Front. Biol. 2013, 8(5): 533–548
DOI 10.1007/s11515-013-1278-7

REVIEW

Proteogenomic review of the changes in primate apoC-I during evolution

Donald L. PUPPIONE (✉)[1], Julian P. WHITELEGGE[1,2]

[1] *The Molecular Biology Institute, University of California, Los Angeles, CA 90095, USA*
[2] *The Pasarow Mass Spectrometry Laboratory, The Jane & Terry Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA*

Abstract **Apolipoprotein C-I has evolved more rapidly than any of the other soluble apolipoproteins. During the course of primate evolution, the gene for this apolipoprotein was duplicated. Prompted by our observation that the two resulting genes encode two distinct forms of apoC-I in great apes, we have reviewed both the genomic and proteomic data to examine what changes have occurred during the course of primate evolution. We have found data showing that one of the duplicated genes, known to be a pseudogene in humans, was also a pseudogene in Denisovans and Neandertals. Using genomic and proteomic data for primates, we will provide in this review evidence that the duplication took place after the divergence of New World monkeys from the human lineage and that the formation of the pseudogene took place after the divergence of the bonobos and chimpanzees from the human lineage.**

Keywords mass spectrometry, apoC-IA, apoC-IB, apes, Old World monkeys, New World monkeys

Introduction

Apolipoprotein C-I (apoC-I), the smallest of the soluble apolipoproteins, is also one of the most positively charged proteins in the circulation. Based on comparative analyses of the rate of non-synonymous amino acid substitution, it was concluded that this apolipoprotein was evolving more rapidly than any of the other soluble apolipoproteins (Luo et al., 1999). During the course of primate evolution, the gene for this apolipoprotein was duplicated. One of the duplicated genes later became a pseudogene. Using genomic and proteomic data for primates, we will provide in this review evidence that the duplication took place after the divergence of New World monkeys from the human lineage and that the formation of the pseudogene occurred after the divergence of bonobos and chimpanzees from the human lineage. The review will be divided into four major parts: A) Apolipoprotein nomenclature; B) Early proteomic and genomic studies; C) Differences in the apoC-I gene of primates; D) Differences in the apolipoprotein C-I of primates.

Received July 1, 2013; accepted August 9, 2013
Correspondence: Donald L. PUPPIONE
E-mail: puppione@chem.ucla.edu

Apolipoprotein nomenclature

Prior to their sequences being determined, human apolipoproteins were designated in a variety and, more often than not, a confusing manner. Although different laboratories ended up characterizing the same apolipoproteins, the nomenclature took a couple of years before it was simplified (Scanu, 1972). For apoC-I, these were some of the designations used initially: 1) apo VLDL VD_1; 2) apo VLDL C_1; or 3) either D_1 or apoLp-Ser or R-Ser. In the case of the first two examples, VLDL (very low density lipoproteins) was used to indicate the apolipoproteins had been derived from the triacylglycerol-rich lipoproteins recovered ultracentrifugally in the very low density lipoprotein class. The use of D_1 in the first and third example indicates that the protein, following delipidation of the lipoprotein, was the first of the C apolipoproteins to elute from a DEAE column. The designations, apoLp-Ser or R-Ser, indicated that the C-terminal amino acid of the protein was serine. Finally, the term, VD_1, in the first example indicated that this apolipoprotein was to be distinguished from HD_1, the latter being obtained from ultracentrifugally isolated HDL (high density lipoproteins). As it turned out, they were the same apolipoprotein. In general, all of the soluble apolipoproteins have been detected in association with both VLDL and HDL. The ultracentrifugal lipoprotein classes with the

corresponding density intervals are listed in Table 1. Further information regarding these early apolipoprotein studies can be found in a review by Scanu (1972). It was therefore fortunate that the designation proposed by Alaupovic et al. (1972) soon became accepted, because comparative studies have shown that the orthologs of the various apolipoproteins do not have the same terminal amino acid and in some cases do not have the same chromatographic profile. The molecular weights and calculated pI values of the nine soluble apolipoproteins (Table 2) are listed with the Alaupovic nomenclature that will be used throughout the remainder of this review.

Early proteomic and genomic studies

Early protein studies

With the development of the preparative ultracentrifuge, it became possible to separate the various plasma lipoproteins into different density classes based on their hydrated density and to develop protocols for their physicochemical characterization (Glazier et al., 1954; Havel et al., 1955; Lindgren, 1975). The soluble apolipoproteins were found to be associated with both the HDL and the triacylglycerol-rich lipoproteins, *viz.* chylomicra and VLDL (Gustafson et al., 1966; Alaupovic et al., 1972). However, early studies by Lee and Alaupovic (1970) and more recent mass spectrometry studies on LDL (Karlsson et al., 2005) indicate that trace amounts of the soluble apolipoproteins are also associated with these lipoproteins as well.

Following delipidation of the lipoproteins, the apolipoproteins were separated chromatographically. HDL studies initially showed that two apolipoproteins, apoA-I and apoA-II, comprising 95% of the protein moiety were present in a molar ratio of 3 to 1 (Shore V and Shore B, 1968; Scanu and Edelstein, 2008). The remaining 5% consisted of apoC-I, C-II and C-III (Gustafson et al., 1966). The triglyceride-rich VLDL were found to have a higher percentage content of the C apolipoproteins. In their study of these apolipoproteins, Brown et al. (1969) reported that in contrast to the apoC-II and C-III, apoC-I was positively charged. Not too much later, the publication of the primary sequence showed human apoC-I to consist of 57 amino acids, with 21% of the amino acids being positively charged (Shulman et al., 1975).

Comparing the primary sequences of four of the apolipoproteins, apoA-I, A-II, apoC-I and apoC-III, Barker and Dayhoff (1977) identified a prominent 11-residue motif repeated throughout the length of the proteins. They went on to conclude that the soluble apolipoproteins evolved through a series of duplications, insertions and deletions from a common ancestor gene. At about the same time, Segrest et al. (1974) and Fitch (1977) proposed that these 11- or 22-amino acid stretches formed a series of amphipathic α-helices, enabling the apolar residues to interact with the cholesterol-phospholipid monolayer encapsulating the neutral lipids in the lipoprotein core. Using NMR, Rozek et al. (1995) reported on the location of two such helices in apoC-I.

Early genomic and proteomic studies

Knott et al. (1984) were the first to report that the gene for human apoC-I was located on chromosome 19 and subsequently found that in addition to the 57 amino acids of the mature protein, the gene encoded a 26 amino acid signal sequence. Later, Scott et al. (1985) and Lusis et al. (1986) reported that the gene for apoC-I is located in a cluster along with the genes for apoE and apoC-II on the long arm of human chromosome 19. Studies by Davidson et al. (1986) reported that the loci for the apoC-I gene was located 4 kb from the apoE gene and they indicated that there might be another genes for apoC-I on chromosome 19, as later shown

Table 1 Density intervals of plasma lipoproteins

Lipoprotein class	Density interval	Major core lipid
Chylomicron & very low density	< 1.0063 g/mL	Triacylglycerol
Low density	1.020–1.063 g/mL	Cholesteryl esters
High density	1.063–1.210 g/mL	Cholesteryl esters

Table 2 Properties of the nine soluble apolipoproteins

Apolipoprotein	Number of amino acids	Molecular mass (Da)	Estimated pI value
A-I	243	28078.3	5.27
A-II	77	8690.9*	5.05
A-IV	376	43402.5	5.18
A-V	343	38904.8	5.99
C-I	57	6630.5	7.93
C-II	79	8914.9	4.66
C-III	79	8764.6	4.72
C-IV	100	11695.4*	9.30
E	299	34236.6	5.52

* Caculated average molecular mass was corrected for N-terminal pyroglutamic acids, but the indicated pI value was obtained with the N terminus being glutamine.

to be the case by Lauer et al. (1988). Interestingly, this second gene was a pseudogene located 3 kb downstream from the other apoC-I gene. It is a pseudogene due to a point mutation in which the codon for glutamine, CAG, encoding the penultimate amino acid of the signal sequence is converted to a stop codon, TAG. As had been reported for the organization of the genes of apoA-I, A-II and C-III, Lauer et al. found that both genes had the same organization, namely four exons separated by three introns, with the first exon being non-coding (Lauer et al., 1988). For both apoC-I gene and the pseudogene, the second intron is located within the codon for glycine of the signal peptide region and the third intron within the codon either for arginine in the case of apoC-I or for tryptophan in the sequence of the pseudogene. A fifth gene was later found in this cluster located between the gene for apoC-II and the pseudogene (Allan et al., 1995; Dang and Taylor, 1996; Zhang et al., 1996). This gene encodes apoC-IV, an apolipoprotein that had not been previously reported. All five genes are in the same transcriptional orientation and are arranged as shown in Fig. 1.

A comparison of the apoC-I gene with the pseudogene revealed that the second and third introns had a variable number of Alu sequences. As indicated in Table 3, showing initial and final coordinates of each sequence, there are ones that are common to both genes. These sequences most likely were also present in the precursor gene that underwent duplication. The others that are unique to one or the other gene would have played a role in their evolution. It has also been noted that when the total number Alu sequences in the two genes are combined, there are approximately 18.5 of them in a 16-kb region, compared with the average of 1 Alu sequence/8-kb in the human genome (Luo et al., 1989). Examining the exonic sequences, the first and second exons are identical (Luo et al., 1989). The fourth exons differ at three residues, resulting in 95% identity. The major differences are in the third exons. In addition to the replacement of the codon for the penultimate residue in the signal sequence with a stop codon, 22 additional nucleotides are different. Interestingly, as noted above, the apoC-I gene encodes a positively charged protein; however, the pseudogene encodes a virtual protein that is negatively charged.

Comparing synonymous and non-synonymous changes in the pseudogene and its functional counterpart, Luo et al. (1989) concluded that the pseudogene probably was formed more than 35 mya prior to the divergence of New World primates from the human lineage. Raisonnier (1991), looking at the differing substitutions among the shared Alu sequences, estimated that the gene duplication occurred 39 ± 4 mya. Analyzing both genes and the adjacent hepatic control region, Freitas et al. (2000) estimated the time of duplication to be 37 mya.

Prior to the assembly of genomes, only a few studies had been done on the apoC-I gene of other primates, namely baboons and macaques. Herbert et al. (1987) were the first to characterize apoC-I in a non-human primate. These authors noted that the cynomolgus monkey had two forms of apoC-I, with one lacking the dipeptide, Asp Pro at the N terminus. They also reported that like human apoC-I, the protein was positively charged. Pastorcic et al. (1992) found the gene for baboon apoC-I to be like humans, with four exons and three introns. The gene also encoded a 26 amino acid signal sequence and a 57 amino acid mature protein (Pastorcic et al., 1992). Comparing the baboon gene to the human pseudogene, they concluded that the number of non-synonymous substitutions were consistent with an ancient inactivation of the gene. They estimated that the pseudogene was formed approximately 25 mya.

Differences in the apoC-I gene of primates

Apo C-I genes of primates

The genomic data in the following sections were obtained from the databases of the National Center of Biologic Information (NCBI), the University of California at Santa Cruz (UCSC) and the Swiss Institute for Bioinformatics (SIB) in Lausanne Switzerland. Alu sequences were identified using CENSOR, a tool avail online on the webpage of the Genetic Information Research Institute, Mountain View, CA.

Now that the genomes of several primates have been assembled, it is possible to make further comparison of genes between species. Although the non-coding exon 1 is not always apparent in some of these assemblies, the three coding exons have been located in most cases. In Table 4 are the coordinates of the first base for the methionine codon in exon 2 and the last base of the stop codon in exon 4 for the primates for which the complete genes have been identified. Exon 2 of all primates contains 58 nucleotides; however, the exons 3 of New World monkeys are slightly larger and those of prosimians slightly smaller. Moreover, there is variation in the overall size of the various C-I genes, with those of prosimians and New World monkeys being between 2.5 and 4.4 kb whereas human and the other great ape genes are around 5 kb. These size differences are due in large part to the variation in the content of Alu elements in the intronic regions.

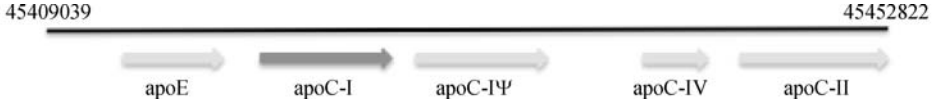


Figure 1 The apolipoprotein gene cluster on chromosome 19 in humans. The direction of the arrows indicates that the genes have the same transcriptional orientation. The symbol, C-IΨ, indicates the relative position of the pseudogene on chromosome 19.

Table 3 Coordinates of Alu elements in the apoC-I gene and the pseudogene of humans

Alu elements intron 2	Start	End
ApoC-I		
AluYd2	45418598	45418854
AluSz	45418942	45419226
AluS	45419246	45419394
Pseudogene		
AluSz	45430736	45430989

Alu elements intron 3	Start	End
ApoC-I		
AluSx	45419793	45420067
AluJo	45420107	45420239
AluJr	45420573	45420835
AluSx	45420954	45421228
AluSg	45421248	45421525
AluY	45421636	45421843
AluSz	45421957	45422214
Pseudogene		
AluSx	45431386	45431679
AluJo	45431710	45431809
AluSg7	45431828	45432106
AluSp	45432179	45432489
AluJ	45432839	45433109
AluSp	45432008	45433487
AluSz	45433761	45434030
AluY	45434449	45434739

During the course of the past 65 million years of primate evolution, Alu elements have retrotransposed throughout their genomes (Deininger and Batzer, 2002). It is estimated that one million copies of Alu elements comprise 11% of the human genome (Li et al., 2001). Clusters of Alu elements from different subfamilies have been described at several loci as a result of retropositional events which occurred in the same chromosomal location during different periods of primate evolution (Rowold and Herrera, 2000).

Prosimian

Prosimians diverged from the human lineage 74 mya.The genomic data for prosimian apoC-I in the UCSC database are limited to the bushbaby (*Otolemur garnettii*) and the gray mouse lemur (*Microcebus murinus*). In the case of the mouse lemur, only exons 2 and 3 have been identified (scaffold_10120). All three coding exons of the bushbaby are present. Based on the coordinates in Table 4A, the bushbaby gene is approximately 2.5 kb. Exon 2 contained 58 nucleotides; however, both exons 3 and 4 contained 126 and 52 nucleotides, smaller than corresponding human exons. No Alu elements were detected in intron 2, however several variations of Alu2 series were found in intron 3. It should also be noted that the terminal amino acid encoded by exon 3 of the mouse lemur would be glycine rather than arginine, as is the case in all the other primates.

New World monkeys

Entries containing the apoC-I gene of New World monkeys that diverged from the human lineage 42.6 mya can be found in various databases. The UCSC entry indicates that the gene of the marmoset (*callithrix jacchus*) is on chromosome 22. The other four entries can be found in the SIB and the NCBI databases. They are in the order shown in Table 4A: Ma's night monkey (*Aotus nancymaae*), Dusky titi (*Callicebus moloch*), Squirrel monkey (*Saimiri boliviensis boliviensis*) and Spider monkey (*Ateles geoffroyi*). Their genes range between 2.7 kb, in the case of the Ma's night monkey and the Dusky titi to 4.4 kb for the marmoset. It should be noted that there are major sequencing gaps in intron 3 of the marmoset gene that if filled may change its size in the future. In each case, both exons 2 and 4 contained 58 nucleotides and exon 3 contained 145 nucleotides, slightly larger than the corresponding human exon. The Alu elements detected in both introns 2 and 3 belong to the AluS series.

Old World monkeys

In contrast to New World monkeys, there are limited data on apoC-I exons of Old World monkeys. Complete apoC-IB

Table 4A Coordinates of the apoC-I genes of Prosimians and New World monkeys

Primates	Accession No.	Start of Ex2	End of Ex4
Prosimian			
Bushbaby	GL873671	2149640	2152116
New World monkeys			
Ma's night monkey	AC146520.2	47233	49965
Dusky titi	AC146285.3	141287	143979
Squirrel monkey	AC151887.2	159235	162244
Spider monkey	AC188244.1	113479	116522
Marmoset Chr. 22*	–	36965985	36970402

*Accession number is omitted if the gene has been located on a chromosome.

Table 4B Coordinates of apoC-IB and apoC-IA genes of Old World monkeys and apes

	Access No.	apoC-IB		apoC-IA	
Old World monkeys					
Colobus	AC148222.2	168001	172552	180044	185060
Olive baboon	JH682906	88330	84609	–	–
Hamadryas baboon	AC145523.3	145351	149539	–	–
Lesser apes					
Gibbon	AC146473.1	20697	26762	–	–
Great apes					
Orangutan Chr. 19	–	–	–	46168619	46173056
Gorilla Chr. 19	–	–	–	42137059	42141439
Chimpanzee Chr. 19	–	50105028	50109366	50116490	50120477
Human Chr. 19	–	45418149	45422487	45430233*	45434314*

*Pseudogene coordinates on chromosome 19.

genes are available only for the Olive baboon (*Papio anubis*) in the UCSC database, for Hamadryas baboon (*Papio hamadryas*) in the NCBI database and for the Mantled guereza (*Colobus guereza*) or colobus monkey in the SIB database. The exonic sequences of the two baboons are identical. Entries for exons 2 and 3, but not 4, on chromosome 19 of the rhesus can also be found for the Rhesus (*Macaca mulatta*) in the UCSC database (not shown in Table 4B). As Table 4B indicates, a second apoC-I gene, apoC-IA, is present in the colobus.

With all three encoding exons for both genes being located, the most complete data are for the colobus monkey. The upstream gene encodes a basic protein, designated apoC-IB. On the other hand, the downstream gene encodes an acidic protein, designated apoC-IA. This is very similar to what was said about apoC-I and the pseudogene in humans, with the important difference being the absence of a stop codon in exon 3 of the apoC-I gene. As will be seen in the section on protein sequence, a glutamine is present at the penultimate residue of the signal sequence of all primates for both apoC-IA, as well as apoC-IB.

Comparing the exons of the baboons and those of the colobus in the upstream gene to the exons of human apoC-I, there is only one nucleotide difference in exon 2, resulting in a non-synonymous change at position −10 in the signal sequence. In exon 4, there are three non-synonymous variations in the sequence, each resulting from a single nucleotide change. Of the 10 codon variations in colobus exon 3, 4 of these are synonymous, 6 non-synonymous, with 3 being conserved, with one of the non-conserved resulting from a change in two nucleotides. Exon 3 codons of the baboons vary at 15 sites, with 5 synonymous, 8 non-synonymous, with 4 being conserved. Among both the conserved nonsynonymous and the non-conserved nonsynonymous changes, a two nucleotide variation in a single codon was observed in each.

Making a similar comparison between the apoC-IA gene and the human pseudogene, there is again a single non-synonymous change at position −10 of the signal sequence encoded by exon 2. In the case of exon 3, there are 11 non-synonymous changes, with three being conserved. Two of the non-conserved variations resulted from changes in two nucleotides in the respective codons. For exon 4, there is a single variation in the first 22 nucleotides, immediately followed by a deletion. The resulting change in the reading frame results in an early stop codon. There are 31 nucleotides in exon 4, including the stop codon, compared with the canonical 58. In the olive baboon entry, two exons, 2 and 4, detected downstream from the apoC-IB gene, were consistent with the presence of an apoC-IA gene; however, due to a major gap in the sequence, exon 3 is currently unavailable for analysis.

On chromosome 19 of the rhesus, there are two exons downstream from the apoC-IB gene with exons 2 similar to that of the colobus apoC-IA gene, but with six nucleotides deleted from the middle region. Exon 3 is much different with a deletion of a single nucleotide that results in a change in the reading frame. This change also leads to a stop codon at the end of the exon. However, a recent entry (JH292475) in the UCSC database provides an alternate version for rhesus apoC-IA. Exon 2 of this second version is identical to the one mentioned above for the first 52 nucleotides, but there is a second potential splice site 6 nucleotides downstream. Exon 3 is almost identical to that of the colobus apoC-IA, with just two nucleotide differences.

In regards to the Alu elements in OWM apoC-I gene, the AluS series were detected in intron 2 of both apoC-IA and B. In addition to the AluS series, AluJ and AluY were detected in intron 3. Interestingly, AluMacYa3, thought to be specific for macaques, were detected only in intron 3 of apoC-IA and B of the mantled guereza.

Lesser apes

Gibbons, so-called lesser apes, diverged from the human lineage 20.4 mya. Currently, there is one entry for a complete apoC-I gene of a gibbon in the SIB database. This entry is for the apoC-IB gene of the Mentawai gibbon (*Hylobates*

klossii). In addition, downstream exons, corresponding to exon 2 and 3 of apoC-IA, can also be found. The exons 2 of both genes are identical and the variation between their human counterparts results in a non-synonymous change at position –10 in the signal sequence. Comparing exons 3 of human apoC-I and gibbon apoC-IB, there are 2 synonymous changes and 4 non-synonymous changes. A similar comparison between gibbon apoC-IA and the pseudogene reveals 1 synonymous and 3 non-synonymous changes, again with one of the latter changes being a codon for glutamine instead of a stop codon. Between Exon 4 of gibbon C-IB and human apoC-I, there are three non-synonymous changes. All of the above changes involve a difference of a single nucleotide.

As was the case for OWM, only the AluS series were detected in intron 2 of both apoC-IA and B. In intron 3 of apoC-IB, Alu series S, J and Y were detected.

Great apes

With the exception of the bonobo, genomic data can be found in UCSC database for the other great apes. The respective divergence time from the human lineage are: 15.7 mya for orangutans (*Pongo abelii*), 8.8 mya for gorilla (*Gorilla gorilla*) and 6.4 mya for bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*). Current entries enable a comparison to be made between exons 2 and 3 of the chimpanzee and the orangutan apoC-IB gene and those of human apoC-I gene. Like their human counterpart, these genes are located on chromosome 19. Exons 2 of humans and chimpanzee are identical and that of the orangutan differs at a single nucleotide, resulting in non-synonymous change at residue –10 of the signal sequence. A single nucleotide difference in exon 3 results in a non-synonymous variation between human and chimpanzee. A similar comparison with exon 3 of the orangutan apoC-IB gene reveals two synonymous and 4 non-synonymous changes.

Comparing exon 4 of the human and chimpanzee gene, there is a single synonymous change. Exon 4 of the orangutan apoC-IB gene has not yet been identified and there are no sequence data for the gorilla apoC-IB gene.

However, it is possible to compare all three coding exons of the apoC-IA genes of each of the great apes to the corresponding exons of the human pseudogene. A single nucleotide change results in exon 2 of the all the great apes differing from the pseudogene at the codon encoding the – 10 residue of the signal sequence. Another nucleotide change in the exon 2 of the orangutan results in a difference at residue –21. In each case, these changes are nonsynonymous.

In addition to the non-synonymous change in exon 3 resulting in a stop codon in the pseudogene, there are one synonymous and two non-synonymous changes in the orangutan gene, three non-synonymous changes in the gorilla gene and one non-synonymous change in the chimpanzee gene. In exon 4, there are 2 non-synonymous changes in the orangutan, with one conserved and the other non-conserved. The latter non-synonymous change is also present in the exons of the chimpanzee and the gorilla.

In terms of the Alu sequences that are present in the intronic regions, AluS series is found in intron 2 of both the apoC-IB and apoC-IA genes as well as the AluY series in intron 2 of the apoC-IB genes of humans and chimpanzees. In intron 3 of both apoC-IB and apoC-IA, Alu series S, J and Y were detected.

Evidence for both apoC-I and the pseudogene in Neandertals and Denisovans

Using the coding exons of human apoC-I and those of the pseudogene, it was possible to locate in the UCSC database the apoC-I genes and the pseudogenes of the Neandertals and the Denisovans.

The Neandertal sequences were obtained from three sets of fossils discovered in the Vindija cave in Croatia, Vi33.16 (54.1% genome coverage), Vi33.25 (46.6% genome coverage) and Vi33.26 (45.2% genome coverage). Together they provide evidence of the presence of an apoC-I gene as well as the pseudogene in Neandertals. The resulting alignments are shown in Tables 5 and 6. Examining the data in Table 5, only 3 variations were detected in the exons, one in exon 3 at position 45419535 and two in exon 4 at positions 45419467 and 45422486. The second of these was the third base in an incomplete codon and the other two were synonymous changes. In Table 6, there are also three variations in the exonic region of the pseudogene. The first in exon 2 at position 45430280 would potentially be a conserved non-synonymous change with a methionine being replaced with an isoleucine. The second in exon 3 at position 45431164 is a synonymous change and the third is the first base of an incomplete codon. The stop codon has been underlined.

The genome sequence of a Denisovan individual was generated from a small fragment of a finger bone discovered in Denisova Cave in southern Siberia in 2008. The alignment of the human exons with the Denisovan genes provided multiple sequence reads that provided complete coverage. However, only those that had the best coverage of the regions of interest were selected. As the data in Tables 7 and 8 indicate, there was only a single variation at position 45431058 in exon 3 of the pseudogene, resulting in a conserved non-synonymous change. However, a shorter sequence read in the same region did not show any variation (data not shown). The stop codon is underlined in Table 8.

Differences in the apolipoprotein C-I of primates

Primary sequence of primate apolipoprotein C-I

Using genomic data, the apoC-I primary sequences of the various primates were obtained. In the next three tables, sequences are divided into three segments, with the signal

Table 5 Alignment of exons of human apoC-I gene with Neandertal sequences*

<table>
<tr><th>Source</th><th>Neandertal apoC-I</th></tr>
<tr><td></td><td>Exon 2 alignment</td></tr>
<tr><td>Vi33. 16</td><td>00000001 TTTCGGCCTCGCCATGAGGCTCTTCCTGTCGCTC 00000034
>>>>>>>> | ||||||||||||||||||||||||||||||| >>>>>>>>
45418136 ctccggcctcgccatgaggctcttcctgtcgctc 45418169</td></tr>
<tr><td></td><td>Exon 3 alignment</td></tr>
<tr><td>Vi33. 25</td><td>00000001 AAAAAAAAAAAAAAACAAATTTTGAACCCCTGCCCATCTTCCTGGCAGGCCCAGCCC 00000057
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419399 aaaaaaaaaaaaaaacaaattttgaacccctgcccatcttcctggcaggcccagccc 45419455</td></tr>
<tr><td>Vi33. 16</td><td>00000001 GCCCAGGGGACCCCAGACGTCTCCAGTGCCTTGGATAAGCTGAAGGAG 00000048
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419458 gcccaggggaccccagacgtctccagtgccttggataagctgaaggag 45419505</td></tr>
<tr><td>Vi33. 26</td><td>00000001 ACCCCAGACGTCTCCAGTGCCTTGGATAAGCTGAAGGAGTTTGGAAACACACTGGAGGACAAGGCTCGAGAA 00000072
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| >>>>>>>>
45419467 accccagacgtctccagtgccttggataagctgaaggagtttggaaacacactggaggacaaggctcgggaa 45419538</td></tr>
<tr><td></td><td>Exon 4 alignment</td></tr>
<tr><td>Vi33. 16</td><td>00000001 TAGAGACATTTCAGAAAGTGAAGGAGAAACTCAAGATTGACTCATAAGGACCT 00000053
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||| ||||||| >>>>>>>>
45422441 cagagacatttcagaaagtgaaggagaaactcaagattgactcatgaggacct 45422493

GAAAGGTGACATCCCAGAAAAGACC 00000078
||| ||||||||||||| | ||| >>>>>>>>
gaagggtgacatcccaggaggggcc 45422518</td></tr>
</table>

*The Neandertal sequences are on top (upper case), with coding sequences in red. The human sequences are shown beneath (lower case), listed with the initial and final coordinates on chromosome 19.

Table 6 Alignment of pseudogene exons with Neandertal sequences*

<table>
<tr><th>Source</th><th>Neandertal pseudogene</th></tr>
<tr><td></td><td>Exon 2 alignment</td></tr>
<tr><td>Vi33. 25</td><td>00000001 TGGTCTTGGAAGGTAAAAGTGGGATGGGAGAATTGCGGAGTT 00000042
>>>>>>>> | |||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45430279 tcgtcttggaaggtaaaagtgggatgggagaattgcggagtt 45430320</td></tr>
<tr><td></td><td>Exon 3 alignment</td></tr>
<tr><td>Vi33. 26</td><td>00000001 TCTCATCTTCCTGGCAGGCCCAGCCCCAGCCTAGGGGG 00000038
>>>>>>>> ||||||||||||||||||||||||||||||||||||| >>>>>>>>
45431035 cctcatcttcctggcaggcccagccccagcctaggggg 45431072</td></tr>
<tr><td>Vi33. 16</td><td>00000001 GCTCCAGAAGTCTCCAACCCCTTTGATGGCCTGGAGGAGTTAGGAAA 00000047
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45431072 gctccagaagtctccaacccctttgatggcctggaggagttaggaaa 45431118

00000001 GACCCTGGAGGACTACACTCGGGAATTCATCAACCGCATCACACAAAGTA 00000050
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||| ||| >>>>>>>>
45431119 gaccctggaggactacactcgggaattcatcaaccgcatcacacagagtg 45431168</td></tr>
<tr><td>Vi33. 25</td><td>00000001 GATGTGGTTAGAACCCTTCCCAGGGCACGGGAGGGCTGGGGTGTGTTTGTGTGGAG 00000056
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45431182 gatgtggttagaacccttcccagggcacgggagggctggggtgtgtttgtgtggag 45431237</td></tr>
<tr><td></td><td>Exon 4 alignment</td></tr>
<tr><td>Vi33. 16</td><td>00000001 AGAGACATTTCGGAAAGTGAAGGAGAAACTCAAGACTGACTCATGAGGA 00000049
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45434269 agagacatttcggaaagtgaaggagaaactcaagactgactcatgagga 45434317</td></tr>
</table>

*The Neandertal sequences are on top in capital letters. Coding sequences are in red and the stop codon, TAG, is underlined. The human sequences are in small letters, listed with the initial and final coordinates on chromosome 19.

sequence on top, the sequence of the mature protein encoded by exon 3 in the middle and the remaining sequence encoded by exon 4 on the bottom. In Table 9, the 54 amino acid apoC-I of the bushbaby is shown with the 60 amino acid apoC-I of New World monkeys. Due to size differences, the sequence gaps have been shaded black. Limited sequence data for mouse lemur are also shown. Among the 26 residues of the signal sequence, there are eight sites that vary. The residues of

Table 7 Alignment of exons of human apoC-I gene with Denisovan sequences*

Denisovan ApoC-I

Exon 2 alignment

```
00000001 TCCGGCCTCGCCATGAGGCTCTTCCTGTCGCTCCCGGTCCTGGTGGTGGTTCTGTCGATCGTCTT 00000065
 >>>>>>>>||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45418137 tccggcctcgccatgaggctcttcctgtcgctcccggtcctggtggtggttctgtcgatcgtctt 45418201

00000066 GGAAGGTAAAAGTGGGATGGGAGAATTGCGGAGTTGGAGATTTGGAAGAGTGAAGGTGGCTACAG 00000130
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45418203 ggaaggtaaaagtgggatgggagaattgcggagttggagatttggaagagtgaaggtggctacag 45418267
```

Exon 3 alignment

```
00000001 GCCCATCTTCCTGGCAGGCCCAGCCCCAGCCCAGGGGACCCCAGACGTCT 00000050
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419430 gcccatcttcctggcaggcccagccccagcccaggggaccccagacgtct 45419479

00000051 CCAGTGCCTTGGATAAGCTGAAGGAGTTTGGAAACACACTG 00000091
>>>>>>>> |||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419480 ccagtgccttggataagctgaaggagtttggaaacacactg 45419520


00000001 GGGACCCCAGACGTCTCCAGTGCCTTGGATAAGCTGAAGGAGTTTGGAAACACACTGGAGGACAA 00000065
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419464 gggaccccagacgtctccagtgccttggataagctgaaggagtttggaaacacactggaggacaa 45419528

00000066 GGCTCGGGAACTCATCAGCCGCATCAAACAGAGTGAACTTTCTGCCAAGATGCGGTT 00000122
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419529 ggctcgggaactcatcagccgcatcaaacagagtgaactttctgccaagatgcggtt 45419585
```

Exon 4 alignment

```
00000001 ACCCCCTTCCTTATTCCTCCCCACAGGGAGTGGTTTTCAGAGACATTTCA 00000050
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45422404 acccccttccttattcctccccacagggagtggttttcagagacatttca 45422453

00000051 GAAAGTGAAGGAGAAACTCAAGATTGACTCATGAGGACCTG 00000091
>>>>>>>> |||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45422454 gaaagtgaaggagaaactcaagattgactcatgaggacctg 45422494
```

*The Denisovan sequences are on top in capital letters. Coding sequences are in red. The human sequences are in small letters, listed with the initial and final coordinates on chromosome 19.

the signal sequence that is removed co-translationally in the lumen of the endoplasmic reticulum are assigned negative numbers. At residue −23, there is a conserved change in the bushbaby and a non-conserved change in the mouse lemur. The other variations are 4 conserved changes (at residues −15, −10, −5, and −1) and 3 non-conserved changes (residues −14 −12, −11). The non-conserved changes are indicated by a change in the background color, whereas the background color is the same for conserved changes, e.g. a lysine replacing an arginine has a blue background. The last 6 amino acids of the signal sequence are separated from the first 20 to indicate that they are encoded by exon 3. In the sequence of the mature protein encoded by exon 3, the 11 amino acid segment found between residues 14 and 24 of New World monkeys is similar to segments in the two prosimian sequences. Single variations of this sequence are found in the dusky titi at residue 20, in the night monkey at residue 19 and in the marmoset at residue 21. Another conserved segment can be found between residues 31 and 42, with only the dusky titi having an isoleucine at residue 32. Because the last two nucleotides of the lemur exon 3 are GG, it is assumed that they are part of an incomplete codon for glycine that awaits completion whenever the sequence of exon 4 is determined. Exon 4 of the bushbaby encodes a 16 amino acid sequence whereas exon 4 of the New World monkeys encodes an 18 amino acid sequence. Aside from the truncation of bushbaby apoC-I, alignment with the corresponding sequences of the New World monkeys shows the penultimate amino acid is a lysine rather than an arginine. For the exon 4 encoded segment of the New World monkeys, there is a single non-conserved variation in the marmoset (G57).

In Table 10, the apoC-IB sequences of a gibbon and three Old World monkeys are compared with human apoC-I and chimpanzee apoC-IB. In each case, the signal sequence contains 26 amino acids and the mature protein 57 amino

Table 8 Alignment of pseudogene exons with Denisovan sequences*

Denisovan apoC-I

Exon 2 alignment

```
00000001 TCCGGCCTCGCCATGAGGCTCTTCCTGTCGCTCCCGGTCCTGGTGGTGGTTCTGTCGATCGTCTT 00000065
 >>>>>>>>||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45418137 tccggcctcgccatgaggctcttcctgtcgctcccggtcctggtggtggttctgtcgatcgtctt 45418201

00000066 GGAAGGTAAAAGTGGGATGGGAGAATTGCGGAGTTGGAGATTTGGAAGAGTGAAGGTGGCTACAG 00000130
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45418203 ggaaggtaaaagtgggatgggagaattgcggagttggagatttggaagagtgaaggtggctacag 45418267
```

Exon 3 alignment

```
00000001 GCCCATCTTCCTGGCAGGCCCAGCCCCAGCCCAGGGGACCCCAGACGTCT 00000050
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419430 gcccatcttcctggcaggcccagccccagcccaggggaccccagacgtct 45419479

00000051 CCAGTGCCTTGGATAAGCTGAAGGAGTTTGGAAACACACTG 00000091
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419480 ccagtgccttggataagctgaaggagtttggaaacacactg 45419520


00000001 GGGACCCCAGACGTCTCCAGTGCCTTGGATAAGCTGAAGGAGTTTGGAAACACACTGGAGGACAA 00000065
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419464 gggaccccagacgtctccagtgccttggataagctgaaggagtttggaaacacactggaggacaa 45419528

00000066 GGCTCGGGAACTCATCAGCCGCATCAAACAGAGTGAACTTTCTGCCAAGATGCGGTT 00000122
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45419529 ggctcgggaactcatcagccgcatcaaacagagtgaactttctgccaagatgcggtt 45419585
```

Exon 4 alignment

```
00000001 ACCCCCTTCCTTATTCCTCCCCACAGGGAGTGGTTTTCAGAGACATTTCA 00000050
>>>>>>>> |||||||||||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45422404 acccccttccttattcctccccacagggagtggttttcagagacatttca 45422453

00000051 GAAAGTGAAGGAGAAACTCAAGATTGACTCATGAGGACCTG 00000091
>>>>>>>> ||||||||||||||||||||||||||||||||||||||||| >>>>>>>>
45422454 gaaagtgaaggagaaactcaagattgactcatgaggacctg 45422494
```

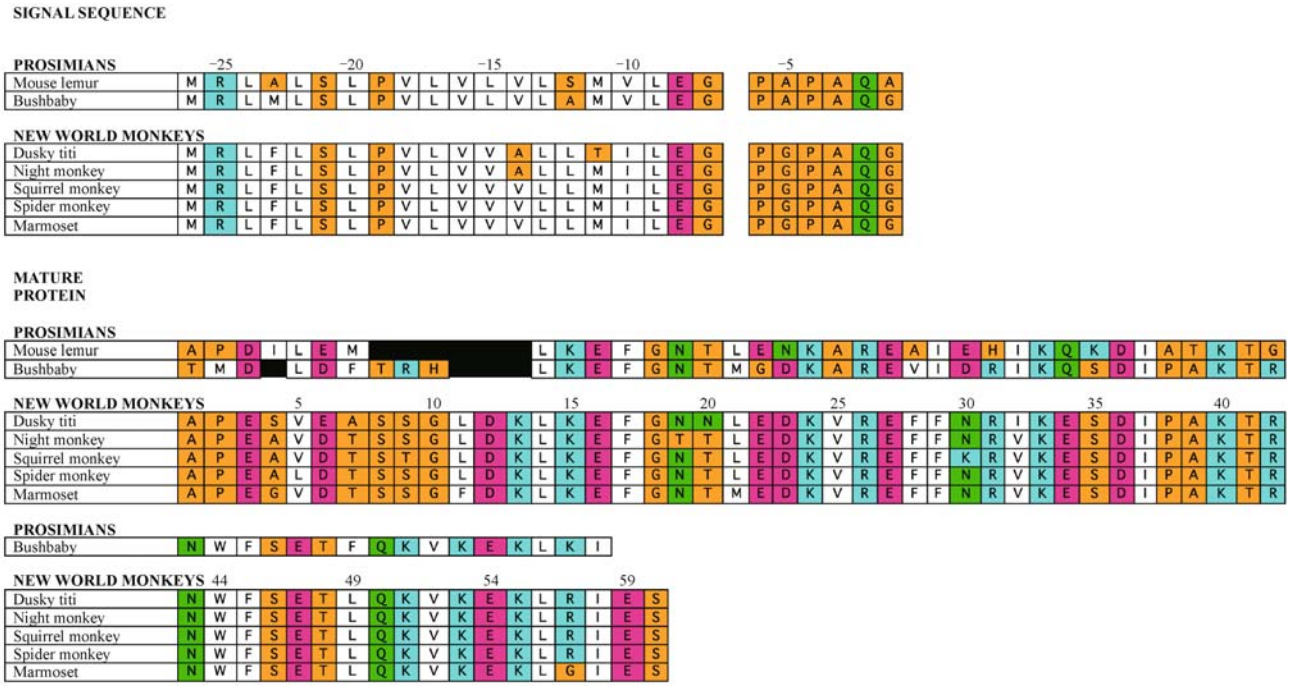
*The Denisovan sequences are on top in capital letters. Coding sequences are in red and the stop codon, TAG, is underlined. The human sequences are in small letters, listed with the initial and final coordinates on chromosome 19.

acids. For the exon2 and 4 encoded segments, the chimpanzee and human sequences are identical. In the sequence of the mature protein encoded by exon 3, there is only a single non-conserved variation in the chimpanzee apoC-IB (N32). For the other primates, the signal sequence varies with a conserved change, a methionine at residue −11. There is also a non-conserved change in the baboon, with a valine at residue −3. The other primates also vary at four sites in the mature protein (A1,V25, N27 and T38). Also macaques and baboons have a tryptophan at residue 23 and the macaques have another conserved variation at residue 9, a glutamate. For the exon 4 encoded segment, there is a conserved change (D40) and two non-conserved changes (R47 and N56).

In Table 11, the apoC-IA sequences of three great apes and two Old World monkeys are compared with the virtual protein encoded by the pseudogene. With the exception of the macaque, the signal sequence contains 26 amino acids. As was noted above in C-2, the rhesus had 6 fewer nucleotides in exon 2 that is reflected in the 2 missing amino acids in the signal sequence. Lacking a stop codon, all the primates have a glutamine at residue −2 as well as a methionine instead of an isoleucine at residue −11. The only other variation occurs in the orangutan sequence with a leucine instead of a serine at residue −21.

Both variations of the rhesus sequences for the first 39 amino acids are shown in Table 11. In one, a deletion of nucleotide results in a considerable change between residues 25 and 33. However, the second rhesus sequence aligns well with that of the colobus monkey. Comparing the first 39 amino acids in colobus apoC-I to the virtual protein encoded by the pseudogene, there are 2 conserved (residues 3 and 14) and 8 non-conserved changes (residues 1, 2, 10, 21, 22, 28, 38, 39). Among the great ape sequences, there is also a conserved variation at residue 3 in both the orangutan and gorilla sequences as well as a conserved difference at residue 25 of the gorilla. In the chimpanzee sequence there are two

Table 9 Alignment of Prosimian and New World monkey apoC-I sequences



The top rows show the 26 amino acid signal sequence; middle rows show the initial sequence encoded by exon 3; the last rows show the terminal sequence encoded by exon 4. The colors correspond to: red for the negative (D and E); blue for the positive (K and R); green for the polar (N and Q); orange for the neutral (A, G, H, P, S, T and Y) and white for the hydrophobic (C, F, L, M and V). The grouping of neutral and hydrophobic amino acids is according to Segrest et al., 1992.

non-conserved changes, one at residue 21 and another at residue 23. Lacking information about rhesus exon 4 in the apoC-IA gene, a comparison of the terminal region of apoC-I is limited to the colobus and the other great apes. Of the last 9 amino acids of the truncated apoC-IA of the colobus, there is a single non-conserved variation at residue 45. For the great apes, there is a non-conserved variation at residue 55. The orangutan also has a conserved variation at residue 47.

Comparing the sequences in Tables 9 and 10, the New World monkey sequences align more closely than with those in Table 11. Based on the sequence data, it was possible to obtain calculated values for the molecular weight and a theoretical pI values of primate apoC-I for which all three coding exons had been identified. These data are shown in Tables 12 and 13. All of the sequences that align with the virtual protein encoded by the pseudogene are indicated to be acidic (Table 13). However, it should be noted that, with the exception of the squirrel monkey, the theoretical pI values of the New World monkeys are acidic (Table 12).

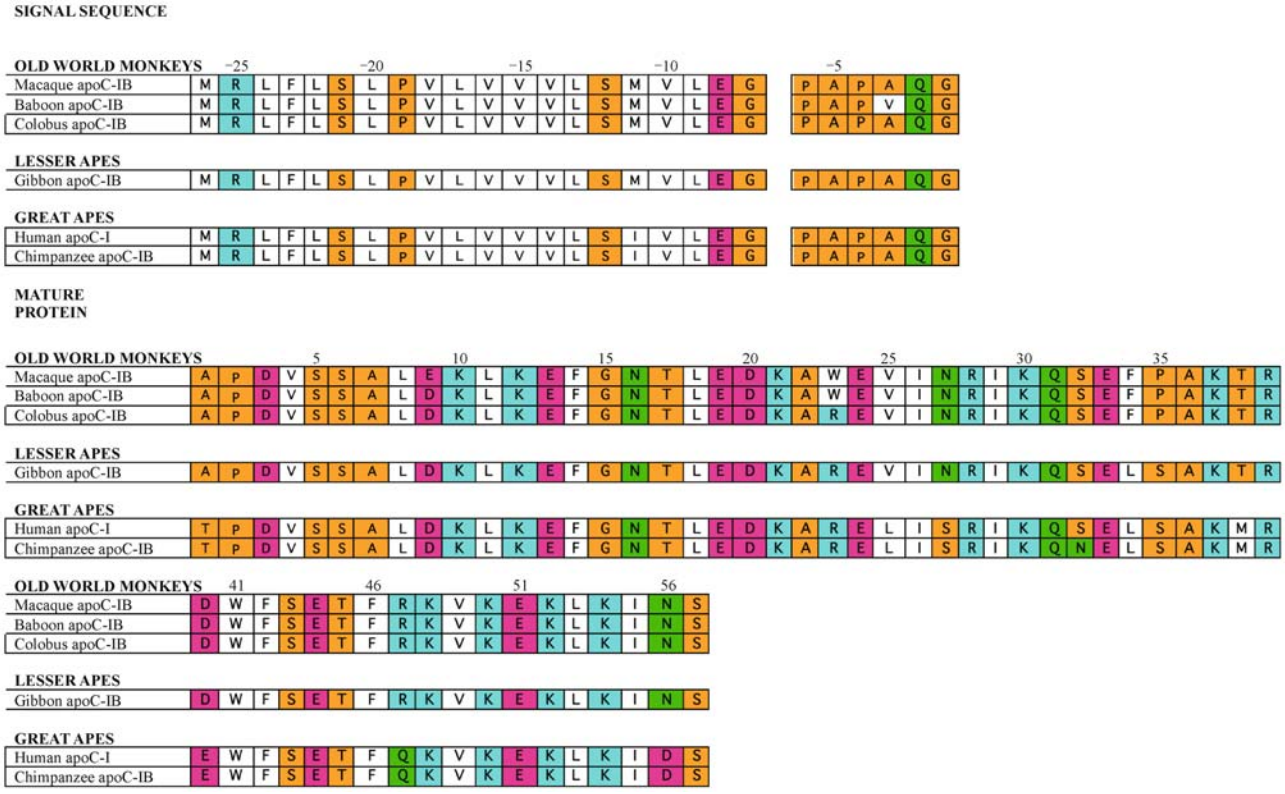
Neandertal and Denisovan sequence

The comparison of the coding exons of the human apoC-I gene and those of the pseudogene to the corresponding sequences in the Neandertal and Denisovan genomes in the UCSC database did not reveal any major differences in the mature protein sequences. The only variation was a single nucleotide difference in two of the Denisovan sequence reads. One of these was in agreement with an alanine being at residue −5 in the signal sequence of virtual protein. Whereas the other read would have replaced the alanine with a threonine.

Mass spectrometric studies of primate apolipoproteins

Using mass spectroscopy, Bondarenko et al. (1999) detected two forms of apoC-I in human plasma. The larger of the two has a molecular mass in agreement with the calculated value shown in Table 12. The other, a truncated form, had a molecular mass of 6518.6 Da due to the loss of the first two amino acids at the N terminus. As noted in section B-2, Herbert et al. (1987) noted in their studies that the cynomolgus monkey had a similar truncated form as well. A novel variation in apoC-I mass has recently been reported in certain patients suffering from coronary heart disease (McNeal et al., 2013). However, it still remains to be determined whether the mass increase of 89 Da is due to a mutation, post-translational modification or a combination of both. Currently, only two publications on naturally occurring apoC-I mutations, resulting in a decrease in molecular mass of 14 Da, have been reported (Wroblewski et al., 2006; Kasthuri et al., 2007). As a result of a point mutation, threonine at residue 45 is replaced by a serine. The first study (Wroblewski et al., 2006) was a preliminary survey that found

Table 10 Alignment of primate apoC-IB with human apoC-I



The top rows show the 26 amino acid signal sequence; middle rows show the initial sequence encoded by exon 3; the last rows show the terminal sequence encoded by exon 4. The colors correspond to: red for the negative (D and E); blue for the positive (K and R); green for the polar (N and Q); orange for the neutral (A, G, H, P, S, T and Y) and white for the hydrophobic (C, F, L, M and V). The grouping of neutral and hydrophobic amino acids is according to Segrest et al., 1992.

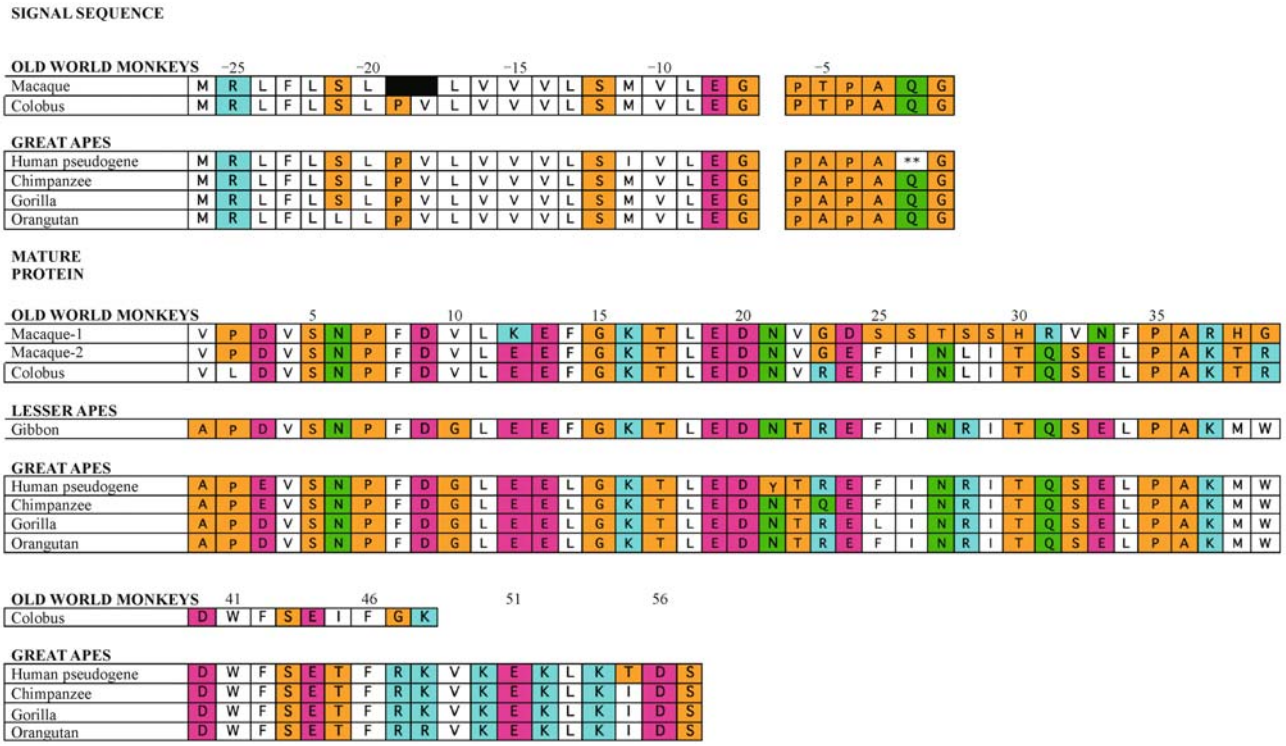
this mutation in two separate groups. One included Native Americans who were members of two nations in regions around Minneapolis, MN. The other group consisted of individuals of Mexican ancestry living in the Minneapolis area. The principal findings were that the truncated form of this apoC-I was preferentially associated with VLDL and that certain individuals with the mutated form of the apolipoprotein also had a high body mass index (BMI) (Kasthuri et al., 2007). The second study reported the presence of the mutated form in members of the Oji-Cree community of Sandy Lake, Ontario (Lahiry et al., 2010). In contrast to some of the individuals in the previous study, individuals with the mutated form of apoC-I had a low BMI. How this variant of apoC-I is involved or associated functionally with these opposing findings has not been determined.

In their mass spectroscopy studies of the apolipoproteins of great ape HDL, Puppione et al. (2010) detected different masses for apoC-I. Both were susceptible to truncation at the N terminus. One showed a high degree of identity to the mature and truncated forms of human apoC-I. The other was homologous to the virtual protein and its truncated form that are encoded by the human pseudogene. The larger mass agreed with the chimpanzee data shown in Table 13. The masses of the apolipoprotein were identical in both the bonobo and the chimpanzee. Because genomic data were not available for the bonobo, top down sequencing was carried out on the isolated protein. In agreement with the mass spectroscopy data, the resulting sequence was identical to that for the mature protein of the chimpanzee shown in Table 11. In other words, they were apoC-IA, encoded by the gene orthologous to the pseudogene. The molecular masses of apoC-IB were consistent with the data in Table 12.

Discussion

In this review, we have used the available genomic and proteomic data to discuss the variations in primate apoC-I. It was prompted by our mass spectrometry study of the HDL apolipoproteins of great ape in which we detected proteins encoded by a gene, previously thought to have converted to a pseudogene approximately 40 mya (Puppione et al., 2010). At the time of its initial discovery, it was proposed that the pseudogene was formed at the time the apoC-I was duplicated prior to the divergence of New World monkeys from the human lineage (Luo et al., 1989). The question whether the pseudogene might also be detected in New World monkeys was also raised at the time (Luo et al., 1989).

Table 11 Alignment of primate apoC-IS with the human pseudogene



The top rows show the 26 amino acid signal sequence; middle rows show the initial sequence encoded by exon 3; the last rows show the terminal sequence encoded by exon 4. The colors correspond to: red for the negative (D and E); blue for the positive (K and R); green for the polar (N and Q); orange for the neutral (A, G, H, P, S, T and Y) and white for the hydrophobic (C, F, L, M and V). The grouping of neutral and hydrophobic amino acids is according to Segrest et al., 1992.

Table 12 Molecular weights and pI values of human and other primate apoC-I

Primate		Molecular weight	pI value	Accession number
Bushbaby	apoC-I	6442.4	9.52	H0XNB8
Dusky titi	apoC-I	6970.8	6.52	P0DKV4
Night monkey	apoC-I	6943.8	6.52	P0DKV2
Squirrel monkey	apoC-I	6997.9	8.28	P0DKV5
Spider monkey	apoC-I	6970.8	6.52	P0DKV3
Marmoset	apoC-I	6895.7	5.39	F7F732
ColobusB	apoC-IB	6640.5	9.46	P0DKU8
MacaqueB	apoC-IB	6670.5	9.11	H9FNA4
GibbonB	apoC-IB	6596.5	9.46	AC146473
ChimapanzeeB	apoC-IB	6657.6	7.93	P0CE38
Human	apoC-I	6630.5	7.93	P02654

Table 13 Molecular weights and pI values of the virtual protein and the apoC-IA of other primates

Primate		Molecular weight	pI value	Accession number
Colobus	apoC-IA	5588.2	4.32	P0DKU7
Orangutan	apoC-IA	6729.5	4.99	P0CE39
Gorilla	apoC-IA	6653.4	4.65	P0CF78
Chimpanzee	apoC-IA	6715.5	4.82	P0CE37
Pseudogene	VP*	6752.5	4.82	

*Sequence of the virtual protein was obtained from exons 3 and 4 located on chromosome 19. See Table 4B for the coordinates of the pseudogene.

In examining the genomic data for five New World monkeys, there was no evidence of any apoC-I gene other than the one for which the coordinates appear in Table 4. The apoC-I and its pseudogene are located in a gene cluster. If the pseudogene had been present, it would have been located somewhere between the terminus of the apoC-I gene and the start of apoC-IV gene, another gene in this cluster (See Fig. 1). This being the case and using the various accession numbers listed in Table 4, it is possible to locate downstream the coordinates of the start of exon 2 for the apoC-IV gene of each of the New World monkeys. However, no exons were detected in this region between the genes for apoC-I and apoC-IV. Although the New World monkey apoC-I genes encode proteins that align well with apoC-IB of the other primates, the apolipoproteins is designated apoC-I, not apoC-IB. In introducing the nomenclature, apoC-IA and apoC-IB (Puppione et al., 2010), the intention was to apply these terms whenever there was evidence of two active genes, with one encoding a basic protein and the other an acidic protein. This is not the case for New World monkeys. Moreover, four of the apoC-I listed in Table 4 were acidic, not basic, based on the theoretical pI value.

Examining the data for the Old World monkeys, the colobus was the only one with complete genes for apoC-IA and C-IB. In the other Old World monkeys, an apoC-IA gene may be present in macaques depending on whether an exon 4 is ever identified. In the olive baboon (JH682906), exon 2 (between 769485 and 77049) and exon 4 (between 73815 and 73872) were located in the UCSC database. However, due to a major gap in the middle of the gene, exon 3 is missing.

An apoC-IB gene is present in both the chimpanzee and the gibbon. The data for the gorilla and the orangutan are incomplete for the apoC-IB gene. Exon 2 and 3 are present in the orangutan, but only exon 2 was detected in the gorilla. Concerning the orangutan gene, it should be pointed out that the coordinates listed in the Ensembl database are incorrect. That entry for apoC-IB lists the coordinates on chromosome 19 as being between 46155811 and 46173053, resulting in a slightly larger than a 17 kb gene. Upon examination, it was found that the first two exons of the apoC-IB gene were combined with the last exon of the apoC-IA gene. However, exons 2 and 3 of the apoC-IA gene can be located on chromosome 19 with coordinates between 46168619 and 46168676 and between 46169439 and 46169574, respectively. As indicated in Table 4B, the chimpanzee and the gorilla also have a gene for apoC-IA. There is both an exon 2 and 3 downstream from the gibbon apoC-IB, but exon 4 is missing.

Based on these data, we propose that duplication events giving rise to the two forms of apoC-I took place after the divergence of Old and New World monkeys. Also the detection of the AluY series in the Old World monkey and ape genes and not in New World monkey genes provides additional support for this conclusion (Deininger, 2011). Evidence of the pseudogene also was found in the analysis of the Denisovan and Neandertal genomes. This indicates that the point mutation causing the codon for glutamine, CAG, to be converted to a stop codon, TAG, took place sometime between the divergence of the bonobo and the chimpanzee from the human lineage and the appearance of the Denisovans.

Several metabolic roles attributed to apoC-I have been reviewed elsewhere (Shachter, 2001). Although not discussed in our review, one of these roles, the inhibition of cholesteryl ester transfer protein (CETP) should be mentioned in light of the seminal studies that were conducted on cholesterol-fed baboons (McGill et al., 1986; Kushwaha et al., 1993). To keep a balance on the distribution of cholesteryl esters among the lipoproteins, CETP transfers primarily polyunsaturated cholesteryl esters from HDL to the less dense lipoproteins. Otherwise, the less dense lipoproteins would become enriched in high melting cholesteryl esters, derived from the liver. Interestingly, certain baboons responded to a diet enriched in lard and cholesterol by developing a high concentration of large, less dense HDL, designated HDL_1 (McGill et al., 1986). It was later found that a truncated form of apoC-I, containing the first 39 amino acids, was inhibiting CTEP (Kushwaha et al., 1993). This prevented the exchange of cholesteryl esters between HDL and the less dense lipoprotein. As a result, the HDL accumulated more cholesteryl esters in their core and became larger and less dense. The nature of the mechanism that gave rise to this truncated form of apoC-I was never explained (Kushwaha et al., 1993).

In this review, we have presented data showing that during the course of evolution, the primary sequences of primate apoC-I changed in terms of the number and types of amino acids. Nevertheless, aligning the sequences one can find segments that are conserved. Using the N-terminal amphipathic α-helix between residues 7 and 29 detected by Rozek et al. (1995) as a guide, conserved sequences can be found. For this 23 amino acid sequence of human apoC-I, the theoretical pI is 6.32. This pI value is also found for the other sequences in Table 10, except for the baboon and macaque sequences. The presence of the tryptophan at residue 23 drops the pI to 5.01. The other variations were a valine at 25 and an asparagine at 27 in the gibbon as well as in the Old World monkeys. Making a similar alignment with the sequences between residues 10 and 32 in Table 9, there are a few more variations between human apoC-I and the New World monkey data. There are four conserved changes (G10, F28, F29 and V32) and also two non-conserved changes (V25 and N30). The squirrel monkey is an exception with an isoleucine at 32 and a lysine at 30. The presence of this lysine resulted in a theoretical pI of 8.43 for this portion of the squirrel monkey sequence, where the corresponding value for the other monkeys was 6.28. A non-conserved change also was at residue 20 with the presence of an asparagine, rather than a threonine. The opposite switch of amino acids occurs in the night monkey sequence at residue 19. There were also

conserved changes (F11 and M21) in the marmoset sequence. Because of the smaller size of the bushbaby apoC-I, the best alignment was seen between residues 10 and 28, with identity for 16 of the 29 amino acids. There were two conserved changes (M17 and V24) and two non-conserved changes (G18 and D26).The resulting theoretical pI value of this segment of the bushbaby sequence is 6.77.

For the C-terminal helix, the 14 amino acid sequence between residues 34 and 47 was selected because previous studies had shown that the phospholipid-bound structure of apoC-I is consistent with the C-terminal helix spanning this region (Gursky, 2001; Benjwal et al., 2007). The gibbon sequence differed with two non-conserved changes (T38M and N40D). In addition to these same two changes, Old World monkeys had two conserved changes (F34L and P35S). Aligning residues 37 to 50 of the New World monkey sequences, there are three conserved changes (I37, P38 and L49) and three non-conserved (T41, N43 and Q50). Because the two polar residues aligned with E40 and R43 in the human sequence, these differences did not result in a change in the theoretical pI value of 8.75. The bushbaby sequence aligns between residues 33 and 46. There are two conserved changes (I33 and P34) and three non-conserved changes (T37, N39 and Q46). Again with the aligning of the polar amino acids with oppositely charged amino acids, there was not a change in the pI value.

Comparing the same residues in the apoC-IA sequences of the great apes and colobus with the virtual protein encoded by the pseudogene, the N-terminal helices differed only slightly. All of them had an asparagine instead of a tyrosine at residue 21. The only other differences were seen in the gorilla sequence, with a glutamine at residue 23 and a leucine at residue 25. The colobus sequence had a total of five variations (V10G, F14L, N21Y, V22T and L28R). These sequences were even more acidic than seen in the comparison with human apoC-I, with the gorilla and the colobus having a pI value of 4.08 and the others having a value of 4.36.

For the C-terminal helix, there were no differences between the great ape apoC-IA sequences and that of the virtual protein. For the colobus apoC-IA sequence, there were four changes (T38M, R39W, I45T and G47R). In contrast to this region being basic in apoC-IB, these C-terminal segments each has a theoretical pI of 6.07.

Based on comparisons of the N-terminal and C-terminal helices outlined above, there are differences in the distribution of the charged amino acids. How these differences alter the function of the apolipoprotein is unknown at this time. Myers et al. (2013), using mutations at three residues, examined the protein interaction with lipids. The alignments of the primate N-terminal helices indicate that only one of these sites was conserved, namely G15. In New World monkeys, this would be G18 and in the bushbaby G14. In their study, the glycine was replaced with either an alanine or a proline. The other site in the N-terminal helix was R23 that was replaced with a proline. In Table 11 R23 is found in all the primates except the macaques and baboons that have a tryptophan at this residue. In New World monkeys the aligned site is R26 and in the bushbaby R22. The last site, M38, was in the C-terminal helix. It was replaced with a proline. Except for the chimpanzee that also has an M38, all the other primates have a threonine at residue 38. Their study led them to conclude that changes leading to higher helical propensity result in more extensive protein lipid interactions (Myers et al., 2013). In another study, Smith et al. (2013) carried out a series of experiments in which the helices were studied separately as well as bi-helical constructs in which the helices were reversed. They found that the C-terminal helix was better in solubilizing phospholipids, but both helices were required for the ATP binding cassette transporter A1 (ABCA1) mediated cholesterol efflux from cells. Because previous studies had indicated that acidic residues in the C-terminal helices of apoA-I were involved in the action of ABCA1 (Natarajan et al., 2004), the possibility of this being true for the glutamates in apoC-I was explored by studying the difference between the wild type apo and constructs with a single mutation in these two different assays (Smith et al., 2013). The constructs were prepared replacing the glutamates at residues 33, 40, 44 and 51 with lysines. The results were essentially the same as for the wild type. Considering the high percentage of positively charged amino acids in apoC-I, one might wonder if they have any functional role. Comparing the data in Tables 9 and 10, most of the lysines and arginines are invariant across the different species. In only a few sites are there any major changes. As the data in Table 11 indicate, some of cationic sites did change with gene duplication. A comparison of the sequences of chimpanzee apoC-IA and apoC-IB indicates that all the changes are located in the sequence encoded by exon 3, 4 lysines (K10G, K12E, K21N and K30T) and an arginine (R39W). In the segment encoded by exon 4, the four lysines and the single arginine align. Whether these differences result in distinct functions for apoC-IA remains to be seen. However, it should also be noted that other residues in both forms have remained invariant during the course of evolution.In the 23 amino acid N-terminal sequence there are 13 invariant residues in both types of chimpanzee apoC-I and in the C-terminal helix 11 of the 14 are the same.

Future studies might find out what great apes and perhaps other primates are doing with two distinct forms of apoC-I. Currently it has been estimated that there are over 14000 pseudogenes in the human genome (Pei et al., 2012). How many of them, like the gene for apoC-IA, are in fact active genes in other primates? Once answered, the major problem will be the identification and characterization of the proteins derived from these active genes. These studies could very well provide new insight into the evolution and physiology of humans.

Acknowledgements

The authors wish to thank Leona G. Chemnick and Dr. Oliver A.

Ryder of the San Diego Zoo's Institute for Conservation Research for Endangered Species for providing us with primate plasmas for our mass spectral studies. The senior author (DLP) is particularly grateful to Dr. Jerzy Jurka and Dr. Kenji Kojima of the Genetic Information Research Institute for providing genomic data for the apolipoproteins of New World monkeys. Finally we wish to thank Prof. Kym Faull, Director of the Pasarow Mass Spectrometry Laboratory at UCLA as well as our other colleagues in the Pasarow Laboratory, including Dr. C.M. Ryan, Dr. L. Della Donna, S. Bassilian and P. Souda.

Compliance with ethics guidelines

Donald L. Puppione and Julian P. Whitelegge declare that they have no conflict of interest.

This manuscript is a review and does not involve a research protocol requiring approval by the relevant institutional review boards or ethics committees.

References

Alaupovic P, Kostner G, Lee D M, McConathy W J, Magnani H N (1972). Peptide composition of human plasma apolipoproteins A, B and C. Expos Annu Biochim Med, 31: 145–160

Alaupovic P, Lee D M, McConathy W J (1972). Studies on the composition and structure of plasma lipoproteins. Distribution of lipoprotein families in major density classes of normal human plasma lipoproteins. Biochim Biophys Acta, 260(4): 689–707

Allan C M, Walker D, Segrest J P, Taylor J M (1995). Identification and characterization of a new human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus. Genomics, 28(2): 291–300

Barker W C, Dayhoff M O (1977). Evolution of lipoproteins deduced from protein sequence data. Comp Biochem Physiol B, 57(4): 309–315

Benjwal S, Jayaraman S, Gursky O (2007). Role of secondary structure in protein-phospholipid surface interactions: reconstitution and denaturation of apolipoprotein C-I:DMPC complexes. Biochemistry, 46(13): 4184–4194

Bondarenko P V, Cockrill S L, Watkins L K, Cruzado I D, Macfarlane R D (1999). Mass spectral study of polymorphism of the apolipoproteins of very low density lipoprotein. J Lipid Res, 40(3): 543–555

Brown W V, Levy R I, Fredrickson D S (1969). Studies of the proteins in human plasma very low density lipoproteins. J Biol Chem, 244(20): 5687–5694

Dang Q, Taylor J (1996). *In vivo* footprinting analysis of the hepatic control region of the human apolipoprotein E/C-I/C-IV/C-II gene locus. J Biol Chem, 271(45): 28667–28676

Davison P J, Norton P, Wallis S C, Gill L, Cook M, Williamson R, Humphries S E (1986). There are two gene sequences for human apolipoprotein CI (apo CI) on chromosome 19, one of which is 4 kb from the gene for apo E. Biochem Biophys Res Commun, 136(3): 876–884

Deininger P (2011). Alu elements: know the SINEs. Genome Biol, 12 (12): 236–247

Deininger P L, Batzer M A (2002). Mammalian retroelements. Genome Res, 12(10): 1455–1465

Fitch W M (1977). Phylogenies constrained by the crossover process as illustrated by human hemoglobins and a thirteen-cycle, eleven-amino-acid repeat in human apolipoprotein A-I. Genetics, 86(3): 623–644

Freitas E M, Gaudieri S, Zhang W J, Kulski J K, van Bockxmeer F M, Christiansen F T, Dawkins R L (2000). Duplication and diversification of the apolipoprotein CI (APOCI) genomic segment in association with retroelements. J Mol Evol, 50(4): 391–396

Glazier F W, Tamplin A R, Strisower B, Delalla O F, Gofman J W, Dawber T R, Phillips E (1954). Human serum lipoprotein concentrations. J Gerontol, 9(4): 395–403

Gursky O (2001). Solution conformation of human apolipoprotein C-1 inferred from proline mutagenesis: far- and near-UV CD study. Biochemistry, 40(40): 12178–12185

Gustafson A, Alaupovic P, Furman R H (1966). Studies of the composition and structure of serum lipoproteins. Separation and characterization of phospholipid-protein residues obtained by partial delipidization of very low density lipoproteins of human serum. Biochemistry, 5(2): 632–640

Havel R J, Eder H A, Bragdon J H (1955). The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum. J Clin Invest, 34(9): 1345–1353

Herbert P N, Bausserman L L, Lynch K M, Saritelli A L, Kantor M A, Nicolosi R J, Shulman R S (1987). Homologues of the human C and A apolipoproteins in the Macaca fascicularis (cynomolgus) monkey. Biochemistry, 26(5): 1457–1463

Karlsson H, Leanderson P, Tagesson C, Lindahl M (2005). Lipoproteomics I: mapping of proteins in low-density lipoprotein using two-dimensional gel electrophoresis and mass spectrometry. Proteomics, 5(2): 551–565

Kasthuri R S, McMillan K R, Flood-Urdangarin C, Harvey S B, Wilson-Grady J T, Nelsestuen G L (2007). Correlation of a T45S variant of apolipoprotein C1 with elevated BMI in persons of American Indian and Mexican ancestries. Int J Obes (Lond), 31(8): 1334–1336

Knott T J, Eddy R L, Robertson M E, Priestley L M, Scott J, Shows T B (1984). Chromosomal localization of the human apoprotein CI gene and of a polymorphic apoprotein AII gene. Biochem Biophys Res Commun, 125(1): 299–306

Kushwaha R S, Hasan S Q, McGill H C Jr, Getz G S, Dunham R G, Kanda P (1993). Characterization of cholesteryl ester transfer protein inhibitor from plasma of baboons (*Papio* sp.). J Lipid Res, 34(8): 1285–1297

Lahiry P, Cao H, Ban M R, Pollex R L, Mamakeesick M, Zinman B, Harris S B, Hanley A J G, Huff M W, Connelly P W, Hegele R A (2010). APOC1 T45S polymorphism is associated with reduced obesity indices and lower plasma concentrations of leptin and apolipoprotein C-I in aboriginal Canadians. J Lipid Res, 51(4): 843–848

Lauer S J, Walker D, Elshourbagy N A, Reardon C A, Levy-Wilson B, Taylor J M (1988). Two copies of the human apolipoprotein C-I gene are linked closely to the apolipoprotein E gene. J Biol Chem, 263 (15): 7277–7286

Lee D M, Alaupovic P (1970). Studies of the composition and structure of plasma lipoproteins. Isolation, composition, and immunochemical characterization of low density lipoprotein subfractions of human plasma. Biochemistry, 9(11): 2244–2252

Li W H, Gu Z, Wang H, Nekrutenko A (2001). Evolutionary analyses of the human genome. Nature, 409(6822): 847–849

Lindgren F T (1975) Preparative ultracentrifugal laboratory procedures and suggestion for lipoprotein analysis. E.G Perkins (Ed.), Analysis of Lipids and Lipoproteins, American Oil Chemists Society, Champaign, IL pp. 204–224

Luo C C, Li W H, Chan L (1989). Structure and expression of dog apolipoprotein A-I, E, and C-I mRNAs: implications for the evolution and functional constraints of apolipoprotein structure. J Lipid Res, 30(11): 1735–1746

Lusis A J, Heinzmann C, Sparkes R S, Scott J, Knott T J, Geller R, Sparkes M C, Mohandas T (1986). Regional mapping of human chromosome 19: organization of genes for plasma lipid transport (APOC1, -C2, and-E and LDLR) and the genes C3, PEPD, and GPI. Proc Natl Acad Sci USA, 83(11): 3929–3933

McGill H C Jr, McMahan C A, Kushwaha R S, Mott G E, Carey K D (1986). Dietary effects on serum lipoproteins of dyslipoproteinemic baboons with high HDL1. Arteriosclerosis, 6(6): 651–663

McNeal C J, Chatterjee S, Hou J, Worthy L S, Larner C D, Macfarlane R D, Alaupovic P, Brocia R W (2013). Human HDL containing a novel apoC-I isoform induces smooth muscle cell apoptosis. Cardiovasc Res, 98(1): 83–93

Meyers N L, Wang L, Gursky O, Small D M (2013). Changes in helical content or net charge of apolipoprotein C-I alter its affinity for lipid/water interfaces. J Lipid Res, 54(7): 1927–1938

Natarajan P, Forte T M, Chu B, Phillips M C, Oram J F, Bielicki J K (2004). Identification of an apolipoprotein A-I structural element that mediates cellular cholesterol efflux and stabilizes ATP binding cassette transporter A1. J Biol Chem, 279: 24044–24055

Pastorcic M, Birnbaum S, Hixson J E (1992). Baboon apolipoprotein C-I: cDNA and gene structure and evolution. Genomics, 13(2): 368–374

Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu X J, Harte R, Balasubramanian S, Tanzer A, Diekhans M, Reymond A, Hubbard T J, Harrow J, Gerstein M B (2012). The GENCODE pseudogene resource. Genome Biol, 13(9): R51

Puppione D L, Ryan C M, Bassilian S, Souda P, Xiao X, Ryder O A, Whitelegge J P (2010). Detection of two distinct forms of apoC-I in great apes. Comp Biochem Physiol Part D Genomics Proteomics, 5 (1): 73–79

Raisonnier A (1991). Duplication of the apolipoprotein C-I gene occurred about forty million years ago. J Mol Evol, 32(3): 211–219

Rowold D J, Herrera R J (2000). Alu elements and the human genome. Genetica, 108: 57–72

Rozek A, Buchko G W, Cushley R J (1995). Conformation of two peptides corresponding to human apolipoprotein C-I residues 7-24 and 35-53 in the presence of sodium dodecyl sulfate by CD and NMR spectroscopy. Biochemistry, 34(22): 7401–7408

Scanu A M (1972). Structural studies on serum lipoproteins. Biochim Biophys Acta, 265(4): 471–508

Scanu A M, Edelstein C (2008). HDL: bridging past and present with a look at the future. FASEB J, 22(12): 4044–4054

Scott J, Knott T J, Shaw D J, Brook J D (1985). Localization of genes encoding apolipoproteins CI, CII, and E to the p13——cen region of human chromosome 19. Hum Genet, 71(2): 144–146

Segrest J P, Jackson R L, Morrisett J D, Gotto A M Jr (1974). A molecular theory of lipid-protein interactions in the plasma lipoproteins. FEBS Lett, 38(3): 247–258

Segrest J P, Jones M K, De Loof H, Brouillette C G, Venkatachalapathi Y V, Anantharamaiah G M (1992). The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. J Lipid Res, 33(2): 141–166

Shachter N S (2001). Apolipoproteins C-I and C-III as important modulators of lipoprotein metabolism. Curr Opin Lipidol, 12(3): 297–304

Shore V, Shore B (1968). Some physical and chemical studies on two polypeptide components of high-density lipoproteins of human serum. Biochemistry, 7(10): 3396–3403

Shulman R S, Herbert P N, Wehrly K, Fredrickson D S (1975). Thf complete amino acid sequence of C-I (apoLp-Ser), an apolipoprotein from human very low density lipoproteins. J Biol Chem, 250(1): 182–190

Smith L E, Segrest J P, Davidson W S (2013). Helical domains that mediate lipid solubilization and ABCA1-specific cholesterol efflux in apolipoproteins C-I and A-II. J Lipid Res, 54(7): 1939–1948

Wroblewski M S, Wilson-Grady J T, Martinez M B, Kasthuri R S, McMillan K R, Flood-Urdangarin C, Nelsestuen G L (2006). A functional polymorphism of apolipoprotein C1 detected by mass spectrometry. FEBS J, 273(20): 4707–4715

Zhang L H, Kotite L, Havel R J (1996). Identification, characterization, cloning, and expression of apolipoprotein C-IV, a novel sialoglycoprotein of rabbit plasma lipoproteins. J Biol Chem, 271(3): 1776–1783