

Complete nucleotide sequence of the gene encoding the rat apolipoprotein E

Chikafusa Fukazawa, Akiyo Matsumoto and John M. Taylor¹

Genetic Engineering Laboratory, National Food Research Institute, Ministry of Agriculture, Forestry and Fisheries, Yatabe, Tsukuba, Ibaraki 305, Japan and ¹Gladstone Foundation Laboratories, Cardiovascular Research Institute, University of California, San Francisco, CA 94140, USA

Submitted 4 November 1986

Apolipoprotein (apo) E is a Mr = 34,000 glycoprotein that serves as a major plasma apolipoprotein found in both very low density and high density lipoproteins. Recent studies suggest that apo E plays a central role in the transport and removal of cholesterol-laden lipoproteins from the circulation (1). A recombinant clone, which covers complete apo E gene in a single insertion, has been isolated from a recombinant bacteriophage library prepared from a rat liver DNA, using the corresponding cDNA as a probe (2). The gene spans 2776-base pair from the apparent start site of transcription (+1) to the 3'-terminal nontranslated region adjacent to poly (A) tail. The 3' mRNA end may occur at any of the three residues shown as arrows. The TATA sequence (-30, -24) and the AATAAA polyadenylation signal (+2756-2761) are underlined. The gene is interrupted by three intervening sequences (indicated by small letters) and both the lipid-binding region and a potential receptor binding domain are encoded in exon 4. Detailed analyses of the protein are presented elsewhere (3).

```

cggggatggggagagttcaccgtggcagaggaatcaacaagaggggcccagggttaaaactcagtttccaccagaaagctgagcctcttaacagatagagccccaccatttccat -295
ttaaagctccagccctttccgtacaacgaatgcaccagaccgccgggaaggggagaagcagttactcagtgcccccagcaccaggcctggaattattcaatgaggagtcagctctttt -175
CAT-box
tgggggggggggggtcacaaggtcaaaaactccacccttttctctgcctctgtgaaggggggggagagacaccgccctcgtgacaggggggtggcccagccgcccttagccctgag -55
CAT-box          w          (exon 1)
gagggggcgggcaggggggagtcctataattggacaggtctgggatccggtcccCTGCTCAGACCCCGGAGGCTAAGGAGTGTGTTGCGGAAGGAGCTGtaagacaagcttgggctggcg 66
TATA-box          (+1) cap site
attcaccagggggcttggtaagactggggcagggaccctgaaatccgctggagtcaggaacacagccacaattattagaaaagcaggaagcccgatagaagacttaggggggggga 186
gacaacactaagactcgtgaggttaactagcctttggccgggtcagaacagatggagggagatggaggggtcatctccagagactcgtggtgcacggcctacggcctcaactagaaggtagt 306
atcaactcttaagcaaggaccgcgatgaccttaggttgcgtttatgtaaatcactactgatttccatcacagctcccaaaattaccctggccagtttcaaaagcgaagaccctcatgcc 426
tgagctccgaagttctggggctcgggggggtaccacttcgaggatggagggacgattaaagcttacattagctactaaaccagagaccggcctagcaggggaaggtagaaggaatg 546
taaatggaccacggcgtgctcctctgctgtggaattgaaccctaggactggagcctggaatttttggcagcgggtccaccgggggtcgtagatagagattgagggggggagtaa 666
atagacccttgggcagatgctctattgtggagatgtttgtgatgactcacagaccctgagaagggaaataaggtgggggctctttgggtggcccgagtttctccaccctaccactgggt 786
          -18          -10          -4
          MetLysAlaLeuIrpAlaLeuLeuValProLeuLeuThrG
          (exon 2)
ttcaaaagacagcttttcttccgcagACTGGCCAATCACCACTGGGAAGATGAAGGCTCTGTG66GCCCTGCTGTGGTCCCATTGTCTGACAGtattggggcaaggggtggctttccctgc 906
tggcaggttgggttggggcagttctgagacctctgagatcaaatagccctgagcagctgttttacatgagccctgtcttccctatcccaatagagagattgatgctcagaggg 1026
taaaggtctgtctataaagctcgaagacttctctcggagcccaacatggtggaagagagagcggattctctctagtgtctctgacctcaagggtacacacagacacacaatta 1146
aaaagttgaagttgtaattctctagtcgcttttaaaggggggtatacatttcaaccctaccagcctagctctcagctctgttctacataggaatactagacagttgactgtgtag 1266
          -3          1
          1yCysLeuAla Glu
tatctaaacagactccacgactgaactccagagcactgttcttcaagagaccacaagcactgtaggctgaccagccttaaaacttactactctcacagagATGGCTGGCC GAG 1385
          (exon 3)          10          20          30          40
GlyGluLeuGluValThrAspGlnLeuProGlyGlnSerAspGlnProIrpGluGlnAlaLeuAsnArgPheIrpAspIyrLeuArgIrpValGlnThrLeuSerAspGlnValGlnGlu
GGAGAGCTGGAGGTGACAGATCAGCTCCCAAGGCAAAAGCCACCAACCTGGGAGCAGGCCCTGAACCCCTTCTGGGATTAACCTGGCTGGGTGGAGCGCTTCTGACCAGGTCCAGGAA 1505
    
```

50 53

GluLeuGlnSerSerGlnValThrGlnGluLeuTh
GAGCTGGAGAGCTCCCAAGTACACACAGAACTGACGtgagtgctcagcgcttcaccctccgacctgctgagtatccagatccaggggttcctctatctgggcaacctacactctcttg 1625
tttccttctccatgagtggtggggcagggtggctgaactctcaactctctgcttttcttgacctctgatactgggatgacagggcattcattatgttcttggctgaatggctt 1745
ttggcttttgagcaggatccaactcaacttaagctggcttcgaagggctgcaattctctctcagctctcaactctctgggaatacaagcgagtaccaccgcaacctcgctctgt 1865

54 (exon 4) 60

rValLeuMetGluAspThrMetThrGluValLysAlaIylrLys
ggttctctccaccctaattctgacctttttgtctctgcatctcctcttctgttctctctgggctcgagGGTACTGATGGAGGACACTATGACTGAAGTAAAGGCATACAAA 1983

70 80 90 100

IysGluLeuGluGluGlnLeuGlyProValAlaGluGluThrArgAlaArgLeuAlaLysGluValGlnAlaAlaGlnAlaArgLeuGlyAlaAspMetGluAspLeuArgAsnArgLeu
AAGGAGCTGGAGAACAGCTG66CCCAGTGG6CGAGGAGACAG66CCAGGCTGACTAAAGAGGTGCAGGCG6CACAGG6CCGCTCTGGGAGCTGACATGGAGGATCTACGCAACCGACTC 2103

110 120 130 140

GlyGlnIyrArgAsnGluValAsnThrMetLeuGlyGlnSerThrGluGluLeuArgSerArgLeuSerThrHisLeuArgLysMetArgLysArgLeuMetArgAspAlaAspAspLeu
GGGAGTACCGCAACGAGGTAAACACCATGCTGG6CCAGAGCACAGAGGAGCTGCGGTCGGGCTCTCCACACCTGCGCAAGATGCGCAAGCGCTGATGCG6GATGCGGATGATCTG 2223

150 160 170 180

GlnLysArgLeuAlaValIyrLysAlaGlyAlaGlnGluGlyAlaGluArgGlyValSerAlaIleArgGluArgLeuGlyProLeuValGluGlnGlyArgGlnArgThrAlaAsnLeu
CAGAAGCGCTTAGCGGTGTACAAG6CCGG6GACAGAGGG6CGCCGAGCGG6GTGAGTCTATCCGTGAGCGCTG66G6CCACTGGAGCAGG6TGTGAGCAG6GATGCGGATGATCTG 2343

190 200 210 220

ArgIrpArgArgProAlaProArgAspArgAlaGlnAlaLeuSerAspArgIleArgGlyArgLeuGluGluValGlyAsnGlnAlaArgAspArgLeuGluValArgGluGlnMet
CGCTGGCGCCGCCAGCCCGCGGATCGCGCCAGGCTTTGAGTGACCGCATCCGAGGG6GCTGGAGGAAGTGG6CAACAG6CCGAGCCGCTAGAGAGG6TGCCTGAGCAGATG 2463

230 240 250 260

GluGluValArgSerLysMetGluGluGlnThrGlnGlnIleArgLeuGlnAlaGluIlePheGlnAlaArgIleLysGlyIrpPheGluProLeuValGluAspMetGlnArgGlnIrp
GAGGAGCTCGCTCCAAGATGGAGGAGCAGACCCAGAGATACGCTTCAGG6CCGAGATCTTCCAG6CCCGCATCAAG6GCTGGTTCGAGCCGCTGATGGAGAGCATGACG6CCAGCTGG 2583

270 280 290 293

AlaAsnLeuMetGluLysIleGlnAlaSerValAlaThrAsnSerIleAlaSerThrThrValProLeuGluAsnGln***
GCAAACTAATGGAGAGATACAG6CCTCTGTGGTCCCACTCCATTGCTCCAGCACAGT6CCCTGGAGAAATCAATGATCATCCCTCACCTACG6CCTGCGGCAACATCCATGACCA 2703
GCTAGGTG6CCTGTCCCAAGCACCCTCTG6CCCTCTG6TGGCCCTTGTCTAATAAAGATTCTCCANGCAGTctctgagctctgtgagtgattccacacagctcagcctcagtttat 2823
poly(A) signal ↑↑↑

References

1. Mahley, R.W. (1982) in Medical clinics of North America ; Lipid Disorders (Havel, R.J., ed) Vol. 66, pp. 375-402, W.B. Saunders Co., Philadelphia, PA
2. Mclean, J.W., Fukazawa, C. and Taylor, J.M. (1983) J. Biol. chem. **258**, 8993-9000.
3. Matsumoto, A. and Fuakazawa, C. Submitted