Vol. 148, No. 1, 1987 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1987 Pages 485-492

CHICKEN APOLIPOPROTEIN A-I: cDNA SEQUENCE, TISSUE EXPRESSION AND EVOLUTION

Lucy Byrnes†, Chi-Cheng Luo¶, Wen-Hsiung Li¶,
Chao-yuh Yang§, and Lawrence Chan†§

Departments of †Cell Biology and §Medicine,
Baylor College of Medicine, Houston, TX 77030
¶Center for Demographic and Population Genetics,
University of Texas, Houston, Texas 77030

Received September 4, 1987

Using an antibody against chicken apolipoprotein (apo) A-I, we identified multiple cDNA clones for the protein in two intestinal cDNA libraries in λgt11. The complete nucleotide sequence of chicken apoA-I cDNA was determined. The sequence predicts a mature protein of 240 amino acids, a 6-amino acid propeptide and an 18-amino acid signal peptide. Using a ^{32}P-cDNA probe, we detected the presence of apoA-I mRNA in 21 day old chicken intestine, liver, kidney, spleen, breast muscle and brain. The primary sequence of apoA-I contains numerous tandem repeats of 11 and 22 residues in a manner similar to the mammalian proteins. Our analysis of apoA-I sequences from human, rabbit, dog, rat, and chicken indicates that the rate of amino acid substitution is considerably faster in the rat lineage than in other mammalian lineages.

Apolipoprotein (apo) A-I is the major apoprotein in high density lipoproteins (HDL). There is an inverse relationship between the propensity to develop atherosclerosis and the plasma HDL and apoA-I concentrations (1,2). ApoA-I is a necessary co-factor for the activation of the enzyme lecithin-cholesterol acyltransferase (3,4), and is implicated in the transport of cholesterol from the peripheral tissues to the liver for disposal (3).

Avian apoA-I is an interesting molecule for a number of reasons. First, the protein is synthesized in numerous peripheral tissues (in addition to the liver and intestine), and serves as a model for cellular cholesterol efflux (5). Second, a pro-segment was found in avian proapoA-I, and the protein has been used as a model for proapoA-I processing (6). Third, avian apoA-I structure is of interest because another apolipoprotein, apoE, is undetectable in this species, though it is almost ubiquitous in distribution in mammals (7-9). There is speculation that avian apoA-I might incorporate some structural features of mammalian apoE. Finally, Fitch (10) and McLahlan (11) first noted

0006-291X/87 $1.50

the presence of multiple internal repeats in human apoA-1. Similar repeats have since been found in all the other mammalian apolipoproteins (12-16). The analysis of possible repeat structure in avian apoA-1 will be useful in our understanding of the structure-function relationships and evolution of the apolipoprotein multigene family.

In this study, we have determined the complete nucleotide sequence of chicken apoA-I cDNA. We identified internal repeats in the deduced amino acid sequence and inferred the rates of evolution of apoA-I among mammals. We have also studied the expression of the mRNA in various chicken tissues.

MATERIALS AND METHODS

ApoA-I cDNA Cloning and Sequencing

Chicken intestinal cDNA libraries were constructed in λgt11 by the method of Young and Davis (17). They were screened by using a goat monospecific anti-chick apoA-I antiserum (18). Plaque DNAs were purified as described (19). They were subcloned into the M13 vector, mp19, before sequencing by the dideoxynucleotide chain termination technique of Sanger et al. (20). Both strands were sequenced in entirety; both M13 primers and synthetic oligonucleotide primers were used in the reactions.

Northern Blot Analysis of Chicken total RNA

Total RNAs were extracted from 21-day old chicken intestine, liver, kidney, breast muscle, spleen and brain by the guanidinium isothiocyanate technique (19). They were electrophoresed on 1.5% formaldehyde-agarose gels, transferred to nitrocellulose membrane (Schleicher & Schuell), and hybridized to ^{32}P-dATP labeled nick-translated λcA1-2 cDNA insert as described by Tsai et al. (21). The membranes were exposed to Kodak X-ray film, XAR-5, for 18 hours.

Estimation of the Number of Amino Acid Substitutions

The number of amino acid substitutions between homologous protein sequences is estimated by Kimura's method (22), which makes a correction for multiple substitutions at the same residue site.

RESULTS AND DISCUSSION

ApoA-I cDNA Cloning and Sequencing

Ninety-seven apoA-I cDNA clones were identified in two chicken intestinal cDNA libraries in λgt11 after 4 x 10^5 recombinants were screened with a specific antiserum (18). The longest clone λcA1-2 was completely sequenced (Figure 1). It contains 963 nucleotides including 21 nucleotides in the 5'-untranslated region, 792 nucleotides of coding sequences, 150 nucleotides in the 3'-untranslated region. The putative polyadenylation signal AATAAA is 20 nucleotides from the 3' end of the clone. No polyA tail is identified. The coding

```
                                30                                  60                                  90
    gagacgccggttcacgcgaagATG AGA CGC GTG CTG GTG ACC CTC GCT GTG CTC TTC CTG ACG GCC ACC CAG GCC|CGC TCC TTC TGG CAC CAC|
                       Met Arg Gly Val Leu Val Thr Leu Ala Val Leu Phe Leu Thr Gly Thr Gln Ala Arg Ser Phe Trp Gln His

                               120                                 150                                 180
   GAT GAG CCC CAG ACG CCC CTG GAC CGC ATT CGG GAT ATG GTG GAC GTC TAC CTG GAG ACG GTG AAG GCC AGC GGC AAG GAT GCC ATC GCC
 1 Asp Glu Pro Gln Thr Pro Leu Asp Arg Ile Arg Asp Met Val Asp Val Tyr Leu Glu Thr Val Lys Ala Ser Gly Lys Asp Ala Ile Ala  30

                               210                                 240                                 270
   CAG TTC GAG TCC TCT GCT GTC GGC AAA CAG CTT GAC CTG AAG CTG GCT GAC AAC CTG GAC ACG CTG AGT GCC GCG GCT GCC AAG CTG CGT
31 Gln Phe Glu Ser Ser Ala Val Gly Lys Gln Leu Asp Leu Lys Leu Ala Asp Asn Leu Asp Thr Leu Ser Ala Ala Ala Ala Lys Leu Arg  60

                               300                                 330                                 360
   GAG GAC ATG GCT CCC TAC TAC AAG GAG GTG CGC GAG ATG TGG CTG AAG GAC ACC GAG GCT CTG CGT GCT GAG CTC AGC AAG GAC CTG GAG
61 Glu Asp Met Ala Pro Tyr Tyr Lys Glu Val Arg Glu Met Trp Leu Lys Asp Thr Glu Ala Leu Arg Ala Glu Leu Thr Lys Asp Leu Glu  90

                               390                                 420                                 450
   GAG GTG AAG GAG AAG ATC CGG CCC TTC CTG GAC CAG TTC TCC GCC AAG TGG ACG GAG GAG CTG GAG CAG TAC CGC CAG CGC CTG ACG CCC
91 Glu Val Lys Glu Lys Ile Arg Pro Phe Leu Asp Gln Phe Ser Ala Lys Trp Thr Glu Glu Leu Glu Gln Tyr Arg Gln Arg Leu Thr Pro 120

                               480                                 510                                 540
   GTG GCT CAG GAG CTG AAC GAG CTC ACC AAG CAG AAG GTC GAG CTG ATG CAG CCC AAC CTG ACC CCC GTG CCT GAG GAG CCC CGC GAT CGT
121 Val Ala Gln Glu Leu Lys Glu Leu Thr Lys Gln Lys Val Glu Leu Met Gln Ala Lys Leu Thr Pro Val Ala Glu Glu Ala Arg Asp Arg 150

                               570                                 600                                 630
   CTG CGT GGG CAC GTG GAG GAG CTG CGC AAG AAC CTG GCG CCA TAC AGC GAT GAG CTG CGG CAG AAG CTG AGC CAG AAG CTG GAG GAC ATC
151 Leu Arg Gly His Val Glu Glu Leu Arg Lys Asn Leu Ala Pro Tyr Ser Asp Glu Leu Arg Gln Lys Leu Ser Gln Lys Leu Glu Glu Ile 180

                               660                                 690                                 720
   CGT GAG AAG GGC ATC CCC GAG GCT TCC GAG TAC CAG GCC AAG GTG ATG GAG CAG CTC AGC AAC CTG CGT GAG AAG ATG ACG CCT CTG GTG
181 Arg Glu Lys Gly Ile Pro Gln Ala Ser Glu Tyr Gln Ala Lys Val Met Glu Gln Leu Ser Asn Leu Arg Glu Lys Met Thr Pro Leu Val 210

                               750                                 780                                 810
   CAG GAA TTC AGG GAG CGC CTC ACC CCC TAT GCT GAG AAC CTC AAG AAC CGC TTG ATC TCC TTC CTG GAT GAA CTC CAG AAG TCC GTG GCC
211 Gln Glu Phe Arg Glu Arg Leu Thr Pro Tyr Ala Glu Asn Leu Lys Asn Arg Leu Ile Ser Phe Leu Asp Glu Leu Gln Lys Ser Val Ala 240

                               848                                 888                                 928
   TGAgctgccggcccagggactgaccccaggccatgctggcctccgggagtcctcggggacccctccttaatctcctctccccccccgacccggagtccgtctcagctttgccattcttt
   ...

   tgtcaaaataaacatgacttaattattggagc
```

Figure 1. Nucleotide sequence of cloned chicken aopA-I cDNA and deduced amino acid sequence of the corresponding protein. The 2 vertical arrows mark the boundaries of the pro-segment. The amino acid residues are numbered by denoting the first residue of the mature peptide as number 1.

region of the sequence can be translated into 264 amino acids which include an 18-amino acid signal peptide, a 6-amino acid prosegment, and a 240-amino acid mature protein. The sequence of the propeptide ArgSerPheTrpGlnHis differs by 2 amino acids from that previously reported by Banerjee et al. (6) using radiolabeling techniques. Our sequence has been confirmed by direct sequencing of purified chicken plasma proapoA-I (Chao-yuh Yang, unpublished results).

Northern Blot Analysis and Tissue Expression of Chicken ApoA-1 mRNA

By Northern blot analysis, we observed the presence of ~ 1 kilobase long apoA-1 mRNA in total RNAs extracted from the 6 tissues examined, with the highest amount in the intestine, followed, in decreasing amounts, by the liver, kidney, spleen, breast muscle and brain, as reflected by the relative intensities of the bands on the X-ray (Figure 2). A previous study indicates the widespread synthesis of apoA-I by peripheral tissues in the chicken (5). Our data corroborate and extend this observation. This distribution of apoA-I

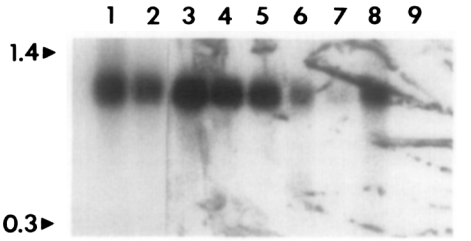

Figure 2. Autoradiograph of chicken RNA blots. Total or polyA RNAs were fractionated on 1.5% formaldelyde-agarose gels and transferred to nitrocellulose paper. Lanes 1 and 2, 1.0 μg and 0.5 μg intestinal polyA RNA, lanes 3 and 4, 5.6 μg and 2.8 μg liver RNA, lane 5, 3.8 μg kidney RNA, lane 4, 8.7 μg muscle RNA, lane 6, 4.8 μg brain RNA, lane 7, 12.4 μg speen RNA, and lane 8, 20 μg yeast tRNA. The numbers on the left denote the size in kilobases of RNA standards run on the same gel (RNA ladder, Bethesda Research Laboratories). For estimation of RNA concentration, total intestinal RNAs were used in other experiments for comparison of the intensity of the bands.

in the chicken contrasts that in the mammal in which apoA-I mRNAs are detected almost exclusively in the liver and intestine (23). The essentially ubiquitous presence of the mRNA is consistent with the postulated role of apoA-I in cholesterol efflux from these organs (5). On the other hand, other alternative function(s) must be postulated for apoA-I production in chicken brain because the protein probably does not cross the blood-brain barrier.

Structural Features and Internal Repeats of Chicken ApoA-I

An interesting structural feature among mammalian apoA-I, A-II, A-IV, C-I, C-II, C-III and E is that they share a common block of 33 residues in the mature peptide region (16). The block consists of 3 repeats of 11 residues and is located at the end of the third exon of the genes encoding these proteins. Figure 3a shows an alignment of chicken apoA-I with human, dog, rabbit and rat A-I. Each boxed column represents hydrophobic, acidic, or basic residues that have been well conserved in evolution among mammalian apolipoproteins A-I, A-II, A-IV, C-I, C-II, C-III and E. Obviously, most of these residues have also been conserved in chicken apoA-I. In addition, many other residues have also been conserved in all five sequences. Since birds and mammals have separated for about 270 million years, the conservation of this common block suggests that it is structurally important for the function of apoA-I.

Another common structural feature among the above mentioned mammalian apolipoproteins is that the part of the mature peptide encoded by exon 4 (exon

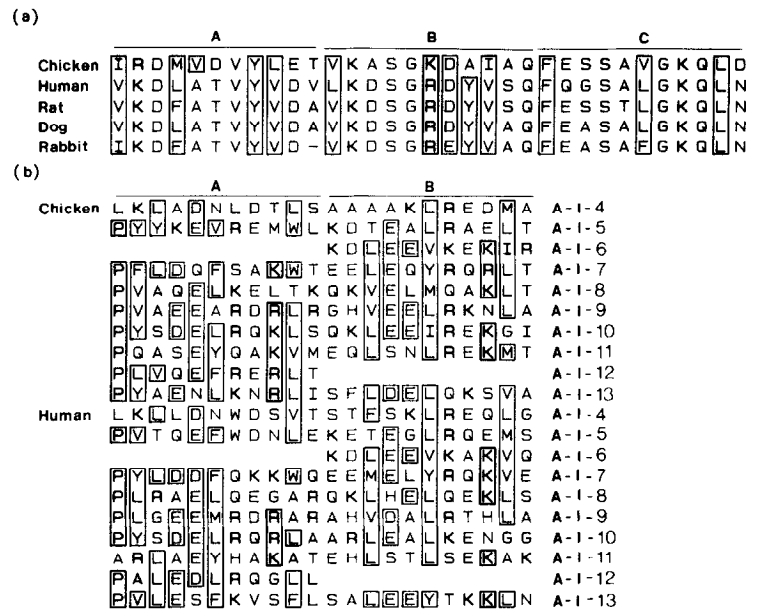

Figure 3. Internal repeats in chicken apoA-I in single letter codes. In the figure, D and E are acidic residues, R and K are basic residues, M,V,L,I,F,Y and W are hydrophobic residues. P is considered separately while the remaining residues (G,A,S,T,N,Q,H and C) are indifferent. (a) The last 33 amino acids of exon 3 can be divided into 3 repeats of 11 amino acids. The boxed columns indicate that the amino acids with the particular characteristics occupy that position in the majority (>50%) of the sequences of all the following apolipoproteins (i.e. A-I, A-II, A-IV, C-I, C-II, C-III and E; see Luo et al (16) for details). (b) The repeats in the last exon are 22-mers, each of which is made up of 2 11-mers, and the other repeats are 11-mers. Criteria for including particular residues in a box are the same as for exon 3. We refer to the first repeat in exon 4 as the fourth repeat in each case; the repeats are denoted as A-I-4, A-I-5, etc.

3 in apoA-IV) is almost completely made up of 22-residue repeats, each of which is a tandem array of two 11-mers (12,13,16). As can be seen from Figure 3b, this is also true for chicken apoA-I. As in the case of human apoA-I, most of the 22-mers in chicken apoA-I start with proline --- only one exception in chicken apoA-I, but two exceptions in human apoA-I. Moreover, in both groups A and B, columns 3, 6 and 10 are mostly occupied by hydrophobic residues, columns 4 and 5 by acidic residues, column 9 by basic residues, while column 11 does not have a predominant pattern. This conservation of repeat pattern in the two distantly related species suggests that the repeats may serve some important structural function for this major HDL apolipoprotein.

Sequence Comparison and Evolution of A-I Apolipoproteins

The chicken, human, rat, rabbit and dog apoA-I sequences can be aligned easily (partial alignment shown in Figure 3). It is obvious that chicken

Table 1. Proportion of Amino Acid Differences (above diagonal) and Estimated Number of Amino Acid Substitutions Per Site (below diagonal) Between Vertebrate ApoA-I Sequences

Species	Chicken	Human	Rat	Rabbit	Dog
Chicken		0.53	0.56	0.53	0.51
Human	0.75±0.07		0.38	0.22	0.18
Rat	0.83±0.07	0.48±0.05		0.36	0.34
Rabbit	0.75±0.07	0.25±0.03	0.45±0.05		0.19
Dog	0.72±0.07	0.20±0.03	0.41±0.05	0.21±0.03	

apoA-I is homologous to the mammalian proteins. We note that the pattern of internal repeats has been well conserved in mammals and chicken (Figure 3), though the homology is below 50% (Table 1).

We have also compared the chicken apoA-I sequence to the human apoE sequence (data not shown). We found that the homology between the two sequences was not any greater than between other mammalian apoA-I and apoE. Further, the consensus sequence in the receptor binding site of apoE, Arg-X-X-ArgLysArg-X-X-Arg/Lys, was missing from the avian apoA-I sequence. Thus, despite the similarity in tissue distribution between avian apoA-I and mammalian apoE, the former is not an apoE-like molecule.

As noted by Luo et al. (16) apoA-I is not a conservative protein. It has evolved considerably faster than β-globin, which evolves at the average rate of 35 mammalian proteins (24). Interestingly, as is evident from Table 1, the rate of amino acid substitutions seems to be faster in the rat lineage than in the human, rabbit and dog lineages. In the Table, the numbers (K) of amino acid substitutions per site between sequences were estimated from Kimura's formula (22). If we use chicken apoA-I as a reference, then we find that the K value between chicken and rat apoA-I is 0.83 - 0.75 = 0.08 higher than that between chicken and human apoA-I. Since the K value between rat and human apoA-I is 0.48, the K value in the rat lineage is (0.48 + 0.08)/2 = 0.28 and that in the human lineage is 0.48 - 0.28 = 0.20. Therefore, the rate in the rat lineage is 0.28/0.20 = 1.4 times higher than that in the human lineage. However, this is probably an underestimate because Kimura's formula would tend

to underestimate the K values between distantly related species. A better reference would be dog apoA-I, for dog is generally thought to be more distantly related to human than rat is (25). Using dog apoA-I as a reference and the same computation procedure, one can show that the K values along the human and rat lineages are 0.135 and 0.345, respectively, and that the rate in the rat lineage is ~ 2.6 times higher than that in the human lineage. This estimate is similar to Datta et al.'s (26) estimate that apoC-III has evolved 3 times faster in the rat lineage than in the human lineage. On the other hand, both apoA-I and apoC-III seem to have evolved at about the same rates in the dog and human lineages (Table 1 and [26]). It is not clear why apoA-I and apoC-III should have evolved much faster in the rat lineage than in other mammalian lineages. When additional sequences become available, it will be interesting to see whether this difference in the rate of evolution between the rat and other mammalian lineages also extends to the other apolipoproteins.

ACKNOWLEDGMENTS

We thank Donald McDonnell for construction of the cDNA libraries and S. Mascola for typing the manuscript. This work was supported by National Institutes of Health grants (HL-16512 to L.C. and GM-30998 to W.H.L.).

REFERENCES

1. Barr, D.P., Russo, E.M., & Eder, H.A. (1951) Am. J. Med. 11, 480-493.
2. Miller, G.J., & Miller, N.E. (1975) Lancet i, 16-19.
3. Glomset, J.A. (1968) J. Lipid Res. 9, 155-167.
4. Fielding, C.J., Shore, V.G., & Fielding, P.E. (1972) Biochem. Biophys. Res. Commun. 46, 1493-1498.
5. Blue, M.-L., Ostapchuck, P., Gordon, J.S., & Williams, D.L. (1982) J. Biol. Chem. 257, 11151-11159.
6. Banerjee, D., Mukherjee, T.K., & Redman, C.M. (1985) J. Cell Biol. 101, 1219-1226.
7. Basu, S.K., Brown, M.S., Ho, Y.K., Havel, R.J., & Goldstein, J.L. (1981) Proc. Natl. Acad. Sci. USA 78, 7545-7549.
8. Blue, M.-L., Williams, D.L., Zacker, S., Khan, S.A., & Blum, C.B. (1983) Proc. Natl. Acad. Sci. USA 80, 283-287.
9. Elshourbagy, N.S., Liao, W.S., Mahley, R.W., & Taylor, I.M. (1984) Proc. Natl. Acad. Sci. USA 82, 203-207.
10. Fitch, W.M. (1977) Genetics 86, 623-644.
11. McLachlan, A.D. (1977) Nature 267, 465-466.
12. Karathanasis, S.K., Zannis, V.I., & Breslow, J.L. (1983) Proc. Natl. Acad. Sci. USA 80, 6147-6151.
13. Boguski, M.S., Elshourbagy, N., Taylor, J.M., & Gordon, J.I. (1984) Proc. Natl. Acad. Sci. USA 81, 5021-5025.

14. Karathanasis, S.K., Yunis, I., & Zannis, V.I. (1986) Biochemistry 25, 3962-3970.
15. Rajavashisth, T.B., Kapstein, J.S., Reue, K.L., & Lusis, A.J. (1985) Proc. Natl. Acad. Sci. USA 82, 8085-8089.
16. Luo, C.C., Li, W.H., Moore, M.N., & Chan, L. (1986) J. Mol. Biol. 187, 325-340.
17. Young, R.A., & Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194-1198.
18. Jackson, R.L., Lin, H.Y., Chan, L., & Means, A.R. (1976) Biochim. Biophys. Acta 420, 342-349.
19. Maniatis, T., Fritsch, Z.F., & Sanbrook, J. (1982) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Sanger, F., Coulson, A., Barrell, B., Smith, A., & Roe, B. (1980) J. Mol. Biol. 143, 161-178.
21. Tsai, M.-J., Tsai, S.Y., Baez, M., Simmen, F.A., Scott, M., Sargan, D.R., Elbrecht, A., & O'Malley, B.W. (1987) In: Laboratory Methods Manual for Hormone Action and Molecular Endocrinology. 11th edition. Ed. Schrader, W.T., & O'Malley, B.W., Chapter 13, pp. 13.1-13.6.
22. Kimura, M. (1983) The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge, England.
23. Cheung, P., & Chan, L. (1983) Nucl. Acids Res. 11, 3703-3715.
24. Li, W.-H., Wu, C.-I., & Luo, C.C. (1985) Mol. Biol. Evol. 2, 150-174.
25. Romer, A.S. (1966) Vertebrate Palentology. University of Chicago Press, Chicago.
26. Datta, S., Li, W.-H., Ghosh, I., Luo, C.-C., & Chan, L. (1987) J. Biol. Chem. 262, 10588-10593.