Nucleotide sequence of cloned cDNA of human apolipoprotein A-I

Peter Cheung and Lawrence Chan

Departments of Cell Biology and Medicine, Baylor College of Medicine, Houston, TX 77030, USA

Received 8 February 1983; Revised 29 April 1983; Accepted 6 May 1983

ABSTRACT

ApoA-I is the major human HDL apoprotein. By oligonucleotide hybridization, we have isolated 5 cDNA clones to human hepatic apo A-I mRNA. One of these clones (pA1-3) was completely sequenced. It has 878 bp plus a poly A tail of 48 and includes all the coding and 3'-untranslated regions of the mRNA and part of the 5'-untranslated region. It predicts a peptide sequence of 267 amino acids (including the 24 amino acid prepropeptides) which is very similar to the sequence reported by Brewer *et al.*, (1978) *Biochem. Biophys. Res. Commun.* 80:623-630. The predicted signal peptide sequence is highly homologous to the rat apoA-I signal peptide. There is no evidence for any internally repeated segments in apoA-I either at the amino acid or at the DNA level. Using pA1-3 as a probe, we have detected on Northern gels apo A-I mRNA sequences of ~1100 nucleotides in human hepatic and baboon hepatic and intestinal RNAs, but not in RNAs from baboon skeletal muscle, kidney or spleen. The demonstration of apo A-I mRNA sequences in specific organs is important to our concept of "reverse cholesterol transport".

INTRODUCTION

There has been renewed interest in the high density lipoproteins (HDL) since the recent rediscovery that HDL are negatively correlated with the development of atherosclerosis (1-4). The molecular basis for this observation is unclear. It has been postulated that HDL might be important in "reverse cholesterol transport", i.e., the delivery of cholesterol from the peripheral tissues to the liver for disposal (5,6). There are *in vitro* experiments involving isolated red cells and cultured cells that support this contention (6-9). Metabolic studies in man have suggested an inverse correlation between the size of total body cholesterol pools and the plasma HDL level (10). It is also possible that HDL might simply be a marker of a certain pattern of lipid transport and metabolism which confers protection against the development of atherosclerosis (11).

Apo A-I is the major apoprotein in HDL. Jackson *et al.* (12) demon-

strated that apo A-I was a necessary component in mixtures with phospholipid to successfully remove cholesterol from ascite cell membranes. Apo A-I is also a potent activator of lecithin-cholesterol acyltransferase, a plasma enzyme catalyzing the conversion of cholesterol and phosphatidylcholine to cholesterol esters and lysophosphatidylcholine (13,14). The liver and intestine are the major sites of synthesis of apo A-I (15-21). However, recently Blue *et al.* (22) reported that in the rooster, significant amounts of the protein were also produced by a variety of other tissues including the kidney, aorta, peripheral arteries, veins and skeletal muscle.

The amino acid sequence of human apo A-I has been reported by two laboratories (23-25). Major differences (in 25 positions) in the reported sequence have been noted. The resolution of these differences is important since apo A-I has been used as a model apoprotein system in many physicochemical studies.

We have successfully cloned the cDNA for human hepatic apo A-I mRNA by the technique of oligonucleotide hybridization. Several of the apo A-I clones we have identified include the complete coding region of the mRNA plus some 3' and 5'-untranslated regions. In this communication, we report the complete nucleotide sequence of one of these clones. We have also used the cloned cDNA as a probe to assay for the presence of apo A-I mRNA from various human and baboon tissues. The derived amino acid sequence from our clone is almost identical to one of the previously reported sequences (25). Furthermore, contrary to observations in the avian species, the baboon liver and small intestine appear to be the only major organs harboring apo A-I mRNA sequences.

MATERIALS AND METHODS

cDNA Library

A human cDNA library was a generous gift of Drs. W.E. O'Brien and A.L. Beaudet (26). It was constructed from a mRNA preparation from adult liver tissue. The ds cDNA was inserted into the Pst I site of pBR322 by the GC-tail technique. The host was *Escherichia coli* K12 RRI.

Oligonucleotide Synthesis and Characterization

A mixture of eight oligonucleotides (17-mers) with the sequence 5'd (TCC AA_C^I TC_C^I TC_C^I TGCCA) (Figure 1) was custom synthesized by P-L Biochemicals by the phosphotriester technique (27). The sequences of the

oligonucleotides were confirmed by nucleotide sequencing by the Maxam-Gilbert technique (28).

Labeling of Oligonucleotides

Oligonucleotides were labeled at the 5' end by transfer of ^{32}P from $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ by using bacteriophage T4 polynucleotide kinase (Boehringer-Mannheim) as described (28). Labeled oligonucleotides were purified by DE52 on exchange chromatography (29).

Screening of cDNA Clones

Recombinant cDNA clones were picked and put on fresh tetracycline (25 ug/ml) plates in an ordered array in duplicate. The bacterial colonies were allowed to grow overnight, transferred to Whatman 541 filter paper, amplified with chloramphenicol, and prepared for hybridization as described by Gergen *et al.* (30). The filters were prehybridized, hybridized and washed as described by Suggs *et al.* (31) and Wallace *et al.* (30). Temperature of hybridization was 46°C as recommended by Wallace *et al.* (30).

Restriction Mapping and Nucleotide Sequence Determination

Inserted cDNA sequences were recovered by digestion with Pst 1. Pst 1 sites were regenerated in clones pAl-3 and pAl-2 but not in pAl-4. Restriction mapping was performed by standard techniques under conditions recommended by the suppliers. The entire sequence of pAl-3 and partial sequences of pAl-2 and pAl-4 were determined by the chemical method of Maxam and Gilbert (28). Fragments of cloned cDNA were labeled at the 5' ends by the sequential action of bacterial alkaline phosphatase and polynucleotide kinase plus $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (28). In some fragments with overhanging 5' ends, reverse transcriptase was used to fill in the termini in the presence of $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$ (32).

Blot Hybridization of mRNA

Total RNA was prepared from human liver, baboon liver, intestine, kidney, muscle, and spleen as previously described (33). Poly A RNA was isolated by two cycles of oligo dT cellulose chromatography (34). Poly A RNA was glyoxylated and separated by electrophoresis on 1.5% agarose gels and transferred to Genescreen paper (New England Nuclear) by the method of Thomas (35). RNA molecules were detected by hybridization to nick-translated (36) ^{32}P -labeled 926 kb insert from pAl-3.

	108					113
Amino Acid	Trp	Gln	Glu	Glu	Met	Glu
mRNA	5'(UGG	CA ^A _G	GA ^A _G	GA ^A _G	AUG	GA ^A _G)3'
Oligonucleotide	3'd(ACC	GT ^T _C	CT ^T _C	CT ^T _C	TAC	CT)5'

Figure 1. Sequence of oligonucleotide used as hybridization probe. The last AG (or TC) wobble for Glu 113 was dropped to reduce the number of oligonucleotides from 16 to 8.

RESULTS AND DISCUSSION

Previous studies in our laboratory indicate that apo A-I mRNA accounts for a major fraction (~1%) of the total human hepatic mRNA activity (18). We therefore screened for apo A-I cDNA clones in an unfractionated cDNA library. Selection for apo A-I cDNA sequence by colony hybridization was performed with a mixture of 8 oligonucleotides (Figure 1) which represented all the possible codons for amino acids number 108-113. The hybridization conditions were those of Wallace *et al.* (29). Under these conditions, after screening 2000 colonies, six were strongly positive. By restriction mapping, five of these showed identical patterns. Of these five, two (pAl-2 and pAl-4) were partially and one (pAl-3) completely sequenced.

The partial restriction map and sequencing strategy of pAl-3 are shown in Figure 2. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites on the map were accounted for by sequence analysis, and all of the restriction sites used to end-label DNA fragments were sequenced across by labeling a neighboring restriction site. The nucleotide sequence of

Figure 2. Partial restriction map and sequencing strategy for pAl-3 DNA. Horizontal arrows specify the direction of sequencing from the ³²P label at the indicated restriction sites (small vertical bar).

```

      10      20      30      40      50
1  TCCCCACGG CCCTTCAGGA TGAAGCTGC GGTGCTGAC TTGGCGTGC
51 TCTTCTGAC GGGGAGCCAG GCTGGGATT TCTGGCAGCA AGATGAACCC
101 CCCGAGAGC CTTGGGATCG AGTGAAGGAC CTGGCCACTG TGTACGTGGA
151 TGTGCTCAGA GACAGCGGCA GAGACTATGT GTCCCACTTT GAAGGCTCCG
201 CTTTGGGAAA ACAGCTAAAC CTAAGCTCC TTGACAACCTG GGACAGCGTG

      260      270      280      290      300
251 ACCTCCACCT TCAGCAAGCT GCGCGAACAG CTCGGCCCTG TGACCCAGGA
301 GTTCTGGGAT AACCTGGAAA AGGAGACAGA GGGCCTGAGG CAGGAGATGA
351 GCAGGATCT  GGAGGAGTG  AAGGCCAAGG  TGCAGCCCTA  CCTGGACGCA
401 TTCCAGAGCA AGTGGCAGGA GGAGATGGAG CTCTACCGGC AGAGGCTGGA
451 GCGCTGCGC  GCAGAGCTCC  AAGAGGGCGC  GCGCCAGAG  CTGACGAGC

      510      520      530      540      550
501 TGCAAGAGAA GCTGAGCCCA CTGGGCGAGG AGATGGGCA CCGCGCGGCG
551 GCCATGTGG  ACGCGCTGCG  CACGATCTG  GCGCCCTACA  GCGACGAGCT
601 GCGCCAGCGC TTGGCCCGCG GCCTTGAGGC TCTCAGGAGG AACGGCGGCG
651 CCAGACTGCG CGAGTACCAC GCCAAGGCCA CCGAGCATCT GAGGACGCTC
701 AGCGAGAGG  CCAAGCCCGC  GCTCGAGGAC  CTCGCGCAG  GCCTGCTGCC

      760      770      780      790      800
751 CGTGTGGAG  AGCTTCAGG  TCAGCTTCCT  GAGCGCTCTC  GAGGAGTACA
801 CTAGAGAGCT CAACACCCAG TGAGGCGCCC GCGCGCGCCC CCCTTCCCGG
851 TGCTCAGAT  AACCTTTCC  AAGTGGGAA  AAAAAAAAA  AAAAAAAAA
901 AAAAAAAAA  AAAAAAAAA  AAAAAA

```

Figure 3. Complete nucleotide sequence of pAl-3 DNA.

pAl-3 is given in Figure 3.

This apo A-I cDNA clone consists of 878 bp plus 48 nucleotides in the poly A tail. The sequence starts 19 nucleotides upstream from the initiation codon. It consists of the codons for the entire signal peptide and pro-segment (16), as well as the codons for the 243 amino acids in the plasma protein. It also includes the termination codon TGA, plus the 3'-untranslated segment of 55 nucleotides and the poly A tail. The putative polyadenylation signal AATAAA starts 21 nucleotides upstream from the first residue of the poly A.

The coding segment of pAl-3 and the derived amino acid sequence are shown in Figure 4. There are several interesting features of this portion of the sequence. The codon usage for the various amino acids is extremely skewed (Figure 5). There is a marked preference for GMP (121) and CMP (106) over AMP (21) and UMP (20). This is especially evident when some of the more frequently represented amino acids are examined. For example, there are 41 residues of Leu in the sequence. Twenty-two of them are encoded by CUG whereas none of them are encoded by UUA. Such extreme bias in the preference for specific codons is also evident in some other amino acids including Phe, Val, Pro, Ala, Gln, Asn, Lys, Asp, Glu, Arg, Ser and Gly (Figure 5).

	22	25	28	31	34	37	40	43	46	49	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97
20	ATG	AAA	GCT	GCG	GTG	CTG	ACC	TTC	GCC	GTG	CTC	TTC	CTG	ACC	GCG	AGC	ONG	GCT	GCG	CAT	TTC	TGG	CAG	CAG	GAT	GAA
MET	LYS	ALA	ALA	VAL	LEU	THR	LEU	ALA	VAL	LEU	PHE	LEU	THR	GLY	SER	GLN	ALA	ARG	HIS	PHE	TRP	GLN	ASP	GLU		
	100	103	106	109	112	115	118	121	124	127	130	133	136	139	142	145	148	151	154	157	160	163	166	169	172	175
98	CCC	CCC	CAG	AGC	CCC	TGC	GAT	CGA	CTG	ANG	GAC	CTG	GCC	ACT	GTG	TAC	GTG	GAT	GTG	CTC	AAA	GAC	AGC	GCC	AGA	GAC
PRO	PRO	GLN	SER	PRO	TRP	ASP	ARG	VAL	LYS	ASP	LEU	ALA	THR	VAL	TYR	VAL	ASP	VAL	LEU	LYS	ASP	ASP	GCC	GLY	ARG	ASP
	178	181	184	187	190	193	196	199	202	205	208	211	214	217	220	223	226	229	232	235	238	241	244	247	250	253
176	TAT	GTG	TCC	CAG	TTT	GAA	GCC	TCC	GCC	TTC	GGA	AAA	ONG	CTA	ANG	CTA	ANG	CTC	CTT	GAC	ANG	TGC	GAC	ACC	GTG	ACC
TYR	VAL	SER	GLN	PHE	GLU	GLY	SER	ALA	LEU	GLY	LYS	GLN	LEU	ASN	LEU	ASN	LEU	LEU	ASP	ASN	TRP	ASP	SER	VAL	THR	
	256	259	262	265	268	271	274	277	280	283	286	289	292	295	298	301	304	307	310	313	316	319	322	325	328	331
254	TCC	ACC	TTC	AGC	ANG	CTG	GCC	GAA	CAG	CTC	GCC	CCT	GTG	ACC	CAG	CAG	TTC	TGC	GAT	AAC	CTG	GAA	ANG	GAG	ACA	GAG
THR	PHE	SER	LYS	LEU	ARG	GLU	GLN	LEU	GLN	LEU	GLY	PRO	VAL	THR	GLN	GLU	PHE	TRP	ASP	ASN	LEU	GLU	LYS	GLU	THR	GLU
	334	337	340	343	346	349	352	355	358	361	364	367	370	373	376	379	382	385	388	391	394	397	400	403	406	409
332	CCC	CTG	AGC	CAG	ATG	ACC	ANG	GAT	CTG	GAG	GAG	CTG	ANG	GCC	ANG	GTG	CAG	CCC	TAC	CTG	GAC	GAC	TTC	CAG	ANG	
GLY	LEU	ARG	GLN	GLU	MET	SER	LYS	ASP	LEU	GLU	GLU	VAL	LYS	ALA	LYS	VAL	GLN	PRO	TYR	LEU	ASP	ASP	PHE	GLN	LYS	
	412	415	418	421	424	427	430	433	436	439	442	445	448	451	454	457	460	463	466	469	472	475	478	481	484	487
410	AAG	TGC	CAG	GAG	ATG	GAG	CTC	TAC	CGC	CAG	ANG	GTG	GAG	CGC	CTG	CGC	GAA	GAG	CTC	CAG	GAG	GCC	GCC	CAG	ANG	
LYS	TRP	GLN	GLU	GLU	MET	GLU	LEU	TYR	ARG	GLN	LYS	VAL	GLU	PRO	LEU	ARG	ALA	GAG	LEU	GLN	GLU	GLY	ALA	ARG	GLN	
	490	493	496	499	502	505	508	511	514	517	520	523	526	529	532	535	538	541	544	547	550	553	556	559	562	565
488	ANG	CTG	CAG	CAG	CTG	CAG	ANG	CTG	AGC	CAG	CTG	GCC	GAG	GAG	ATG	CGC	GAG	CGC	GCC	GCC	GCC	GAT	GTG	GAG	GCC	
LYS	LEU	HIS	GLU	LEU	GLN	GLU	LYS	LEU	SER	PRO	LEU	GLY	GLU	GLU	MET	ANG	ASP	ARG	ARG	ALA	ANG	ALA	HIS	VAL	ASP	ALA
	568	571	574	577	580	583	586	589	592	595	598	601	604	607	610	613	616	619	622	625	628	631	634	637	640	643
566	CTG	CGC	ACG	CAT	CTG	CGC	CCC	TAC	AGC	GAG	CAG	CTG	CGC	CAG	CGC	TTC	GCC	GCC	GCC	CTT	GAG	GCT	CTC	ANG	GAG	ANG
LEU	ARG	THR	HIS	LEU	ALA	PRO	TYR	SER	ASP	SER	GLU	LEU	ARG	GLN	ARG	LEU	ALA	ALA	ALA	ARG	LEU	GLU	ALA	LEU	LYS	GLU
	646	649	652	655	658	661	664	667	670	673	676	679	682	685	688	691	694	697	700	703	706	709	712	715	718	721
644	GCC	GCC	GCC	AGA	CTG	GCC	GAG	TAC	CAG	GCC	ANG	GCC	ACC	GAG	CAT	CTG	AGC	ACC	CTC	AGC	CAG	ANG	GCC	ANG	CCC	GCC
GLY	GLY	ALA	ARG	LEU	ALA	GLU	TYR	HIS	ALA	LYS	ALA	THR	GLU	HIS	LEU	SER	THR	LEU	SER	GLU	LYS	ALA	LYS	PRO	ALA	
	724	727	730	733	736	739	742	745	748	751	754	757	760	763	766	769	772	775	778	781	784	787	790	793	796	799
722	CTC	GAG	GAC	CTC	GCC	CAG	GCC	CTG	CTG	CGC	CTC	GAG	ACC	TTC	ANG	CTC	AGC	TTC	CTC	AGC	GCT	CTC	CAG	GAG	TAC	
LEU	GLU	ASP	LEU	ARG	GLN	GLY	LEU	LEU	PRO	VAL	LEU	GLU	SER	PHE	LYS	VAL	SER	PHE	LEU	SER	ALA	LEU	GLU	GLU	TYR	
	802	805	808	811	814	817	820	823																		
800	ACT	ANG	ANG	CTC	ANG	ACC	CAG	TGA																		
THR	LYS	LYS	LEU	ASN	THR	GLN																				

Figure 4. Amino acid sequence derived from coding region of cloned pAl-3 cDNA.

Comparison of the signal peptide and pro-segment derived from pAl-3 to the rat sequence as determined by *in vitro* translation and automated radio-sequencing (16) reveals very strong homology between the two sequences (Figure 6). Of the 18 amino acids in the signal peptide, 13 are identical. The 5 residues which differ in the two sequences can all be accounted for by single-base changes. In the 6 amino acid pro-segment, only 4 were sequenced in the rat and three of these are identical to the human sequence. This homology of the pre- and pro-segments of apo A-I between the human and the rat is remarkable and the conservation of the neutral dipeptide Gln-Gln at the cleavage site of the pro-segment might indicate that it serves as part of the recognition site for a specific

FIRST POS'N		SECOND POS'N								THIRD POS'N	
		T		C		A		G			
T	PHE	1	SER	0	TYR	1	CYS	0	T		
	PHE	7	SER	3	TYR	6	CYS	0	C		
	LEU	0	SER	0	TAA	0	TGA	1	A		
	LEU	3	SER	0	TAG	0	TRP	5	G		
C	LEU	2	PRO	1	HIS	4	ARG	0	T		
	LEU	12	PRO	7	HIS	2	ARG	12	C		
	LEU	2	PRO	1	GLN	4	ARG	1	A		
	LEU	22	PRO	1	GLN	15	ARG	1	G		
A	ILE	0	THR	2	ASN	0	SER	0	T		
	ILE	0	THR	6	ASN	5	SER	13	C		
	ILE	0	THR	1	LYS	3	ARG	2	A		
	MET	4	THR	3	LYS	19	ARG	1	G		
G	VAL	0	ALA	4	ASP	5	GLY	0	T		
	VAL	1	ALA	12	ASP	11	GLY	9	C		
	VAL	0	ALA	1	GLU	4	GLY	1	A		
	VAL	14	ALA	6	GLU	26	GLY	1	G		

Figure 5. Codon usage in apo A-I mRNA as derived from the nucleotide sequence of pA1-3.

protease, as suggested by Gordon et al. (16). It will be interesting to study the nucleotide sequence homology of the entire secreted protein between these species when the rat sequence becomes available.

The amino acid sequence derived from pAl-3 is almost identical to that reported by Brewer et al. (25). The only three differences are at amino acids 34, 146 and 147 which were all predicted to be Glu rather than Gln as reported by Brewer et al. (25). While this study was in progress, Shoulders and Baralle (37) reported the cloning of apo A-I DNA. Only short clones were obtained and sequence information provided could only account for 16 amino acids of the protein. Since the comple-

Human	Met	Lys	Ala	Ala	⁵ Val	Leu	Thr	Leu	Ala	¹⁰ Val	Leu	Phe	Leu	Thr	¹⁵ Gly	Ser	Gln	Ala	²⁰ Arg	Glu	Phe	Trp	Gln	Gln
Rat	—	—	—	—	—	—	Ala	Val	—	Leu	Val	—	—	—	—	Cys	—	—	X	His	—	X	—	—

Figure 6. Comparison of the pre- and pro-segments of human and rat apo A-I. The human sequence is derived from the nucleotide sequence of pA1-3, the rat sequence is derived from in vitro translation and automated radio-sequencing as reported by Gordon et al. (16). The first 18 amino acids constitute the signal peptide, the last 6 amino acids form the pro-segment (16), -, identical residues, X, residue unknown.

tion of our DNA sequence, Breslow et al. reported the cloning and sequence analysis of a partial apo A-I cDNA (38). The Breslow apoA-I clone consists of the codons for carboxy-terminal portion of the molecule and is missing the coding sequence for the initial 117 (93 + 24 residues in the prepropeptide) amino acids. Furthermore, the preliminary partial sequence presented was derived from sequencing on a single strand only, with minimal overlap of fragments. Comparison of the pAl-3 sequence with the sequence of the carboxy-terminal portion of the molecule presented by Breslow et al. reveals four differences, two of which are in the coding, and two in the 3'-untranslated region. The differences are: (i) residue 146 is Gln (CAG) in the Breslow sequence and Glu (GAG) in our sequence, (ii) Tyr 192 is TAT in the Breslow sequence and TAC in our sequence, (iii) there is an A between two G's 3 bases downstream from the termination codon in the Breslow sequence but not in our sequence, and (iv) there are only two G's preceding the poly A tail in the Breslow sequence compared to the 3 G's in our sequence (such that it reads -TGGGA_n). The reasons for these discrepancies are not clear. To ensure that our sequence is correct, we have repeated the sequence analysis of regions that showed the discrepancies. We are confident about our sequence since in contrast to the Breslow sequence, we have sequenced both strands of DNA at these regions: for Glu 146, by 3'-labeling of the neighboring Dde I site and 5'-labeling of the Bgl I site next to it and for Tyr 192, the site was sequenced across in both directions in three different end-labeled restriction fragments (Bgl I, Dde I and Taq I). Tyr 192 (TAC) is also part of a restriction site for Rsa I which confirms that this region of the sequence is GTAC rather than GTAT. These regions of the sequence are also identical in the three separate clones (pAl-3, pAl-2 and pAl-4) we have studied. If the Breslow sequence is confirmed by sequence determination on the complementary strands, the observed differences (substitution, insertion and deletion) might represent true sequence heterogeneity of the mRNAs. Differences, or polymorphism, at the amino acid level have been observed in another apolipoprotein, apo E (39). If the observed "polymorphism" of apo A-I at the nucleotide level proves real, it would introduce another order of complexity to our understanding of sequence heterogeneity of the apolipoproteins. It should be noted that the differences between Breslow's sequence and our present sequence result in only a single amino acid difference (Glu vs Gln) in the carboxy-terminal two-

Figure 7. Dot-matrix analysis of apoA-I. ApoA-I sequence is represented both on the horizontal axis and the vertical axis such that any internally homologous segment will show up as a solid line at 45°. Left: Analysis of the DNA sequence-derived amino acid sequence of apoA-I at the single amino acid level. Right: Analysis of the nucleotide sequence of the coding region of pA1-3 at the dinucleotide level.

thirds of the secreted protein. Since Breslow *et al.* did not present any sequence data on initial 117 amino acids, we do not know whether there would be other differences in that region of the sequence.

ApoA-I is an interesting protein both in terms of its function and its structure. On the basis of the Baker amino acid sequence (23), McLachlan (40) and Fitch (41) contended that part of apoA-I consists of regular 11- or 22-residue repeat units. Their analyses were however not very stringent: instead of being classified according to their probable codons (i.e., identical or single-base changes), the amino acids were subgrouped according to their charge behavior (i.e., nonpolar, negatively-charged or positively charged). They postulated from such evidence that apoA-I might have evolved by internal gene duplication, and that the protein consists of a set of segments arising by simple unequal crossing over (40,41). We have re-examined both the DNA-derived amino acid sequence of apoA-I (Figure 7 left) and the nucleotide sequence of cloned apoA-I cDNA (pA1-3) (Figure 7 right) by dot-matrix analyses. It is evident from such analyses that there is no pattern of internally repeated segments in apoA-I, either at the amino acid or at the DNA level. We conclude that the so-called internal repeats are the result

Figure 8. Autoradiograph of Genescreen paper following hybridization to ^{32}P -pA1-3 DNA. Poly A RNAs were glyoxylated and fractionated on 1.5% agarose gel. They were transferred to Genescreen paper and hybridized to ^{32}P -pA1-3 DNA by the method of Thomas (35). Lane 1, human liver RNA (0.1 ug), lane 2, baboon intestine RNA (0.1 ug), lane 3, baboon kidney RNA (10 ug), lane 4, baboon spleen RNA (10 ug), lane 5, baboon liver RNA (0.1 ug), lane 6, baboon skeletal muscle RNA (10 ug).

of the presence of functional segments of amphipathic helices which may be important for lipid binding (42,43). In such functional helical units, the regular interspersions of nonpolar and charged amino acids is not unexpected (42,43).

HDL have been postulated to play a critical role in "reverse cholesterol transport" (5,6,44). Information on the major tissues that synthesize apo A-I is of special interest. Such information is especially important since Blue *et al.* (22) demonstrated translatable mRNA from the liver, intestine, kidney and skeletal muscle in the rooster. The authors postulated that the apo A-I synthesized *in situ* might be important in the local regulation of cholesterol efflux independent of the circulating HDL. While the available evidence in the mammalian species indicates that apo A-I is synthesized exclusively in the liver and the small intestine (21), it is important to determine if apo A-I mRNA sequences are present in significant amounts in other tissues in such animals. We have therefore assayed for apo A-I mRNA sequences in RNA extracted from various tissues of the baboon, since the baboon apo A-I mRNA cross-hybridizes extremely well to the corresponding human cDNA. Poly A RNA from these tissues were glyoxylated and frac-

tionated on a 1.5% agarose gel. They were transferred to Genescreen paper (New England Nuclear) by the method of Thomas (35), and hybridized to ^{32}P -pA1-3 insert DNA labeled by nick-translation. As is evident in Figure 8, mRNA from human liver, baboon liver and intestine hybridized strongly to the apo A-I cDNA probe with a mobility consistent with approximately 1100 nucleotides. mRNA preparations from skeletal muscle, kidney and spleen did not show any detectable apo A-I sequences by this technique despite the use of 100 fold higher mRNA concentrations. It can be estimated that if apo A-I mRNA sequences existed in those baboon tissues at all, their concentration would be less than .001% that present in the baboon liver or intestine.

It is thus evident that in the baboon, the liver and intestine constitute the major organs of synthesis of apo A-I mRNA. The cloned apo A-I cDNA can be used as a probe for the molecular basis of expression in such tissues. Furthermore, it can be used to study the basic defects in a number of genetically transmitted disorders involving apo A-I structure, function or expression (19,45-49).

ACKNOWLEDGEMENTS

We thank Drs. M. Moore, and Y.C. Lin-Lee for helpful discussion, and J. Colmenero, J.P. Moore and G. Clem for excellent technical assistance. The work described in this report was supported by Health, Education and Welfare grant HL-27341, a grant from the March of Dimes Birth Defects Foundation, Fellowship grant F32 HL06635 to P.C., and Training Grant T32AM07348.

REFERENCES

1. Barr, D.P., Russ, E.M., and Eder, H.A. (1951) *Am. J. Med.* 11:480-493.
2. Gofman, J.W., Young, W., and Tandy, R. (1966) *Circulation* 34:679-697.
3. Miller, G.J., and Miller, N.E. (1975) *Lancet* 1:16-19.
4. Gordon, T., Castelli, W.P., Horthland, M.C., Kannel, W.B., and Dawber, T.R. (1977) *Am. J. Med.* 62:707-714.
5. Glomset, J.A. (1968) *J. Lipid Res.* 9:155-167.
6. Glomset, J.A., and Norum, K.R. (1973) *Adv. Lipid Res.* 11:1-65.
7. Stein, O., and Stein, Y. (1973) *Biochim. Biophys. Acta.* 326:232-244.
8. Stein, O., Vanderhoek, J., and Stein, Y. (1976) *Biochim. Biophys. Acta.* 431:347-358.
9. Bates, S.R., and Rothblat, G.H. (1974) *Biochim. Biophys. Acta.* 360:38-55.

10. Miller, N.E., Nestel, P.J., and Clifton-Bligh, P. (1976) *Atherosclerosis* 23:535-547.
11. Havel, R.J., Goldstein, J.L., and Brown, M.S. (1980) in *Metabolic Control and Disease*, Bondy, P.K., and Rosenberg, L.E., Eds., pp. 393-494, W.B. Saunders, Philadelphia.
12. Jackson, R.L., Gotto, A.M., Stein, O., and Stein, Y. (1975) *J. Biol. Chem.* 250:7204-7209.
13. Fielding, C.J., Shore, V.G., and Fielding, P.E. (1972) *Biochem. Biophys. Res. Commun.* 46:1493-1498.
14. Fielding, C.J., Shore, V.G., and Fielding, P.E. (1972) *Biochim. Biophys. Acta.* 270:513-518.
15. Lin-Su, M.-H., Lin-Lee, Y.-C., Bradley, W.A., and Chan, L. (1981) *Biochemistry* 20:2470-2475.
16. Gordon, J.I., Smith, J.P., Andy, R., Alpers, D.H., Schonfeld, G., and Strauss, A.W. (1982) *J. Biol. Chem.* 257:971-978.
17. Tanaka, Y., Lin-Lee, Y.-C., Lin-Su, M.-H., and Chan, L. (1982) *Metabolism* 31:861-865.
18. Chan, L., Antalis, T., Lin-Su, M.-H., Cheung, P., and Lin-Lee, Y.-C. (1983) in *Atherosclerosis VI*, (in press), Springer-Verlag, New York.
19. Zannis, V.I., Lee, A.M., Lees, R.S., and Breslow, J.L. (1982) *J. Biol. Chem.* 257:4978-4986.
20. Noel, S.D., and Rubinstein, D. (1974) *J. Lipid Res.* 15:301-308.
21. Wu, A.-L., and Windmueller, H.G. (1979) *J. Biol. Chem.* 256:7316-7322.
22. Blue, M.-L., Ostapchuck, P., Gordon, J.S., and Williams, D.L. (1982) *J. Biol. Chem.* 257:11151-11159.
23. Baker, H.N., Gotto, A.M., Jr., and Jackson, R.L. (1975) *J. Biol. Chem.* 250:2725-2738.
24. Delahunty, T., Baker, N., Gotto, A.M., Jr., and Jackson, R.L. (1975) *J. Biol. Chem.* 250:2718-2724.
25. Brewer, H.B., Jr., Fairwell, T., LaRue, A., Ronan, R., Houser, A., and Bronzert, T.J. (1978) *Biochem. Biophys. Res. Commun.* 80:623-630.
26. Matteson, K.J., Su, T.-S., O'Brien, W.E., and Beaudet, A.L. (1982) *DNA* 1:201.
27. Itakura, K., and Riggs, A.D. (1980) *Science* 209:1401-1407.
28. Maxam, A.M., and Gilbert, W. (1980) in *Methods in Enzymology*, Moldave, K., and Grossman, L., Eds., Vol. 65, pp. 499-560, Academic Press, New York.
29. Wallace, R.B., Schold, M., Johnson, M.J., Dembek, P., and Itakura, K. (1981) *Nucleic Acids Res.* 9:3647-3656.
30. Gergen, J.P., Stern, R.H., and Wensink, P.C. (1979) *Nucleic Acids Res.* 7:2115-2136.
31. Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H., Itakura, K. (1981) *Proc. Natl. Acad. Sci.* 78:6613-6617.
32. Davies, R.W. (1982) in *Gel Electrophoresis of Nucleic Acids: A Practical Approach*, Rickwood, D., and Hames, B.D., pp. 117-172. IRL Press, Oxford.
33. Chan, L., Bradley, W.A., and Means, A.R. (1980) *J. Biol. Chem.* 255:10060-10063.
34. Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* 69:1408-1412.
35. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* 77:5201-5205.
36. Chan, L., Dugaiczky, A., and Means, A.R. (1980) *Biochemistry* 19:5631-5637.

37. Shoulders, C.C., and Baralle, F.E. (1982) *Nucleic Acids Res.* 10:4873-4882.
38. Breslow, J.L., Ross, D., McPherson, J., Williams, H., Kurnit, D., Nussbaum, A.L., Karathanasis, S.K., and Zannis, V.I. (1982) *Proc. Natl. Acad. Sci. USA* 79:6861-6865.
39. Rall, J.C., Weisgraber, K.H., Innerarity, T.L., and Mahley, R.W. (1982) *Proc. Natl. Acad. Sci. USA* 79:4696-4700.
40. McLachlan, A.D. (1977) *Nature* 267:465-466.
41. Fitch, W.M. (1977) *Genetics* 86:623-644.
42. Segrest, J.P., Jackson, R.L., Morrisett, J.D., and Gotto, A.M. (1974) *FEBS Lett.* 38:247-253.
43. Segrest, J.P., and Feldmann, R. (1977) *Biopolymers* 16:2053-2065.
44. Basu, S.K., Brown, M.S., Ho, Y.K., Havel, R.J., and Goldstein, J.L. (1981) *Proc. Natl. Acad. Sci. USA* 78:7545-7549.
45. Weisgraber, K.H., Bersot, T.P., and Mahley, R.W. (1980) *J. Clin. Invest.* 66:901-907.
46. Utermann, G., Feussner, G., Franceschini, G., Haas, J., and Steinmetz, A. (1982) *J. Biol. Chem.* 257:501-507.
47. Schaefer, E.J., Heaton, W.H., Wetzel, M.G., Brewer, H.B., Jr., (1982) *Arteriosclerosis* 2: 16-26.
48. Norum, R.A., Lakier, J.B., Goldstein, S., Angel, A., Goldberg, R.B., Block, W.D., Noffze, D.K., Dolphin, P.J., Edelglass, J., Bogorad, D.D., and Alaupovic, P. (1982) *New Engl. J. Med.* 306:1513-1519.
49. Weisgraber, K.H., Rall, S.C., Jr., Bersot, T.P., Mahley, R.W., Franceschini, G., and Sirtori, C.R. (1983) *J. Biol. Chem.* 258:2508-2513.