Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 11992–11997, October 1997
Evolution

Convergent evolution of apolipoprotein(a) in primates and hedgehog

(lipoprotein(a)/plasminogen/gene duplication)

RICHARD M. LAWN*†, KAREN SCHWARTZ*, AND LASZLO PATTHY‡

*Falk Cardiovascular Research Center, Stanford University School of Medicine, Stanford, CA 94305-5246; and ‡Institute of Enzymology, Hungarian Academy of Sciences, Budapest, P.O. Box 7, H-1518 Hungary

Edited by Charles G. Sibley, Santa Rosa, CA, and approved August 26, 1997 (received for review June 27, 1997)

ABSTRACT **Apolipoprotein(a) [apo(a)] is the distinguishing protein component of lipoprotein(a), a major inherited risk factor for atherosclerosis. Human apo(a) is homologous to plasminogen. It contains from 15 to 50 repeated domains closely related to plasminogen kringle four, plus single kringle five-like and inactive protease-like domains. This expressed gene is confined to a subset of primates. Although most mammals lack apo(a), hedgehogs produce an apo(a)-like protein composed of highly repeated copies of a plasminogen kringle three-like domain, with complete absence of protease domain sequences. Both human and hedgehog apo(a)-like proteins form covalently linked lipoprotein particles that can bind to fibrin and other substrates shared with plasminogen. DNA sequence comparisons and phylogenetic analysis indicate that the human type of apo(a) evolved from a duplicated plasminogen gene during recent primate evolution. In contrast, the kringle three-based type of apo(a) evolved from an independent duplication of the plasminogen gene approximately 80 million years ago. In a type of convergent evolution, the plasminogen gene has been independently remodeled twice during mammalian evolution to produce similar forms of apo(a) in two widely divergent groups of species.**

Lipoprotein(a) [Lp(a)] represents one of the major inherited risk factors for cardiac heart disease and cerebral stroke (1–3). It is a particle whose function and evolutionary history remain a mystery. Lp(a) consists of a low density lipoprotein (LDL) particle, containing cholesterol, phospholipid, and apolipoprotein (apo) B-100, plus the sole distinguishing feature of the covalent attachment of apo(a). Cloning and sequencing of human apo(a) cDNA revealed an unexpected close similarity to plasminogen (4). Plasminogen is a protease zymogen containing five distinct kringle domains and a catalytic subunit. Upon activation by plasminogen activators, plasmin cleaves fibrin to initiate clot lysis, and activates certain growth factors and matrix proteases by its endopeptidase activity. The plasminogen kringles share less that 50% identity and have several distinctive differences from one another. Human apo(a) contains global sequence similarity to plasminogen. It lacks the amino-terminal domains of the plasminogen but contains a variable number of copies (from 12 to 50 copies) of a domain with ≈75% identity to kringle four of plasminogen, a single kringle five-like domain, and a proteolytically inactive relative of the protease domain of plasminogen (4, 5). Although its function remains obscure, notable features of apo(a) are the highly repeated nature of its kringle domains, retention of the lysine binding site of plasminogen kringle four in at least one of its kringles, addition of a seventh unpaired cysteine residue in only one kringle through which it forms covalent linkage to apoB-100, and loss of proteolytic activity of its protease-like domain. These features allow apo(a) to circulate as part of a lipoprotein particle, bind to fibrin and other known substrates of plasminogen, inhibit plasminogen activation, and in elevated concentrations inhibit fibrinolysis and increase the risk of atherosclerosis and cerebral stroke. An estimated 20% of the human population is at increased risk of myocardial infarction and stroke due to their high plasma concentration of Lp(a). Although its adaptive function remains obscure, an apo(a)-like gene has apparently evolved twice during the course of mammalian evolution.

Lp(a) was once thought to be confined to a subset of primates. Apo(a) mRNA and protein has been detected in only Old World monkeys, apes, and humans, consistent with the hypothesis that the apo(a) gene arose from a duplicated copy of the plasminogen gene during primate evolution (see ref. 6 and references therein). The report of an Lp(a)-like particle in hedgehogs complicated the scenario of apo(a) evolution (7, 8). Insectivores such as hedgehogs are among the most distant relatives of primates among placental mammals. Cloning and sequencing of hedgehog apo(a) cDNA demonstrated a remarkable example of "parallel gene evolution" (6). The sequenced hedgehog apo(a) cDNA contains 31 repeats that contain approximately 80% identity to hedgehog kringle 3, and no kringle 4-like, 5-like, or protease domains whatsoever. (Kringles 3 and 4 of plasminogen are clearly distinct, and the highly repeated domains of human apo(a) are unambiguously related to kringle 4.) Yet, like human apo(a), only one of the hedgehog apo(a) kringles contains a seventh cysteine residue (in a location distinct from the human form) and forms a covalent linkage to apoB-100, allowing it to circulate as a lipoprotein particle. Hedgehog apo(a) also retains the amino acids that constitute the plasminogen lysine binding site, which in this species occurs in plasminogen kringle 3 (in contrast to human plasminogen kringle 4) and binds to lysine and fibrin. By apparent remodeling of a plasminogen-like gene, hedgehog and human ancestors independently evolved an apo(a) protein with multiple kringle domains that covalently links to apoB-containing lipoproteins, binds fibrin, lacks proteolytic activity, and competitively inhibits plasminogen activation (6).

Previous analyses have not reached agreement as to the likely time of gene duplications and creation of the primate (kringle 4-like) and hedgehog (kringle 3-like) apo(a) genes or established the likely intermediates in their formation (4, 9, 10). Among the issues to be resolved are the antiquity of the

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

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: Lp(a), lipoprotein(a); apo, apolipoprotein; LDL, low density lipoprotein; Myr, million years; UPGMA, unweighted pair group method using arithmetic averages.
Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AF012297, for Wallaby plasminogen cDNA sequence).
†To whom reprint requests should be addressed. e-mail: richard.lawn@forsythe.stanford.edu.

two forms of apo(a) and the expectation that they may occur in a wider range of species than yet appreciated. The original estimate of divergence time using a comparison of the 3′ untranslated sequences of human apo(a) and plasminogen cDNAs and the substitution rate of untranslated sequences in primate globin genes yielded a value of 40 million years (Myr) ago for the separation of these two paralogs (4). After rhesus monkey sequences became available, an analysis by Pesole *et al.* (9) gave an estimate of 93 Myr. This disparity leads to widely different predictions about the species distribution of the apo(a) gene. By determining the sequence of a marsupial plasminogen cDNA, we have been able to reanalyze the likely evolution of these two incarnations of apo(a).

METHODS

Wallaby Cloning and Sequencing. A liver cDNA library derived from *Macropus eugenii* was a generous gift of Gerhard Schreiber (University of Melbourne). Plasminogen clones were identified by hybridization with human plasminogen cDNA and sequenced by using vector and plasminogen sequence-specific oligonucleotides and the Sequenase version 2.0 DNA kit (U.S. Biochemical). Areas of ambiguous sequence were resolved by deoxyinosine triphosphate and 7-deaza-GTP reactions by using conditions recommended by the vendor.

Sequence Comparisons and Phylogenetic Analyses. The nucleic acid sequences of hedgehog (U33171), mouse (J04766), cow (X79402), rhesus (J04697), and human (M74220) plasminogens and hedgehog (U33170), rhesus (J04635), and human (X06290) apo(a)s were from GenBank.

The amino acid sequences and nucleic acid sequences of plasminogens and apo(a)s were aligned with CLUSTAL W (11) or PILEUP of the Genetics Computer Group Wisconsin Package, Version 9. The protease domains of plasminogens and apo(a)s were defined as those homologous to the light chains of plasmin; the boundaries of kringle domains were the first and last cysteines of these disulfide bonded structures.

Dendrograms for amino acid sequences and nucleic acid sequences were produced by CLUSTAL W and PILEUP, which use the unweighted pair group method using arithmetic averages (UPGMA) of Sneath and Sokal (12). Phylogenetic trees were also constructed with maximum likelihood and maximum parsimony methods using the DNAMLK and DNAPARS programs of PHYLIP (phylogeny inference package) Version 3.57c developed by Joseph Felsenstein (University of Washington, Seattle) (13). SEQBOOT and CONDENSE programs of the PHYLIP package were used in bootstrap analysis of 100 resamplings of the data sets.

The DIVERGE program of the Genetics Computer Group Wisconsin Package, based on the method of Li (14), was used to estimate the pairwise number of synonymous and nonsynonymous substitutions per site between aligned nucleic acid sequences of kringles of plasminogens and apo(a)s and to determine the K_s and K_a values.

Divergence times for hedgehog and primate apo(a)s were calculated according to Li and Graur (15), by using the divergence times for the rhesus/human split (25 Myr ago), rodent/primate split (75 Myr), and cow/primate split (80 Myr) of Li *et al.* (16). The divergence time for the marsupial/placental split (130 Myr) is from Janke *et al.* (17).

RESULTS AND DISCUSSION

Evolution of Primate apo(a). A cDNA library derived from the liver of the Tammar wallaby (*Macropus eugenii*) was provided by Gerhard Schreiber. Human plasminogen cDNA probes were used to identify wallaby plasminogen cDNA clones. They contain global similarity with the human ortholog, with 73.7% overall nucleotide and 70.9% amino acid identity (GenBank accession no. AF012297). The wallaby plasminogen sequence then was used to analyze the sequence relationships of plasminogen to the "kringle 4-based" human (primate type) apo(a) and the "kringle 3-based" hedgehog (insectivore type) apo(a).

A dendrogram of the nucleotide sequences of the kringle 4-like domains of apo(a) and plasminogens was constructed by using the UPGMA distance method (data not shown). If we assume that kringle 4 domains of plasminogen and apo(a) diverged at a constant rate (see below), this analysis would suggest that the plasminogen/apo(a) split occurred within placental mammals, whereas multiplications of the kringle domains of apo(a) are a more recent event. The extreme example are the majority of human apo(a) kringles that have 100% nucleotide identity and form the basis for the heterogeneity in size of apo(a) proteins in the human population. Alleles of apo(a) in the human population contain 12–50 kringle repeats, causing a range in molecular mass of 250–800 kDa (5). This variability exhibits polarity, as often is seen with short hypervariable arrays. The 5′ apo(a) kringles that are identical in nucleotide sequence show allelic variation in copy number, whereas kringle 4 types 30–37 that diverge in sequence do not commonly vary between individuals (18). Variation in the number of identical kringles may be caused by out-of-register homologous recombination during meiosis. This process is still occurring in present human populations, because rare cases have been reported of kringle number change from parent to offspring (19). The one exception to the clustering of apo(a) kringles in this analysis is that human apo(a) kringle 4–37 is more closely linked to human and rhesus plasminogen than to the other apo(a) kringles. Interestingly, this is the one kringle of human apo(a) that most closely retains the lysine/fibrin binding pocket of plasminogen kringle 4 and is responsible for most of the fibrin binding affinity of Lp(a) (20–22).

It is likely that the forces of selection are unequal for apo(a) and plasminogen kringle 4 domains; therefore, the UPGMA dendrogram might overestimate the distance of apo(a) and plasminogen kringle 4 domains. For example, plasminogen kringles not only possess lysine binding sites but contain surface elements that are known to interact with other domains of plasminogen. Plasminogen maintains a relatively compact conformation that alters in response to various effectors. The interactions among kringle domains in plasminogen are critical to its biological functions and affect protease activation and substrate binding (23, 24). In contrast, electron microscopic and velocity sedimentation studies suggest that apo(a) has a relatively "open" confirmation, resembling a coil of repeated domains (25), although solvent conditions may alter its conformation (26). It might be suggested that apo(a) kringles contain more amino acid positions that are liberated for divergence than does plasminogen kringle 4 and may diverge at a faster rate. Therefore, to attempt to eliminate the influence of differences in functional/structural constraints, we have compared only those positions that are conserved both in plasminogen and all apo(a) kringle four-like domains and used only the nucleotides encoding these conserved amino acids in a UPGMA analysis (Fig. 1*A*). This analysis also supports the hypothesis that the apo(a) precursor arose within the group of placental mammals, not predating mammalian radiation.

The time of divergence of human apo(a) from a plasminogen ancestor [i.e., the time of gene duplication that gave rise to apo(a)'s ancestor] was calculated according to Li and Graur (15) by using synonymous substitutions (K_s) in aligned nucleic acid sequences of kringle four and five domains of various mammalian plasminogens and human and rhesus apo(a). In the calculations, the divergence times proposed by Li *et al.* (16) for the rhesus/human split (25 Myr ago), rodent/primate spilt (75 Myr), and cow/primate split (80 Myr) were used, and the divergence time for the marsupial/placental split (130 Myr) is taken from Janke *et al.* (17). As noted by Janke *et al.* (17), the

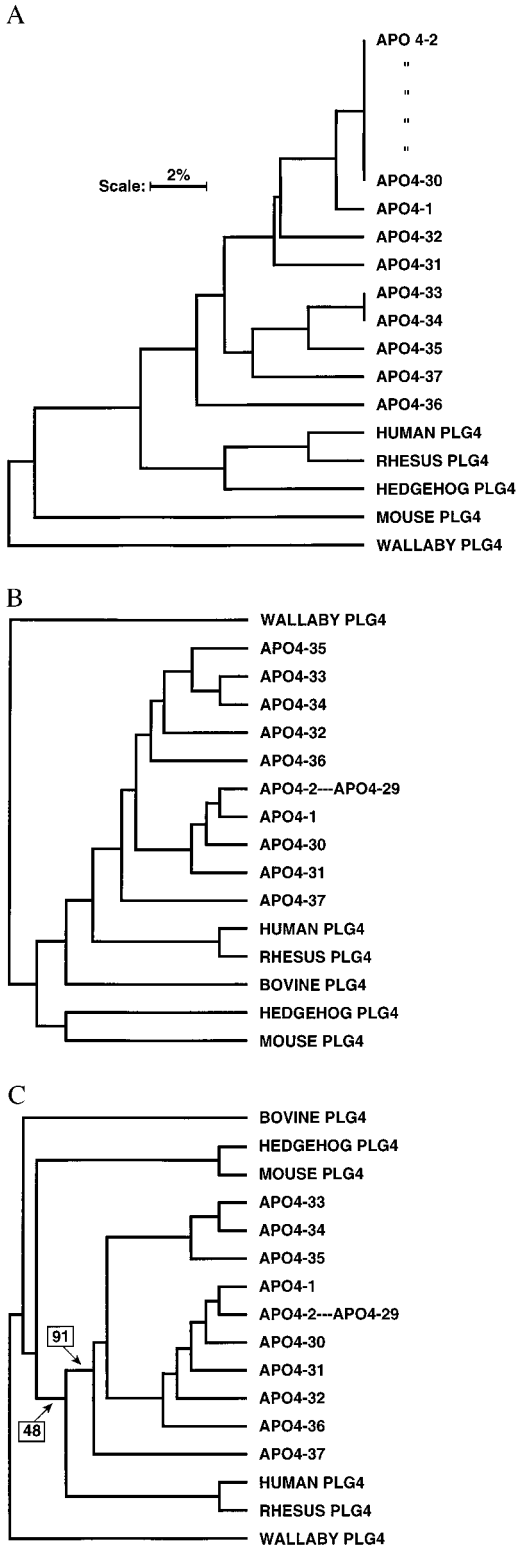

FIG. 1. Relationship of the kringle 4 domains of plasminogen and primate apo(a). (*A*) UPGMA dendrogram of kringle 4 nucleotide sequences of human (*Homo sapiens*), rhesus monkey (*Macaca mulatta*), bovine (*Bos tauris*), mouse (*Mus musculus*), hedgehog (*Erinacaeus europeus*), wallaby (*Macropus eugenii)* plasminogens and of human apo(a). Only the conserved amino acid positions of kringle 4 domains were considered, which are CY–G–Y-G–T–G–CQ-W–M-PH-H—P–L–NYCRNPD–P-C-T-DP–RWE-CNL–C. The complete sequence of human apo(a) kringles 4–2 through 4–29, inclusive, are identical, whereas in this analysis, the considered portions of kringle 30 are also identical to the preceding ones and kringles 4–33 and 4–34 are identical to one another. DNAMLK (DNA maximum likelihood with molecular clock) (*B*) and DNAPARS (DNA maximum parsimony) (*C*) trees of kringle 4 nucleotide sequences of plasminogen and primate apo(a) are shown. Percentage bootstrap support is shown in boxes for selected branches.

position of *Lypotyphla* (hedgehog) is basal to other placental mammals including myomorph rodents and ferungulates. Consistent with this, our present calculations give ≈86 Myr ago for the divergence of hedgehog from other placental mammals. From the above species-divergence times, our different calculations gave ≈33 Myr ago for the split of primate apo(a) from its plasminogen ancestor. This is very similar to the estimate based upon the analysis of the 3′ untranslated regions and to the generally accepted time of divergence of Old World (*Catarrhini*) and New World (*Platyrrhini*) monkeys (35–45 Myr), which demarks the group of species that possess from those that lack apo(a) protein, based on current biochemical evidence.

Additional dendrograms were constructed for nucleotide sequences encoding the kringle 5 and protease domains of human apo(a) and available mammalian plasminogen genes (data not shown). Again the dendrograms place apo(a) divergence well within the group of placental mammals. Each of these analyses contradict the scenario that the primate type of apo(a) diverged before mammalian radiation. The bulk of evidence is in favor of the suggestion that the gene duplication giving rise to the primate type of apo(a) occurred within the primate line. However, it must be emphasized that sequence analysis and direct biochemical evidence imply that the vast multiplications of kringles that are a hallmark of present apo(a) occurred more recently, likely postdating the divergence of Old World and New World monkeys.

Two other forms of analysis were employed to model the evolutionary history of apo(a). Fig. 1*B* uses DNAMLK analysis (DNA maximum likelihood with molecular clock). The cluster of apo(a) kringles branches off the primate line, suggesting that the gene duplication occurred in the line leading to primates after its separation from other placental mammals. This again contradicts the earlier conclusion of Pesole *et al.* (9), based on the comparison of human and rhesus apo(a) and plasminogen sequences, that the gene duplication took place before mammalian radiation (see below). Fig. 1*C* shows the DNAPARS (DNA parsimony) analysis of the sequence data. Again, the branching pattern for plasminogen kringle 4 domains generally harmonizes with the accepted species tree and shows the apo(a) kringles clustering in the line leading to primates. Robust bootstrap support (91%) exists for the contention that apo(a) arose after mammalian radiation. The suggestion that it arose in the primate lineage is supported by a bootstrap value of 48%.

Evidence for "Recent" Divergence of Primate apo(a) from Plasminogen. When human apo(a) cDNA was first cloned and sequenced, comparison of the untranslated regions of human apo(a) and plasminogen cDNAs led to the proposal that the two genes first diverged about 40 Myr ago (4). This estimate was based on the substitution rate for untranslated sequences of primate globin genes. With the subsequent availability of rhesus monkey apo(a) and plasminogen cDNA sequences (27), Pesole *et al.* (9) revised the estimate of divergence to ≈90 Myr, well before the radiation of mammals. This implied "that apo(a) should be present not only in primates but in other mammals." Our current analysis is more consistent with the apparent confinement of apo(a) to primates (6). In their comparison of rhesus and human apo(a) 3′ untranslated sequences, Pesole *et al.* (9) calculated a nucleotide substitution rate of 3.28 substitutions per site per 10^9 years. In fact, if it had been accumulating substitutions at this rate since its formation, the apo(a)/plasminogen split would have indeed occurred at about 36 Myr ago, close to the time of the divergence of Old

and New World monkeys. In addition, these authors pointed out that the protease domain of apo(a) has accumulated nonsynonymous vs. synonymous mutations as if it were a functionally unconstrained pseudogene. This is consistent with apo(a)'s lack of plasmin-like proteolytic activity. In this case, the UPGMA comparison of protease domains (data not shown), which places the divergence after the rodent/primate split (75 Myr ago), would be an overestimate, also consistent with the more recent time of divergence from the primate line. Pesole *et al.* (9) also pointed out that the most 3′ of the apo(a) kringles, kringle 4–31 to kringle 4–37, were more similar between human and rhesus monkey than to one another in each of the species and had substitution rates in nonsynonymous and synonymous sites that were comparable to plasminogen. Hence, this domain ". . . is under the same evolutionary pressure as the homologous domain in plasminogen and thus can be used as a reliable molecular clock" (9). In contrast, the upstream kringles have a higher intraspecies similarity and likely reflect recent duplications or molecular drive mechanisms occurring since the rhesus/human divergence. This is also consistent with the observation of switches in kringle numbers within human families, as noted above. It is the comparison of these "molecular clock-like" kringle 4–31 to kringle 4–37 sequences of human and rhesus apo(a) vs. plasminogen kringle 4 that led to the estimate of a 93 Myr ago divergence. However, this conclusion depends on the assumption that the paralogs have always been diverging at the same rate in these domains. That this is not the case is supported by the differences between the phenetic UPGMA tree (which assumes rate constancy) and the cladistic DNAPARS tree. It is a reasonable postulate that as the function of apo(a) kringles (lysine binding for only kringle 4–37, covalent binding to apoB-100 for kringle 4–36, a role in assembly of the lipoprotein particle for kringles 4–32 to 4–36, and unknown functions of kringles 4–1 through 4–30) diverged from the function of plasminogen kringle 4 (lysine binding, interaction with plasminogen kringles 1–3, and conformational changes upon proteolytic activation), the rate of sequence substitutions was appreciably higher than it became once its new function had been established. The molecular clock can only be presumed to run at a constant rate in the case of genes whose function is not experiencing any significant change. Therefore, we submit that the weight of evidence predicts that the divergence of plasminogen and apo(a) genes occurred more recently than the radiation of mammals, probably during the evolution of primates, the only animals known to possess this type of apo(a).

Evolution of Insectivore apo(a). The hedgehog apo(a)-like gene apparently was derived by remodeling the kringle 3 domain of plasminogen, in a process distinct from the derivation of the human apo(a) gene from plasminogen kringle 4, kringle 5, and protease domains (6). Both the UPGMA alignments using total nucleotide sequences (data not shown) and using only the absolutely conserved residues (Fig. 2*A*) suggest that the ancestor of hedgehog apo(a) arose in the line leading to placental mammals after the separation from marsupials, with the latter analysis placing it closest to hedgehogs. An analysis of K_s values for the nucleic acid sequences of kringle 3 domains of plasminogen and hedgehog apo(a) yields an estimate of 80 Myr ago for the split, close to the calculated time of divergence of hedgehogs from other placental orders (86 Myr). Again, the gross multiplication of kringles seems to have taken place much later and to be confined to a subset of these species. DNAMLK analysis of plasminogen kringle 3 and hedgehog apo(a) (Fig. 2*B*) places the cluster of these apo(a) kringles closest to the insectivore line, suggesting that the gene duplication occurred in that line after its separation from other placental mammals. However, the cladogram based on DNAPARS analysis with high bootstrap confidence (Fig. 2*C*) does not support this clustering and would suggest a gene duplication that predated divergence of marsupial and placental mammals. It should be noted, however, that if the tree were rooted for wallaby plasminogen kringle 3, it would yield a topology similar to that shown in Fig. 2*B* for DNAMLK. Thus, the evidence suggests that the gene duplication leading to hedgehog apo(a) occurred shortly after the divergence of placental mammals and possibly within the insectivore line. A systematic search for kringle 3-based forms of apo(a) in other species is called for to clarify this scenario, given the discrepancies between some of the analytic predictions and the fact that the predicted time of divergence is close to a value that would predate the separation of many orders of placental mammals.

Existence of Additional apo(a)-Like Genes. The gene duplications that led to the plasminogen and apo(a) genes produced other paralogous genes and pseudogenes. Human chromosome 6 contains a cluster of apo(a), plasminogen, and two additional genes or pseudogenes containing ≈90% DNA sequence identity at this 5′ ends, and at least three close paralogs occur on other chromosomes (28–30). Although transcripts have been reported for two of these genes, no convincing detection of the corresponding proteins or their functions has been reported, although it has been speculated that they could compete for plasminogen binding and modulate its function (31, 32). Humans also contain two more distantly related proteins, hepatocyte growth factor (or scatter factor) and hepatocyte growth factor-like protein (or macrophage stimulating protein). These two proteins are composed of four kringles and an inactive protease-like domain with approximately 38% amino acid identity to plasminogen. They are recognized by high-affinity receptors and affect cell proliferation and motility (33–35).

Hedgehogs also contain several plasminogen-like genes, which may not have counterparts in humans. Rouy (8) found two proteins smaller than plasminogen in hedgehog plasma that also bound to lysine-Sepharose. We have sequenced their corresponding cDNAs. They each encode seven kringles, have a lower degree of identity to hedgehog plasminogen than does hedgehog apo(a), and may represent more ancient derivatives of the plasminogen gene (G. Lindahl, R.M.L., L.P., and K.S., unpublished observations). Kringle motifs apparently have been fertile grounds for proteins that bind to fibrin as well as other shared substrates.

CONCLUSIONS

Our data are consistent with the scenario that human and hedgehog apo(a) arose independently by remodeling different domains of plasminogen or plasminogen-like genes. In each case, intragenic multiplication to create dozens of similar domains has occurred more recently, in a limited number of species. In humans there is evidence that this process is still occurring. In addition to containing 12–50 repeated kringle domains, human and hedgehog apo(a)-like genes have modified different parts of the plasminogen gene to achieve fibrin binding and covalent binding to apoB-100, while losing proteolytic activity by either nucleotide substitution (human) or domain deletion (hedgehog). The evidence indicates that the human apo(a) gene arose from a duplicated plasminogen-like gene during primate evolution, consistent with its currently known confinement to Old World monkeys, apes, and humans. The hedgehog apo(a)-like gene appears to be a product of an earlier independent gene duplication and might be found within a greater range of mammals than is currently reported. This does not favor an alternate scenario whereby a duplication of the plasminogen gene occurring before the radiation of mammals and containing kringles 3,4,5 and the protease domain spawned the hedgehog and primate forms of apo(a) (10).

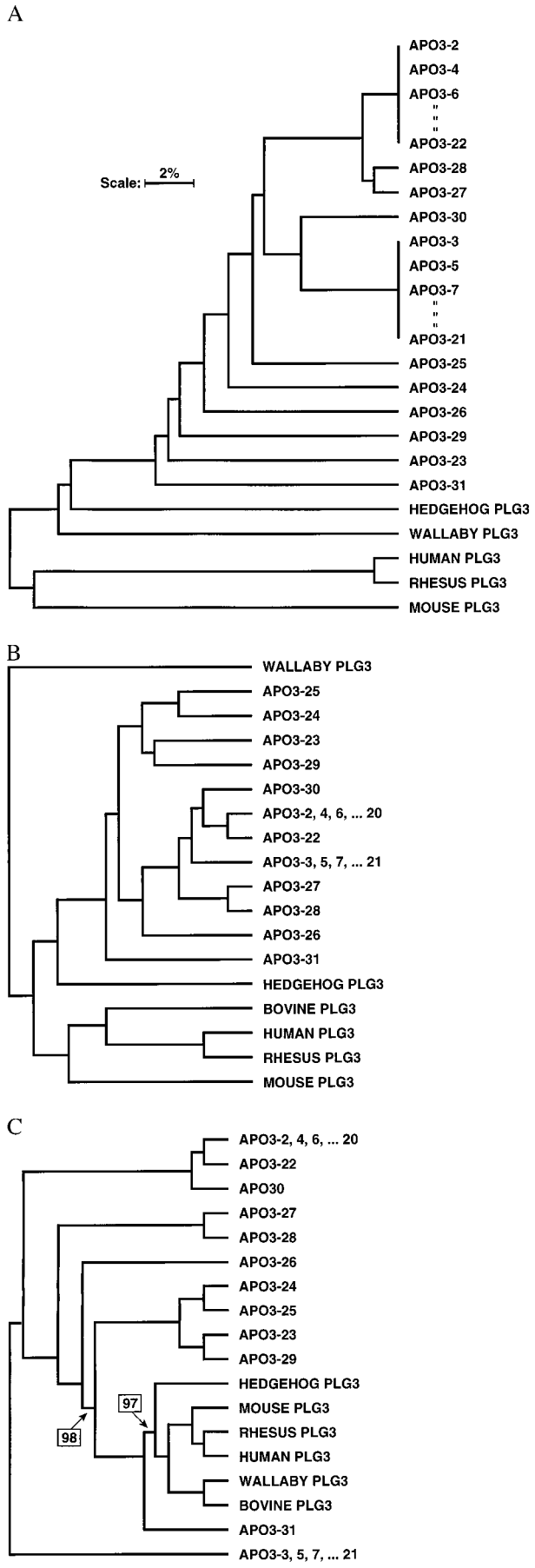

FIG. 2. Relationship of the kringle 3 domains of plasminogen and hedgehog apo(a). (*A*) UPGMA dendrogram for the nucleotide sequences of only the conserved amino acid positions of the kringle 3 domains, CL–G–Y–G–T–S–C–W–Q–PH–H–T–P–L–N–CRN–G–PWC–TT–R–E–C–IP–C. Many hedgehog apo(a) kringles have apparently duplicated as a cassette of two adjacent kringles, so that the sequence of kringle 3–2 = 3–4 = 3–6 . . . = 3–22, and kringle 3–3 = 3–5 = 3–7 . . . = 3–21. The sequence of kringle 3–1 is incomplete and is not included in the analysis. DNAMLK (*B*) and DNAPARS (*C*) trees of kringle 3 nucleotide sequences of plasminogens and hedgehog apo(a) are shown. Percentage bootstrap support is shown in boxes for selected branches.

It is intriguing to note that, despite its recurrent evolution, we do not as yet know the adaptive function of apo(a). The fact that an apo(a)-like gene with shared properties has evolved twice in addition to the fact that it has retained intact an extremely long ORF argue that it is not simply the residue of a random duplication of DNA that has not been extinguished. Yet, individuals with elevated plasma concentration of apo(a) have increased risk for atherosclerosis, restenosis, and cerebral stroke. A large body of research indicates that apo(a) can contribute to the development of these diseases by binding to plasminogen substrates and thus inhibiting fibrinolysis and the plasmin-dependent activation of transforming growth factor β, by otherwise promoting thrombosis and by delivering cholesterol to the vessel wall (1, 36–38). Positive functions remain speculative. It has been proposed that having a form of LDL that can bind to fibrin might be a selective advantage in wound healing to deliver cholesterol to sites of tissue damage for new cell membrane biosynthesis (39). Although this is consistent with the immunolocalization of Lp(a) in wounded tissues, there is as yet no experimental confirmation of the efficacy of Lp(a) in wound-healing models (40). The discovery that a kringle containing fragment of plasminogen, called angiostatin, is an inhibitor of angiogenesis (41) prompted the testing of apo(a) in that regard. However, we found that apo(a) fails to modulate angiogenesis in an *in vivo* model system (R.M.L. and L. Fajardo, unpublished results). There remains the more general idea that apo(a) was selected in some way to modulate fibrinolysis. Although many of these proposals are destined to remain speculative, they should ideally consider not only the vast multiplicity of apo(a) kringles and its ability to bind fibrin and other extracellular and cell surface binding sites of plasminogen but also its loss of proteolytic activity and its covalent linkage to an LDL-like particle. [It should be noted that the covalent linkage of apo(a) to apoB-100 solidifies a noncovalent affinity of the two proteins and that plasminogen itself has a much weaker, but measurable, noncovalent affinity to LDL (42)]. We are now testing the idea that apo(a) might play an adaptive role in protection against certain infectious diseases. This proposal is based on the observation that a number of invasive pathogens express plasminogen activators or receptors, which may aid in the dissolution of extracellular barriers to their migration from the site of entry to the bloodstream (43). In addition, lipoproteins are capable of binding certain microorganisms or their surface components. Hence, it might have been adaptive to combine a circulating competitor of plasminogen with a lipoprotein sink. As a test case, we are investigating the interaction of Lp(a) and the plague bacillus *Yersinia pestis*. This pathogen contains a surface protease with plasminogen activating activity. Deletion of the protease gene causes a millionfold increase in the median lethal dose to mice, but only when the infection occurs at a superficial site, not when the bacteria are delivered to deeper tissues (44, 45). In preliminary studies, we have found that Lp(a) inhibits the activity of the *Y. pestis* protease *in vitro* and are investigating the infectivity of the bacterium in Lp(a) transgenic mice (J. Goguen and R.M.L., unpublished results).

The recurring evolution of apo(a) represents a novel example of parallel gene evolution. A few examples are known where identical amino acid substitutions have independently

occurred in different species to achieve optimization of function. This includes the occurrence of several uniquely shared amino acids in lysozyme of ruminants and colombine monkeys, to accommodate low pH foregut fermentation in these species that have independently evolved this mode of digestion (46). Unfortunately, such a clear functional link between primates and hedgehogs remains elusive. An intriguing recent example of convergent evolution is the independent evolution of similar antifreeze glycoproteins in Arctic and Antarctic fish. In this case, it appears that different ancestral genes were duplicated and modified to encode proteins with similar repeating peptide triplets (47, 48). The evolutionary history of apo(a) represents a case in which a novel gene and a relative with global sequence similarity and biochemical properties have arisen independently by modification of the plasminogen gene (or an unknown paralog). Although the odds are stacked against truly "convergent" evolution of genes with global sequence similarity that do not arise from a common precursor, we might expect to find other examples where selection has favored the independent "remodeling" of a common precursor gene to produce highly similar novel genes in disparate species.

We thank collaborators Luis Fajardo, Jon Goguen, and Gisela Lindahl for permission to cite unpublished data and thank Gerhard Schreiber for use of his wallaby cDNA library. This work was supported by National Institutes of Health Grant HL-50590 to R.M.L. and Grants OTKA 14642 and 22949 to L.P.

1. Scanu, A. M. & Fless, G. M. (1990) *J. Clin. Invest.* **85,** 1709–1715.
2. Bostom, A. G., Cupples, L. A., Jenner, J. L., Ordovas, J. M., Seman, L. L., Wilson, P. W. F., Schaefer, E. J. & Castelli, W. P. (1996) *J. Am. Med. Assoc.* **276,** 544–548.
3. Assmann, G., Schulte, H. & von Eckardstein, A. (1996) *Am. J. Cardiol.* **77,** 1179–1184.
4. McLean, J. W., Tomlinson, J. E., Kuang, W. J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M. & Lawn, R. M. (1987) *Nature (London)* **330,** 132–137.
5. Lackner, C., Boerwinkle, E., Leffert, C., Rahmig, T. & Hobbs, H. H. (1991) *J. Clin. Invest.* **87,** 2153–2161.
6. Lawn, R. M., Boonmark, N. W., Schwartz, K., Lindahl, G. E., Wade, D. P., Byrne, C. D., Fong, K. J., Meer, K. & Patthy, L. (1995) *J. Biol. Chem.* **270,** 24004–24009.
7. Laplaud, P. M., Beaubatie, L., Rall, S. J., Luc, G. & Saboureau, M. (1988) *J. Lipid Res.* **29,** 1157–1170.
8. Rouy, D., Laplaud, P. M., Saboureau, M. & Angles-Cano, E. (1992) *Arterioscler. Thromb.* **12,** 146–154.
9. Pesole, G., Gerardi, A., de Jeso, F. & Saccone, C. (1994) *Genetics* **136,** 255–260.
10. Lawn, R. M., Patthy, L., Pesole, G. & Saccone, C. (1997) *J. Mol. Evol.* **44,** 234–236.
11. Higgins, D. G. & Sharp, P. M. (1989) *Comput. Appl. Biosci.* **5,** 151–153.
12. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy*, (Freeman, San Francisco), pp. 230–234.
13. Felsenstein, J. (1989) *Cladistics* **5,** 164–166.
14. Li, W.-H. (1993) *J. Mol. Evol.* **36,** 96–99.
15. Li, W.-H. & Graur, D. (1991) *Fundamentals of Molecular Evolution*, (Sinauer, Sunderland, MA), pp. 117–118.
16. Li, W.-H., Tanimura, M. & Sharp, P. M. (1997) *J. Mol. Evol.* **25,** 330–342.
17. Janke, A., Xu, X. F. & Arnason, U. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 1276–1281.
18. van der Hoek, Y. Y., Wittekoek, M. E., Beisiegel, U., Kastelein, J. J. & Koschinsky, M. L. (1993) *Hum. Mol. Genet.* **2,** 361–366.
19. Lackner, C., Cohen, J. C. & Hobbs, H. H. (1993) *Hum. Mol. Genet.* **2,** 933–940.
20. Ernst, A., Helmhold, M., Brunner, C., Petho-Schramm, A., Armstrong, V. W. & Muller, H. J. (1995) *J. Biol. Chem.* **270,** 6227–6234.
21. Scanu, A. M., Pfaffinger, D., Lee, J. C. & Hinman, J. (1994) *Biochim. Biophys. Acta* **1227,** 41–45.
22. Hoover-Plow, J. L., Boonmark, N., Skocir, P., Lawn, R. & Plow, E. F. (1996) *Arterioscler. Thromb. Vasc. Biol.* **16,** 656–664.
23. Vali, Z. & Patthy, L. (1982) *J. Biol. Chem.* **257,** 2104–2110.
24. Cummings, H. S. & Castellino, F. J. (1985) *Arch. Biochem. Biophys.* **236,** 612–618.
25. Phillips, M. L., Lembertas, A. V., Schumaker, V. N., Lawn, R. M., Shire, S. J. & Zioncheck, T. F. (1993) *Biochemistry* **32,** 3722–3728.
26. Fless, G. M., Furbee, J., Snyder, M. L. & Meredith, S. C. (1996) *Biochemistry* **35,** 2289–2298.
27. Tomlinson, J. E., McLean, J. W. & Lawn, R. M. (1989) *J. Biol. Chem.* **264,** 5957–5965.
28. Ichinose, A. (1992) *Biochemistry* **31,** 3113–3118.
29. Magnaghi, P., Citterio, E., Malgaretti, N., Acquati, F., Ottolenghi, S. & Taramelli, R. (1994) *Hum. Mol. Genet.* **3,** 437–442.
30. Byrne, C. D., Schwartz, K., Meer, K. J., Cheng, J. F. & Lawn, R. M. (1994) *Arterioscler. Thromb.* **14,** 534–541.
31. Weissbach, L. & Treadwell, B. V. (1992) *Biochem. Biophys. Res. Commun.* **186,** 1108–1114.
32. Byrne, C. D., Schwartz, K. & Lawn, R. M. (1995) *Arterioscler. Thromb. Vasc. Biol.* **15,** 65–70.
33. Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K. & Shimuzu, S. (1989) *Nature (London)* **342,** 440–443.
34. Han, S., Stuart, A. & Friezner-Degen, S. J. (1991) *Biochemistry* **30,** 9768–9780.
35. Yoshimura, T., Yuhki, N., Wang, M. H., Skeel, A. & Leonard, E. J. (1993) *J. Biol. Chem.* **268,** 15461–15468.
36. Utermann, G. T. (1989) *Science* **246,** 904–910.
37. Miles, L. A. & Plow, E. F. (1990) *Thromb. Haemostasis* **63,** 331–335.
38. Grainger, D. J., Kemp, P. R., Liu, A. C., Lawn, R. M. & Metcalfe, J. C. (1994) *Nature (London)* **370,** 460–462.
39. Brown, M. S. & Goldstein, J. L. (1987) *Nature (London)* **330,** 113–114.
40. Yano, Y., Shimokawa, K., Okada, Y. & Noma, A. (1997) *J. Histochem. Cytochem.* **45,** 559–568.
41. O'Reilly, M. S., Holmgren, L., Shing, Y., Chen, C., Rosenthal, R. A., Moses, M., Lane, W. S., Cao, Y., Sage, E. H. & Folkman, J. (1994) *Cell* **79,** 315–328.
42. Trieu, V. N., Zioncheck, T. F., Lawn, R. M. & McConathy, W. J. (1991) *J. Biol. Chem.* **266,** 5480–5485.
43. Boyle, M. D. & Lottenberg, R. (1997) *Thromb. Haemostasis* **77,** 1–10.
44. Brubaker, R. R., Beesley, E. D. & Surgalla, M. J. (1965) *Science* **149,** 422–424.
45. Sodeinde, O. A., Subrahmanyam, Y. V., Stark, K., Quan, T., Bao, Y. & Goguen, J. D. (1992) *Science* **258,** 1004–1007.
46. Stewart, C. & Wilson, A. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 891–899.
47. Chen, L., DeVries, A. & Cheng, C.-H. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 3811–3816.
48. Chen, L., DeVries, A. L. & Cheng, C.-H. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 3817–3822.