*RI sites of pBR322. Recombinant clones were selected by testing for antibiotic resistance and hybridization to the apoC-II cDNA probe. Plasmid DNA was isolated as described [6].

Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Laboratories. Reactions were performed under the conditions specified by the suppliers.

2.3. *DNA sequencing and primer extension of apoC-II mRNA*

DNA fragments were end labelled as reported [6] and sequenced by the chemical cleavage method of Maxam and Gilbert [16] or the dideoxynucleotide chain termination method of Sanger et al. [17] following subcloning into the M13 vectors mp18 and mp19.

A 30 base-long synthetic oligonucleotide complementary to amino acid residues −9 to +1 (Val-Gln-Gly-Thr-Gln-Gln-Pro-Gln-Gln-Asp) of apoC-II was obtained from OCS Laboratories and used as a primer for 5′-extension of apoC-II mRNA. 24 ng of primer and 15 μg of human liver polyadenylated RNA were utilized for 5′-extension by the method described in [18]. The extended product was sequenced by the Maxam and Gilbert procedure [16] and then analyzed on 8% urea polyacrylamide sequencing gels.

3. RESULTS AND DISCUSSION

Eight apoC-II genomic clones were identified by screening 1.5×10^6 phage from each library with the apoC-II cDNA probe. Two clones were grown in large scale and analyzed following restriction endonuclease digestion by Southern blot. These clones contained a 14 and an 18 kb fragment of human genomic DNA inserted into the *Bam*HI cloning site of Charon 28 and EMBL3, respectively. A 5 kb *Bam*HI and a 3.7 kb *Eco*RI fragment containing the apoC-II gene were purified by agarose gel electrophoresis and subcloned into the *Bam*HI and *Eco*RI sites of pBR322, respectively.

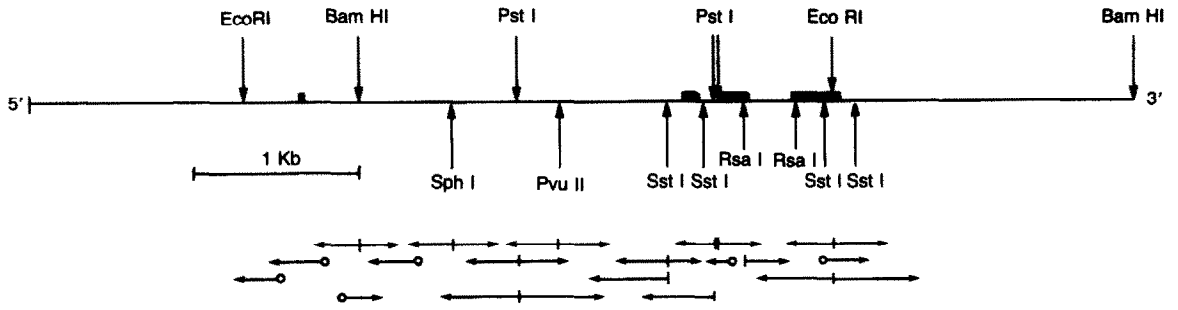

Fig.1. Restriction map of the preproapoC-II gene. The four exons are indicated by the solid bars. The strategy for sequencing the preproapoC-II gene is illustrated by the horizontal arrows. Open circles indicate the use of synthetic oligonucleotides as primers for sequencing. Only enzymes used to generate fragments for sequencing are indicated.

One recombinant of each was then used for subsequent restriction enzyme and sequence analysis.

Fig.1 illustrates the restriction endonuclease cleavage map and sequence analysis strategy of the apoC-II genomic clones. A combination of the Maxam and Gilbert [16] and Sanger et al. [17] sequencing procedures was utilized.

The complete nucleotide sequence of the preproapoC-II is shown in fig.2. A total of 3840 nucleotides were sequenced including the entire apoC-II gene, 200 bp in the 5′-flanking region and 273 bp in the 3′-flanking region. The apoC-II gene is composed of 4 exons interrupted by 3 introns. All 3 introns begin and end with the dinucleotides GT and AG, respectively, and conform to the consensus sequence around exon-intron splice junctions [19]. Primer extension studies established that the 5′-untranslated region of apoC-II mRNA consists of 38 bp and contains a guanine residue at the mRNA start site (nucleotide no.230) which is the second most commonly reported nucleotide at the mRNA start site [19]. Thirty bases upstream from this guanine residue is a typical Goldberg-Hogness box [19] at nucleotides 200–206. Another putative promoter region, the CAAT-box, has been detected in several eukaryotic genes 70–80 bp upstream from the mRNA start site [19,20]. In the apoC-II gene, however, we were unable to identify CAAT-box-like sequences 200 nucleotides upstream from the TATA box.

The first exon consists of 25 nucleotides, is 30 bp downstream from a typical TATA box structure, and encodes most of the 5′-untranslated region of apoC-II mRNA. It is followed by an unusually long first intron which contains 2391 bp and 4 Alu type repetitive sequences which are labelled Alu 1–4 in fig.2. Alu sequences are also present in the intervening sequences of other apolipoprotein genes including apoA-II [14,21], apoC-III [22] and apoE [23,24]. The significance of the Alu repeats found within mammalian genes is unclear, but it has been suggested that they may play a role in DNA replication or gene regulation [25]. We have also identified a 22 dinucleotide sequence of GT repeats followed by a 7 dinucleotide sequence of GA repeats between nucleotides 372 and 429 of the first intron. A similar region of GT repeats is present in the second intron of the apoA-II gene [14]. GT repeats are fairly abundant in the mammalian genome [26], and it has been suggested that these sequences may be associated with an increased frequency of recombination [27]. Fig.3 is an autoradiograph of the sequencing gel of the DNA strand complementary to this GT-repeat sequence.

The second exon of the apoC-II gene is 68 bp long and encodes the rest of the 5′-untranslated region of the apoC-II mRNA and amino acid −28 (methionine) to −11 (phenylalanine) of the signal peptide. The second and third exons are interrupted by the second intron which consists of 167 bp and interrupts the codon for amino acid −11 of the signal peptide. The third (160 bp) and fourth (229 bp) exons encode amino acids −10 (Glu) to +44 (Arg) and amino acids +45 (Asp) to the polyadenylation site at the 3′-end of the apoC-II gene, respectively. The third intron is 298 bp long and contains a 38 bp long sequence that is repeated 6 times within the intron between nucleotides 3091 and 3312, suggesting that duplication of this region of the apoC-II gene occurred at some point during evolution. Nucleotides 3220–3262 of this repeat sequence show 90% homology with nucleotides 424 to 471 of the 5′-flanking region of the apoE gene [23].

A comparison of the complete preproapoC-II genomic sequence obtained from our human placental library (shown in fig.1) and our white blood cell library reveals 4 base differences between them. These include the deletion of a T and a G at positions 63 and 858, and the insertion of AA at position 1347 within intron 1 as shown in fig.1. However, we identified 150 separate nucleotide differences from the previously published apoC-II sequence of Wei et al. [13] which was determined from a human peripheral blood lymphocytes genomic library. These differences are illustrated in fig.2. Most of the base changes are located within the first intron or the 3′-untranslated region of the gene whereas no differences were found in the exons. Major differences include a 22 versus an 8 dinucleotide sequence of GT repeats between bp 372 and 414 of our sequence (see fig.3) and the presence of an extra *Sst*I restriction enzyme site at position 2539–2544 within the 1st intron of the apoC-II gene. The reason for the variability of the apoC-II sequence from different genomic libraries is unclear at this time.

We believe that the availability of genomic and

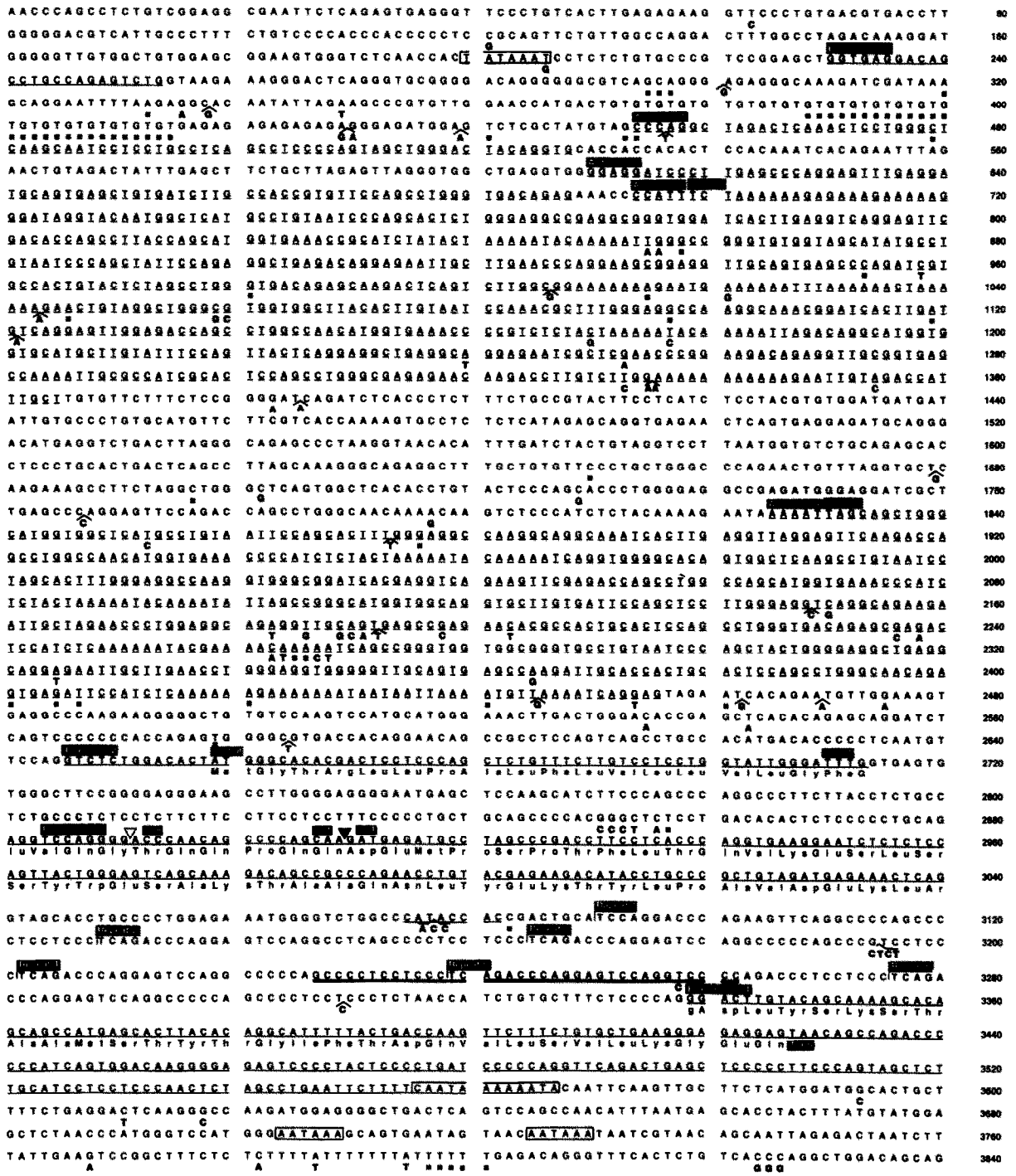

Fig.2. Complete nucleotide sequence of the preproapoC-II gene. TATA-box and polyadenylation signals are boxed. Exons are underlined and numbered. Alu sequences are underlined by an interrupted line and numbered. The beginning and end of the 6-times repeated sequences within intron III are indicated by vertical bars and numbered. The 6 bp of the core enhancer sequence and the 42 bp homologous to the apoE 5′-flanking sequence are both underlined with

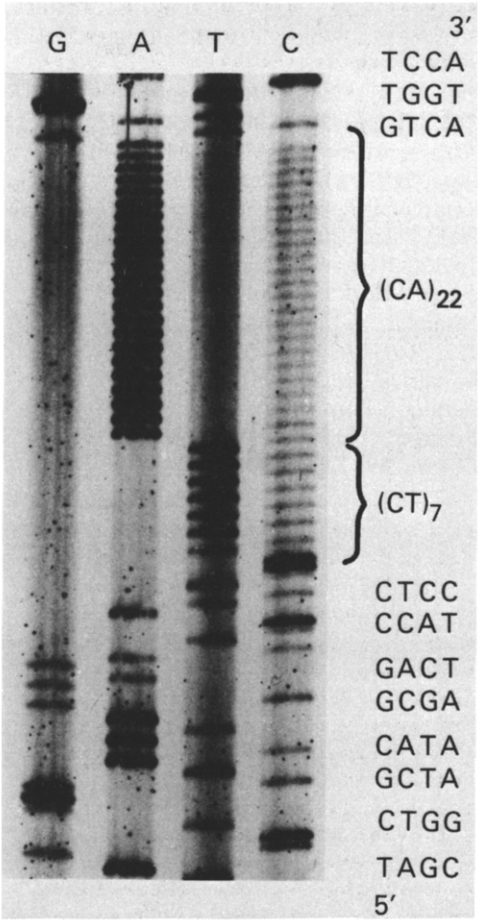

Fig.3. Autoradiograph of the sequencing gel of the DNA strand complementary to the GT-GA repeat sequence within the first intron. Twenty-two dinucleotide GT repeats and 7 dinucleotide GA repeats are illustrated. Sequencing was performed by the dideoxynucleotide chain termination method of Sanger.

cDNA clones of preproapoC-II will permit the investigation of the molecular mechanisms involved in the regulation of apoC-II biosynthesis. In addition, the complete genomic sequence of normal preproapoC-II will provide the basis for understanding the underlying defects in patients with apoC-II deficiency and other dyslipoproteinemias.

ACKNOWLEDGEMENTS

We would like to thank Dr Steven Grant for his help with the sequencing procedures and Mrs Imogene Surrey for assistance with this manuscript.

REFERENCES

[1] LaRosa, J.C., Levy, R.I., Herbert, P., Lux, S.E. and Fredrickson, D.S. (1970) Biochem. Biophys. Res. Commun. 41, 45–62.
[2] Havel, R.J., Shore, J.G., Shore, B. and Bier, D.M. (1970) Circ. Res. 27, 597–600.
[3] Smith, L.C., Pounall, H. and Gotto, A.M. (1978) Annu. Rev. Biochem. 47, 751–777.
[4] Hospattankar, A.V., Fairwell, T., Ronan, R. and Brewer, H.B. jr (1983) J. Biol. Chem. 259, 318–322.
[5] Fojo, S.S., Taam, L., Fairwell, T., Ronan, R., Bishop, C., Meng, M.S., Hoeg, J.M., Sprecher, D.L. and Brewer, H.B. jr (1986) J. Biol. Chem. 261, 9591–9594.
[6] Fojo, S.S., Law, S.W. and Brewer, H.B. jr (1984) Proc. Natl. Acad. Sci. USA 81, 6354–6357.
[7] Jackson, C.L., Bruns, G.A. and Breslow, J.L. (1984) Proc. Natl. Acad. Sci. USA 81, 2945–2949.
[8] Myklebost, O., Williamson, B., Marsham, A.F., Myklebost, S.R., Rogers, J., Woods, D.E. and Humphries, S.E. (1984) J. Biol. Chem. 259, 4401–4404.
[9] Sharpe, C.R., Sideli, A., Shelley, C.S., Lucero, H.A., Shoulders, C.C. and Baralli, F.E. (1984) Nucleic Acids Res. 12, 3917–3932.
[10] Fojo, S.S., Law, S.W., Brewer, H.B. jr, Sakaguchi, A.Y. and Naylor, S.L. (1984) Biochem. Biophys. Res. Commun. 122, 687–693.
[11] Fojo, S.S., Law, S.W., Sprecher, D.L., Gregg, R.E., Baggio, C. and Brewer, H.B. jr (1984) Biochem. Biophys. Res. Commun. 124, 308–313.
[12] Humphries, S.E., Williams, L., Myklebost, O., Stalenhoef, A.F.H., Demacken, P.N.M., Baggio, G., Crepaldi, G., Galton, D.J. and Williamson, R. (1984) Hum. Genet. 67, 151–155.
[13] Wei, C.-F., Tsao, Y.-K., Robberson, D.L., Gotto, A.M., Brown, K. and Chan, L. (1985) J. Biol. Chem. 260, 15211–15221.
[14] Lackner, K.J., Law, S.L. and Brewer, H.B. jr (1985) Nucleic Acids Res. 13, 4597–4608.

double lines. The open triangle indicates the cleavage site for the leader sequence. The solid triangle indicates the cleavage site between proapoC-II and mature apoC-II. Sequence differences between the apoC-II sequence reported in this paper and that determined by Wei et al. [13] are illustrated as nucleotide deletions (■), insertions (∧) or substitutions and are shown in bold letters.

[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Handbook, Cold Spring Harbor Laboratory, NY.
[16] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[17] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[18] Law, S.W., Gray, C. and Brewer, H.B. jr (1983) Biochem. Biophys. Res. Commun. 112, 257–264.
[19] Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383.
[20] Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucleic Acids Res. 8, 127–142.
[21] Tsao, Y.-K., Wei, C.-F., Robberson, D.L., Gotto, A.M. jr and Chan, L. (1985) J. Biol. Chem. 260, 15222–15231.
[22] Protter, A.A., Levy-Wilson, B., Miller, J., Benson, G., White, T. and Seilhamer, J.J. (1984) DNA 3, 449–456.
[23] Paik, Y.K., Chang, D.J., Reardon, C.A., Davies, C.E., Mahley, R.W. and Taylor, J.U. (1985) Proc. Natl. Acad. Sci. USA 82, 3445–3449.
[24] Das, H.K., McPherson, J., Bruno, G.A.P., Karathanasis, S.K. and Breslow, J.L. (1985) J. Biol. Chem. 260, 6240–6247.
[25] Taylor, J.H. and Watanabe, S. (1981) ICN-UCLA Symp. Mol. Cell. Biol. 22, 597.
[26] Nishioka, Y. and Leder, P. (1980) J. Biol. Chem. 255, 3691–3694.
[27] Kilpatrick, M.W., Klysik, J., Singleton, C.K., Zarling, D.A., Jovin, T.M., Hanan, L.H., Erlanger, B.F. and Wells, R.D. (1984) J. Biol. Chem. 259, 7268–7274.