Biol. Chem. Hoppe-Seyler
Vol. 373, 187–193, April 1992

Mouse Apolipoprotein AI

cDNA-Derived Primary Structure, Gene Organisation and Complete Nucleotide Sequence

Wilhelm STOFFEL, Rolf MÜLLER, Erika BINCZEK and Kay HOFMANN

Institut für Biochemie, Medizinische Fakultät, Universität zu Köln

(Received 14 February 1992)

Summary: Apolipoprotein AI, the dominant protein component of serum high density lipoprotein, is intimately involved in cholesterol homeostasis. Apo AI activates the lecithin-cholesterol acyltransferase within the HDL particle and functions as ligand for a putative HDL receptor – two properties, which render this apolipoprotein a key mediator in reversed cholesterol transport.

A functional analysis of the apo AI gene demands the isolation of the mouse apo AI gene for expression as transgenes in different mutant forms in the mouse.

Here we describe the isolation of a full length apo AI-specific mouse liver cDNA clone with the human cDNA (892 bp) and the derived amino-acid sequence coding a polypeptide of 264 amino-acid residues. It showed a 70.7% homology to the rat and 66% to the human apo AI sequence.

With this cDNA as probe the mouse apo AI gene was isolated and its organization analysed. Four exons, three of which are coding sequences, are aligned similarly to the human gene.

The gene embraces 1825 bp between the transcription start, and the poly(A)+ tail attached 62 bp downstream of the stop codon. The complete nucleotide sequence of the four exons and three introns of the mouse apo AI gene was determined and its homology compared with that of the rat and human gene. Extensive deletions and a strongly reduced homology of the three introns of the two genes are obvious.

Maus-Apolipoprotein AI: cDNA-abgeleitete Apolipoprotein-Primärstruktur und vollständige Genstruktur

Zusammenfassung: Apolipoprotein AI ist die Hauptproteinkomponente der Serum-High-Density-Lipoproteine, die direkt für die Cholesterin-Homeostase verantwortlich sind. Apo AI aktiviert die Lecithin-Cholesterin-Acyltransferase und bildet den Liganden für den putativen HDL-Rezeptor, zwei Eigenschaften, die diesem Apoprotein eine Schlüsselfunktion im reversen Cholesterintransport verleihen.

Eine funktionale Analyse des Apo-AI-Gens erfordert die Isolierung des Maus-Apo-AI-Gens für seine Expression in Transgene von verschiedenen mutierten Formen des Gens in der Maus.

Wir beschreiben hier die Isolierung eines vollständigen Apo-AI-spezifischen Maus-Leber-cDNA-Klons und der davon abgeleiteten Aminosäuresequenz. Der Klon umfaßt ein Leseraster von 792 bp, die ein

---

Enzymes:
Calf alkaline-intestinal phosphatase, orthophosphoric monoester phosphohydrolase (alkaline medium) (EC 3.1.3.1);
DNA ligase, poly(deoxyribonucleotide): poly(deoxyribonucleotide) ligase (AMP-forming) (EC 6.5.1.1);
Lecithin-cholesterol acyltransferase, recommended name: phosphatidylcholine:sterol *O*-acyltransferase (EC 2.3.1.43);
Polynucleotide kinase, ATP:5′-dephosphopolynucleotide 5′-phosphotransferase (EC 2.7.1.78);
Taq DNA polymerase, deoxynucleoside-triphosphate:DNA deoxynucleotidyltransferase (DNA-directed) (EC 2.7.7.7).

Abbreviations:
apo AI, apolipoprotein AI of serum high density lipoprotein; cDNA complementary DNA; dsDNA, double-stranded DNA; HDL, high density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; PCR, polymerase chain reaction; pfu, plaque-forming units.

Apo AI Polypeptid mit 264 Aminosäureresten codieren. Die Sequenz weist eine Homologie von 70.7% zur Ratten- und 66% zur menschlichen Apo-AI-Sequenz auf.
Mit Hilfe dieser humanen cDNA als Sonde wurde das Mäuse-Apo-AI-Gen isoliert und sein Aufbau untersucht. Vier Exons, von denen drei kodierende Sequenzbereiche enthalten, sind ähnlich wie beim menschlichen Gen angeordnet.
Das Gen umfaßt 1825 bp zwischen dem Transkriptionsstart und der poly(A)$^+$-Sequenz, die sich 62 bp 3′-wärts des Stopcodons befindet. Die komplette Nucleotidsequenz der vier Exons und drei Introns des Mäuse-Apo-AI-Gens wurde ermittelt und die Sequenz mit der der Ratte und des Menschen verglichen. Ausgedehnte Deletionen und eine stark verringerte Homologie der drei Introns der beiden Gene sind offensichtlich.

Key terms: Amino-acid sequence, apo AI gene organization, complete genomic sequence, human-rodent genomic sequence homology.

The functional analysis of the main apolipoproteins of serum high density lipoproteins, particularly of apoprotein AI, has focussed on their gene expression in order to learn about its contribution to cholesterol homeostasis and to define the molecular rationale of phenomena such as reversed cholesterol transport in the blood and the proposed inverse relationship to the development of lipid deposits in aterosclerotic plaques of blood vessels and coronary heart disease in general. Although the primary structure of mature apolipoprotein AI is well known[1] our understanding of lipid-apolipoprotein interactions for the integrity of the HDL particle and the activator function of apo AI for the lecithin-cholesterol-acyltransferase (LCAT) is very limited.

The available recombinant DNA techniques allow us to address these questions on a molecular level, not only in experiments in vitro but also in vivo by introducing the mutant apolipoprotein gene as transgene into the mouse. The experiments so far introduced the human apo AI gene randomly into the mouse genome which is then expressed in addition to the healthy mouse apo AI gene[2].

Here we describe the isolation, organization and the complete nucleotide sequence of the mouse apolipoprotein AI gene as a prerequisite for gene-targeting experiments for the deletion or substitution by mutated apo AI. The primary structure of mouse apo AI was first derived from a full length apo AI-specific cDNA isolated from a mouse liver cDNA library. With the use of this cDNA the full length genomic mouse apo AI DNA has been isolated from a mouse leucocyte genomic library, the apo AI gene organisation into four exons and three intervening sequences and its complete nucleotide sequence has been established. The conserved genomic organization and homology within the coding sequences of the human and rodent (mouse and rat) apo AI gene is discussed.

Materials and Methods

Restriction endonucleases, T_4 DNA ligase, Taq DNA polymerase and polynucleotide kinase were perchased from BRL; [^{35}S]dATP, [^{32}P]ATP and [^{32}P]dATP from Amersham Buchler (Braunschweig, Germany), calf alkaline intestinal phosphatase and the random priming kit from Boehringer (Mannheim, Germany), the mouse liver cDNA library in λgt11 5′ stretch and the mouse genomic library from leucocytes in EMBL-3 SP6/T7 from Clontech Laboratories (Palo Alto, USA).

All molecular biological techniques were carried out according to Ausubel et al.[3] if not otherwise stated.

Human apo AI full length cDNA used for hybridization analyses has been described previously[4].

Full length apolipoprotein AI-specific cDNA clones were isolated from an oligo(dT) and randomly-primed mouse liver cDNA library in λgt11. 3.5×10^5 plaques on *E. coli* Y1090 strain were screened with synthetic 30mer oligonucleotides coding the N- and C-terminal ten amino-acid residues of mature human apo AI labelled with ^{32}P in the polynucleotide-kinase reaction or cDNA fragment, randomly primed with [^{32}P]dATP using the random primer labelling kit, Boehringer. These probes were also used for screening the genomic library and for Southern blot hybridization analysis.

DNA restriction enzyme fragments and PCR fragments were purified with Clean-A-Gene DNA purification kit (Renner GmbH, 6701 Dannstadt, Germany). Nucleotide sequence analysis was carried out with the dideoxy chain termination method of Sanger et al.[5] in its modification for double strand sequencing[6]. Oligonucleotides were synthesized with an automatic DNA synthesizer Applied Biosystems, Model 380A.

Nucleotide sequence homology analyses were performed with the aid of computer programs.
The nucleotide sequence of the mouse apo AI cDNA and the mouse apo AI gene have been entered into the EMBL data library under accession numbers cDNA X64262 and gene X64263.

Results

1) Characterization of full length mouse apo AI-specific cDNA

Apo AI-specific cDNA clones were isolated from a mouse liver λgt11 cDNA library with ^{32}P-labelled oligonucleotides as probes which encoded the nine N-terminal amino-acid residues of the signal sequence (5′ oligo) and C-terminal (3′ oligo) respectively of the human apolipoprotein sequence (Brewer et al.[1]).

```
  1  .ac cca gac tgt cgg aga gct ccg ggg agg tca ccc aca ccc ttc agg ATG AAA GCT GTG   60
                                                                  Met Lys Ala Val

 61  GTG CTG GCC GTG GCT CTG GTC TTC CTG ACA GGG AGC CAG GCT TGG CAC GTA TGG CAG CAA  120
     Val Leu Ala Val Ala Leu Val Phe Leu Thr Gly Ser Gln Ala Trp His Val Trp Gln Gln

121  GAT GAA CCC CAG TCC CAA TGG GAC AAA GTG AAG GAT TTC GCT AAT GTG TAT GTG GAT GCG  180
     Asp Glu Pro Gln Ser Gln Trp Asp Lys Val Lys Asp Phe Ala Asn Val Tyr Val Asp Ala

181  GTC AAA GAC AGC GGC AGA GAC TAT GTG TCC CAG TTT GAA TCC TCC TCC TTG GGC CAA CAG  240
     Val Lys Asp Ser Gly Arg Asp Tyr Val Ser Gln Phe Glu Ser Ser Ser Leu Gly Gln Gln

241  CTG AAC CTG AAT CTC CTG GAA AAC TGG GAC ACT CTG GGT TCA ACC GTT AGT CAG CTG CAG  300
     Leu Asn Leu Asn Leu Leu Glu Asn Trp Asp Thr Leu Gly Ser Thr Val Ser Gln Leu Gln

301  GAA CGG CTG GGC CCA TTG ACT CGG GAC TTC TGG GAT AAC CTG GAG AAA GAA ACA GAT TGG  360
     Glu Arg Leu Gly Pro Leu Thr Arg Asp Phe Trp Asp Asn Leu Glu Lys Glu Thr Asp Trp

361  GTG AGA CAG GAG ATG AAC AAG GAC CTA GAG GAA GTG AAA CAG AAG GTG CAG CCC TAC CTG  420
     Val Arg Gln Glu Met Asn Lys Asp Leu Glu Glu Val Lys Gln Lys Val Gln Pro Tyr Leu
         EcoRI
421  GAC GAA TTC CAG AAG AAA TGG AAA GAG GAT GTG GAG CTC TAC CGC CAG AAG GTG GCG CCT  480
     Asp Glu Phe Gln Lys Lys Trp Lys Glu Asp Val Glu Leu Tyr Arg Gln Lys Val Ala Pro

481  CTG GGC GCC GAG CTG CAG GAG AGC GCG CGC CAG AAG CTG CAG GAG CTG CAA GGG AGA CTG  540
     Leu Gly Ala Glu Leu Gln Glu Ser Ala Arg Gln Lys Leu Gln Glu Leu Gln Gly Arg Leu

541  TCC CCT GTG GCT GAG GAA TTT CGC GAC CGC ATG CGC ACA CAC GTA GAC TCT CTG CGC ACA  600
     Ser Pro Val Ala Glu Glu Phe Arg Asp Arg Met Arg Thr His Val Asp Ser Leu Arg Thr

601  CAG CTA GCG CCC CAC AGC GAA CAG ATG CGC GAG AGC CTG GCC CAG CGC CTG GCT GAG CTC  660
     Gln Leu Ala Pro His Ser Glu Gln Met Arg Glu Ser Leu Ala Gln Arg Leu Ala Glu Leu

661  AAG AGC AAC CCT ACC TTG AAC GAG TAC CAC ACC AGG GCC AAA ACC CAC CTG AAG ACA CTT  720
     Lys Ser Asn Pro Thr Leu Asn Glu Tyr His Thr Arg Ala Lys Thr His Leu Lys Thr Leu

721  GGC GAG AAA GCC AGA CCT GCG CTG GAG GAC CTG CGC CAT AGT CTG ATG CCC ATG CTG GAG  780
     Gly Glu Lys Ala Arg Pro Ala Leu Glu Asp Leu Arg His Ser Leu Met Pro Met Leu Glu

781  ACG CTT AAG ACC AAA GCC CAG AGT GTG ATC GAC AAG GCC AGC GAG ACT CTG ACT GCC CAG  840
     Thr Leu Lys Thr Lys Ala Gln Ser Val Ile Asp Lys Ala Ser Glu Thr Leu Thr Ala Gln

841  tga ggt gcc cgc ttc cac tcc cca ccc ccg cat tgg ctt tct tac aat aaa cct ttc caa  900
901  aat gga aaa aaa aaa aaa aaa aaa                                                  924
```

Fig. 1. Nucleotide sequence of mouse apo AI cDNA and the derived amino-acid sequence displayed above the nucleotide sequence.

5′ and 3′ oligonucleotide primers, the *Eco*RI site and the polyadenylation signal are underlined.

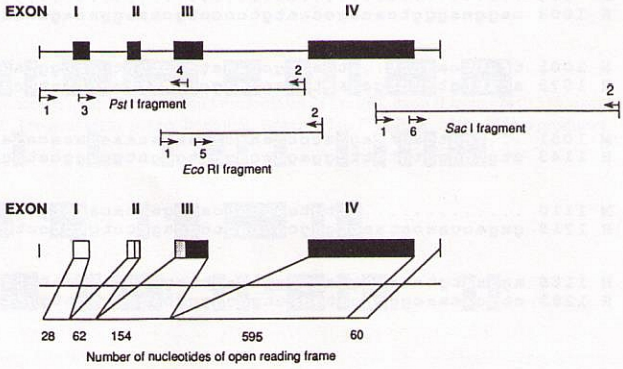

Fig. 2. A) Restriction map of mouse genomic apo AI clone EMBL-3 AI and strategy of sequence analysis.

The short and long EMBL-3 arms are depicted as thin lines. The region containing the apo AI gene with the four exons is blown up as filled boxes. Arrows show sequencing primers. The relevant restriction sites are: P, *Pst*I; E, *Eco*RI; S, *Sac*I. The gap between the 3′ end of *Eco*RI fragment and the 5′ end of the *Sac*I fragment as part of exon IV is deduced from the cDNA sequence of the mouse cDNA clone described in the previous chapter.

The stippled areas in exons II and III are coding sequences of the prepro sequences.

```
M    1 ctgcagccaccccctgctgccccccactgaacccttgatcccagctcggcagccccgcagcttcctgtt
H    1 ctgcagcactcccctccgcccccca.ctgaacccttgacccctgccctgcacgccccgcagcttgctgtt

M   71 tgcccact.ctctttgcccagcctcaggaacagagctgatccttgaactc.taagttccacatcgccagc
H   70 tgcccactcctatttgcccagtcccagggacagagctgatccttgaactctaagttccacattgccagg

M  139 aaaagtaagcagtggcagggcca.ggctgagcttatcagtctcccaagtcccccagcccctgcccacacac
H  140 accagtgagcagcaacagggccagggctgggcttatcagcctccca...gcccagaccctggctgcagac

M  208 atatatagaccagggaagaagagctggacACCCAGACTGTCCGGAGAGCTCCGGGGAGGtaagtgctgcta
H  207 ataaatagggccctg...caagagctggctGCTTAGAGACTGCGAGA......AGGAGGtgcgtcctgctg

M  278 cctgccttcggtggctctggctccatagcgcctccagttgatgctccactgtccaaagtcccaaatcaat
H  268 cctgcc..cggcactctggctccccag...ctcaaggttcaggcttgccccaggccgggcctctgggt

M  348 accgttggatatctcgcacctttagccattctagccaatgcttccatgggcttgaattgtgtgtggagcc
H  333 acctgaggtcttctccgctctgtgccctttct........cctcacctggctgcaatgagtggggagca

M  418 tttccatgacaatgcctccctttcagcccagcccctcctcctccttcttcttcttcagGTCACCCAC.ACCC
H  395 cggggcttctgcatgctgaaggcacccactcagccaggccctttcttctcctccagGTCCCCCACGCCCC

M  487 TTCAGGATGAAAGCTCTGGTGCTCGCCCGTGGCTCTGGTCTTCCTGACAGgtaggtgcctcttgacctgcg
H  465 TTCAGGATGAAAGCTCCGGTGCTGACCTTGGCCGTGCTCTTCCTGACGGgtaggtgtcccctaacctagg

M  557 tg.....ggactacctcctgggcacagagaacagaattc
H  535 agccaaccatcgggggctttctccctaaatccccgtggcccacccctcctgggcagaggcagcaggttcc

M  592 ccactgttctcttc.....cctgactccgagtctaacctaa
H  605 tcactggcccccctcccccacctccaagcttggcctttcggctcagatctcagcccacagctggcctga

M  628 catgggtctcccct.ccatcccagGGAGCCAGGCTTGGCACGTATGGCAGCAAGATGAA...CCCCAGT
H  675 tctgggtctcccctcccacccctcagGGAGCCAGGCTCGGCATTTCTGGCAGCAAGATGAACCCCCCCAGA

M  694 CCCAATGGGACAAAGTGAAGGATTTCGCTAATGTGTATGTGGATCCGGTCAAAGACAGCGGCAGAGACTA
H  745 GCCCCTGGGATCGAGTGAAGGACCTGGCCACTGTGTACGTGGATCTGCTCAAAGACAGCGGCAGAGACTA

M  764 TGTGTCCCAGTTTGAATCCTCCTCCTTGGGCCAACAGCTGAAgtaaggaaagacctagcgtggggcctgg
H  815 TGTGTCCCAGTTTGAAGGCTCCGCCTTGGGAAAACAGCTAAAgtaag....gacccagccctggggttgag

M  834 agcaggtcaagggctgcccatccagggtgggcagagagaccagtgagaagatgctggaactgagctggct
H  881 ggcagg.....ggcagggggcagaggcctgtggggatgatgttgaagccagactggcc

M  904 agcccttcacgggctttcctaccagctgggcaccatggcaggttccagtggaggactagggatgggattc
H  933 gagtcctcacctaatatctgatgagctgggcccccacagatggtctggatggagaaaccggaatggatctc

M  974 .....atctggctgttgggtaacacagcccc
H 1003 caggcagggtcacagcccatgtcccctgcaaaggacagaccagggctgcccgatgcgtgatcacagagcc

M 1001 tcattcagcct..atgagtgccaaatccctttttcctttggtaacccccagtac.....
H 1073 acattgtgcctgcaagtgtagcaagcccctttccctttcttcaccacctcctctgctcctgcccagcaaga

M 1051 ...tggtcagacagcacccaaaacaaaacaaaacaaaacaaaaacaaaaaacgggactggc.....
H 1143 ctgtgggctgtcttcggagaggagaatgcgctggaggcatagaagcgaggtcctcaagggcccactttg

M 1110 .....cttgtaaccagcaccgaccacatctgtgttcttggtgccttccg.ccatgtgtacag
H 1213 gagaccaacgtaactgggcaccagtcccagctctgtctcctttttagctcctctctgtgcctcggtccag

M 1166 aggaatgtaaagggaaggcagtgagttagagggtgcatgttcgggggaaactagga...ccatagcaact
H 1283 ctgcacaacgggcatggctggcgggcagggggtgttggttgagagtgtactggaaatgctaggccact
```

```
M 1233 gcacattagggacaggtggcacccagctatcatgtgcatggatctgcagaccagggcagcgcatgatg
H 1353 gcacctccgcggacaggtgtcacccagggctcaccctgataggctggggcgctggg.............

M 1303 cctgggctcgtctctcagccgctctcttcccctctagCCTGAATCTCCTGGAAAACTGGGACACTCTGG
H 1410 ...aggcagccctcaacccttctgtctcaccctccagCCTAAAGCTCCTTGACAACTGGGACAGCGTGA

M 1373 GTTCAACCGTTACTCAGCTGCAGGAACGGCTGGGCCCATTGACTCGGGACTTCTGGGATAACCTGGACAA
H 1477 CCTCCACCTTCACCAAGCTGCGCGAACAGCTCGGCCCTGTGACCCAGGAGTTCTGGGATAACCTGGAAAA

M 1443 AGAAACAGATTGGGTGAGACAGGAGATGAACAAGGACCTAGAGGAAGTGAAACAGAAGGTGCAGCCCTAC
H 1547 GGAGACAGAGGGCCTGAGGCAAGAGATGAGCAAGGATCTGGAGGAGGTGAAGGCCAAGGTGCAGCCCTAC

M 1513 CTGGACGAATTCCAGAAGAATGGAAAGAGGATGTGGAGCTCTACCGCCAGAAGGTGGCGCCTCTGCGCG
H 1617 CTGGACGACTTCCAGAAGAAGTGGCAGGAGGAGATGGAGCTCTACCGCCAGAAGGTGGAGCCGCTGCGCG

M 1583 CCGAGCTGCAGGAGAGCGCGCGCCAGAAGCTGCAGGAGCTGCAAGGGAGACTGTCCCCTGTGGCTGAGGA
H 1687 CAGAGCTCCAAGAGGGCGCGCGCCAGAAGCTGCACGAGCTGCAAGAGAAGCTGAGCCCACTGGGCGAGGA

M 1653 ATTTCGCGACCGCATGCGCACACACGCTAGACTCTCTGCGCACACAGCTAGCGCCCCACAGCGAACAGATG
H 1757 GATGCGCGACCGCGCGCGCGCCCATGTGGAGCGCTGCGCACGCATCTGGCCCCCTACAGCGACGAGCTG

M 1723 CGCCAGAGCCTGGCCCAGCGCCTGGCTGAGCTCAAG......AGCAACCCTACCTTGAACGAGTACCACA
H 1827 CGCCAGCGCTTGGCCGCGCGCCTTGAGGCTCTCAAGGAGAACGGCGGCGCCAGACTGGCCGAGTACCACG

M 1787 CCAGGGCCAAAACCCACCTGAAGACACTTGGCGAGAAAGCCAGACCTGCGCTGGAGGACCTGCGCCATAG
H 1897 CCAAGGCCACCGAGCATCTGAGGCACGCTCAGCGAGAAGGCCAAGCCCGCGCTCGAGGACCTCCGCCAAGG

M 1857 TCTGATGCCCCATGCTGGAGACGCTTAAGACCAAAGCCCAGAGTGTGATCGACAAGGCCAGCGAGACTCTG
H 1967 CCTGCTGCCCCGTGCTGGAGAGCTTCAAGGTCAGCTTCCTGAGCGCTCTCGAGGAGTACACTAAGAAGCTC

M 1927 ACTGCCCAGtgaggtgcccgcttccactccccaccccgcattggcttcttacaataaaccctttccaaa
H 2037 AACACCCAGtgaggcgcccgcggccgcccccttcccgg........tgctcagaataaacgtttccaaa

M 1997 atgg..aatagcttctttcttgggggacatagggcgggcgctaagggggacatcaagggacgtgagaaca
H 2099 gtggaagcagcttctttctttgggagaatagagggggtgcgggggacatccgggggagcccgggagggg

M 2065 tggtgccgcactgggggattcctttgtacgcgacatctcagc.tcttaacggtcactcaagctgggcacct
H 2169 gccttggccctggagcagggactcctgccggatctcaacaactccgtgcccagactggacgtcttagg

M 2134 ggctggttcagggtatgagacagaatccttctaagaaggctcc
H 2239 gccaagatcgacgttggag..gacctgctggacgcntggctgc
```

Fig. 2.B) Nucleotide sequences of the mouse apo AI gene and sequence homologies to human genomic nucleotide sequence.

Sequencing primers:
1: M13 universal primer
2: M13 reverse primer
3: 5′-TAGACCAGGGAAGAAGA-3′
4: 5′-CGTGCCAAGCCTGGCTC-3′
5: 5′-TGTGTATGTGGATGCGC-3′
6: 5′-TGCCGCCATAGTCTGATG-3′.

Exon sequences are given in capitals, intron sequences in small letters. Exon I nucleotides 237 to 264, exon II from 474 to 535, exon III from 652 to 805 and exon IV from 1341 to 1935. The average sequence homology of exons I to IV is 76%, of the 175 bp promoter region 85% and 44% for introns I to III.

Approximately 3.5 × 10^5 pfu plated on competent *E. coli* Y1090 were screened and the λ-DNA of plaque-purified clones isolated and the insert released with *Eco*RI. Two *Eco*RI fragments 425 and 580 bp long were isolated. The 425-bp fragment hybridized with the 5′-oligonucleotide probe and the 580 bp fragment with the 3′-oligonucleotide. The two fragments were subcloned into pBluescript II vector previously restricted with *Eco*RI and dephosphorylated for dideoxy chain termination sequencing of the dsDNA[6] with specific oligonucleotide primers. Using the antisense and sense 30mer oligonucleotides now resembling those of the mouse apo AI cDNA (underlined in Fig. 1) as primers in the PCR with the mouse λ-apo AI clone as template a blunt-ended 900-bp PCR fragment was produced for double-strand sequencing again after subcloning into the *Sma*I site of the pBluescript II vector. The complete sequence of the full length mouse apo AI cDNA (Fig. 1) was obtained using oligonucleotides as primers according to the growing sequence information from the 5′ and 3′ ends of the cDNA insert.

Fig. 1 presents the nucleotide sequence of the mouse cDNA clone and the derived amino-acid sequence. The *Eco*RI site at position 424 explains the two *Eco*RI fragments released from the cDNA insert upon *Eco*RI digestion of the λgt11 clone.

The open reading frame encodes a polypeptide of 264 amino-acid residues, three residues shorter than the human prepro apo AI[7,8]. A comparison of the signal sequence and the following six-amino-acid prosequence with the unusual Gln-Gln sequence at the C-terminus of mouse, rat and human apo AI indicates the strong homology conserved through species and supports the well known common pathway of protein processing of human apo AI[8–13].

2) Characterization of the mouse apolipoprotein AI gene

Screening of 3 × 10^5 recombinants of mouse genomic EMBL-3 SP6/T7 with the full length 892 bp long human apo AI cDNA labelled by random priming yielded one clone with a 15 kb insert which harbored all coding sequences as documented by Southern blot hybridization with 5′- and 3′-labelled synthetic oligonucleotides and the nick-labelled human apo AI cDNA. This clone was selected for further analysis. The restriction map with three enzymes, *Eco*RI, *Pst*I and *Sac*I, is shown in Fig. 2A. *Eco*RI, *Pst*I and *Sac*I fragments were subcloned in pBluescript II for double-strand sequencing using the appropriate primers as indicated in Fig. 2A. Four exons are distributed over 1.9 kb. The nucleotide sequence of the mouse apo AI gene is listed in Fig. 2B.

The comparison of the mouse and human apo AI-genomic sequences clearly underlines the extensive homology and conservative exchanges within the coding exons and the 5′-untranslated region. Coding exons II, III, and IV show 85, 84 and 74% homology, respectively. The 175 bp promoter domain 5′ of the putative transcription start is equally homologous in the apo AI gene of both species. The intron homologies are below 50% (40% on average).

Discussion

The isolation and characterization including the complete nucleotide sequence of a full length cDNA clone and of the apolipoprotein AI gene of the mouse is reported in this paper. The coding region of the prepro apo AI mRNA is 795 nucleotides long. Its cDNA-derived translation product is a 264-amino-acid-residue polypeptide and exhibits three amino-acid deletions within the mature sequence as compared to the human prepro apo AI. Like the human apo AI it carries an N-terminal 18-residue long signal sequence followed by a six-amino-acid prosequence which are processed intracellularly[8–13] and a 240 residues long mature form is attained. Its structural characteristics with two 11- and nine 22-amino-acid-residue-internal repeats forming lipid-binding amphipathic helices resembles those of the human apo AI despite the numerous mutations and thereby refers to identical functional properties.

The homologous amino-acid sequences of mouse, rat and human apo AI is documented in Fig. 3.

The mouse-genomic DNA sequences revealed the four-exon and three-intron organization, which all soluble apolipoproteins except apo AIV have in common[14,15]. The 5′-noncoding region is interrupted by intron I. Intron II interrupts the open reading frame of the signal sequence. The alignment of the mouse and human genomic sequence indicates a 76% homology within exon I but a 75% to 85% homology of the exon II, III, and IV sequences. The overall homology of the introns I, II, and III is only 44%. However, the prómoter region of 175 bp upstream of the putative transcription start is 85% homology, Fig. 2B.

The polyadenylation signal is located 42 bp and the polyadenylation site 63 bp downstream the stop codon TGA of the apo AI gene. No TATA box could be identified within 240 bp upstream of the sequenced 5′ end of our mouse cDNA clone which suggests that the mouse apo AI promoter lacks the TATA box. This closely related structure-function relationship of human and mouse apo AI makes feasible the study of apolipoprotein expression in its normal and mutated

```
Mouse   1   MKAVVLAVALVFLTGSQAWHVWQQDE·PQSQWDKVKDFANVYVDAVKDSGRDYVSQFESS
Rat     1   MKAAVLAVALVFLTGCQAWEFWQQDE·PQSQWDRVKDFATVYVDAVKDSGRDYVSQFESS
Human   1   MKAAVLTIAVLFLTGSQARHFWQQDEPPQSPWDRVKDLATVYVDVTKDSGRDYVSQFEGS

Mouse  60   SLGQQLNLNLLENWDTLGSTVSQLQERLGPLTRDFWDNLEKETDWVRQEMNKDLEEVKQK
Rat    60   TLGKQLNLNLLDNWDTLGSTVGRLQEQLGPVTQEFWANLEKETDWLRNEMNKDLENVKQK
Human  61   ALGKQLNLKLLDNWDSVTSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEVKAK

Mouse 120   VQPYLDEFQKKWKEDVELYRQKVAPLGAELQESARQKLQELQGRLSPVAEEFRDRMRTHV
Rat   120   MQPHLDEFQEKWNEEVEAYRQKLEPLGTELHKNAK....EMQRHLKVVAEEFRDRMRVNA
Human 121   VQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHV

Mouse 180   DSLRTQLAPHSEQMRESLAQRLAELKSNPT··LNEYHTRAKTHLKTLGEKARPALEDLRH
Rat   176   DALRAKFGLYSDQMRENLAQRLTETRNHPT··LIEYHTKAGDHLRTLGEKAKPALDDLGQ
Human 181   DALRTHLAPYSDETRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQ

Mouse 238   SLMPMLETLKTKAQSVIDKASETLTAQ
Rat   234   GLMPVLEAWKAKIMSIDEAKKLNA.
Human 241   GLLPVLESFKVSFISALEEYTKKLNTQ
```

Fig. 3. Homology of amino acid sequences of prepro apo AI of mouse, rat and man.

Homology between mouse and rat is 70.7% and between mouse and man 66%.

forms as transgenes in the mouse model for their functional analysis in vivo.

References

1 Brewer, H.B., Jr., Fairwell, T., LaRue, A., Ronan, R., Houser, A. & Bronzert, T.J. (1978) *Biochem. Biophys. Res. Commun.* **80**, 623–630.

2 Chajek-Shaul, T., Hayek, T., Walsh, A. & Breslow, J.L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6731–6735.

3 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.

4 Stoffel, W. & Binczek, E. (1988) *Biol. Chem. Hoppe-Seyler* **369**, 1055–1063.

5 Sanger, F., Nicklen, S. & Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

6 Chen, E.Y. & Seeburg, P.H. (1985) *DNA* **4**, 165–170.

7 Gordon, J.I., Smith, D.P., Andy, R., Alpers, D.H., Schonfeld, G. & Strauss, A.W. (1982) *J. Biol. Chem.* **257**, 971–978.

8 Stoffel, W., Blobel, G. & Walter, P. (1981) *Eur. J. Biochem.* **120**, 519–522.

9 Stoffel, W., Knyrim, K. & Bode, C. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1631–1640.

10 Stoffel, W., Bode, C. & Knyrim, K. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 439–445.

11 Gordon, J.I., Smith, D.P., Alpers, D.H. & Strauss, A.W. (1982) *J. Biol. Chem.* **257**, 8418–8423.

12 Zannis, V.I., Karathanasis, S.K., Keutmann, H.T., Goldberger, G. & Breslow, J.L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2574–2578.

13 Zannis, V.I., McPherson, J., Goldberger, G., Karathanasis, S.K. & Breslow, J.L. (1984) *J. Biol. Chem.* **259**, 5495–5499.

14 Karathanasis, S.K., Zannis, V.I. & Breslow, J.L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6147–6151.

15 Luo, C.C., Li, W.H., Moore, M.N. & Chan, L. (1986) *J. Mol. Biol.* **187**, 325–340.

W. Stoffel*, R. Müller, E. Binczek and K. Hofmann, Institut für Biochemie, Medizinische Fakultät der Universität zu Köln, Joseph-Stelzmann-Str. 52, W-5000 Köln, Germany.

* To whom correspondence should be addressed.