Mouse Apolipoprotein A-IV Gene: Nucleotide Sequence and Induction by a High-Lipid Diet

SIMON C. WILLIAMS,[1] SALLY M. BRUCKHEIMER,[1] ALDONS J. LUSIS,[2] RENEE C. LeBOEUF,[3]
AND ALAN J. KINNIBURGH[1]*

*Department of Human Genetics, Roswell Park Memorial Institute, Buffalo, New York 14263[1]; Department of Medicine
and Microbiology and the Molecular Biology Institute, University of California, Los Angeles, California 90024[2]; and
Research Service, Veterans Administration, Wadsworth Medical Center, Los Angeles, California 90073[3]*

Apolipoprotein A-IV (apo A-IV) functions in conjunction with other apolipoproteins to form lipoprotein particles which are involved in lipid homeostasis. In this report we present the nucleotide sequence of the mouse apo A-IV gene and demonstrate its induction in the liver by chronically high dietary lipid. The apo A-IV gene consists of three exons and two introns. The introns separate evolutionarily conserved and functional polypeptide domains. Intron 1 divides most of the apo A-IV signal peptide from the amino terminus of the mature plasma protein. The second intron separates a highly evolutionarily conserved, variant amphipathic peptide repeat from the remainder of the mature apo A-IV protein. The 5' flanking region has several interesting features. The apo A-IV gene has variant TATA and CAT box sequences, TTTAAA and CCAACG, respectively. There are five G-rich direct repeats of 10 nucleotides and a short inverted repeat in the 5' flanking region. We speculate that these sequence elements in the 5' flanking region may be involved in the regulation of apo A-IV gene expression. We also show that chronically high dietary lipid induces liver apo A-IV levels 10-fold in C57BL/6 mice, a strain susceptible to atherosclerotic lesions, while we observed no induction in nonsusceptible BALB/c and C3H mice.

The apolipoproteins are polypeptide carriers of cholesterol, triglycerides, and phospholipids in the circulation. They are involved in both exogenous (dietary) and endogenous (de novo synthesized) lipid transport. Eight major apolipoproteins are associated with the various circulating lipoproteins. Seven of these (A-I, A-II, A-IV, C-I, C-II, C-III, and E) appear to be members of a dispersed gene family. Apolipoprotein A-IV (apo A-IV) is a component of three lipoprotein particles; chylomicrons, which are involved in intestinal adsorption and transport of lipids; very low density lipoproteins, which are secreted by the liver when serum cholesterol or lipid is insufficient for cellular needs; and high-density lipoproteins, which bind cholesterylesters in the serum and transfer these to low-density lipoprotein particles for uptake and catabolism by the liver (for reviews, see Brown et al. [6] and Mahley et al. [16]). The levels of various lipoproteins in the circulation are strongly correlated with risk of coronary heart disease (for an example, see Castelli et al. [9]). The precise functions of apo A-IV are not known, although it can activate lecithin cholesterol acyltransferase.

The apo A-IV protein is composed of 11- or 22-amino-acid repeats that have the potential to form amphipathic α-helices. These repeats comprise approximately 85% of the apo A-IV polypeptide (2; this report) and are thought to be responsible for lipid binding (24). This repeating motif is believed to have arisen by intragenic duplication and is found in other apolipoprotein genes (for a review, see Breslow [5]). The remainder of the apo polypeptide consists of a short 11-amino-acid segment at the NH$_2$ terminus and a 66-amino-acid peptide region at the COOH terminus. These peptide regions most likely contain the domains necessary for lipoprotein particle formation and interaction (2; this report).

We have previously isolated a mouse apo A-IV cDNA clone on the basis of its induction in porphyric liver (7). Here we report the nucleotide sequence of the genomic apo A-IV sequence and the deduced amino acid sequence. We also demonstrate that a high-lipid diet induces liver apo A-IV mRNA levels in inbred mice susceptible to diet-induced atherosclerosis. In contrast, two inbred strains that are not susceptible to atherosclerosis show no liver apo A-IV mRNA induction when fed this high-lipid diet.

MATERIALS AND METHODS

Enzymes and isotopes. Restriction endonucleases and modification enzymes were obtained from New England Biolabs, Boehringer Mannheim, Bethesda Research Laboratories, and Pharmacia Fine Chemicals. Enzymes were used under the conditions specified by the manufacturer. Radiolabeled nucleotides were obtained from Amersham Corp. and New England Nuclear Corp.

cDNA and genomic cloning. Full-length cDNA clones were isolated by screening a porphyric mouse cDNA liver library (a gift of Peter Curtis, Wistar Institute, Philadelphia, Pa.) with a 456 base pair (bp) cDNA probe homologous to the 3' end of the mouse apo A-IV gene (7). The probe was labeled with [^{32}P]dATP by the random primer method (14).

Genomic clones were isolated by screening a λL47.1 library (1) made by partial digestion of SWR/J mouse genomic DNA with MboI (a gift of Steven Weaver, University of Illinois, Chicago) with the 3' cDNA probe described above.

DNA sequencing and computer analysis. Both genomic and cDNA clones were sequenced by the chemical degradation method of Maxam and Gilbert (17) or the dideoxy chain termination method of Sanger et al. (23). Genomic PstI fragments of 1.80, 0.85, 0.65, 0.50, 0.45, and 0.05 kilobases (kb) were subcloned into pUC13, and both strands of each subclone were sequenced. The sequences across the junc-

```
      -440      -430      -420      -410      -400      -390      -380      -370      -360
gcagccctatgccatctctgggctcggtgccatctctgtagctgatgttcgacaaagttcaggttggtggcagctgtcagactggtggct

      -350      -340      -330      -320      -310      -300      -290      -280      -270
gtctcactggggtggaaagaggagacttggaccttgttctctcagactggcacagacccagggctgccaaccgggcctctggggtccagt
        1                                                                          2

      -260      -250      -240      -230      -220      -210      -200      -190      -180
tctgttcagggactcccctagactcctaggctcattcctcctgaagtttctggctatccttccggcctcttggacagggtggagccaac

      -170      -160      -150      -140      -130      -120      -110      -100       -90
tcaagaagactgcttccctctgctgcctgtgtgctgtcagcttccacgttgtcttggggccactaaggtccaggggggcctcctgggggtg
                                                    3              4              5

       -80       -70       -60       -50       -40       -30       -20       -10
tgtcaccttccaacgtggagtcacactgggaaggaggcggggaggggagggctggaggggctttaaatgggtggctggccttgcctgcagtc

        10        20        30        40        50        60        70        80        90
AATCTGCACAGGGACACAGGTACACCGTTTCTTCTGACTCCGGGAAACATCCAGTGTAGCCGAAACTGTCCCAGCCCAGTGAGGAGCCAG

       100       110       120       130
GATGTTCCTGAAGGCTGCGGTGCTGACCCTGGCCTTGGTGGCCACCGgtgagtagacactgcacttgggaggcagcaagaaaagccagct
MetPheLeuLysAlaAlaValLeuThrLeuAlaLeuValAlaThrG
- - - - - - - - - - - - - - SIGNAL PEPTIDE - -

ctagaactggcggacaaccggggtggcctgtatttgcccagcagctcataggagaacaggcctttgttctccctggcacttgattggcct

gggttatcccagggatggggcaatggtttgggttatccaaactccaacattatcagctcagagctgaggcagaggggccaagagagagat

                                      140       150       160       170       180       190
gatcctcataaagttgcctagaactctctctccccagGCACCCGGGCTGAGGTCACTTCGGACCAAGTGGCCAATGTGGTGTGGGATTAC
                                      lyThrArgAlaGluValThrSerAspGlnValAlaAsnValValTrpAspTyr
                                      - - - - - -MATURE PLASMA PROTEIN                    *

       200       210       220       230       240       250       260
TTTACCCAGCTAAGCAACAATGCCAAGGAGGCTGTAGAACAGTTTCAGAAGACGGATGTCACTCAGCAGCTCAGgtaagtgcgatcagtc
PheThrGlnLeuSerAsnAsnAlaLysGluAlaValGluGlnPheGlnLysThrAspValThrGlnGlnLeuSe

tacaaggcgaggcttgaagtccatagctgacctcagaggtgggacactggctcctggagttcttctgttctcactaaaggagtcttgcct

ccctggaacctagattgtccctgtaaacaggaaagctggaaccagggacccatccctatagttccttcctatctgcatgtgcagagttgg

taatagccaaatcccatcacaaatcactagcaggaagaggaaagagacgtgtcctgtaagcgaggtgcctggcaacaggaaggaggttgg

tatttggggacccactgagcacatgcaaggacatcagaccctgtgccactggaagacatgtgttgcacaatgtgaccctaggagggggagt

gatgccaaagttcagccccattctttatccacaggtaaggaaagcaaacttctcaagtcacatggtgggcagataaggatgcagaggctc

agcagaagccttgcagataaaactccaaagtcaaactaatcctaggagatttcctgaaagcatgacctaccccagggggagggtcaaagga

caagacggaggtctctgttccatggacacactcctgacctaagcagggtatagagctgagtgttctacaagcgtatctaatggtgcttcc

                        270       280       290       300       310       320       330
ttgtctccatcctttccctgaagTACCCTCTTCAAGGACAAACTTGGGGATGCTAGTACGTATGCTGATGGGGTGCACAACAAGCTGGTG
                        rThrLeuPheLysAspLysLeuGlyAspAlaSerThrTyrAlaAspGlyValHisAsnLysLeuVal
                        *
```

FIG. 1. Nucleotide sequence of the mouse apo A-IV gene. The sequence extends from 450 bases upstream of the transcriptional start site to 200 bases downstream of the poly(A) addition site. The transcriptional start site is designated nucleotide 1. The exons (capital letters) are numbered; the introns (lowercase letters) are not. The amino acid sequence is shown below the nucleotide sequence, and the signal peptide, mature apo A-IV protein, and extent of the amphipathic repeats are labeled. An asterisk marks the amino acid at the beginning of each amphipathic repeat; 8 of 14 repeats begin with a proline residue. The TATA and CAT box homologies and the polyadenylation signal are boxed. The tandem repeats in the 5' flanking sequences are denoted by dashed arrows and numbered as in Fig. 6. The inverted repeat is shown by solid arrows.

```
       340       350       360       370       380       390       400       410       420
CCCTTTGTCGTACAGCTGAGTGGGCATCTAGCCAAGGAAACTGAGAGGGTGAAGGAAGAGATCAAGAAGGAGCTGGAGGACCTACGTGAC
ProPheValValGlnLeuSerGlyHisLeuAlaLysGluThrGluArgValLysGluGluIleLysLysGluLeuGluAspLeuArgAsp
  *                                                                    *

       430     · 440       450       460       470       480       490       500       510
CGCATGATGCCCCATGCCAACAAAGTAACCCAGACGTTCGGGGAGAACATGCAGAAGTTGCAGGAGCACCTGAAGCCCTATGCCGTGGAC
ArgMetMetProHisAlaAsnLysValThrGlnThrPheGlyGluAsnMetGlnLysLeuGlnGluHisLeuLysProTyrAlaValAsp
              *                                                                  *

       520       530       540       550       560       570       580       590       600
CTGCAAGATCAGATCAACACACAGACCCAGGAAATGAAGCTCCAGCTGACCCCATACATCCAGCGCATGCAGACCACGATCAAGGAGAAT
LeuGlnAspGlnIleAsnThrGlnThrGlnGluMetLysLeuGlnLeuThrProTyrIleGlnArgMetGlnThrThrIleLysGluAsn
                                                            *

       610       620       630       640       650       660       670       680       690
GTGGACAACCTGCACACCTCGATGATGCCCCTTGCCACCAACTTAAAGGACAAGTTTAACAGGAATATGGAAGAGCTCAAGGGGCACCTA
ValAspAsnLeuHisThrSerMetMetProLeuAlaThrAsnLeuLysAspLysPheAsnArgAsnMetGluGluLeuLysGlyHisLeu
                              *

       700       710       720       730       740       750       760       770       780
ACCCCCCGTGCCAACAGGCTGAAGGCTACGATCGACCAGAACCTGGAGGATCTGCGCCGCAGCCTGGCCCCTCTGACGGTGGGCGTGCAG
ThrProArgAlaAsnArgLeuLysAlaThrIleAspGlnAsnLeuGluAspLeuArgArgSerLeuAlaProLeuThrValGlyValGln
  *                                                                       *

       790       800       810       820       830       840       850       860       870
GAGAAACTCAACCATCAGATGGAGGGCCTGGCCTTCCAGATGAAGAAGAACGCGGAGGAGCTCCAGACCAAGGTCTCCGCAAAAATCGAC
GluLysLeuAsnHisGlnMetGluGlyLeuAlaPheGlnMetLysLysAsnAlaGluGluLeuGlnThrLysValSerAlaLysIleAsp
                                          *

       880       890       900       910       920       930       940       950       960
CAGCTGCAGAAGAATCTGGCCCCGCTGGTGGAAGACGTGCAGAGCAAGGTGAAGGGCAACACGGAAGGGCTGCAGAAGGCTCTGGAAGAC
GlnLeuGlnLysAsnLeuAlaProLeuValGluAspValGlnSerLysValLysGlyAsnThrGluGlyLeuGlnLysAlaLeuGluAsp
                      *                                                    *

       970       980       990      1000      1010      1020      1030      1040      1050
CTGAACAAGGCGCTGGAGCAGCAGGTGGAGGAGTTCCGACGCACTGTGGAGCCCATGGGAGAGATGTTCGGCGGGGCTCTGGTGCAGCAG
LeuAsnLysAlaLeuGluGlnGlnValGluGluPheArgArgThrValGluProMetGlyGluMetPheGlyGlyAlaLeuValGlnGln
                                                            *

      1060      1070      1080      1090      1100      1110      1120      1130      1140
CTGGAACAGTTCAGACAGCAGCTGGGTCCCAATTCGGGGGAGGTGGAAAGCCACTTGAGCTTCCTGGAGAAGAGCCTGAGGGAGAAGGTC
LeuGluGlnPheArgGlnGlnLeuGlyProAsnSerGlyGluValGluSerHisLeuSerPheLeuGluLysSerLeuArgGluLysVal

      1150      1160      1170      1180      1190      1200      1210      1220      1230
AACTCCTTTATGAGCACCCTGGAAAAAAAGGGGAGCCCAGACCAGCCTCAAGCCCTCCCCCTCCGGAGCAGCCCCAGGAGCAGGCTCAG
AsnSerPheMetSerThrLeuGluLysLysGlySerProAspGlnProGlnAlaLeuProLeuProGluGlnAlaGlnGluGlnAlaGln

      1240      1250      1260      1270      1280      1290      1300      1310      1320
GAGCAGGCTCAGGAGCAGGTGCAGCCCAAACCTCTGGAGAGCTGAGCCCTCAGCCCATCACATGCCCTCAGCCCATCACAGCAGCAGACA
GluGlnAlaGlnGluGlnValGlnProLysProLeuGluSer

      1330      1340      1350      1360      1370      1380      1390      1400      1410
CCTGTCCTGCCCCACCACCTGTCTGTCACTCTGTCCCCAGGCACTTCTTGTACCAGCTTGGGACACATGTCCTGTGGGAGGTGAAGCCAC

      1420      1430      1440
ATCTCGCTACTCAATAAAGCAACTGAGAAATTAGCCCatcgggttgccctgtgattccttgggggctggcctgacgtagggaggaaatca
           POLY A SIGNAL

aggcatctggctggggacatgggggtgaggggtggggccggtaatgagccttcgggtgggtcgggggttggtgtgctagagaggagaatg

cagaaaagacgccagtctatcagaaaacaaaacacagatcccgttcacagcctgcag
```

A.

B.

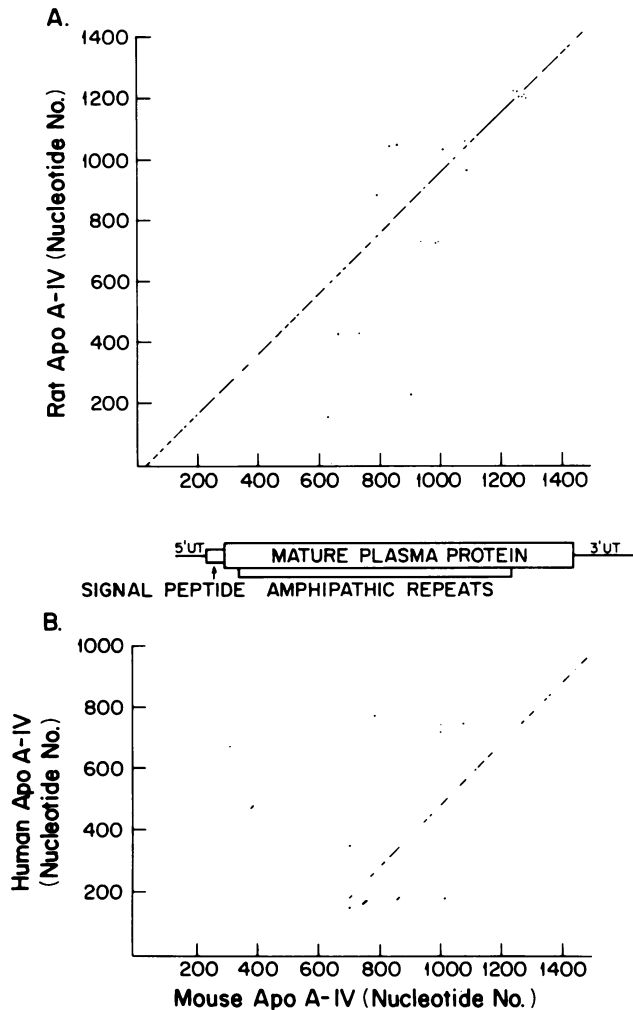Mouse Apo A-IV (Nucleotide No.)

FIG. 2. Dot matrix comparisons of mouse versus rat and mouse versus human apo A-IV exon nucleotide sequences. (A) Comparison of mouse and rat apo A-IV cDNA sequences. (B) Comparison of mouse and human apo A-IV cDNA sequences. The human sequence is missing approximately 480 nucleotides at the 5' end. The sequences were compared with a window of six nucleotides, moving one nucleotide at a time. A match of five out six nucleotides was sufficient for inclusion in the graph. The rat exon sequences are from Boguski et al. (2), and the partial human exon sequences are from Karathanasis (15).

tions of the fragments were determined either from the cDNA clones or by using unique restriction sites in the genomic clone. The sequence shown here does not include the sequence of the 0.85- and 0.50-kb fragments, as they are

FIG. 3. Comparison of mouse (M), rat (R), and human (H) apo A-IV amino acid sequences. For clarity the sequence is divided into signal peptide, mature protein amino terminus, amphipathic repeats, and carboxy terminus. Positions of identity are marked by a dash in the rat and human sequences, amino acid substitutions are indicated, and gaps in the sequences indicate that an amino acid has been inserted or deleted in one sequence. The mouse versus rat and mouse versus human percent homologies are listed at the right of each peptide region. The numbers above the sequence indicate the amino acid residues; those at the left number the amphipathic repeats. The rat and the partial human amino acid sequence are from Boguski et al. (2) and Karathanasis (15), respectively.

Amino Acid No.		-19 Signal Peptide	% Homology
	M	MFLKAAVLTLALVA TGTRA	
	R	----V---V----I---Q-	
		1 Mature Protein	
	M	EVTSDQVANVVW	
	R	----------M-	84%
Amphipathic			
Repeat No.		13	
1	M	DYFTQLSNNAKEAVEQFQKTDVTQQLS	
	R	-----------------L--------N	93%
		40	
2	M	TLFKDKLGDASTYADGVHNKLV	
	R	---Q----NIN----DLQ---	68%
		62	
3	M	PFVVQLSGHLAKETERVKEEIK	
	R	--A-------T-------R---Q	82%
		83	
4	M	KELEDLRDRMM	
	R	---------AN--	82%
	H		
		94	
5	M	PHANKVTQIFGENMQKLQEMLK	
	R	------S-M--D-V------R	77%
	H	-E	
		116	
6	M	PYAVDLQDQINTQTQEMKLQLT	
	R	---T---A---A---D--R---	77%
	H	---DQ-RT-V-T-ABQLRR--T	50%
		138	
7	M	PYIQRMQTTIKENVDNLHTSMM	
	R	----------QD--E--QS--V	73%
	H	-YA---ERVLRE-ADS-QA-LR	41%
		160	
8	M	PLATNLKDKFNRNMEELKGHLI'	
	R	-F-NE--E--Q--G--Q--	68%
	H	-H-DE--A-IDQ-V-E---R--	59%
		182	
9	M	PRANRLKATIDQNLEDLRRSLA	
	R	----E-------------SR--	86%
	H	-Y-DEF-VK---TV-E---RS--	59%
		204	
10	M	PLTVGVQEKLNHQMEGLAFQMK	
	R	--AE---------------	91%
	H	-YAQDT-------L---T----	68%
		226	
11	M	KNAEELQIKVSAKIDGLQKNLA	
	R	------H----TN---------	86%
	H	------KARI-ASAEE-RQR--	50%
		248	
12	M	PLVEDVQSKVKGNTEGLQ	
	R	---------L---------	96%
	H	--A---RGNLR-------	67%
		266	
13	M	KALEDLNKALEQQVEEFRRTVE	
	R	-S-------Q-D----V---A--	77%
	H	-S-AE-GGH-D----E---R--	64%
		288	
14	M	PMGEMFGGALVQQLEQFRQQLG	
	R	-L-DK-NM-----M-K-----	68%
	H	-Y-EN-NK-----M-QL-TK--	64%
		310 Carboxy Terminal	
	M	PNSGEVESHLSFLEKSLREKVN	
	R	SD--D----------N----S	
	H	PHA-D--G------D--D--N	
		332	
	M	SFMSTLEKKGSPDQPQALPLP	
	R	--------Q--------L----	91%
	H	--F--FKE-E-Q-KTLS--EL	61%
		353	
	M	EQAQEQAQEQAQEQVQPK PLES	
	R	-----V---V------- ----	
	H	--Q--H---Q-----MLA----	
		---4 Amino Acid-Repeats	

```
Rat        P L A E D L Q T K L N Q N V E D L R K Q L V      73%
                       •             • •   x •     x

Mouse      P L A E D L Q G K L N Q N M E E L Q E Q L K
                                   E
           •   •   x       x   • •       x • x

Human      P Y A D E L K X K L D Q Q A E E L R R R L X      59%
               E                 V
```

FIG. 4. Alignment of the 22-amino-acid consensus repeat sequences from mouse, rat, and human apo A-IV. Numbers on the right indicate the percent homology between mouse and rat and mouse and human sequences. Note that for the mouse one position, and for the human four positions could not be unambiguously determined. For amino acid substitutions, conservative substitutions are marked with an asterisk and nonconservative substitutions are marked with an ×.

the extreme 5' and 3' fragments. The sequence of the cDNA was determined from a series of overlapping clones in pUC8. Samples were run on 5 or 6% acrylamide–7 M urea gels (0.6 mm by 120 cm by 35 cm). Analysis of DNA sequences was performed on either an Apple IIe or IBM PC microcomputer with the University of Minnesota and Cornell DNA programs and the Bionet system of programs.

Dietary studies. Inbred mice (BALB/cJ, C3H/HeJ, and C57BL/6J) were purchased from Jackson Laboratories, Bar Harbor, Maine. Female mice between the ages of 8 and 12 weeks were maintained for 15 weeks on either a standard diet of mouse chow (Purina 5001) or a high-fat atherogenic diet containing 1.25% cholesterol and 15% fat as previously described (20). Total liver RNA was extracted by a guanidine hydrochloride procedure (10) and quantitated by absorbance at 260 nm.

RNA electrophoresis and blotting. Ten micrograms of total cell RNA was glyoxalated and electrophoresed through a 1.2% agarose gel (18). RNA from the gel was blotted onto a nylon mesh, Zetabind (Cuno Inc.). The RNA was cross-linked to the Zetabind by exposure to 254-nm UV light for 3 to 5 min. DNA was labeled by the random primer method to a specific activity of 1×10^9 to 2×10^9 cpm/μg, and 2×10^6 cpm/ml of hybridization buffer was used for hybridization. Blots were hybridized and washed by the method of Church and Gilbert (12) and exposed to XAR X-ray film (Kodak). Autoradiograms were quantitated by densitometric scanning. Dilutions of various RNA samples demonstrated that the hybridization signal was proportional to the amount of RNA present on the blot.

RESULTS

Sequence of the mouse apo A-IV gene. We isolated a full-length cDNA clone of the mouse apo A-IV gene from a liver cDNA library made from porphyric BALB/cJ mice. A number of independent and overlapping clones were isolated. In addition, an SWR/J mouse genomic λL47.1 library was screened and a complete apo A-IV gene was isolated. The entire apo A-IV gene was subcloned as a 4.3-kb HindIII fragment, and from this clone various subclones were prepared. The apo A-IV cDNA and genomic clones were sequenced, and the amino acid sequence was deduced (Fig. 1). The apo A-IV gene contained two introns; the first intron interrupted the Gly codon at amino acid −4, and the second intron interrupted the Ser codon at amino acid 39. Exon 1 was 137 nucleotides long; it contained a 91-nucleotide 5' untranslated region and encoded 15 residues of the 19-amino-acid hydrophobic signal peptide. Intron 1 was 260 nucleotides long. Exon 2 was 127 nucleotides long and encoded the terminal four amino acids of the signal peptide

and the first 39 amino acids of the mature protein. The second intron was 669 nucleotides long. The third exon was 1,183 nucleotides long and encoded the majority (336 of 394) of the apo A-IV amino acids. The remainder of exon 3 consisted of a 175-nucleotide 3' untranslated region (Fig. 1). The nucleotide where transcriptional initiation occurs (+1) was determined by S1 nuclease mapping (data not shown) and by comparison with a rat apo A-IV sequence (M. Boguski and J. Gordon, personal communication).

Putative promoter nucleotide sequence homologies. There are several notable features in the apo A-IV 5' flanking sequence. First, the canonical TATA box and CAT box sequences were variant. The TATA homology was TT TAAA and the CCAATT homology was an abbreviated CCAA (Fig. 1). The rat genomic DNA sequence also contains these variant sequences (M. Boguski and J. Gordon, personal communication). Other features of the 5' flanking region were as follows. (i) A G-rich 10- or 11-bp direct repeat was present five times. These direct repeats were centered at positions −99, −109, −118, −280, and −353 (Fig. 1). (ii) An inverted repeat of 5 bp containing a four-nucleotide unpaired loop was centered at position −249 (Fig. 1).

Comparison of mouse, rat, and human apo A-IV exon sequences and amino acid sequences. The coding region of the mouse apo A-IV gene was 1,182 nucleotides long and encoded a protein of 394 amino acids. The 5' and 3' untranslated regions were 91 and 175 bases long, respectively. The mouse exons showed 89% homology to the rat nucleotide sequence and 81.5% homology at the amino acid level. Compared with an incomplete human sequence, the mouse apo A-IV exons were 73% homologous to the nucleotide sequence and 58.6% homologous to the amino acid sequence. Regions of greater than 80% nucleotide sequence homology were found scattered between the exons of mouse and rat or mouse and human apo A-IVs. Dot matrix comparisons clearly showed the patchwork homology of the rat and human apo A-IV exon sequences with the mouse sequences (Fig. 2). In the mouse and rat apo A-IV genes, these regions of nonhomology were clustered point mutations or small deletions or insertions and did not correlate with peptide domains such as the amphipathic repeats.

The mouse and rat amino acid sequences showed that the mouse protein was three amino acids longer. This was due to the following differences between the two sequences. (i) Three nucleotides were deleted in the mouse sequence relative to the rat sequence between nucleotides 133 and 134 (Fig. 1). This resulted in the loss of an Ile residue at this position. (ii) In the mouse sequence, the 48 nucleotides beginning at position 1208 and ending at 1255 encoded four repeats of the sequence Glu-Gln-X-Gln, where X is either Ala or Val. In the rat, one of these repeats was deleted by the

removal of the 12 nucleotides corresponding to positions 1208 to 1219 in the mouse. Neither of these changes resulted in a change of reading frame.

The mouse and human sequences show a one-to-one correspondence of amino acids from amino acids 114 to 371 in the mouse. There was one additional amino acid in the human protein after amino acid 371 (this assumes that the second stop codon identified by Karathanasis [15] is correct, as amino acid homologies among the three species appear to indicate). Thus, both the mouse and the human proteins contain four of the Glu-Gln-X-Gln repeats. However, in the human protein, X is either Gln or Val.

Analysis of amino acid repeats. All of the apolipoproteins, with the exception of apo B, have now been sequenced. Apo A-I, apo A-IV, and apo E are made up of repeating peptides of 22 amino acids (with some variation), often beginning with a proline residue (3, 21). The mouse apo A-IV has a similar pattern of 22 amino acid repeats (Fig. 3). Mouse apo A-IV showed a high degree of amino acid homology to a rat (2) and a partial human sequence (15). That these 22-amino-acid repeats probably arose initially from the duplication of an ancestral 11-amino-acid repeat is shown by the occurrence of one repeat of 11 amino acids as the fourth repeat and the similarities between the first and second group of 11 amino acids in each repeat (Fig. 3).

This similarity is seen more clearly in a consensus derived by taking the most frequently occurring amino acid at each position (Fig. 4). Although some positions cannot be decided unambiguously from the amino acid sequence, these positions can be decided by taking the most frequently used base at each position in that codon (Fig. 4). A comparison of the mouse, rat, and human consensus amphipathic amino acid repeats showed that the mouse consensus peptide was 73% homologous to the rat and 59% homologous to the human peptide. These values agree well with the homologies seen for the complete proteins. However, a close examination of the positions where amino acids were changed revealed that four of six changes for the mouse-rat comparison, five of nine changes for the mouse-human comparison, and six of nine changes for the rat-human comparison were conservative substitutions (this does not include two positions in the human consensus to which no residue was assigned) (Fig. 4). Hydropathy curves for each consensus sequence (11) showed an almost identical pattern of alternating hydrophilic and hydrophobic amino acids (data not shown). It therefore appears that apo A-IV proteins have a highly conserved secondary structure that is not evident in comparisons of primary structure.

Effect of dietary fat on apo A-IV expression. We investigated the effect of dietary cholesterol and fats on the expression of the apo A-IV gene. Mice were fed a high-lipid diet for 15 weeks. Liver RNAs were prepared from these mice and from normal controls. Nothern blotting of these RNAs was performed, and the blot was probed with our full-length mouse apo A-IV cDNA. Comparison of the basal (uninduced) level of apo A-IV mRNA revealed a strain difference (Fig. 5). C3H and BALB/c mice had a three to four times higher basal level of apo A-IV mRNA than C57BL/6 mice. After a 15-week high-fat diet, the C3H and BALB/c apo A-IV mRNA levels were unchanged, but the C57BL/6 apo A-IV mRNA level was 11 times higher than the basal level. Thus, there were two separate differences between C57BL/6 mice and C3H and BALB/c mice. First, the basal level of apo A-IV mRNA is lower by three- to fourfold in C57BL/6 than in the other strains. Second, in response to high levels of dietary lipid and fat, there was a larger
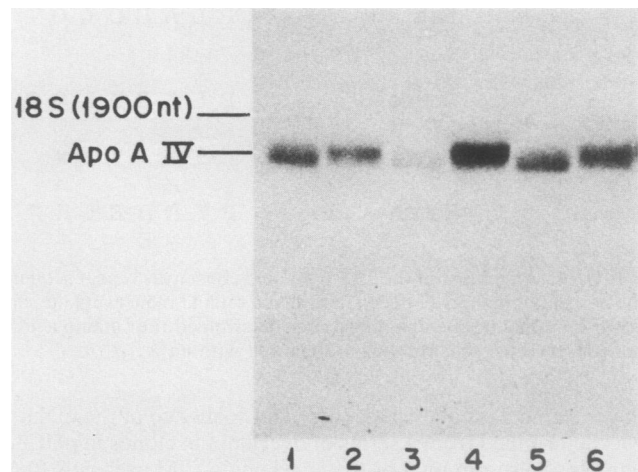
FIG. 5. Liver RNA levels of apo A-IV in mice fed normal and high-fat diets as determined by Northern blotting. Liver RNAs are from (lane 1) BALB/c mouse on a normal diet, (lane 2) BALB/c mouse on a high-fat diet, (lane 3) C57BL/6 mouse on a normal diet, (lane 4) C57BL/6 mouse on a high-fat diet, (lane 5) C3H mouse on a normal diet, and (lane 6) C3H mouse on a high-fat diet. The faint band above the apo A-IV mRNA is due to nonspecific hybridization with 18S mRNA. nt, Nucleotides.

induction of apo A-IV mRNA in C57BL/6 than in BALB/c and C3H mice. Many mice have been examined with similar results. Preliminary results from recombinant inbred mouse lines indicate that these two factors may be genetically separable (unpublished data).

DISCUSSION

We have sequenced the mouse apo A-IV gene and found several interesting features. First, the 5' flanking region of the apo A-IV gene contains five G-rich direct repeats of 10 or 11 nucleotides with the following consensus sequence: TC CTGGGG$^{TG}_{CC}$ (Fig. 6). Short direct repeats such as these have been shown to be important in the induction of other eucaryotic genes, particularly those for which small effector molecules increase transcription (8, 25, 26). Liver apo A-IV gene expression has been shown to be regulated by porphyria (7), glucocorticoids, insulin (13), and, in certain inbred mice, dietary lipid (this report). These G-rich direct repeats may be involved in apo A-IV gene regulation by all or some of these inducers. The apo A-IV 5' flanking region also contains a TATA box-like sequence, TTTAAA, and an abbreviated CAT box element, CCAACG. These variant canonical sequences are conserved between mouse and rat apo A-IV (M. Boguski and J. Gordon, personal communication). In the future, studies by in vitro mutagenesis of the apo A-IV gene coupled with introduction of these altered genes into hepatoma cells in culture will hopefully allow the determination of which 5' flanking nucleotide sequence elements are necessary for apo A-IV gene regulation.

The apo A-IV sequence also revealed that the three apo A-IV exons roughly separate functional protein domains. Intron I divides most of the apo A-IV leader peptide (amino acids −19 to −5) from the remaining four amino acids of the leader peptide (amino acids −4 to −1) and the amino-terminal region of the mature protein (Fig. 1 and 3). The second intron separates a highly conserved, variant amphipathic peptide repeat (27 amino acids versus the canonical 11 or 22 amino acids) from the remainder of the

Nucleotide Sequence Repeat No.		% Homology with Consensus Sequence
1	CACTGGGGTG	80
2	CTCTGGGGTC	80
3	TCTTGGGGCC	90
4	TCCAGGGGCC (G)	80*
5	TCCTGGGGTG (G)	90*

```
Consensus                        T G
Sequence            T C C T G G G G + +
                                   C C

  %
Agreement           <60><80><   100    >
```

FIG. 6. Comparison of G-rich direct repeats in the apo A-IV 5' flanking nucleotide sequence (see Fig. 1). The asterisk indicates that 10% homology has been subtracted for the deletion of one base.

mature apo A-IV protein. This variant 27-amino-acid repeat may therefore have a distinct evolutionary origin or a distinct physiological function.

Several interesting features of the apo A-IV protein were revealed by a comparison of the mouse apo A-IV amino acid sequence and the rat and a partial human apo A-IV sequences (Fig. 3). First, amino acid homologies were clustered. Within the repeating amphipathic domains, the repeats numbered 10 and 12 were 91 and 96% homologous between mouse and rat, respectively, and 68 and 67% homologous between mouse and human, respectively. Repeats 9 and 11 were both 86% homologous between mouse and rat, but only 59 and 50% homologous between mouse and human, respectively. Other highly homologous regions were repeats 1 and 4, which are two variant repeats of 27 and 11 amino acids, respectively. Repeat number one is 93% homologous between rat and mouse and repeat four is 82% homologous between these two species. Perhaps these regions have specific functions in lipid binding or lecithin cholesterol acyltransferase activation and therefore are highly evolutionarily conserved. The second apo A-IV polypeptide region of interest is the carboxy terminus. There was a hydrophilic repeat of Glu-Gln-Ala-Gln in mouse and rat between amino acids 353 and 368. In the human sequence the repeat is Glu-Gln-Gln-Gln. The terminal repeat is the variant Glu-Gln-Val-Gln, but it is conserved among mouse, rat, and human (Fig. 3). Mouse and human have four repeats and rat has three. This region is the most highly conserved of the 66 amino acids between the 14 amphipathic repeats and the carboxy terminus. This terminal region may therefore be involved in lipoprotein particle formation or lipoprotein particle receptor recognition.

Finally, we have demonstrated that chronically high dietary lipid can induce liver apo A-IV mRNA levels in certain inbred mouse strains. C57BL/6 mice showed a greater than 10-fold induction of liver apo A-IV mRNA, whereas BALB/c and C3H mice did not show induction. Moreover, a polymorphism in basal levels of apo A-IV mRNA was observed,

with C57BL/6 mice exhibiting a three- to fourfold-lower level of apo A-IV mRNA than the other strains examined when maintained on a normal chow diet. It is noteworthy that these strains also differed in susceptibility to atherosclerosis when maintained on the same high-lipid diet. Strain C57BL/6 develops numerous large lesions in the aorta and cardiac arteries, while strains C3H and BALB/c are almost completely resistant (19, 20). The strains also exhibit quantitative variations of circulating lipoproteins when fed a high-lipid diet (4; K. L. Reue, Ph. D. dissertation, University of California, Los Angeles, 1985; A. J. Lusis and R. C. LeBoeuf, Methods Enzymol., in press). It will be of interest to determine the role of apo A-IV expression in such genetic variations. Since apo A-IV mRNA is induced by several physiologically distinct mechanisms, whereas other related proteins such as apo A-I are not, apo A-IV may play a crucial role in lipid homeostasis.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Benton, W. E., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to plaques in situ. Science 196:180–182.
2. Boguski, M. S., N. Elshourbagy, J. M. Taylor, and J. I. Gordon. 1984. Rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-amino acid segment with amphipathic helical potential. Proc. Natl. Acad. Sci. USA 81:5021–5025.
3. Boguski, M. S., N. Elshourbagy, J. M. Taylor, and J. I. Gordon. 1985. Comparative analysis of repeated sequences in rat apoli-

poproteins A-I, A-IV, and E. Proc. Natl. Acad. Sci. USA 82:992–996.

4. Breckenridge, W. C., A. Roberts, and A. Kukisis. 1985. Lipoprotein levels in genetically selected mice with increased susceptibility to atherosclerosis. Arteriosclerosis 5:256–264.

5. Breslow, J. L. 1985. Human apolipoprotein molecular biology and genetic variation. Annu. Rev. Biochem. 54:699–727.

6. Brown, M. S., P. T. Kovanen, and J. L. Goldstein. 1981. Regulation of plasma cholesterol by lipoprotein receptors. Science 212:628–635.

7. Buchberg, A. M., and A. J. Kinniburgh. 1985. Induction of liver apolipoprotein A-IV mRNA in porphyric mice. Nucleic Acids Res. 13:1953–1963.

8. Carter, A. D., B. K. Felber, M. Walling, M.-F. Jubier, C. J. Schmidt, and D. H. Hamer. 1984. Duplicated heavy metal control sequences of the mouse metallothionein-I gene. Proc. Natl. Acad. Sci. USA 81:7392–7396.

9. Castelli, W. D., J. T. Doyle, T. Gordon, C. G. Hames, M. C. Hjortland, S. B. Hulley, A. Kagan, and W. J. Zukel. 1977. HDL cholesterol and other lipids in coronary heart disease. Circulation 55:767–772.

10. Chirgwin, J. M., A. E. Przybyla, R. J. McDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–5302.

11. Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. 47:251–276.

12. Church, G., and W. Gilbert. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA 81:1991–1995.

13. Elshourbagy, N. A., M. S. Boguski, W. S. L. Liao, L. S. Jefferson, J. I. Gordon, and J. M. Taylor. 1985. Expression of rat apolipoprotein A-IV and A-I genes: mRNA induction during development and in response to glucocorticoids and insulin. Proc. Natl. Acad. Sci. USA 82:8242–8246.

14. Feinberg, A. D., and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.

15. Karathanasis, S. K. 1985. Apolipoprotein multigene family: tandem organization of human apolipoprotein A-I, CIII and

16. Mahley, R. W., T. L. Innerarity, S. L. Rall, and K. H. Weisgraber. 1984. Plasma lipoproteins: apolipoprotein structure of function. J. Lipid Res. 25:1277–1294.

17. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499–560.

18. McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA 74:4835–4838.

19. Morrisett, J. D., H.-S. Kim, J. P. Patsch, S. K. Datta, J. J. Trentin. 1982. Genetic susceptibility and resistance to diet-induced atherosclerosis and hyperlipoproteinemia. Arteriosclerosis 2:312–324.

20. Paigen, B., A. Morrow, C. Brandon, D. Mitchell, and P. Holmes. 1985. Variation in susceptibility to atherosclerosis among inbred strains of mice. Atherosclerosis 57:65–73.

21. Rajavashisth, T. B., J. S. Kaptien, K. L. Reue, and A. J. Lusis. 1985. Evolution of apolipoprotein E: mouse sequence and evidence for an 11-nucleotide ancestral unit. Proc. Natl. Acad. Sci. USA 82:8085–8089.

22. Roberts, A., and J. S. Thompson. 1976. Inbred mice and their hybrids as an animal model for atherosclerosis research. Adv. Exp. Med. Biol. 67:313–327.

23. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.

24. Segrest, J. P., R. L. Jackson, J. D. Morrisett, and A. M. Gotto. 1974. A molecular theory of lipid-protein interactions in the plasma lipoproteins. FEBS Lett. 38:247–253.

25. Stuart, G. W., P. F. Searle, H. Y. Chen, R. L. Brinster, and R. D. Palmiter. 1984. A 12-base-pair DNA motif that is repeated several times in metallothionein gene promoters confers metal regulation to a heterologous gene. Proc. Natl. Acad. Sci. USA 81:7318–7322.

26. Zinn, K., D. DiMaio, and T. Maniatis. 1983. Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell 34:865–879.

A-IV genes. Proc. Natl. Acad. Sci. USA 82:6374–6378.