Volume 14 Number 23 1986 Nucleic Acids Research

Complete nucleotide sequence of the gene encoding the rat apolipoprotein E

Chikafusa Fukazawa, Akiyo Matsumoto and John M.Taylor[1]

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

Submitted 4 November 1986

Apolipoprotein (apo) E is a M_r = 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).

```
cggggatggggagagttcacccgtggcagaggaatcaacaaagaggggccagggctaaaactcagttttcaccccagaagctgagcctcttaacagatagagccccccacctatttccat -295
ttaagcctccagcccttttccgtacaacgaatgcaccagaccccgcgggaaaggggagaagcagtactcagtgccccagcaccaaggcctggattattcaatgaggagtcagctcctttt -175
                                                                                                      CAT-box
tgggggggggggtcacaaggcatcaaaactccacctctttcctctgccctgctgtgaagggggggggagagcaacccgcctcgtgacaggggttggcccagcccgccctagccctgag  -55
                    CAT-box                                  ▼          (exon 1)
gagggggcgggcagggggggagtcctataattggacaggtctgggatccggtcccCTGCTCAGACCCCGGAGGCTAAGGAGTTGTTTCGGAAGGAGCTGgtaagacaagcttgggctggcg   66
                         TATA-box                      (+1) cap site
attcacccagggggcttggtaagacgtgggcaggggaccttgaaatccgctggagtccaggaaacaggcacaattattagaaaagcaggaagcccgatagaagacttagggtggcgggga  186
gacaaactaagatcgtgaggtaactagcctttgccggggtcagaacagatggaggagatggaggggtcatctccaggagtactgtcggtgcacggcatacgggctcaactagaaggtagt  306
atcaatctttagcgaaggacccgcgatgagccttaggttgcgtttatgtaaatacctactgatttccatcacagtccccaaaattaccctggccagtttcaaagcgaagaccctcatgcc  426
tgagcgtccgaagttctggggtcgcgggagggtaccacttcgcagggatggaggacgattaaaagcttacattagtactaaaccgagagcccggcttagcagggaaggtagaaaggaatg  546
taaattggaccaccggctgtccctctgcgttgtggaatttgaaccctaggactgggagctggaatttttggcagcgggtccaccccggggtgctgagatagagataggaggggggaggtaa  666
atagacctttgggcagatgctctattgtggagatgtttgtgatgactcacagacctgagaagggaattaaggtgggggctctttgggtggcccgagtttctccacccctacccactggtg  786
                                                       -18                      -10              -4
                              (exon 2)                 MetLysAlaLeuTrpAlaLeuLeuLeuValProLeuLeuThrG
ttcaaagacagcttttccttccgcagACTGGCCAATCACAACTGGGAAGATGAAGGCTCTGTGGGCCCTGCTGTTGGTCCCATTGCTGACAGgtatggggcaaggggtggctttccctgc  906
tggcaggttggggttggggcagttctgagacctctgagattcaatagccctgagcagctgttttacatgcaggccttgtctttccctatccataatatgagaggattgatgctcagaggg 1026
taaaggtgcttgtcataaagctcgaagacttatctcggagcccaacatggtggaaggagagagcggattctctctagttgtcctctgacctccaagggtacacacacagacacacaatta 1146
aaaagttgaagtgtgtaatttcttagtcgcttttaaaggggaggtatacattttctaaccccataccagcctagtctcagctctgttctacataggatactagacagtgtatgcttgtag 1266
                                                                                                                -3     1
                                                                                                          lyCysLeuAla Glu
tatctaaacagactccacgactgacttccagacgcactgttgcttcaaagagacccaagagcactgtaggtcctgacccagccttaaacttactactctacacagGATGCCTGGCC GAG 1385
       (exon 3)           10                             20                             30                             40
GlyGluLeuGluValThrAspGlnLeuProGlyGlnSerAspGlnProTrpGluGlnAlaLeuAsnArgPheTrpAspTyrLeuArgTrpValGlnThrLeuSerAspGlnValGlnGlu
GGAGAGCTGGAGGTGACAGATCAGCTCCCAGGGCAAAGCGACCAACCCTGGGAGCAGGCCCTGAACCGCTTCTGGGATTACCTGCGCTGGGTGCAGACGCTTTCTGACCAGGTCCAGGAA 1505
```

© IRL Press Limited, Oxford, England.

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