Volume 224, number 2, 261–266 November 1987

# The complete amino acid sequence of proapolipoprotein A-I of chicken high density lipoproteins

Chao-yuh Yang*, Zi-Wei Gu*, Wolfgang Patsch*, Shi-ai Weng*, Tae W. Kim* and Lawrence Chan*+

*Departments of *Medicine* and +*Cell Biology, Baylor College of Medicine and The Methodist Hospital, Houston, TX 77030, USA*

Received 21 September 1987

The complete amino acid sequence of proapolipoprotein (proapo) A-I of chicken high density lipoproteins was determined by sequencing overlapping peptides produced by trypsin, *S. aureus* V8 protease, and cyanogen bromide cleavage. There are 240 amino acid residues in mature chicken apoA-I. By direct sequence analysis of a cyanogen bromide peptide, we also determined the sequence of a 6-amino-acid prosegment which is present at approx. 10% the molar amount of the mature peptide in chicken plasma. Sequence comparison among apoA-I from chicken, human, rabbit, dog and rat, and secondary structure analysis indicate that while the degree of sequence homology is only moderate (< 50% between chicken and man), there is good conservation of apoA-I secondary structure, especially in the N-terminal two-thirds of the protein in these widely separated species.

Lipoprotein; Atherosclerosis; Evolution; Lipid binding; Protein processing

## 1. INTRODUCTION

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 [1–3]. ApoA-I is implicated in the transport of cholesterol from the peripheral tissues to the liver for disposal [4,5]. The protein was found to be a necessary co-factor for the activation of the enzyme lecithin-cholesterol acyltransferase [6,7]. Mammalian apoA-I is secreted from the liver as a mixture of proapoA-I and mature apoA-I [8–10]. ProapoA-I can be processed to the mature protein by an enzyme present in plasma HDL, mesenteric lymph, and lymph chylomicrons [11].

In the chicken, previous studies suggest that HDL apoA-I is present exclusively as the mature protein [12]. The proapoA-I was thought to be efficiently processed to its mature form in the Golgi apparatus and the prosegment was reported to have an identical sequence to the human apoA-I propeptide [12]. Furthermore, chicken apoA-I is synthesized in numerous peripheral tissues and serves as a model for cellular cholesterol efflux [13]. For these reasons, the primary structure of chicken apoA-I and its propeptide is of interest in our understanding of apoA-I function and metabolism.

Correspondence address: C. Yang, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## 2. MATERIALS AND METHODS

### 2.1. *Chicken apoA-I purification, enzymatic cleavage and peptide isolation*

Laying White Leghorn hen plasma was obtained from Texas Animal Specialties. ApoA-I was purified from the plasma high density lipoproteins as described [14]. The protein (5 mg/ml, 0.1 M

*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/87/$3.50 

$NH_4HCO_3$, pH 8.0) was completely digested with trypsin or *S. aureus* V8 protease (enzyme/substrate, 1:50) at room temperature. Enzymatically or chemically cleaved peptides were purified on either a Waters or a Beckman HPLC system. The peptides were initially separated on a Vydac $C_{18}$ reversed-phase HPLC column (250 × 4.6 mm) at 50°C at a flow rate of 1.5 ml/min. The tryptic and *S. aureus* V8 protease digests were separated using 0.1% trifluoroacetic acid (TFA) in water as the starting buffer A, and 0.08% TFA in 95% acetonitrile and 5% water as buffer B. A linear gradient of buffer B was increased from 0 to 60% at a rate of 1%/min. Peptides were detected by their absorbance at 220 nm. The eluent under each peak was collected manually in a tube and dried by a Speedvac (Savant Instruments, NY). The purity of each peak fraction was checked by the manual sequencing method. The pure peptides were then sequenced; the partially purified fractions were subjected to rechromatography with a Hypersil $C_{18}$ column (250 × 4.6 mm) using a phosphate buffer system. Buffer A was 0.005 M sodium phosphate, pH 6.0; buffer B was 10% buffer A plus 90% acetonitrile [15].

### 2.2. *Amino acid and sequence analyses*

Chicken apoA-I or pure peptides were first hydrolyzed in the gas phase with 6 M HCl in a Waters hydrolysis apparatus for 1 h at 150°C. After hydrolysis, the samples were dried and prepared for amino acid analysis using precolumn derivatization with phenyl isothiocyanate [16].

Sequence determination of shorter peptides (<20 residues) was performed according to Chang et al. [17] using a modified Edman reagent, dimethylaminoazobenzene isothiocyanate, and TLC [17,18] or HPLC [19] methods. The sequences of the protein and larger peptides were determined by a gas-phase sequenator with an online PTH amino acid analyzer from Applied Biosystems.

### 2.3. *C-terminal analysis with carboxypeptidase Y*

The protein (2 nmol) was dissolved in 0.1 N pyridine acetate, pH 6.0, and then digested with 100 μl of Pierce immobilized carboxypeptidase Y (10–14 U/ml) at room temperature for 1, 2, 4 and 24 h. To check for autodigestion by carboxypeptidase Y, the same amount of enzyme was incubated and sampled simultaneously. Samples were dried and coupled with PITC for PTC amino acid analysis using a Waters Pico-Tag amino acid analyzer.

### 2.4. *Secondary structure predictions*

Algorithms for secondary structure predictions were those described by Eisenberg et al. [20] for helical hydrophobic moment, and Kyte and Doolittle [21] for relative hydrophobicity.

## 3. RESULTS

The amino acid composition of chicken apoA-I reveals that there is no cysteine in chicken apoA-I. Therefore, the structural analysis of the protein using enzymatic or chemical cleavage was performed directly.

### 3.1. *Tryptic peptides* (fig.1)

The tryptic peptide mixtures of chicken apoA-I were purified by a Waters HPLC system on a Vydac $C_{18}$ column with the TFA buffer system. The fractions which required rechromatography with the sodium phosphate buffer system for purification were T27, T36, T1, T20, T29 and T30. 36 of the tryptic peptides were purified and characterized. These peptides covered most of the chicken apoA-I sequence except for four dipeptides, namely TK (position 100,101), QR (position 122,123), ER (position 221,222) and NR (position 232,233) which might have eluted within the void volume. Rechromatography of fractions coming out of 16% B with the sodium phosphate buffer system showed that T1, T20, T29 and T30 had been purified.

### 3.2. S. aureus *V8 protease peptides and cyanogen bromide peptides* (fig.1)

Most of the *S. aureus* V8 peptides were purified to homogeneity by chromatography on HPLC with the TFA buffer system on a Vydac $C_{18}$ column. 28 peptides were sequenced. They provided sequence information which allowed the alignment of the tryptic peptides. 5 mg of the protein were used for this cleavage with 50-times (by wt) of cyanogen bromide. Peptides CN1 ⟶ CN4 provided additional sequence and overlaps for chicken apoA-I.

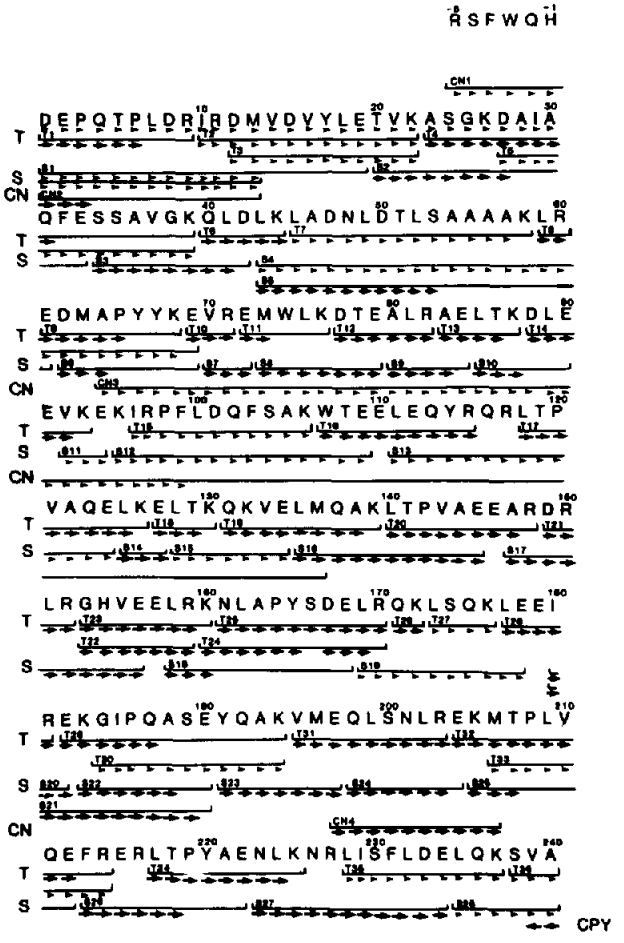


Fig.1. Complete amino acid sequence of chicken apoA-I. T, tryptic peptides; S, peptides from digestion with *S. aureus* V8 protease; CN, cyanogen bromide peptides; , gas-phase sequencing; →, modified Edman manual sequencing; CPY, carboxypeptidase Y analysis.

### 3.3. *Alignment of peptides and complete amino acid sequence of chicken proapoA-I*

The amino acid sequence of chicken apoA-I was determined primarily by aligning overlapping peptides of the three digestions. Based on the results of cyanogen bromide cleavage, chicken plasma apoA-I consists of 90% mature apoA-I and 10% proapoA-I. The sequence of the propeptide is Arg-Ser-Phe-Trp-Gln-His. The N-terminus of mature chicken apoA-I was confirmed as Asp by direct sequencing of the first 30 residues of the intact protein. The C-terminus of chicken apoA-I was identified as Ala by carboxypeptidase Y cleavage of the whole protein. Chicken apoA-I has 240 amino acid residues with the sequence shown in fig.1. This sequence is similar to the primary structure deduced from a recently published cDNA sequence, except for Glu-124 which was predicted to be Lys in the DNA-deduced sequence [22]. We have also sequenced a complete chicken apoA-I cDNA clone which reads GAG for codon 124 predicting a Glu in this position (Byrnes, L. and Chan, L., unpublished).

## 4. DISCUSSION

### 4.1. *Amino acid sequence of chicken proapoA-I*

We have determined the complete amino acid sequence of chicken proapoA-I by the direct sequencing of overlapping peptide fragments. The 6-amino-acid propeptide sequence determined in this study, Arg-Ser-Phe-Trp-Gln-His, differs by two residues from that of Banerjee et al. [12], Arg-His-Phe-Trp-Gln-Gln, which was determined by radiolabeling and immunochemical techniques [12]. Interestingly, the latter sequence is identical to that in man [23]. Our propeptide sequence has also been confirmed by that predicted from cloned apoA-I cDNA (Byrnes, L. and Chan, L., unpublished). The N-terminal 20 amino acid residues of the mature chicken apoA-I have been reported previously by two laboratories [12,24]. Our sequence matches that presented by Shackelford and Lebherz [24], but differs from that by Banerjee et al. [12] in that Tyr-17 was reported as Thr-17 in the latter sequence. Whether the apparent discrepancies between our data and those of Banerjee et al. [12] in the prosegment and in the N-terminal region represent true sequence heterogeneity is unclear at present.

### 4.2. *ProapoA-I in chicken plasma*

A previous study on chicken apoA-I synthesis suggests that in the liver, intracellular immunoreactive apoA-I is a mixture of proapoA-I and mature apoA-I, whereas the plasma protein is exclusively in the form of the mature protein [12]. Our current analysis by direct peptide sequencing indicates that proapoA-I constitutes approx. 10% of the plasma protein. The different observations between the two studies could be explained by: (i) the more sensitive nature of the direct sequencing of CN peptides and/or (ii) the proapoA-I being

derived mainly from the intestine; and/or (iii) the different source of animals used, though leghorns were used in both laboratories. We conclude that at least in our animals, the cellular processing of proapoA-I is not 100% efficient. There is, however, no information on the relative stabilities of chicken plasma mature apoA-I vs proapoA-I which would influence the steady-state concentrations of these protein species.

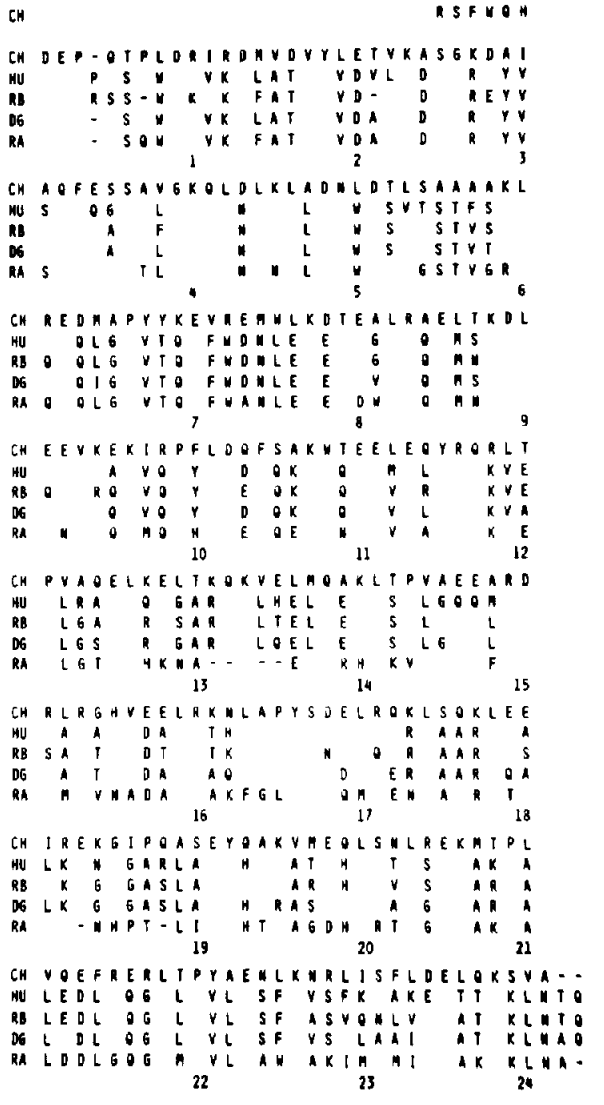

Fig.2. Comparison of the amino acid sequence of chicken apoA-I with apoA-I sequences from other vertebrates. CH, chicken; HU, human [30]; RB, rabbit [31]; DG, dog [32]; RA, rat [33].

In both human and rat proapoA-I, the 6-amino-acid prosegment ends with a C-terminal Gln-Gln dipeptide [8–10,25], unusual amino acids for protein precursors that are processed proteolytically. In the chicken, the Gln-Gln pair is replaced by Gln-His, indicating that the requirement for Gln in the cleavage site is not absolute. Furthermore, the prosegment is efficiently processed in the chicken liver ([12] and present study), whereas in the rat and human, it is almost totally uncleaved as the nascent polypeptide chain is secreted from the liver [8,26,27]. Whether the difference in structures of the substrate (proapoA-I) accounts for some of the difference in processing efficiency is unclear. Comparative studies using the mammalian vs chicken enzyme(s) and mammalian vs chicken proapoA-I as substrates will shed some light on the mechanism of proapoA-I processing in the liver and intestine.

The presence of a prosegment and of the specific processing enzyme(s) in the chicken indicates that the processing pathway evolved before the divergence of the avian species, some 270 million years ago. While the exact function of proapoA-I

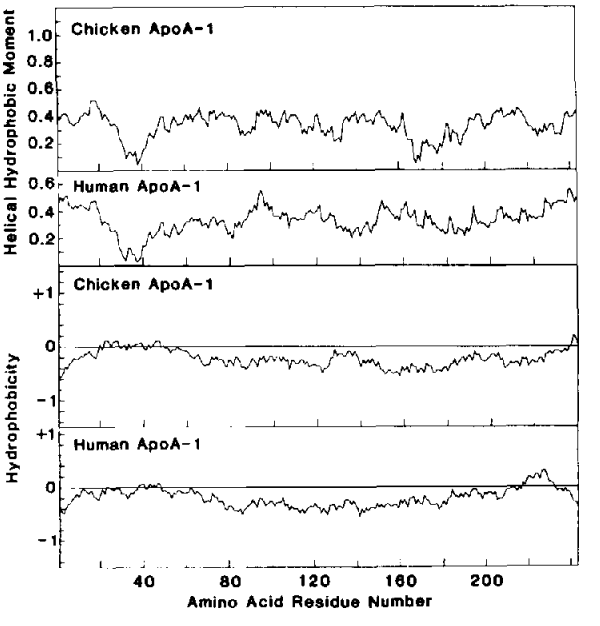


Fig.3. Secondary structure analysis of chicken and human apoA-I. The helical hydrophobic moment was determined by the algorithm of Eisenberg et al. [20], and the hydrophobicity according to Kyte and Doolittle [21].

is unknown, the conservation of the pathway suggests that it may be important to the overall function of apoA-I in HDL metabolism.

### 4.3. *Sequence alignment and structure of chicken apoA-I*

The amino acid sequence of chicken apoA-I can be readily aligned to those of the human, rabbit, dog and rat proteins (fig.2). As expected, the degree of homology between avian and mammalian apoA-I is less than that among the mammalian proteins themselves. ApoA-I is an interesting molecule because the human protein contains tandem repeats of 11 and 22 amino acid residues [28,29]. Despite the considerable divergence of the avian sequence, similar repeats of 11- and 22-mers can also be identified in the chicken (not shown, see also [22]). When predictive algorithms are used to analyze human and chicken apoA-I, it is evident that the secondary structures are generally well conserved between these species as indicated by the similar patterns in the helical hydrophobic moment [20] and hydrophobicity profiles [21] (fig.3). The conservation appears to be better in the N-terminal two-thirds of the protein. The avian and mammalian lineages have diverged for ~270 million years, and the sequence homology between the human and avian proteins is less than 50%. In the face of this degree of divergence in primary sequence, it is remarkable that there is such a well-conserved secondary structure.

## ACKNOWLEDGEMENTS

We thank S. Mascola for assistance in the preparation of this manuscript. This work was supported in part by a National Research and Demonstration Center on Arteriosclerosis, grant HL-27341, individual NIH grants (HL-16512 to L.C. and HL-34457 to W.P.), and grants from the American Heart Association (870863 to C.-Y.), and the Muscular Dystrophy Association to the Jerry Lewis Neuromuscular Disease Research Center at Baylor College of Medicine.

## REFERENCES

[1] Barr, D.P., Russo, E.M. and Eder, H.A. (1951) Am. J. Med. 11, 480–493.
[2] Miller, G.J. and Miller, N.E. (1975) Lancet i, 16–19.
[3] Heiss, G., Johnson, N.J., Reiland, S., Davis, C.E. and Tyroler, H.A. (1980) Circulation 62 (suppl.IV), 116–136.
[4] Glomset, J.A. (1968) J. Lipid Res. 9, 155–167.
[5] Glomset, J.A. and Norcum, K.R. (1973) Adv. Lipid Res. 11, 1–65.
[6] Fielding, C.J., Shore, V.G. and Fielding, P.E. (1972) Biochem. Biophys. Res. Commun. 46, 1493–1498.
[7] Fielding, C.J., Shore, V.G. and Fielding, P.E. (1972) Biochim. Biophys. Acta 270, 513–518.
[8] Gordon, J.I., Smith, D.P., Andy, E., Alpers, D.H., Schonfeld, G. and Strauss, A.W. (1982) J. Biol. Chem. 257, 971–975.
[9] Stoffel, W., Kruger, E. and Deutzmann, R. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 227–237.
[10] Stoffel, W., Bode, C. and Knyrim, K. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 439–445.
[11] Edelstein, C., Gordon, J.I., Toscas, K., Sims, H.F., Strauss, A.W. and Scanu, A.M. (1983) J. Biol. Chem. 258, 11430–11433.
[12] Banerjee, D., Mukherjee, T.K. and Redman, C.M. (1985) J. Cell Biol. 101, 1219–1226.
[13] Blue, M.-L., Ostapchuck, P., Gordon, J.S. and Williams, D.L. (1982) J. Biol. Chem. 257, 11151–11159.
[14] Jackson, R.L., Lin, H.-Y., Chan, L. and Means, A.R. (1976) Biochim. Biophys. Acta 420, 342–349.
[15] Yang, C.Y., Pauly, E., Kratzin, H. and Hilschmann, N. (1981) Hoppe-Seyler's Z. Physiol. Chem. 362, 1131–1146.
[16] Yang, C.Y. and Sepulveda, F.I. (1985) J. Chromatogr. 346, 413–416.
[17] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[18] Yang, C.Y. (1979) Hoppe-Seyler's Z. Physiol. Chem. 360, 1673–1675.
[19] Yang, C.Y. and Wakil, S.J. (1984) Anal. Biochem. 137, 54–57.
[20] Eisenberg, D., Weiss, R.M. and Tergwillager, T.C. (1982) Nature 299, 371–374.
[21] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[22] Rajavashisth, T.B., Dawson, P.A., Williams, D.L., Shackelford, J.E., Lebherz, H. and Lusis, A.J. (1987) J. Biol. Chem. 262, 7058–7065.
[23] Cheung, P. and Chan, L. (1983) Nucleic Acids Res. 11, 3703–3715.
[24] Shackelford, J.E. and Lebherz, H.G. (1983) J. Biol. Chem. 258, 7175–7180.
[25] Poncin, J.E., Martial, J.A. and Gielen, J.E. (1984) Eur. J. Biochem. 140, 493–498.

[26] Zannis, V.I., Karathanasis, S.K., Keutmann, H.T., Goldberger, G. and Breslow, J.L. (1983) Proc. Natl. Acad. Sci. USA 80, 2574.
[27] Stoffel, W., Bode, C. and Knyrim, K. (1983) Hoppe-Seyler's Z. Physiol. Chem. 364, 439–445.
[28] Fitch, W.M. (1977) Genetics 86, 623–644.
[29] McLachlan, A.D. (1977) Nature 267, 465–466.
[30] Brewer, H.B. jr, Fairwell, T., LaRue, A., Ronan, R., Hauser, A. and Bonzert, T.J. (1978) Biochem. Biophys. Res. Commun. 80, 623–630.
[31] Yang, C.Y., Yang, T.M., Pownall, H.J. and Gotto, A.M. jr (1986) Eur. J. Biochem. 160, 427–431.
[32] Chung, H., Randolph, A., Reardon, I. and Heinrikson, R.L. (1982) J. Biol. Chem. 257, 2961–2967.