*Proc. Natl. Acad. Sci. USA*
Vol. 81, pp. 5021–5025, August 1984
Biochemistry

# Rat apolipoprotein A-IV contains 13 tandem repetitions of a 22-amino acid segment with amphipathic helical potential

**(full-length cDNA cloning/lipid-binding domains/lecithin:cholesterol acyltransferase activation/gene structure and evolution)**

MARK S. BOGUSKI*, NABIL ELSHOURBAGY†, JOHN M. TAYLOR†, AND JEFFREY I. GORDON*‡

Departments of *Biological Chemistry and ‡Medicine, Washington University School of Medicine, St. Louis, MO 63110; and †Gladstone Foundation Laboratories, Department of Physiology, Cardiovascular Research Institute, University of California, San Francisco, CA 94140

*Communicated by David M. Kipnis, May 2, 1984*

ABSTRACT **Apolipoprotein (apo) A-IV is a 46,000 $M_r$ protein that is a major component of rat high density lipoproteins and chylomicrons. We have isolated, from a rat intestinal cDNA library, a full-length apo A-IV clone of 1423 base pairs and determined its nucleotide sequence. The 1173-nucleotide coding region specifies a protein of 391 amino acids, which includes a 20-amino acid signal peptide. The portion of the cDNA sequence representing the mature plasma protein contains a 66-nucleotide subsequence that is repeated at least 13 times. Although each repeated unit shows some sequence variation, base changes in comparably positioned codons generally conserve the chemical type, if not the identity, of the corresponding amino acids. Nine of the thirteen 22-amino acid repeat units in the derived protein sequence begin with proline and most of these docosapeptides are predicted to have a high content of α-helix according to Chou–Fasman rules. When hydration potentials of individual residues are considered, many of the helices are shown to be amphipathic and may thus constitute lipid-binding domains with the ability to activate lecithin:cholesterol acyltransferase. The structure and organization of the repeat units of rat apo A-IV bear a striking similarity to a corresponding repeated sequence block in human apo A-I. This finding raises the possibility of close tandem linkage of the apo A-I and A-IV genes and suggests unequal crossing-over as a mechanism for the evolution of these genes.**

Apolipoprotein (apo) A-IV is one of the major proteins associated with rat chylomicrons and high density lipoproteins (HDL) (1) and is also a constituent of human and canine chylomicrons (2). Unlike other apolipoproteins, about 50% of plasma apo A-IV does not appear to be associated with classical lipoprotein density classes (3, 4). Furthermore, apo A-IV may undergo extensive redistribution among circulating lipoproteins and a free pool (4). In this regard, it has been suggested that the cholesteryl ester content of HDL may regulate apo A-IV binding to plasma HDL (5). The metabolic function of apo A-IV remains unknown.

Processing of the primary translation product of apo A-IV mRNA differs from the other apo A class proteins, apo A-I and A-II. Both of these HDL-associated proteins are initially synthesized as prepropolypeptides and therefore undergo both co- and post-translational proteolysis (6–8), whereas the initial product of apo A-IV translation contains a signal peptide but not a prosegment (9). Nevertheless, the signal peptides of rat apo A-I and A-IV have a remarkable degree of sequence homology (9).

Similarities in the amino acid sequences of the mature plasma forms of apo A-I, A-II, C-I and C-III have led to the speculation that all of these proteins evolved from a common ancestral gene (10). If this hypothesis is true, the apolipoproteins would constitute a gene family. Supporting this view is the recent demonstration that the apo A-I and C-III genes are closely linked in humans (11).

The purpose of the present study was to determine the structure of apo A-IV with the hope that this information would provide some insight into the function and evolution of this protein. Our results demonstrate that apo A-IV has multiple potential lipid-binding domains and is very closely related to apo A-I, much more so than any of the other apolipoproteins. We suggest that apo A-IV may be important in lipid binding and also that this protein may be a significant activator of lecithin:cholesterol acyltransferase (LCAT). Furthermore, it is likely that the apo A-IV gene is linked to the gene for apo A-I.

## MATERIALS AND METHODS

**Isolation of cDNA Clones.** mRNA was prepared from the proximal small intestine of Sprague–Dawley rats as described (6). Intestinal cDNA libraries were constructed according to Land *et al.* (12). Plasmids containing apo A-IV cDNA were identified by hybridization selection and translation of the mRNA, followed by immunoprecipitation of the translation product with specific antibodies (13).

**DNA Sequencing.** Restriction fragments of two apo A-IV cDNA clones were subcloned into various M13 vectors as described in Fig. 1. Single-stranded phage DNAs served as templates for sequencing by the dideoxy chain-termination method (14). Recent improvements in gel electrophoresis technology (15) were implemented for analysis of the sequencing reactions with the exception that deoxyadenosine 5′-[α-$^{32}$P]triphosphate (Amersham) was used instead of deoxyadenosine 5′-[α-[$^{35}$S]thio]triphosphate. The large (Klenow) fragment of *Escherichia coli* DNA polymerase I was used in the sequencing reactions, except when it was necessary to sequence through homopolymer tails. In this case, reverse transcriptase was used with appropriate alterations in the ratios of deoxy- to dideoxynucleotides (16).

**Computer Analysis.** A semi-automated sequence gel reader was used for data collection. This gel reader and assorted nucleic acid and protein sequence analysis software were obtained from CompuGene (St. Louis, MO). Additional programs were supplied by Dean Goddette (Department of Biological Chemistry, Washington University) and David Lipman (17). The latter programs were modified for use on a VAX 11/780 (Digital Equipment).

## RESULTS

**Apo A-IV DNA Sequence and Derived Protein Sequence.** Several cloned apo A-IV cDNAs were isolated from our intestinal libraries. One of these plasmids, pAIV303 (Fig. 1), contained a full-length transcript of apo A-IV mRNA that had been estimated, by RNA transfer blot analysis, to be 1.7

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HDL, high density lipoprotein(s); apo, apolipoprotein(s); LCAT, lecithin:cholesterol acyltransferase.

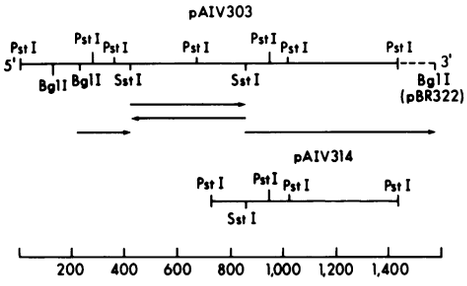


FIG. 1. Restriction map of apo A-IV cDNA clones and summary of sequence analysis. The bottom scale is in nucleotides. Initially, a bank of *Pst* I fragments from each clone was prepared by subcloning a total *Pst* I digest of each plasmid into M13 vector MWB2348 (18). Sequencing was carried out on recombinant phage DNA derived from randomly selected plaques and continued until a complete set of *Pst* I subclones was identified and sequenced on both strands. Restriction sites identified in this first round of sequencing were used to develop maps of both original cDNA clones. Specific fragments of pAIV303 were then subcloned into M13 mp10 or mp11 and sequences across all *Pst* I junctions confirmed by sequencing in the direction of the arrows.

kilobases in size, including poly(A) (13). Clone pAIV303, and an independent isolate pAIV314, were sequenced completely.

Analysis of the nucleotide sequence (Fig. 2) revealed only one long open reading frame, beginning at a methionine codon (nucleotide 1) and ending at the stop codon TGA (nucleotide 1174). The cDNA insert also contained 5′ and 3′ nontranslated regions of 91 and 158 nucleotides, respectively. The latter region contains the polyadenylylation signal, A-A-T-A-A-A, beginning at nucleotide 1318.

The derived amino acid sequence contains 391 residues with a calculated $M_r$ of 44,465. This value is in good agreement with the 46,000 $M_r$ estimated for rat apo A-IV by gel electrophoresis (1). We had obtained previously (9) partial amino acid sequences of the signal peptide and the amino terminus of the mature plasma protein (underlined in Fig. 2). These data agree with the sequence derived from the cDNA. Furthermore, the amino acid composition of the derived protein sequence is in accord with previously published compositions (1, 2).

**Identification and Analysis of Internal Repeats.** Inspection of the rat apo A-IV amino acid sequence revealed an unusually even distribution of proline residues in the molecule (marked by asterisks in Fig. 2). When the amino acid sequence was arranged in blocks starting, whenever possible, with proline, an 11.5-cycle, 22-amino acid repeat unit appeared (residues 82–330 in Fig. 3). Residues 33–81 are included in Fig. 3 according to the analysis described below. This repeated sequence block was conspicuous in its correspondence to a 22-amino acid repeated sequence block found previously to exist in human apo A-I (20, 21). The tandemly arrayed docosapeptides in apo A-IV are not exact duplications. However, most amino acid substitutions are conservative in that they generally lead to replacement of one residue with another of similar physical-chemical properties. Of the nonconservative substitutions, about half are replacements by the small, neutral amino acids glycine, serine, and threonine.

These repeated docosapeptides in apo A-IV were quantitatively characterized by the method of Kubota *et al.*, in which protein sequence homology can be expressed in terms of correlation coefficients (22). For each amino acid, a value is assigned based upon a combination of quantitative physi-

```
      -81       -71       -61       -51       -41       -31       -21       -11        -1        10        20        30
ATCCTCACAGCGACACAGTGAACCATTTCTTCTGACGCTAGGAAACATCCAGTGTAGCTGAGGCTGTCCCAACCCAGTGAGGAGCCCAGGATGTTCCTGAAGGCTGTGGTGCTGACCGTG
                                                                                    MetPheLeuLysAlaValValLeuThrVal
                                                                                    ----------------------- ---
                                                                                     ^signal peptide  .  .  .  .

       40        50        60        70        80        90       100       110       120       130       140       150
GCCCTGGTGGCCATCACCGGGACCCAGGCTGAGGTCACTTCCGACCAGGTGGCCAATGTGATGTGGGACTACTTCACCCAGCTAAGCAACAATGCCAAGGAGGCTGTGGAACAACTGCAG
AlaLeuValAlaIleThrGlyThrGlnAlaGluValThrSerAspGlnValAlaAsnValMetTrpAspTyrPheThrGlnLeuSerAsnAsnAlaLysGluAlaValGluGlnLeuGln
------------ ------------ ---------------------------------- ------ ---
. . . . . ^mature plasma protein                ^beginning of repeated sequences

      160       170       180       190       200       210       220       230       240       250       260       270
AAGACAGATGTCACTCAACAGCTCAATACCCTCTTCCAGGACAAACTTGGGAACATTAACACCTATGCCGATGACCTGCAGAACAAGCTGGTGCCCTTCGCCGTTCAACTGAGTGGACAT
LysThrAspValThrGlnGlnLeuAsnThrLeuPheGlnAspLysLeuGlyAsnIleAsnThrTyrAlaAspAspLeuGlnAsnLysLeuValProPheAlaValGlnLeuSerGlyHis
                                                                                                          *

      280       290       300       310       320       330       340       350       360       370       380       390
CTAACCAAGGAAACAGAGAGGGTGAGGGAAGAGATCCAGAAGGAGCTCGAGGACCTACGGGCCAACATGATGCCCCATGCCAACAAAGTGAGCCAGATGTTCGGGGACAACGTGCAGAAG
LeuThrLysGluThrGluArgValArgGluGluIleGlnLysGluLeuGluAspLeuArgAlaAsnMetMetProHisAlaAsnLysValSerGlnMetPheGlyAspAsnValGlnLys
                                                                                *

      400       410       420       430       440       450       460       470       480       490       500       510
TTGCAGGAACACCTGAGGCCCTATGCCACGGACCTGCAAGCTCAGATCAACGCACAGACCCAGGATATGAAACGCCAGCTGACCCCCTACATCCAGCGCATGCAGACCACAATACAAGAC
LeuGlnGluHisLeuArgProTyrAlaThrAspLeuGlnAlaGlnIleAsnAlaGlnThrGlnAspMetLysArgGlnLeuThrProTyrIleGlnArgMetGlnThrThrIleGlnAsp
          *                                                                                       *

      520       530       540       550       560       570       580       590       600       610       620       630
AATGTGGAAAACCTGCAGTCCTCCATGGTGCCCTTTGCCAACGAGCTAAAGGAAAAGTTTAACCAGAATATGGAAGGGCTCAAGGGGCAACTAACCCCCCGTGCCAACGAGCTGAAGGCC
AsnValGluAsnLeuGlnSerSerMetValProPheAlaAsnGluLeuLysGluLysPheAsnGlnAsnMetGluGlyLeuLysGlyGlnLeuThrProArgAlaAsnGluLeuLysAla
                    *                                                                                 *

      640       650       660       670       680       690       700       710       720       730       740       750
ACGATCGACCAGAACCTGGAAGACCTGCGCAGCAGACTGGCCCCTCTCGCGGAGGGTGTGCAGGAAAAACTCAACCATCAGATGGAAGGCTTGGCCTTCCAGATGAAGAAGAACGCTGAG
ThrIleAspGlnAsnLeuGluAspLeuArgSerArgLeuAlaProLeuAlaGluGlyValGlnGluLysLeuAsnHisGlnMetGluGlyLeuAlaPheGlnMetLysLysAsnAlaGlu
                                  *

      760       770       780       790       800       810       820       830       840       850       860       870
GAGCTCCATACCAAAGTCTCCACAAATATCGACCAGCTGCAGAAGAATCTGGCCCCGCTGGTGGAAGATGTGCAAAGCAAGCTGAAAGGCAACACGGAAGGACTGCAGAAGTCTCTGGAA
GluLeuHisThrLysValSerThrAsnIleAspGlnLeuGlnLysAsnLeuAlaProLeuValGluAspValGlnSerLysLeuLysGlyAsnThrGluGlyLeuGlnLysSerLeuGlu
                                                *

      880       890       900       910       920       930       940       950       960       970       980       990
GACCTGAACAAGCAGCTGGACCAGCAGGTGGAGGTGTTCCGGCGTGCTGTGGAGCCCCTGGGGGATAAGTTCAACATGGCTCTGGTGCAGCAGATGGAGAAGTTCAGGCAGCAGCTGGGC
AspLeuAsnLysGlnLeuAspGlnGlnValGluValPheArgArgAlaValGluProLeuGlyAspLysPheAsnMetAlaLeuValGlnGlnMetGluLysPheArgGlnGlnLeuGly
                                                *                                                 end of repeated sequences^

     1000      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100      1110
TCCGATTCGGGGGACGTGGAAAGCCACTTGAGCTTCCTGGAGAAGAACCTGAGGGAAAAGGTCAGCTCCTTTATGAGCACCCTGCAAAAAAAGGGGAGCCCAGACCAGCCCCTAGCCCTC
SerAspSerGlyAspValGluSerHisLeuSerPheLeuGluLysAsnLeuArgGluLysValSerSerPheMetSerThrLeuGlnLysLysGlySerProAspGlnProLeuAlaLeu

     1120      1130      1140      1150      1160      1170      1180      1190      1200      1210      1220      1230
CCCCTCCCGGAGCAGGTTCAGGAACAGGTCCAGGAGCAGGTGCAGCCCAAACCTCTGGAGAGCTGAGCTGTCCCTGTGTCCTCGGCCCATCACAGCAGCAGACACCTGCCCTGCCCCACC
ProLeuProGluGlnValGlnGluGlnValGlnGluGlnValGlnProLysProLeuGluSer

     1240      1250      1260      1270      1280      1290      1300      1310      1320      1330      1340      1350
ACCTGTCTGTCACTCTGTTCCCAAGCACTTCTCGTACCAGCTTGAGGACACATGTCCTGTGGGTGACGATACCTCCTCTCGTTACTCAATAAAGCATCTGAG
                                                                                  Poly(A)
                                                                                   signal
```

FIG. 2. Nucleotide sequence of rat apo A-IV cDNA. Portions of the amino acid sequence determined previously (9) are underlined. The core repeated sequence block, as described in the text, extends from nucleotides 244 to 777. Proline residues in the repeated sequence block are marked with an asterisk (*).

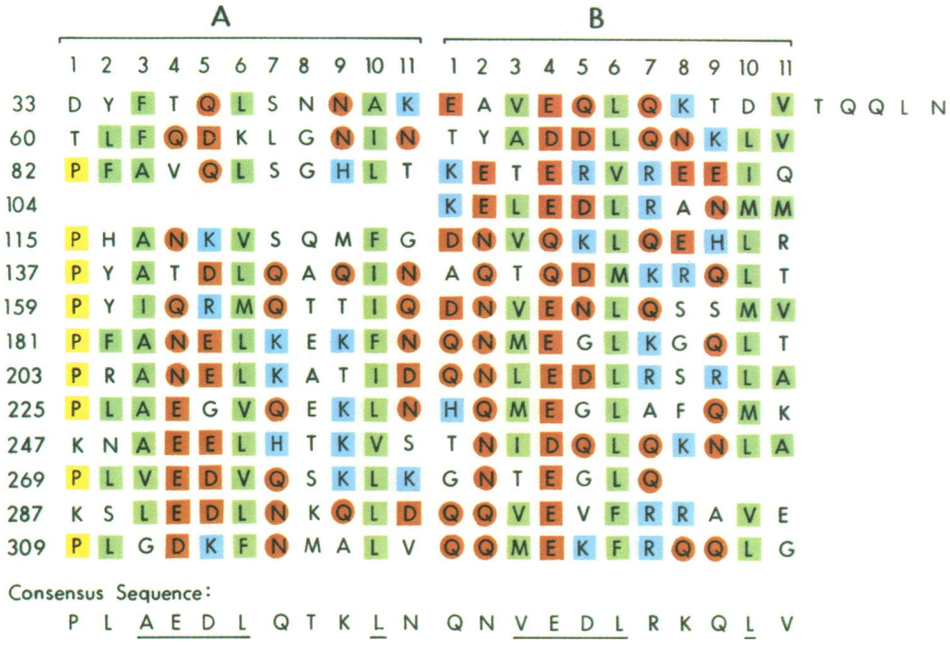


FIG. 3. Optimal alignment of 22-amino acid repeat units in rat apo A-IV. Numbers along the left margin are residue numbers for the preprotein. Gaps have been introduced to optimize homology based on the results in Fig. 4*A*. Amino acids have been assigned to groups based on their hydropathy index (19) and charge. Hydrophobic residues (A, alanine; M, methionine; F, phenylalanine; L, leucine; V, valine; I, isoleucine) have hydropathy indices from 1.8 to 4.5 and are represented by green. Acidic residues (E, glutamic acid; D, aspartic acid) have the same hydropathy index (−3.5) as their amide derivatives (Q, glutamine; N, asparagine) and all four amino acids are colored red to indicate this fact. However, since glutamine and asparagine are uncharged, they are represented by circles. The basic amino acids, H, histidine; K, lysine; and R, arginine, have hydropathy indices of −3.2, −3.9, and −4.5, respectively, and are indicated in blue. G, glycine; T, threonine; and S, serine have hydropathy indices near zero and are uncolored. Column A8 has been left uncolored since it lacks a predominant type of residue. (A threonine was used in the consensus sequence for the purpose of the correlation analysis in Fig. 4*A*.) Also uncolored are residues that are inconsistent with the predominant type in their column. Tyrosine (Y), despite its aromatic side chain, has a hydropathy index of −1.3, which places it among the more hydrophilic residues in the Kyte–Doolittle scale (19). Proline (P) is colored yellow to emphasize the regularity of its ocurrence in the first position of the repeat units as well as its absence from other locations. Portions of the consensus sequence are underlined to indicate highly conserved regions and also to emphasize the tandem symmetry in adjacent 11-amino acid segments.

cal, chemical, and evolutionary properties such as hydrophobicity, α-helical potential, and mutability (22). To apply this method to apo A-IV, we first aligned the 11 docosapeptides that were obvious repeats on visual inspection and derived a 22-amino acid consensus sequence based upon the most frequent amino acid, or type of amino acid, per column (Fig. 3). This 22-mer was then used as a comparison sequence to study the magnitude and extent of the periodicity in pre-apo A-IV. The results are displayed in Fig. 4*A*. To assess the specificity of the technique, the 20-amino acid signal peptide of pre-apo A-IV was also used, as a negative control (Fig. 4*B*).

It is apparent from Fig. 4*A* that statistically significant peaks occur with a period of 22 residues. This periodicity is rigorously conserved in that portion of the sequence beginning at residue 93 and ending at residue 268. In portions of the sequence flanking this block, insertions or deletions have disrupted the harmonic structure. For example, the repeat units beginning at residues 33 and 269 consist of 27 and 18 amino acids, respectively, rather than 22 (Figs. 3 and 4*A*). It is interesting to note that the discontinuity occurring at residue 104 is a displacement of precisely 11 amino acids. This finding in addition to the occurrence of prominent secondary oscillations with a period of 11 (Fig. 4*A*) demonstrate that each 22-mer is actually a tandem array of two related 11-mers. The two halves of each of these tandem symmetry groups are labeled A and B in Fig. 3. Fitch has noted that, for apo A-I (19), most 11-unit segments are more similar to the 11-unit segment once removed than to the adjacent 11-mer. This is also true for apo A-IV. Fig. 4*B* demonstrates that the Kubota method is specific and that the signal sequence of apo A-IV is unrelated to the repeated docosapeptides.

The human apo A-I sequence has an 11-amino acid interruption in the basic pattern of 22-amino acid repeat units, just as apo A-IV does at position 104 (Fig. 3). Because of this matching pattern, it is possible to assign a one-to-one correspondence for residues 82–259 of pre-apo A-IV (Fig. 3) to residues 66–243 in the mature plasma form of apo A-I. Thus, the repeated sequence block of rat apo A-IV is expanded relative to that of human apo A-I in both the amino- and carboxyl-terminal directions—that is, there are additional repeats flanking the core-repeated sequence block in apo A-IV that are not present in apo A-I. Independent analysis, based upon optimal alignments (17) of the complete nucleotide sequences, resulted in the same conclusion. In the core-repeated sequence block (nucleotides 244–777 in Fig. 2), rat apo A-IV and human apo A-I (23) are 46% homologous at the DNA sequence level.

**Structure and Function Predictions.** Amphipathic α-helices have been implicated as the structures responsible for lipid binding (24, 25) and, in some cases, LCAT activation (26, 27). Thus, we employed the method of Kubota *et al.* to search the apo A-IV sequence for regions of physical-chemical correlation with several natural and synthetic amphipathic sequences with known abilities to bind phospholipids (24, 25) and activate LCAT (26). Although all test sequences were homologous to some degree, homologies of the greatest magnitude and extent were obtained with a synthetic docosapeptide designed to be an LCAT activator (26). We found multiple overlapping regions of significant homology (correlation coefficients of 0.4 or greater) that generally corresponded to the repeated sequences as defined in Fig. 3 (data not shown). Each of these regions was examined in greater detail as described below.

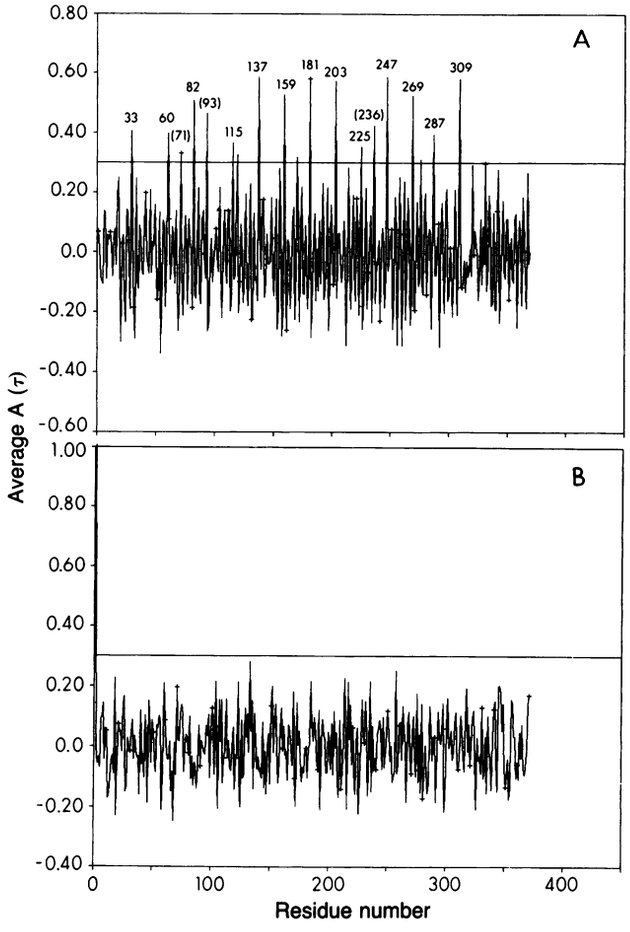


FIG. 4. Correlograms for test sequences with pre-apo A-IV. Numbers above the peaks correspond to the residue numbers in Fig. 3. Numbers in parentheses indicate homologs displaced 11 residues from the primary oscillation. The small horizontal lines perpendicular to and intersecting some peaks are spacing marks that occur every 5 residues. The horizontal line that intersects the vertical axis at 0.3 represents a level of statistical significance based on the results of Kubota *et al.* for several proteins of demonstrated periodicity (22). (*A*) Consensus sequence from Fig. 3. (*B*) apo A-IV signal peptide (see Fig. 1).

Rules for empirical prediction of protein secondary structure (28) were used to determine if the repeated sequences in apo A-IV had the potential to form α-helical conformations. apo A-IV was predicted to have an overall α-helical content of about 53%, which is in good agreement with results obtained from circular dichroic spectra of the protein (1). Results for the region spanning residues 181–202, identified as a potential "LCAT activator" sequence, are presented diagrammatically in Fig. 5*A*. It is evident that this docosapeptide has a predominant α-helical potential. That the predicted structure is not entirely α-helix is not of great concern since it has been demonstrated experimentally that binding to a phospholipid surface increases the content of α-helix in lipid-binding proteins and peptides (30).

α-Helices can be examined for amphipathic characteristics after representation as helical "wheels" (29). The Edmundson wheel representation of the docosapeptide in Fig. 5*A* is shown in Fig. 5*B*. (For the sake of clarity, only the first 18 residues, representing five turns of an α-helix, are shown.) The structure conforms to the definition of amphipathicity since it has a hydrophobic face (Phe-182, Leu-186, Leu-197, Phe-190, Ala-183, Met-194) and an acidic face (Glu-195, Glu-188, Glu-185) separated from the hydrophobic face by basic residues (Lys-189, Lys-187, Lys-198). The definition of an amphipathic structure (24) does not specify an axial length

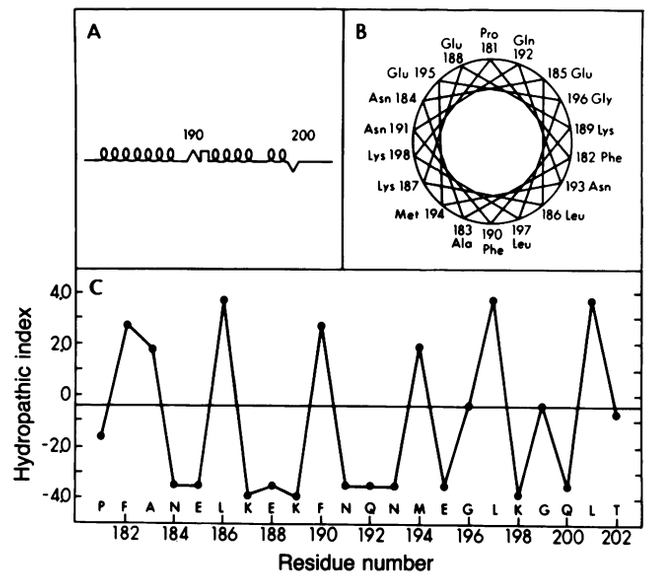


FIG. 5. Amphipathic helix analysis of the rat apo A-IV docosapeptide repeat unit beginning at residue 181 (Figs. 3 and 4*A*). (*A*) Prediction of secondary structure based upon Chou–Fasman (28) rules. Ω, α-Helix; ⎍, β-turn; ⌄, β-pleated sheet. Each symbol represents one amino acid residue. Flat lines indicate areas where no prediction could be made. (*B*) Edmundson wheel (29) diagram. Side chain positions along the helix are projected onto a plane. Successive residues are separated by 100° of arc yielding 3.6 residues per turn. (*C*) Kyte–Doolittle (19) plot of hydropathy values.

for the helix. The average length of an α-helical domain in globular proteins corresponds to 11 residues or three helix turns (31). Thus, as McLachlan has pointed out (21), repeat units of 11 or 22 residues are particularly favorable in helical domains.

Recent studies (32) have indicated that visualizing helices as Edmundson wheels may lead to bias in interpreting these structures as being amphipathic. For this reason, and also because of the high frequency of asparagine and glutamine residues where acidic ones usually predominate (Figs. 3 and 5*B*), we applied the method of Kyte and Doolittle (19) to analyze such structures quantitatively. In this technique, a hydropathy index is assigned to each amino acid based upon its free energy of transfer from water to ethanol and also upon the statistical frequency of its position in proteins of known crystal structure. When the sequence spanning residues 181–202 is plotted as a function of hydropathic index (Fig. 5*C*), a pattern that alternates regularly between hydrophobicity and hydrophilicity is clearly evident. Similar alternating patterns were also observed throughout the entire repeated sequence domain of the apo A-IV molecule (data not shown).

In the hydropathy scale of Kyte and Doolittle (19), the index values for aspartate, glutamate, asparagine, and glutamine are identical, all having a value of −3.5. This observation is very revealing in terms of the high frequency of asparagine and glutamine residues in charged regions of the repeat units (Figs. 3 and 5*B*). This suggests that, in terms of structural constraints on the evolution of these sequences, the conservation of relative hydrophobicity has been at least as important as the conservation of charge. It is possible that, in an amphipathic structure, aspartate and glutamate are functionally interchangeable with their amide derivatives.

## DISCUSSION

The striking similarities between the structures of human apo A-I and rat apo A-IV have major implications for the metabolic function of apo A-IV as well as for the structure, location, and evolution of the apo A-IV gene.

Human apo A-I is known to contain repeated sequences (10, 20, 21). An 11-residue periodicity, related to the apo A-I repeat unit, has also been noted in human apo A-II, C-II, and C-III by Barker and Dayhoff, who proposed that all four of these proteins diverged from a single ancestral gene (10). McLachlan, however, has argued against this view (21) and his analysis implies that any similarity between apo A-I and C-III is the result of convergent evolution. It should be noted in this context that the apo A-I and apo C-III genes have recently been shown to be closely linked in humans (11). Fitch has proposed that the repeated segments in apo A-I are the result of a series of intragenic unequal crossovers and has constructed a phylogeny of the segments based on maximum parsimony (20).

Our data on rat apo A-IV leave little doubt that apo A-I and apo A-IV have evolved from divergence of a common ancestral gene. Although individual residues in the core repeated sequence blocks have diverged, the patterns of amino acids are identical. That the core repeated sequence blocks are 46% homologous at the nucleotide level is impressive considering that these are different proteins from different species. We propose that a common A-I/A-IV precursor gene arose by *intragenic* duplication of a 66-nucleotide unit and that subsequent *intergenic* duplication created two separate genes in which the number of internal repeats became either reduced or further amplified. This hypothesis would account for the close correspondence of core repeated sequence blocks in proteins of different size. If apo A-I and apo A-IV are the result of divergence following an unequal crossover, these genes may be closely linked, in tandem. The repeated sequences in human apo A-I constitute the fourth and final exon of the apo A-I gene (23), which raises the possibility of evolution through exon duplication.

In 1974, Segrest *et al.* hypothesized that amphipathic (amphiphilic) helices are functional domains responsible for lipid binding (24). Numerous experimental studies have provided support for this hypothesis (reviewed in ref. 30). For example, Sparrow *et al.* synthesized a series of model amphipathic peptides and demonstrated that the ability to bind phospholipid depends upon the degree of hydrophobicity of the nonpolar face (25). We have found multiple regions in the apo A-IV molecule with the potential to assume amphipathic helical conformations. However, since a major portion of circulating apo A-IV is unassociated with lipoprotein particles, other conformational parameters must contribute to lipid binding.

Kaiser, Kezdy, and co-workers have provided evidence that indicates that the LCAT-activating activity of apo A-I is a function of its multiple 22-residue amphipathic segments (26, 27). From studies employing model synthetic docosapeptides designed to be prototypic amphipathic $\alpha$-helices, they have concluded that the presence of a large enough amphipathic structure is necessary and sufficient for LCAT activation. However, this conclusion may be unwarranted since the LCAT-activating activities reported were a fraction of those achieved with apo A-I and also because apo A-II, C-II, and C-III all have amphipathic segments yet lack significant LCAT-activating activity (33). Nevertheless, we suggest that apo A-IV may be an effective activator of LCAT based upon its striking structural similarity to apo A-I and the multiplicity of its amphipathic segments that may interact cooperatively (34). If this hypothesis proves to be true, apo A-IV may represent an alternative pathway for LCAT activation. Recent studies have provided evidence that the apo A-IV distribution in rat plasma can be influenced by the LCAT reaction (5). DeLamatre *et al.* have shown that apo A-IV moves from the lipoprotein-free fraction to HDL in association with cholesterol esterification (5). Thus, if apo A-IV is able to participate in the LCAT reaction, it could regulate its own binding to lipoprotein particles.

We thank Dean Goddette and Dr. Alan Templeton for helpful discussions, Dr. David Lipman for providing computer programs, and Dr. A. W. Strauss for continued support. This work was supported in part by National Institutes of Health Grants AM 30292 and AM 31615. M.S.B. is supported by Medical Scientist Training Program Grant GM 07200. J.M.T. is the recipient of Research Career Development Award CA00872 from the National Cancer Institute. J.I.G. is a Fellow of the John A. and George L. Hartford Foundation.

1. Swaney, J. B., Braithwaite, F. & Eder, H. A. (1977) *Biochemistry* **16,** 271–279.
2. Weisgraber, K. H., Bersot, T. P. & Mahley, R. W. (1978) *Biochem. Biophys. Res. Commun.* **85,** 287–292.
3. Beisiegel, U. & Utermann, G. (1979) *Eur. J. Biochem.* **93,** 601–608.
4. Fidge, N. H. (1980) *Biochim. Biophys. Acta* **619,** 129–141.
5. DeLamatre, J. G., Hoffmeier, C. A. & Roheim, P. S. (1983) *J. Lipid Res.* **24,** 1578–1586.
6. Gordon, J. I., Smith, D. P., Andy, R., Alpers, D. H., Schonfeld, G. & Strauss, A. W. (1982) *J. Biol. Chem.* **257,** 971–978.
7. Gordon, J. I., Budelier, K. A., Sims, H. F., Edelstein, C., Scanu, A. M. & Strauss, A. W. (1983) *J. Biol. Chem.* **258,** 14054–14059.
8. Stoffel, W., Kurger, E. & Deutzmann, R. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **264,** 227–237.
9. Gordon, J. I., Smith, D. P., Alpers, D. H. & Strauss, A. W. (1982) *J. Biol. Chem.* **257,** 8418–8423.
10. Barker, W. C. & Dayhoff, M. O. (1977) *Comp. Biochem. Physiol. B.* **57,** 309–315.
11. Karathanasis, S. K., Zannis, V. I. & Breslow, J. L. (1983) *Nature (London)* **304,** 371–373.
12. Land, H., Grez, M., Hauser, H., Lindenmaier, W. & Schutz, G. (1981) *Nucleic Acids Res.* **10,** 2251–2266.
13. Gordon, J. I., Smith, D. P., Alpers, D. H. & Strauss, A. W. (1982) *Biochemistry* **21,** 5424–5431.
14. Sanger, F., Nicklen, S. & Colson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3962–3965.
16. Smith, A. J. H. (1980) *Methods Enzymol.* **65,** 560–580.
17. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
18. Barnes, W. M. & Bevin, M. (1983) *Nucleic Acids Res.* **11,** 349–368.
19. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
20. Fitch, W. M. (1977) *Genetics* **86,** 623–644.
21. McLachlan, A. D. (1977) *Nature (London)* **267,** 465–466.
22. Kubota, Y., Takahashi, S., Nishikawa, K. & Ooi, T. (1981) *J. Theor. Biol.* **91,** 347–361.
23. Karathanasis, S. K., Zannis, V. I. & Breslow, J. L. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6147–6151.
24. Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M. (1974) *FEBS Lett.* **38,** 247–253.
25. Sparrow, J. T., Morrisett, J. D., Pownall, H. J., Jackson, R. L. & Gotto, A. M. (1975) in *Peptides: Chemistry, Structure and Biology*, eds. Walter, R. & Meienhofer, J. (Ann Arbor Science, Ann Arbor, MI), pp. 597–602.
26. Yokoyama, S., Fukushima, D., Kupferberg, J. P., Kezdy, F. J. & Kaiser, E. T. (1980) *J. Biol. Chem.* **255,** 7333–7339.
27. Kaiser, E. T. & Kezdy, F. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1137–1143.
28. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
29. Schiffer, M. & Edmundson, A. B. (1976) *Biophys. J.* **7,** 121–125.
30. Morrisett, J. D., Jackson, R. L. & Gotto, A. M. (1977) *Biochim. Biophys. Acta* **472,** 93–133.
31. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York), pp. 69–70.
32. Flinta, C., Heline, G. & Johansson, J. (1983) *J. Mol. Biol.* **168,** 193–196.
33. Soutar, A. K., Garner, C. W., Baker, H. N., Sparrow, J. T., Jackson, R. L., Gotto, A. M. & Smith, L. C. (1975) *Biochemistry* **14,** 3057–3064.
34. Fukushima, D., Yokoyama, S., Kroon, D. J., Kezdy, F. J. & Kaiser, E. T. (1980) *J. Biol. Chem.* **255,** 10651–10657.