INTRODUCTION

Atherosclerosis is the major cause of mortality and morbidity in the Western societies due to the development of clinical events such as myocardial infarction and stroke. The role of lipoprotein particles in coronary heart disease (CHD) is well established. Elevated plasma low-density lipoprotein (LDL)-cholesterol levels are associated with CHD [1], whereas there is a strong inverse relationship between high-density lipoprotein (HDL)-cholesterol levels and the incidence of CHD [2]. Endothelial cells are the first cells of the vessel wall to come in contact with lipoprotein particles in blood and there is growing evidence that these cells have a major effect on lipid metabolism [3]. Endothelial cells not only express a variety of lipoprotein receptors, which suggests a direct interaction with lipoprotein particles, but also can bind many lipases expressed in other tissues at the cell surface that have the ability to hydrolyze lipoprotein triglycerides and phospholipids [3].

We previously isolated APOL3 (CG12-1) cDNA and now describe the isolation of APOL1 and APOL2 cDNA from an activated endothelial cell cDNA library and show their endothelial-specific expression in human vascular tissue. APOL1–APOL4 are clustered on human chromosome 22q13.1, as a result of tandem gene duplication, and were detected only in primates (humans and African green monkeys) and not in dogs, pigs, or rodents, showing that this gene cluster has arisen recently in evolution. The specific tissue distribution and gene organization suggest that these genes have diverged rapidly after duplication. This has resulted in the emergence of an additional signal peptide encoding exon that ensures secretion of the plasma high-density lipoprotein-associated APOL1. Our results show that the APOL1–APOL4 cluster might contribute to the substantial differences in the lipid metabolism of humans and mice, as dictated by the variable expression of genes involved in this process.

Key Words: apolipoprotein L, gene cluster, duplication, evolution, endothelial cells, atherosclerosis, vascular wall, mouse lipid metabolism

RESULTS

Isolation of APOL1–APOL3 cDNAs

Screening of the activated HUVEC cDNA library with a probe from the 3’ end of APOL3 led to the identification of two additional homologous cDNAs. One of these contained the complete coding sequence of APOL1 with an open reading frame of 398 amino acids (Fig. 1). The first 27 amino acids were found to be a secretory signal peptide when analyzed by GCG SPScan. This signal peptide is 15 residues longer than that previously published, due to an additional methionine (start codon) in-frame and upstream of the previously suggested initiating methionine [5]. The amino-terminal sequence of the predicted mature protein, after cleavage of this putative
signal peptide, matches exactly the sequence of the secreted APOL1 that was determined by direct N-terminal microsequencing [5]. Analysis of the chromosomal sequence of APOL1 revealed that all intron–exon boundaries conform to the GT/AG rule (Table 1), except for the sequence of 0.7 kb that lacks from the last exon of our 2.2 kb cDNA of APOL1. As the size of the APOL1 mRNA by northern blot analysis is approximately 3 kb, this sequence probably does not represent a genuine intron, but might have been spliced in our HUVEC cDNA because of the specific genotype of the donor. This apparent polymorphism, leading to a smaller cDNA, will not affect the encoded protein because it lies within the 3' UTR. The composite sequence of APOL1, now being 2.9 kb, is very close to the actual size of the mRNA as observed by northern blot analysis.

The second clone constituted a partial cDNA with an incomplete open reading frame lacking the 5' sequences. Searching dbEST we found an IMAGE consortium cDNA clone (GenBank acc. no. AA531428) [8] containing the 5' portion of the cDNA and the complete coding sequence. Although this clone lacked part of the 3' end due to internal priming, it shared an 878-bp overlap with our incomplete clone. Combining these two sequences revealed the full-length sequence of APOL2 cDNA, which was 2403 bp, encoding a novel protein of 337 amino acids (Fig. 1). The 3' UTR of APOL2 contained two polyadenylation signals (AATAAA) at positions 1971 and 2383, which may lead to two different transcripts. Indeed, both types of transcripts are present in dbEST. APOL1 shows an N-terminal extension compared with the other sequences due to the presence of an upstream start codon, which is in-frame with that used by APOL2 and APOL3 (Fig. 1). This part contains the signal peptide and apparently ensures the secretion of APOL1. Neither APOL2 nor APOL3 contained a consensus signal peptide.

Genomic Organization
All three genes are clustered on chromosome 22q13.1 on a region of 130 kb, between nucleotides 33,178,000 and 33,308,000, as can be viewed at the Genome browser (August 2001 freeze; http://www.genome.ucsc.edu/; Fig. 2). There is an additional highly homologous region of 15.7 kb between APOL2 and APOL3, corresponding to APOL4. Dot-plot analysis of the genomic sequences of the members of this gene cluster revealed a high homology throughout the genomic sequences including the intronic sequences, implying genomic duplication (Fig. 3A). The previously reported APOL5 and APOL6 are only distantly related to the APOL1–APOL4 cluster, as homology is restricted to the cDNA.

<table>
<thead>
<tr>
<th>Intron</th>
<th>Gene</th>
<th>Intron–exon boundaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>APOL1</td>
<td>ATTCCCTTGTaaagtc...tcttaacGAGGAGGC</td>
</tr>
<tr>
<td>2</td>
<td>APOL2</td>
<td>ATCCCTTGTaaagtc...ctttactAGACAAGAG</td>
</tr>
<tr>
<td>3</td>
<td>APOL3</td>
<td>GACCTCTGTaaagtc...ctttactAGACAAGAG</td>
</tr>
<tr>
<td>4</td>
<td>APOL1</td>
<td>TGCCCAAGGTaagct...tcttactGTGACGAG</td>
</tr>
</tbody>
</table>

Capital letters represent exons and small letters are introns. All intron–exon boundaries conform to the gt/ag rule except for the last part of APOL1.
level with no homology in the intronic regions (Fig. 3B). APOL3 spans approximately 21 kb of the genomic DNA, whereas APOL2 and APOL1 span 14 and 15 kb, respectively. This difference is mainly caused by the length of the first intron, which is 11, 6, and 4 kb in APOL3, APOL2, and APOL1, respectively. The first intron of APOL3 is also the only non-homologous region of the genomic sequence, when compared with APOL1 or APOL2, and contains seven Alu repeats.

Expression Profile of the APOL Gene Family

Because of the high homology among the APOL genes, specific probes with low homology were chosen and found by Southern blot analysis to yield less than 1% cross hybridization (data not shown). Northern blot analysis showed APOL1, APOL2, and APOL3 mRNA to be upregulated in HUVEC after prolonged tumor necrosis factor (TNF)-α stimulation (Fig. 4). Northern blot analysis with a probe derived from a putative exon from APOL4 did not give a detectable signal in cultured endothelial cells. When normalized for β-actin expression, APOL1 was more than 10-fold upregulated after 48 hours and APOL2 was upregulated more than sixfold after 6 hours. As previously reported, APOL3 was also upregulated by 10-fold after 6 hours.

To assess the tissue specificity of these genes, we carried out multiple tissue northern blot analysis with the same probes. For all three genes, a single hybridizing band of the expected size was observed (Fig. 5A), which demonstrates both the specificity of these probes and the existence of a single prominent splice form. An additional band (2.7 kb) observed for APOL2 in brain only is in agreement with the presence of two polyadenylation sites in the transcript. Again, a similar experiment with an exon from APOL4 failed to
produce a detectable signal, indicating that no genuine mRNA is transcribed in any of the tissues examined. We quantified the hybridization signals after normalization for the amount of poly(A)+ RNA loaded in each lane (Figs. 5A and 5B). The expression of APOL1 mRNA (3 kb) was high in placenta, lung, and liver, and low in the kidney and heart, but there was virtually no expression in pancreas. The tissue distribution of APOL3 mRNA (2.3 kb) expression was similar to that of APOL1. In contrast, APOL2 mRNA (2.4 kb) was mostly expressed in brain and to a lesser amount in heart, with very low expression levels in other tissues (Fig. 5B).

In situ hybridization on human vascular tissue was carried out to evaluate the cell specificity of the expression of the various APOL1 transcripts in vivo. As a positive control, von Willebrand factor (vWF) was used, which is exclusively expressed by the endothelial cells in the vessel wall [4]. APOL3 mRNA and low levels of APOL2 mRNA were specifically expressed by endothelial cells, with no detectable expression in the neointima or the media (Fig. 6). APOL1, in addition to its endothelial cell expression, was also expressed by isolated cells in the shoulder region of the lesion. The location, morphology, and number of these cells suggested that they are most probably macrophages, a cell type known to also express other apolipoproteins like APOE.

APOL Expression Across Species

Presently, several model organisms are used to study the effects of aberrant plasma lipid profiles on cardiovascular disease, most notably the mouse and rabbit. We searched the NCBI EST databases for expressed APOL homologs in other species, but were unable to find any. Therefore, we determined the genomic presence of this gene family in different species using specific probes for Southern blot analysis, as shown in Fig. 7. As expected, the evolutionary well-conserved α-tubulin gene was detected in all species tested under the low stringency hybridization conditions used. APOL1, APOL2, and APOL3 each show specific bands in humans corresponding to the sizes of the predicted restriction fragments. Unlike the α-tubulin gene, however, the APOL gene family is only detected in humans and monkeys under identical low-stringency hybridization conditions. When less unique sequences from APOL1–APOL3 were used, each probe detected its paralogs in primates, but still no specific bands were detected in other species (data not shown).

**DISCUSSION**

Although endothelial cells are the first cells of the vessel wall to come in contact with lipoprotein particles in blood, their role in local lipid metabolism is not fully understood. It is generally accepted that endothelial injury caused by various agents, such as oxidized LDL, is the first step in the chronic inflammatory process in the vessel wall leading to atherosclerotic plaques. The APOL1 protein was first isolated from HDL particles and was reported to be expressed specifically by the pancreas [5]. We have described here the isolation of the complete cDNA encoding APOL1 by conventional cDNA cloning and shown expression of APOL1 mRNA in liver, heart, lung, and placenta, but virtually no expression in pancreas. In vascular tissue, APOL1 was specifically expressed by endothelial cells. Due to the positive correlation between plasma levels of APOL1 and plasma triglyceride levels, it was suggested that APOL1 might have a role in
lipolysis of triglyceride-rich particles [9]. The site of APOL1 expression is consistent with its possible role in lipolysis because this process takes place at the luminal surface of endothelial cells [3]. Furthermore, the expression of all three members of the APOL gene family was upregulated by TNF-α/H9251, which is a potent proinflammatory cytokine. During infection and inflammation, plasma lipid profiles show characteristic proatherogenic changes such as increased triglyceride levels, which are mostly due to decreased lipolysis [10]. Therefore, the effect of TNF-α on the expression of the APOL gene family could be of importance in regulating the lipid profile in vivo during inflammation.

Two different groups recently reported the PCR cloning of APOL gene family members and their mapping to human chromosome 22q13.1 [6,7]. Using a conventional cloning procedure, we isolated APOL1–APOL3 from an endothelial cell library and show expression of these genes in vivo. We did not detect APOL4 expression in the tissues tested here. However, using RT-PCR, APOL4 expression has been reported by others [6,7]. Hence, the importance of APOL4 expression in vivo remains to be established. Although splice variants of APOL1–APOL4 have been reported [7], we did not detect these, indicating that in vivo the reported splice variants for the APOL gene family are apparently expressed at a very low level, beyond the detection limit of northern blot analysis. One of the APOL3 splice variants (APOL3c), although not detected (Fig. 5A), was also found in our endothelial cell cDNA library. Using an in vitro transcription-translation assay, protein expression was only detected for APOL3 and not for APOL3c (data not shown). These observations cause uncertainties on the biological role of the reported splice variants, which might originate from the recent emergence of this gene cluster.

The APOL gene family consists of the APOL1–APOL4 cluster and the APOL5–APOL6 cluster [6,7]. There is a 400 kb intergenic region between these two clusters that contains non-related genes (genome browser at http://www. genome.ucsc.edu/). Four observations indicate that the APOL1–APOL4 cluster is the result of recent tandem duplication. First, dot-plot analysis of the genomic sequences of the APOL1–APOL4 cluster revealed a sustained high homology among these genes, including the intronic regions (Fig. 3A), whereas APOL5 and APOL6 show homology to this cluster mainly at the protein level. The only non-homologous region in this cluster was the first intron of APOL3, which also contained many additional Alu repeats. Nishio et al. used these repetitive DNA elements to analyze genomic expansion of the albumin gene family on human chromosome 4 [11]. In their case, the vitamin-D-binding protein gene was the earliest gene in the line of evolution and had the highest number of repetitive elements, indicating that such differences accumulate after duplication. Given that APOL3 is both the least homologous of the three family members and has a much

FIG. 4. Northern blot analysis of endothelial cells activated by TNF-α for various time periods. APOL1 shows a late response to TNF-α and APOL2 and APOL3 a delayed early response. The expression was quantified and normalized for β-actin expression levels.

FIG. 5. Multiple tissue northern blot analysis. A single hybridizing band corresponding to the sizes of the isolated cDNAs was observed for APOL1–APOL3 in all tissues, but an additional band for APOL2 was detected in brain only (A). The expression of APOL1–APOL3 in each of the lanes represented in (A) was quantified and normalized for the amount of poly(A)+ RNA to correct for loading differences (B). APOL1 and APOL3 are widely expressed, with high expression in the liver, lung, placenta, and heart. APOL2 has a more restricted expression pattern, being highly represented in the heart and brain.
higher number of repetitive elements, APOL3 may be the ancestor gene in this gene cluster. Second, the overall gene structure (that is, intron–exon boundary structure) has remained almost identical, indicating a recent duplication in the human genome. Third, the original analysis of human chromosome 22 suggested that it contains a high degree of segmental duplications [12], amounting to as much as 10.8% of the finished sequence [13]. Most of these segmental duplications are recent in evolution and even partly human-specific, and have contributed to primate genome evolution. The most recently duplicated sequence lies most proximal to the centromere, whereas more “ancient” duplications lie more distal. Fourth, in accordance with these observations, we indeed could only detect the APOL1–APOL4 cluster in humans and African green monkey. The APOL1–APOL4 gene cluster is located more distal at chromosome 22q13.1 (position 33 Mb) and must have arisen before the divergence of African green monkeys. Finished sequences of different species are needed for a more accurate phylogenetic analysis.

Most duplicated genes experience a brief period of relaxed selection early in their history [14]. The vast majority of gene duplicates are silenced within a few million years, unless they have evolved effectively and survive the selection. In this regard, it is conceivable that APOL4 is in the process of silencing and, although expressed at low levels, has no in vivo relevance [15]. It has been proposed that a multi-gene system provides an efficient mechanism for generating genetic diversity (that is, generating new alleles that are useful for the organism) [16]. The DNA sequence of the coding regions of the APOL1–APOL4 gene cluster members was more than 80% homologous, whereas the amino acid identity was only 42%, indicating that these proteins have evolved diversely after duplication. The most striking difference among the protein-encoding parts of APOL1–APOL4 was the additional exon in APOL1 (exon X), which contained an upstream start codon, in-frame with the one used by the other transcripts. This gives rise to a larger signal peptide encoding N-terminal part. Thus, the introduction of this new exon has directly resulted in the generation of a secreted protein, APOL1, which has been shown to be specifically associated with plasma HDL particles. The diversification is not only in the protein-encoding regions but also in the regulatory elements, leading to the tissue-specific expression of APOL1–APOL3, with APOL2 being the only member highly expressed in the brain.

Although a large proportion of the mouse genome has been sequenced, there is no evidence for the existence of the APOL1–APOL4 gene cluster in mice. Completion of genomic sequencing of different species would be important in the final phylogenetic analysis of this gene cluster. The mouse is
from the genomic duplication of the plasminogen gene that has only occurred in primates [23]. We have shown that the APOL1–APOL4 gene cluster has emerged only in primates and might as such contribute to the differences in the lipid metabolism among species.

**Material and Methods**

Sequencing of clones isolated from an endothelial cell cDNA library. The cDNA library was constructed and screened using a radioactive probe, representing the sequence 1687–2298 of APOL3 (CG12-1; GenBank acc. no. AF070675), as described [4]. Sequencing of the clones was performed on purified plasmid DNA using the AutoRead Sequencing-kit and Cy5-labeled T7- or SP6-oligonucleotides and analyzed on the ALF-express automatic sequencer (materials and protocol: Pharmacia, Uppsala, Sweden). Exon 4 of the APOL1 transcript was amplified from the Marathon-ready placenta cDNA library (Clontech, Palo Alto, CA). An intron-spanning set of primers was designed containing the forward sequence from exon 2b (5′-CCTCGGTGACTGGGCTGCTGGC-3′) and the reverse sequence from exon 4b (5′-GATAATCCACTGCACCTGACGTC-3′). The PCR product was cloned and sequenced. Sequence files from the ALF-express were exported in GCG format and analyzed and stored using the GCG program (Wisconsin Package Version 9.1, Genetics Computer Group (GCG), Madison, WI). Sequence homology was confirmed by BLAST searches at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov). Multiple alignments and dot-plot analysis were performed by MultiPipMaker (http://bio.cse.psu.edu/pipmaker/ [24]).

**RNA isolation and northern blot analysis.** Total RNA was extracted from unstimulated (0 hour) human umbilical vein endothelial cells (HUVEC) and HUVEC exposed to TNF-α (for 1.5, 3, 6, or 20 hours) and blotted to Hybond-N nylon membranes (Amersham, Buckinghamshire, UK) as described [4]. Purified DNA fragments, corresponding to the following sequences, were used as probes: APOL1, 141–1324 (GenBank acc. no. AF305428); APOL2, 698–1194 (GenBank acc. no. AF305429); APOL3, 703–1194 (CG12-1; GenBank acc. no. AF070675); and commercially available β-actin cDNA probe (Clontech). Filters were hybridized, washed as described, and analyzed by autoradiography [4]. Radioactivity was quantified using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). The human multiple tissue northern blot (Clontech) was probed according to the manufacturer’s protocol with the same fragments as described above.

**In situ hybridization.** Human vascular tissue specimens, displaying various stages of atherosclerosis, were collected during organ transplantation from multi-organ donors who did not have a prior history of vascular disease (approved by the AMC Medical Ethical Committee 95/146). All specimens were fixed in formalin within 5 minutes after resection and subsequently parafin-embedded. Paraffin sections (5 μm) were mounted on 3-aminopropyl-triethoxy-silane-coated slides. The following sequences were used to synthesize various riboprobes: APOL1, 647–1041 (GenBank acc. no. AF305428); APOL2, 709–1194 (GenBank acc. no. AF305429); APOL3, 703–1362 (CG12-1; GenBank acc. no. AF070675); and human vWF, 8239–8442. Riboprobes were synthesized by in vitro transcription of cDNA fragments cloned in various pGEM vectors (Promega, Madison, WI), containing T and SP6 RNA polymerase transcription initiation sites [4]. In situ hybridization was performed as described [4].

**Multiprimer Southern blot.** Tissues from different species were obtained from the animal facility (GDA, AMC, Amsterdam). High molecular weight DNA from rabbit, rat, mouse and pig was isolated by incubating approximately 200 mg of minced liver tissue in a buffer containing 100 mM Tris-HCl (pH 8), 5 mM EDTA, 200 mM NaCl, 0.1% (w/v) SDS, and 0.3 mg/ml proteinaseK (Sigma) during 16 hours at 55°C under continuous rotation. Insoluble material was detected in each of the species tested here under identical low-stringency hybridization conditions.
pelleted and the supernatant was incubated for 15 minutes at room temperature with 1 mg/ml RNase A. The supernatant was then extracted three times with equal volumes of phenol, once with a phenol chloroform isomyl alcohol mixture (25:24:1, v/v/v), and once with a chloroform isomyl alcohol mixture (24:1, v/v). The DNA was then precipitated from the aqueous phase and picked up using a glass rod. After washing in 70% (v/v) ethanol and brief drying, the DNA was solubilized in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA. High molecular weight DNA from insect, dog, Chinese hamster, human and African green monkey was isolated from respectively Sf9, MDCK, CHO-I, EC-RF24 [25], and CV1 cell cultures. Approximately 7 × 10^6 cells were incubated during 1 hour at 37°C in a buffer containing 10 mM Tris-HCl (pH 8.0), 100 mM EDTA, 20 µg/ml RNase A, and 0.5% (w/v) SDS. After addition of 100 µg/ml proteinase K, the lysate was incubated for 3 h at 55°C under constant rotation. Extraction and precipitation were performed as described above. The resulting DNA (5µg) was digested to completion with EcoRI and BamHI, run on a 0.7% (w/v) agarose gel, and blotted to Hybond-N+ according to the manufacturer’s instructions. The following fragments were used as probes: APOL1, Srf-I-SphI; APOL2, PstI-PstI; and APOL3, SstI-HindIII. Probes were labeled to high specific activity using a random-oligo labeling kit (Life Technologies) and [γ-32P]dATP (Amersham). Unincorporated nucleotides were removed using the Qiaquick nucleotide removal kit (Qiagen). Hybridization was performed in ExpressHyb hybridization solution (Clontech) at 60°C during 16 hours. After hybridization, filters were washed three times in 2x SSC, 0.05% (w/v) SDS at room temperature and two times in 1x SSC, 0.05% SDS at 45°C. Filters were analyzed by autoradiography using Xomat-AR films (Eastman Kodak).

ACKNOWLEDGMENT
This work was supported by the Molecular Cardiology Program of the Netherlands Heart Foundation (The Hague; grant no. M93.007)

RECEIVED FOR PUBLICATION SEPTEMBER 28, 2001; ACCEPTED JANUARY 7, 2002

REFERENCES

Sequence data from this article have been deposited in the DDBJ/EMBL/GenBank Data Libraries under accession numbers AF305428, AF305429, and AF070675 for APOL1, APOL2, and APOL3 (CG12-1), respectively.