Published in final edited form as:

Comp Biochem Physiol Part D Genomics Proteomics. 2010 March 1; 5(1): 73–79. doi:10.1016/j.cbd.2009.12.003.

Detection of two distinct forms of apoC-I in great apes

Donald L. Puppione^a, Christopher M. Ryan^b, Sara Bassilian^b, Puneet Souda^b, Xinshu Xiao^{a,c}, Oliver A. Ryder^d, and Julian P. Whitelegge^{a,b}

^aThe Molecular Biology Institute; The Pasarow Mass Spectrometry Laboratory, The Jane & Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA 90095, USA

^bDavid Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA

^cDept of Physiological Sciences, University of California, Los Angeles, CA 90095, USA

^dSan Diego Zoo's Institute for Conservation Research, Escondido, CA 92027, USA

Abstract

ApoC-I, the smallest of the soluble apolipoproteins, associates with both TG-rich lipoproteins and HDL. Mass spectral analyses of human apoC-I previously had demonstrated that in the circulation there are two forms, either a 57 amino acid protein or a 55 amino acid protein, due to the loss of two amino acids from the N-terminus. In our analyses of the apolipoproteins of the other great apes by mass spectrometry, four forms of apoC-I were detected. Two of these showed a high degree of identity to the mature and truncated forms of human apoC-I. The other two were homologous to the virtual protein and its truncated form that are encoded by a human pseudogene. In humans, the genes for apoC-I and its pseudogene are located on chromosome 19, the pseudogene being 2.5 kb downstream from the apoC-I gene. Based on the similarity between the apoC-I gene and the pseudogene, it has been concluded that the latter arose from the former as a result of gene duplication approximately 35 million years ago. Interestingly, the virtual protein encoded by the pseudogene is acidic, not basic like apoC-I. In the chimpanzee, there also are two genes for apoC-I, the one upstream encodes a basic protein and the downstream gene, rather than being a pseudogene, encodes an acidic protein (P86336). In addition to reporting on the molecular masses of great ape apoC-I, we were able to clearly demonstrate by “Top-down” sequencing that the acidic form arose from a separate gene. In our analyses, we have measured the molecular masses of apoC-I associated with the HDL of the following great apes: bonobo (*Pan paniscus*), chimpanzee (*Pan troglodytes*), and the Sumatran orangutan (*Pongo abelii*). Genomic variations in chromosome 19 among great apes, baboons and macaques as they relate to both genes for apoC-I and the pseudogene are compared and discussed.

Keywords

Great apes; top down sequencing; mass spectrometry; apolipoprotein C-I; human pseudogene

Corresponding author: Don L. Puppione, Ph.D., Boyer Hall, Molecular Biology Institute, UCLA, Los Angeles, CA 90095, USA.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Early studies of the protein moiety of human VLDL revealed a group of apolipoproteins with N-terminal amino acids, serine and threonine, distinct from those associated with the major apolipoproteins of LDL and HDL (Rodbell, 1958; Gustafson et al, 1964; Shore and Shore, 1964; Gustafson et al, 1966; Granda and Scanu, 1966). It was later found that trace amounts of these VLDL apolipoproteins also were associated with HDL (Shore and Shore, 1969). Electrophoretic analyses demonstrated that the apolipoprotein having N-terminal threonine and C-terminal serine, now known as apoC-I, was basic (Brown et al., 1969). Consistent with this observation, the 57 amino acid protein has since been found to contain 9 lysine and 3 arginine residues (Jackson et al., 1974; Schulman et al., 1975).

The human apoC-I gene is located on chromosome 19 along with the genes for apoE, apoC-IV, and apoC-II and a pseudogene of apoC-I (Lauer et al., 1988). All the genes have the same transcriptional orientation. The pseudogene, which arose from a duplication of apoC-I, is located between the apoC-I and apoC-IV genes (Lauer et al., 1988; Dang and Taylor, 1996). Based on the analyses of the sequence differences between the apoC-I gene and the pseudogene, it has been concluded that the duplication occurred before the divergence of Old and New World monkeys, between 25 and 35 million years ago (Luo et al., 1989; Raisonnier, 1991; Pastorcic et al., 1992).

Using mass spectrometry to obtain molecular values of the soluble apolipoproteins associated with human VLDL, two forms of apoC-I were observed (Bondarenko et al., 1999). In addition to the mature form, a post-translationally modified protein, minus a threonine and a proline from the N-terminus, was detected. The truncated apolipoprotein was designated apoC-I'.

In our studies, we have employed mass spectrometry to obtain accurate molecular masses of the soluble apolipoproteins associated with the HDL of different mammals (Puppione et al., 2005a, 2005b, 2006, 2008). In each of these previous reports, we have been able to demonstrate close agreement between the measured mass values and the calculated molecular weights, using entries in the genomic databases to obtain the protein sequences. In our study of dog apolipoproteins, we also performed "Top down" sequencing of apoC-I (Puppione et al., 2008). Using these same approaches, we have obtained values for the molecular masses of bonobo, chimpanzee and orangutan apoC-I. As was observed in humans, our analyses revealed both a mature and truncated form of apoC-I in these great apes. In the early phase of these studies, two other apolipoproteins with masses close, but not identical to apoC-I, were detected in the bonobo spectra. Upon examining the NCBI database, two mRNA entries for the apoC-I of the chimpanzee, a close relative of the bonobo, were located. One of these entries yields a basic 57 amino acid protein with a calculated molecular weight agreeing with the measured mass values for two of the bonobo apolipoproteins. On examining the second mRNA the resulting primary sequence also consisted of 57 amino acids, but the protein was acidic with a pI value of 4.82. The resulting molecular weight also was different, but it agreed with the other observed masses seen in the bonobo spectrum. Because of their charge differences, we propose that the acidic form be designated apoC-IA and the basic form apoC-IB. Following reverse-phase separation of bonobo apolipoproteins, a fraction enriched in the truncated acidic form, *i.e.* apoC-IA', was obtained. "Top down" sequencing verified that the protein was a product of the second gene for apoC-I. Subsequent analyses also enable us to detect apoC-IA in the chimpanzee and orangutan spectra. In addition to presenting mass and sequence data for the apoC-I of these great apes, we also discuss how the two forms of the mature protein are related to human apoC-I and its pseudogene. Genomic variations in chromosome 19 among great apes, baboons and macaques as they relate to both the human gene for apoC-I and the pseudogene are compared and discussed.

2. Methods

2.1 Source of great ape plasmas

The San Diego Wild Animal Park, a repository for frozen mammalian plasmas, provided samples from the following animals: bonobo (*Pan paniscus*) (male and 3 females); chimpanzee (*Pan troglodytes*) (4 females); and Sumatran orangutan (*Pongo abelii*) (male and female). The frozen plasmas were transported on dry ice to the Los Angeles campus of the University of California. The frozen plasmas (approximately 1.5 mL from each ape) were transported on dry ice to the Los Angeles campus of the University of California. Protocols and analyses, described below, were carried out on individual samples. The samples were not pooled

2.2 Ultracentrifugal isolation of HDL

Solution densities were adjusted as described previously (Schumaker and Puppione, 1986). The lipoproteins were separated using a TLA-100 rotor centrifuged in a Beckman Optima-TLX ultracentrifuge at 80,000 rev/min (240,000 *g* at r_{av}) at 20 °C. Initially, tubes containing 0.18 mL of plasma adjusted to a density of 1.063 g/mL were centrifuged for 4 h. After removing the top 0.060 mL, the infranatants were adjusted to a density of 1.210 g/mL. Following 4 h of centrifugation, the HDL were recovered in the top 0.040 mL. All salt solutions used for ultracentrifugation contained 0.04% Na₂ EDTA and 0.05% NaN₃.

2.3 Size exclusion chromatography separation and analysis by electrospray-ionization mass spectrometry (SEC ESI-MS)

The apolipoproteins were separated by size exclusion chromatography prior to analysis in the mass spectrometer (Whitelegge et al., 1998). HDL fractions were first dialyzed against a NaCl solution (density = 1.0063 g/mL). The dialyzed fractions then were acidified by mixing 0.01 mL with 0.09 mL of 90% formic acid immediately prior to SEC-MS. SEC-MS was performed in CHCl₃/MeOH/1% aqueous formic acid (4/4/1; v/v/v) using a Super SW 2000 column (4.6 × 300 mm, Tosoh Bioscience, Montgomeryville, PA, USA) at 0.250 mL/min and 40°C. Prior to interacting with the electrospray-ionization source, the column effluent was monitored with a UV detector set at 280 nm. Mass spectrometry (ESI-MS) was performed using a triple quadrupole instrument (API III, Applied Biosystems) (Whitelegge et al., 1998). Data were processed using MacSpec 3.3, Hypermass and BioMultiview 1.3.1 software (Applied Biosystems). The resulting molecular mass values were compared with calculated molecular weights derived from genomic entries in various databases that included the National Center for Biological Information (www.ncbi.nlm.nih.gov), the Genome Bioinformatics website of the University of California at Santa Cruz (UCSC) (<http://genome.ucsc.edu/>) and the Ensembl website (www.ensembl.org/index.html). Calculated values for molecular weight and pI were obtained using ProtParam at the proteomic server of the Swiss Institute of Bioinformatics (<http://ca.expasy.org/>).

2.4 Reverse-phase chromatography with online ESI-MS and fraction collection

Separation by reverse-phase chromatography was carried out to obtain fractions relatively enriched in apoC-I. Partial delipidation first was performed by mixing 0.1 mL of the HDL fraction with 1 mL of 80% acetone (− 20 °C, 1 h). Precipitated apolipoproteins were dissolved in 0.1 mL of 90% formic acid and loaded on a reverse phase column (PLRP/S 5 µm, 300 Å, 2 × 150 mm, Polymer Labs, Amherst, MA, USA) previously equilibrated at 95% A, 5% B (A, 0.1% TFA in water and B, 0.05% TFA in a 1:1 mixture of acetonitrile/isopropanol). The column was eluted with a dynamic gradient as described (Whitelegge et al., 2002). Column temperature was 40 °C and the flow rate was 0.1 mL/min. Volumes of the collected fractions were 0.1 mL. The eluent was passed through a UV detector (280 nm) prior to a liquid-flow splitter with fused silica capillaries to transfer liquid to both the ESI source (50cm) and the fraction collector (25

cm). Fractions were kept at -20°C before analysis by mass spectrometry. Analysis of the ESI-MS data obtained as described above enabled the correspondence between fraction number and apolipoprotein elution time to be determined.

2.5 Top-down sequencing of apolipoproteins

Top-down tandem mass spectrometry experiments were performed on a hybrid linear ion-trap 7-T FTICR mass spectrometer (LTQ-FT, Thermo Fisher Corporation, San Jose, CA, USA) equipped with an off-line nanospray source. HPLC fractions were individually loaded into 2 µm i.d. externally-coated nanospray emitters (New Objective Inc., Woburn, MA, USA) and desorbed using a spray voltage of between 1.2 - 1.4 kV (versus the inlet of the mass spectrometer). These conditions produced a flow rate of 20 - 50 nL/min (Whitelegge et al., 2008).

Ion transmission into the linear trap and further to the ion cyclotron resonance cell (ICR) was automatically optimized for maximum ion signal. The ion count targets for the full scan and MS² ICR experiments were 2×10^6 . The m/z resolving power of the instrument was set at 100,000 (defined by $m/\Delta m_{50\%}$ at m/z 400). Individual charge states of the multiply protonated protein molecular ions were selected for isolation and collisionally activated dissociation (CAD) in the linear ion trap followed by the detection of the resulting product ions in the ICR cell. For the CAD studies, the precursor ions were activated using a collision energy sufficient to dissociate greater than 50% of the precursor ion.

2.6 Data analysis with Prosight

All FTICR mass spectra, derived from an average of between 50 - 500 transient signals, were processed using XtractAll (Xcaliber 2.0, Thermo Scientific) to produce monoisotopic mass lists (s/n = 1.1, fit 0%, remainder 0%, modeled using Averagine elemental composition). Prosight PC (version 1.0; Thermo Scientific) software suites were used to first identify and then fully characterize the protein primary structure. The 'sequence tag compiler' function was used to generate short sequence tags for protein identification using the 'sequence tag' function. Once the protein was identified, the 'single protein mode' function was used to fully characterize the primary structure with custom post-translational modifications as required. A threshold of 10 ppm is used for matching in Prosight with typical RMS error on product ions of < 5 ppm. Nomenclature for assignment of peptide/protein fragments was according to Biemann (Bierman and Scoble, 1987; Bierman, 1989).

3. Results

3.1 Genomic entries for great ape apoC-I indicate that there are separate genes for two distinct apolipoproteins

Most soluble apolipoprotein genes consist of four exons and three introns. The first of these exons is non-coding. Exon 2 of human apoC-I gene encodes 20 residues of the 26 amino acid signal sequence and exon 3 encodes the remaining segment of the signal sequence plus the first 39 amino acids of the mature protein. Exon 4 encodes the final segment of a 57 mature apolipoprotein. This is also true for the chimpanzee.

In the NCBI database, there are two mRNA entries for chimpanzee apoC-I. Based on the listed coordinates on chromosome 19, the apolipoprotein encoded by the upstream gene (GI: 743864) would have a calculated molecular weight of 6657.6 Da and a pI of 7.93. For the other entry (GI: 743907), the calculated molecular mass had a value of 6715.5 Da for the mature protein and a pI of 4.82. In the UCSC database, two genes for orangutan apoC-I also can be located on chromosome 19. For the first with coordinates between 46,155,998 and 46,160,271, only exons 2 and 3 were found due to gaps in the DNA sequence. The other downstream gene

(46,168,619-46,173,063) contained three coding exons. This second orangutan gene yielded an apolipoprotein, having a calculated molecular mass of 6729.5 Da and a pI of 4.79.

The resulting sequences from the chimpanzee and orangutan genomic entries for both forms of apoC-I are shown in Table 1. The apolipoproteins are designated apoC-IB and apoC-IA based on their calculated pI values. Comparing the first 39 amino acids in the apoC-IB sequences encoded by exons 3, there are only three non-conserved positions, a serine and asparagine variation at residues 27 and 32 and a methionine and threonine variation at residue 38. The differences in the 57 amino acid sequences of apoC-IA sequences are less, with two conserved changes at residues, 3 and 49. On the other hand, there are 10 non-conserved changes if a comparison is made between apoC-IB and C-IA of the chimpanzee. In terms of the charged amino acids, apoC-IB contains 12 positive and 11 negative residues, whereas apoC-IA contains 9 positive and 12 negative residues. Comparing just the sequences encoding the first 39 amino acids of the orangutan apolipoproteins, there are 8 non-conserved changes.

3.2 Molecular masses of the two forms of great ape apoC-I

The apolipoproteins were analyzed by LC-MS. In the resulting spectra for two chimpanzees, molecular masses of 6658.0 Da and 6461.0 Da were observed. The larger value agrees with the calculated molecular weight of apoC-IB and the smaller one indicates that as reported for human apoC-I (Bondarenko et al., 1999), there was a post-translational modification with the loss of a threonine and a proline from the N-terminus (calculated molecular weight, 6459.3 Da). When the apolipoproteins of three bonobos were analyzed, these same two molecular mass values were noted, along with two others, 6717.0 Da and 6548.0 Da. These latter values agree with the calculated molecular weight of mature and truncated apoC-IA (6547.3 Da) determined from chimpanzee mRNA (GI: 743907). In the analysis of a female orangutan apolipoproteins, three masses could be identified. The minor peak at 6729.6 Da corresponds to the mature apoC-IA and the one at 6562 Da to the truncated form (calculated molecular mass 6561.3 Da). The measured mass of 6747.0 Da could be an oxidized apoC-IA with a methionine converted to methionine sulfoxide. Finally, the 6429.0 Da peak may be a truncated form of orangutan apoC-IB, but a final conclusion must await the future completion of the genomic sequence. Representative spectra are compared in Figure 1.

Taken together, these data indicate that the chimpanzee and bonobo probably have identical sequences for both apoC-IB and C-IA. Moreover, although not detected in the chimpanzee, apoC-IA was observed in both the bonobo and the orangutan spectra.

3.3 Reverse-phase ESI-MS analysis of great ape apoC-I

Using reverse-phase chromatography, it was possible to separate the apolipoproteins based on elution time. For the bonobo, apoC-IB was detected eluting at 41.9 min around the leading edge of a peak, identified with a small, but interfering amount of albumin. The spectrum, shown in Figure 2, shows the peaks associated with the mature and truncated forms of the apolipoprotein. A similar spectrum can be seen for apoC-IA in this figure as well. ApoC-IA eluted later at 58.5 min, associated with the leading edge of the major peak, identified as being apoA-I. In the case of the chimpanzee, both mature and truncated apoC-IA eluted at 53.8 min, but no apoC-IB was detected (Fig. 3). Both apolipoproteins were not detected when the orangutan chromatogram was analyzed.

3.4 Top down sequencing of bonobo apoC-IA

Fractions enriched in apoC-IA were analyzed by high-resolution mass spectrometry allowing us to perform dissociation experiments on the full-length apolipoprotein. After LC fractionation, full-length apoC-IA precursor ions were isolated in the mass spectrometer and subjected to CAD with the mass of the product ions recorded. The data were analyzed with

ProSight PC software (version 1.0; Thermo Scientific) allowing identification and characterization of primary structure until the intact mass was explained.

In our analyses, it was only possible to analyze an ion hypothesized to be bonobo apoC-IA' due to low intensity signals from the chimp ion assigned as apoC-IA'. The measured monoisotopic molecular mass of the bonobo protein, 6543.3188 Da, was within 2 ppm of the value calculated based upon the chimp sequence. Peak lists from CAD MSMS experiments performed on the +5 and +4 charge states (m/z 1310 & 1637) of the bonobo protein were pooled and matched to the chimp sequence with a 10 ppm tolerance. Ten b- and fourteen y- fragment ions matching the chimp sequence (See Fig. 4B) were assigned giving a very high confidence that the sequence of the bonobo protein is identical to that of the chimp (Pscore of $6e-49$ represents the probability of some other random sequence matching the data). However, it should be noted that transposition of amino acids within regions not covered by fragmentation or other isomeric possibilities (isoleucine and leucine are indistinguishable) can not be ruled out.

4. Discussion

Although apoC-I is present in trace amounts on HDL, measurements with the mass spectrometer enabled us to detect two forms of this apolipoprotein in the great apes and also to do top-down sequencing of a truncated acidic form. Our data have demonstrated that both bonobos and chimpanzees have identical masses for the two forms of mature apoC-I. The primary sequences obtained from genomic data for both of the chimpanzee apolipoproteins indicate that one form is basic and the other acidic. Furthermore, top-down sequencing of bonobo apoC-IA' demonstrates that the primary sequence obtained by translating chimpanzee mRNA entry (GI: 743907) are the same.

As already noted, the two chimpanzee mRNA entries indicated that the gene for apoC-IB is upstream from the gene for apoC-IA on chromosome 19. Because on human chromosome 19, the gene for apoC-I also is located upstream from the pseudogene, the apolipoprotein sequences, shown in Table 2, were compared with the human sequences. Ignoring the 26 amino acid signal sequence where the stop codon occurs at the -2 residue, the two human sequences are 82.5% similar, with 38 invariant and 9 conserved residues. However, comparing these human sequences to those of the chimpanzee in Table 1, one observes a high degree of identity. There is a single difference at residue 31, with an asparagine in human apoC-I and a serine in chimpanzee apoC-IB. The virtual protein and chimpanzee apoC-IA differ at two residues. The chimpanzee sequence has an asparagine and an isoleucine instead of a tyrosine and a threonine at residues 21 and 57, respectively.

Although there currently are sizable gaps in the orangutan genome, we were able to locate the three exons encoding apoC-IA on chromosome 19 (46,168,619-46,173,063). The orangutan sequence differs slightly from the chimpanzee apoC-IA with an aspartate instead of a glutamate at residue 3 and an arginine instead of a lysine at residue 48 (Table 1). When we examined chromosome 19 of the rhesus monkey (*Macaca mulatta*) in the UCSC database, we found four exons located between the genes for apoE and apoC-IV. The upstream exons encoded the 26 amino acid signal sequence and the first 39 amino acids of a mature apoC-I. The downstream exons had DNA sequences that were very similar to the pseudogene and the genes for the great apes apoC-IA. The first of these downstream exons aligns quite well with exon 2 of both the pseudogene and the great apes, except the rhesus exon is missing 6 nucleotides (Table 3). Using the signal sequences in Table 1 for comparison, the resulting rhesus signal sequence would be missing a proline and a valine at residues 8 and 9. Comparing the second exon with exon 3 of the pseudogene, the similarity was 85.2%. Whether these rhesus exons are actually involved

with encoding a protein, as we report here for the great apes, remains to be seen. In any case, unlike exon 3 of the human pseudogene, this second rhesus exon does not contain a stop codon.

The duplication event giving rise to the pseudogene has been proposed to have taken place prior to the divergence of Old and New World monkeys. In a previous study of baboons (*Papio anubis*), the structure of the apoC-I gene, including the sequences of the four exons and the three introns, was reported (Pastorcic et al., 1992). Using Southern blots to characterize restriction fragments, Pastorcic and colleagues were able to identify those integral to the apoC-I gene and a few others that were not predicted. Based on these latter observations, these authors concluded that they possibly had detected a gene corresponding to the human pseudogene. Because baboons and macaques are related phylogenetically, instead of a pseudogene they possibly detected exons similar if not identical to the downstream exons shown in Table 3 for the rhesus.

It had previously been suggested that the pseudogene became nonfunctional after its formation 35 million years ago (Luo et al., 1989). However, our observation of two forms of apoC-I in the great apes would indicate that the change in the codon for glutamine to a stop codon in exon 3 of the duplicated gene did not take place until after the bonobo-chimpanzee branch diverged from the human lineage, approximately 6 million years ago (Steiper and Young, 2006).

In other mammals, like mice and dogs, apoC-I is basic and slightly larger with 62 amino acids (Puppione et al., 2006, 2008). In a mouse study, Hoffer and colleagues did not locate any gene comparable to a pseudogene (Hoffer et al., 1993) and there have been no reports of one in dogs. It will be interesting to see if this is also the case when the genome of the marmoset and other New World monkeys have been completed. The fact that other mammals and humans have only a basic form of apoC-I raises the question why do great apes and perhaps, baboons and macaques have an acidic form as well. As can be seen from Tables 1 and 2, a sequence comparison shows very little differences in the regions encoded by exons 2 and 4. In a recent report, James and colleagues have found that in the region encoded by exon 4, the presence of a phenylalanine at residues 42 and 56 plays an important role in the interaction of human apoC-I with synthetic phospholipids (James et al., 2009). Phenylalanine is present in these same positions in not only apoC-IB but also apoC-IA of great apes (Table 1). The major differences between apoC-IB and apoC-IA, however, occur in the region encoded by exon 3. Because transgenic mice expressing human apoC-I develop hypertriglyceridemia (Berbée et al., 2005), it has been suggested that the presence of the stop codon in exon 3 of the human pseudogene arose from selective pressure to control levels of this apolipoprotein (Freitas et al., 2000). But the human gene in these mice encodes a basic protein, not a gene that encodes an acidic apolipoprotein. Based on the difference in their net charge, there is no reason to assume that the acidic form will have the same protein interactions currently ascribed to human apoC-I. Moreover, it also should be noted that even though great apes have two different forms of apoC-I, none of the plasma samples that were analyzed in our study were lactescent.

Great apes are endangered and doing detailed physiological experiments on these magnificent mammals is not possible. Other approaches may provide the answer to the function of the acidic form of apoC-I. Perhaps, in the future a transgenic mouse study could be carried out using the gene for apoC-IA from one of the great apes. The resulting data may very well indicate that the presence of a pseudogene in humans has nothing to do with controlling levels of triglycerides in the blood.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The senior author (DLP) expresses his appreciation to Drs. Duilio Cascio and Luis Arturo Medrano-Soto for many hours of discussion on the comparative genomics of the great apes. The authors thanks Leona G. Chemnick of the San Diego Zoo's Institute for Conservation Research for Endangered Species program for her assistance in the transferring of plasma samples. UCLA support from NIH grants (R21 RR021913-01, P01 NS049134-01, U19 AI067769) is gratefully acknowledged. NSF, the Pasarow Family and the W.M. Keck Foundation also are thanked for funds toward purchase of instruments.

References

- Berbée JFP, van der Hoogt CC, Sundararaman D, Havekes LM, Rensen PCN. Severe hypertriglyceridemia in human APOC1 transgenic mice is caused by apoC-I-induced inhibition of LPL. *J Lipid Res* 2005;46:297–306. [PubMed: 15576844]
- Bieman K, Scoble HA. Characterization by tandem mass spectrometry of structural modifications in proteins. *Science* 1987;237:992–998. [PubMed: 3303336]
- Biemann K. Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom* 1988;16:99–111. [PubMed: 3072035]
- Bondarenko PV, Cockrill SL, Watkins LK, Cruzado ID, Macfarlane RD. Mass spectral study of polymorphism of the apolipoproteins of very low density lipoprotein. *J Lipid Res* 1999;40:543–555. [PubMed: 10064743]
- Brown WV, Levy RI, Fredrickson DS. Studies of the proteins in human plasma very low density lipoproteins. *J Biol Chem* 1969;244:5687–5694. [PubMed: 4981584]
- Dang Q, Taylor J. 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* 1996;271:28667–28676. [PubMed: 8910501]
- Freitas EM, Gaudieri S, Zhang WJ, Kulski JK, van Bockxmeer FM, Christiansen FT, Dawkins RL. Duplication and diversification of the apolipoprotein CI (APOCI) genomic segment in association with retroelements. *J Mol Evol* 2000;50:391–396. [PubMed: 10795830]
- Granda JL, Scanu A. Solubilization and properties of the apoproteins of the very low- and low-density lipoproteins of human serum. *Biochemistry* 1966;5:3301–3308. [PubMed: 4961658]
- Gustafson A, Alaupovic P, Furman RH. Studies of composition and structure of serum lipoproteins: Physical-chemical characterization of phospholipid-protein residues obtained from very-low-density human serum lipoproteins. *Biochim Biophys Acta* 1964;84:767–769. [PubMed: 14266264]
- Gustafson A, Alaupovic P, Furman RH. 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* 1966;5:632–640. [PubMed: 4957397]
- Hoffer MJ, van Eck MM, Havekes LM, Hofker MH, Frants RR. The mouse apolipoprotein C1 gene: structure and expression. *Genomics* 1993;18:37–42. [PubMed: 8276416]
- Jackson RL, Sparrow JT, Baker HN, Morrisett JD, Taunton OD, Gotto AM Jr. The primary structure of apolipoprotein-serine. *J Biol Chem* 1974;249:5308–5313. [PubMed: 4369340]
- James PF, Dogovski C, Dobson RCJ, Bailey MF, Goldie KN, Karas JA, Scanlon DB, O'Hair RAJ, Perugini MA. Aromatic residues in the C-terminal helix of human apoC-I mediate phospholipid interactions and particle morphology. *J Lipid Res* 2009;50:1384–1394. [PubMed: 18984910]
- Lauer SJ, Walker D, Elshourbagy NA, Reardon CA, Levy-Wilson B, Taylor JM. Two Copies of the Human Apolipoprotein C-I Gene Are Linked Closely to the Apolipoprotein E Gene. *J Biol Chem* 1988;263:7277–7286. [PubMed: 2835369]
- Luo CC, Li WH, Chan L. 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* 1989;30:1735–1746. [PubMed: 2515239]
- Pastorcic M, Birnbaum S, Hixson JE. Baboon apolipoprotein C-I: cDNA and gene structure and evolution. *Genomics* 1992;13:368–374. [PubMed: 1612596]
- Puppione DL, Whitelegge JP, Yam LM, Schumaker VN. Mass spectral analysis of pig apo HDL: Identification of pig apoA-II, a dimeric apolipoprotein. *Comp Biochem Physiol B* 2005a;141:89–94. [PubMed: 15820138]

- Puppione DL, Whitelegge JP, Yam LM, Bassilian S, Schumaker VN, MacDonald MH. Mass spectral analysis of domestic and wild equine apoA-I and A-II: detection of unique dimeric forms of apoA-II. *Comp Biochem Physiol B* 2005b;142:369–373. [PubMed: 16230041]
- Puppione DL, Yam LM, Bassilian S, Souda P, Castellani LW, Schumaker VN, Whitelegge JP. Mass spectral analysis of the apolipoproteins on mouse high density lipoproteins. Detection of post-translational modifications. *Biochim Biophys Acta* 2006;1764:1363–1371. [PubMed: 16876491]
- Puppione DL, Yam LM, Bassilian S, Souda, MacDonald MH, Halgand F, Whitelegge JP. Mass spectral analysis of the apolipoproteins on dog (*Canis lupus familiaris*) high density lipoproteins. Detection of apolipoprotein A-II. *Comp Biochem Physiol D* 2008;3:290–296.
- Raisonnier A. Duplication of the apolipoprotein C-I gene occurred about forty million years ago. *J Mol Biol* 1991;32:211–219.
- Rodbell M. N-terminal amino acid and lipid composition of lipoproteins from chyle and plasma. *Science* 1958;127:701–702. [PubMed: 13529038]
- Schumaker VN, Puppione DL. Sequential flotation ultracentrifugation. *Methods Enzymol* 1986;128:155–170. [PubMed: 3724500]
- Segrest JP, Jones MK, de Loof H, Brouillette CG, Venkatachalapathi YV, Ananthramiah GN. The amphipathic helix in the exchangeable apolipoproteins: a review of secondary structure and function. *J Lipid Res* 1992;33:141–166. [PubMed: 1569369]
- Shore B, Shore V. Some physical and chemical properties of the lipoproteins produced by lipolysis of human serum S₂₀₋₄₀₀ lipoproteins by post-heparin plasma. *J Atherosclerosis Res* 1964;2:104–114.
- Shore B, Shore V. Isolation and characterization of polypeptides of human serum lipoproteins. *Biochemistry* 1969;8:4510–4516. [PubMed: 5353112]
- Shulman RS, Herbert PN, Wehrly K, Fredrickson DS. The complete amino acid sequence of C-I (apoLp-Ser), an apolipoprotein from human very low density lipoproteins. *J Biol Chem* 1975;250:182–190. [PubMed: 166984]
- Steiper ME, Young NM. Primate molecular divergence dates. *Mol Phylogenet Evol* 2006;41:384–394. [PubMed: 16815047]
- Whitelegge JP, Gundersen C, Faull KF. Electrospray-ionization mass spectrometry of intact intrinsic membrane proteins. *Protein Sci* 1998;7:1423–1430. [PubMed: 9655347]
- Whitelegge JP, le Coutre J, Lee JC, Engel CK, Privé GG, Faull KF, Kaback HR. Towards the bilayer proteome, electrospray-ionization mass spectrometry of large intact transmembrane proteins. *Proc Natl Acad Sci USA* 1999;96:10695–10698. [PubMed: 10485888]
- Whitelegge JP, Zhang H, Aguilera R, Taylor RM, Cramer WA. Full subunit coverage liquid chromatography electrospray ionization mass spectrometry (LCMS+) of an oligomeric membrane protein: cytochrome b(6)f complex from spinach and the cyanobacterium *Mastigocladus laminosus*. *Mol Cell Proteomics* 2002;1:816–827. [PubMed: 12438564]
- Whitelegge JP, Zabrouskov V, Halgand F, Souda P, Bassilian S, Yan W, Wolinsky L, Loo JA, Wong DT, Faull KF. Protein-Sequence Polymorphisms and Post-translational Modifications in Proteins from Human Saliva using Top-Down Fourier-transform Ion Cyclotron Resonance Mass Spectrometry. *Int J Mass Spectrom* 2007;268:190–197. [PubMed: 19050733]

Abbreviations

apoC-I'	truncated apoC-I minus the first two amino acid from the N-terminus
apoC-IB	the basic form of apoC-I in non-human great apes
apoC-IA	the acidic form of apoC-I in non-human great apes
FTICR-MS	Fourier-transform Ion Cyclotron resonance mass spectrometry

Figure 1.

Following size-exclusion chromatography multiple forms of apoC-I were observed in electrospray-ionization mass spectral profiles of great apes. Mature (6658.0 Da) and truncated (6461.0 Da) chimpanzee apoC-IB (top), mature and truncated bonobo apoC-IB (6658.0 and 6460.0 Da) and C-IA (6717.0 and 6548.0 Da) (middle) and mature and truncated orangutan apoC-IA (6729.6 and 6562.0 Da) and possibly oxidized apoC-I (6747. Da). Ordinary units of intensity are linear and arbitrary; the abscissa units of average molecular mass are in Daltons.

Figure 2.

Following reverse-phase chromatography, mature and truncated apoC-IB (upper) and mature and truncated apoC-IA were observed in electrospray-ionization mass spectral profile of bonobo apolipoproteins. Ordinary units of intensity are linear and arbitrary; the abscissa units of average molecular mass are in Daltons.

Figure 3.

Following reverse-phase chromatography, mature and truncated apoC-IA were observed in electrospray-ionization mass spectral profile of chimpanzee apolipoproteins. Ordinary units of intensity are linear and arbitrary; the abscissa units of average molecular mass are in Daltons.

Figure 4.

Top-down high-resolution mass spectrometry of bonobo apoC-IA'. An HPLC fraction containing a protein of monoisotopic mass 6543.3188 Da was analyzed by nano-electrospray ionization on a hybrid linear ion-trap Fourier-transform ion cyclotron resonance mass spectrometer. A five-charge precursor ion (m/z 1310) was isolated (inset A) and subsequently subjected to CAD in the linear ion-trap prior to transmission of product ions to the cyclotron cell with mass analysis at 100,000 resolution. The complex product ion spectrum was matched to the chimpanzee apoC-IA' sequence using Prosight PC software (version 1; Thermo Scientific). The b-(top left slash) and y- (lower right slash) ion coverage achieved with the peak-list from the tandem mass spectrum (shown in A) combined with the peak-list from another CAD experiment on the four-charge precursor at 1637 Da (not shown) is shown on the sequence map below (B). Selected product ions were annotated on the figure. A full table of assigned product ions is shown in Table S1 with average RMS deviation between measured and calculated mass of less than 4.7 ppm.

Table 2

Sequences of human apoC-I and the virtual protein encoded by the pseudogene

		Signal sequences																																							
apoC-I	M	R	L	F	L	S	L	P	V	L	V	V	V	V	L	S	I	V	L	E	G							P	A	P	A	Q	G								
pseudogene	M	R	L	F	L	S	L	P	V	L	V	V	V	V	L	S	M	V	L	E	G							P	A	P	A	***	G								
apoC-I	T	P	D	V	S	S	A	L	D	K	L	K	E	F	G	N	T	L	E	D	K	A	R	E	L	I	S	R	I	K	Q	S	E	L	S	A	K	M	R		
pseudogene	A	P	E	V	S	N	P	F	D	G	L	E	E	L	G	K	T	L	E	D	Y	T	R	E	F	I	N	R	I	T	Q	S	E	L	P	A	K	M	W		
apoC-I	E	W	F	S	E	T	F	Q	K	V	K	E	K	L	K	K	I	D	S																						
pseudogene	D	W	F	S	E	T	F	R	K	V	K	E	K	L	K	K	T	D	S																						

The top rows show the 26 amino acid signal sequences; middle rows show the initial 39 amino acid sequences of mature protein encoded by exon 3; last rows show the terminal 18 amino acids encoded by exon 4. The colors correspondence is the same as in Table 1. The stop codon in the pseudogene signal sequence is indicated by three asterisks.

Table 3

Alignment of two rhesus exons with exons 2 and 3 of the human pseudogene and of the genes for apoC-IA of chimpanzee and orangutan

Exon 2	
Human:	50122073-50122130
Chimp:	50523956-50524013
Orang:	46168619-46168676
Rhesus:	51354365-51354416
ATGAGGCTCTTCCTGTCGCTCCCGGTCCTGGTGGTGGTTCTGTCGATCGTCTTGGAAAG ATGAGGCTCTTCCTGTCGCTCCCGGTCCTGGTGGTGGTTCTGTCGATGGTCTTGGAAAG ATGAGGCTCTTCCTGTTGCTCCCGGTCCTGGTGGTGGTTCTGTCGATGGTCTTGGAAAG ATGAGGCTCTTCCTGTCGC-----TCCTGGTGGTGGTTCTGTCGATGGTCTTGGAAAG ***** **	
Exon 3	
Human:	50122892-50122951
Chimp:	50524506-50524565
Orang:	46169439-46169498
Rhesus:	51355172-51355231
GCCCAGCCCCAGCC <i>TA</i> GGGGCTCCAGAAGTCTCCAACCCCTTTGATGGCCTGGAGG GCCCAGCCCCAGCCCAGGGGGCTCCAGAAGTCTCCAACCCCTTTGATGGCCTGGAGG GCCCAGCCCCAGCCCAGGGGGCTCCGGACGTCTCCAACCCCTTTGATGGCCTGGAGG GCCCAACCCAGCCAGGGGGTTCCAGACGTCTCCAACCCCTTTGATGTCCTGAAGG ***** **	
Human:	50122952-50123011
Chimp:	50524566-50524625
Orang:	46169499-46169558
Rhesus:	51355232-51355290
AGTTAGGAAAGACCTGGAGGACTACACTCGGGAATTCATCAACCGCATCACACAGAGTG AGTTAGGAAAGACCTGGAGGACAACACTCGGGAATTCATCAACCGCATCACACAGAGTG AGTTAGGAAAGACCTGGAGGACAACACTCGGGAATTCATCAACCGCATCACACAGAGTG AGTTTGAAAGACCTGGAGGACAATG-TTGGGGATTATCCACCTCATCACACAGAGTG ***** *	
Human:	50123012-50123027
Chimp:	50524626-50524641
Orang:	46169559-46169574
Rhesus:	51355291-51355306
AACTTCCTGCCAAGATGTG AACTTCCTGCCAAGATGTG AACTTCCTGCCAAGATGTG AACTTCCTGCAAGACACG *****	

The coordinates on chromosome 19 of each primate are shown immediately to the left of the corresponding DNA sequences. The stop codon in exon 3 of the human pseudogene is indicated with red italicized letters.