Vol. 209, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 6, 1995 Pages 365-371

CHARACTERIZATION OF THE APOLIPOPROTEIN(a) GENE

Akitada Ichinose*

Department of Molecular Pathological Biochemistry, Yamagata University School of Medicine, Yamagata 990-23, Japan

Department of Biochemistry, University of Washington, Seattle, WA 98195

Received March 4, 1995

Summary. Five homologous genes have been isolated and characterized and found to belong to the plasminogen-apolipoprotein(a) gene family. One of these genes is the true apolipoprotein(a) gene because it has typical regulatory elements and exons coding for kringle 4 repeats. Nucleotide sequence analysis revealed that each of 11 types of the kringle 4 repeats is encoded by two exons and that the splice junctions of the exons coding for the (pseudo)serine protease domain of apo(a) were also located in exactly the same positions as those of plasminogen.

Apolipoprotein(a) [apo(a)] is a high molecular weight glycoprotein of 300-700 kD, which is connected to apolipoprotein B-100 in lipoprotein(a) [Lp(a)] by a disulfide bond (1-3). Apo(a) and plasminogen are very similar in amino acid and genomic sequences (4-6). A major difference between these proteins, however, is that apo(a) has from 10 to 50 repeats of kringle 4 of plasminogen. Concentrations of Lp(a) in plasma vary among individuals from <1.0 to >150 mg/dl over a range of 1,000-fold. A population with a higher plasma Lp(a) level than 25-30 mg/dl has a higher incidence of acute myocardial infarction and cerebral infarction (7,8), while low functional levels of plasminogen in plasma due to molecular abnormalities are associated with venous thrombosis (9,10).

In the present study, in order to elucidate the control mechanism of its expression, the gene coding for apo(a) was characterized and nucleotide

*Address correspondence to Akitada Ichinose, M.D., Ph.D. at Yamagata University. Fax: 81-236-24-4534.

0006-291X/95 $5.00

sequences of all the exons and intron/exon boundaries in the apo(a) gene were determined. In addition, differences in the nucleotide sequence of its 5'-flanking region were identified among individuals.

MATERIALS AND METHODS

Venous blood was drawn after informed consent had been obtained from normal individuals and patients with myocardial infarction. Genomic DNA samples were extracted from the leukocytes. PCRable genomic DNAs of normal individuals were also purchased from BIOS Lab. (New Haven, CT). All reagents used in this study were of analytical grades.

Sequence analysis. Three genomic libraries were screened by *in situ* hybridization to obtain the phage clones which contain the apo(a) gene (5). Phage DNA was amplified and prepared by liquid culture lysis. Genomic DNA inserts were isolated by digestion of the phage DNA with *Eco*RI and/or *Sal*I endonuclease(s) and then subcloned into plasmid pUC18 or pUC19. DNA fragments were also subcloned into M13mp18 or M13mp19, then sequenced by the dideoxy termination method (11) employing [^{35}S]dATP and buffer gradient gels. The phage clones (broken lines in Figure 1) were confirmed to contain exons coding for the apo(a) gene. The genomic DNAs were also amplified (solid lines in Figure 1) and sequenced by employing gene-specific primers designed from the cDNA and genomic sequences for apo(a). All regions were examined in more than two individuals and their nucleotide sequences were consistent except for polymorphisms as described below. Other related genes were excluded by nucleotide sequencing.

***In vitro* amplification.** 0.1-1.0 μg of a genomic DNA was amplified in a 50 μl reaction mixture as described (12) employing 2.5-5.0 units of *Thermus aquaticus* DNA polymerase (Taq polymerase, StrataGene). After 30 cycles of amplification, nine μl of each reaction mixture was applied to a 0.8% agarose (International Biotechnologies Inc.) gel containing ethidium bromide (0.5 μg/ml) in 1 X TBE buffer (89 mM Tris base/89 mM boric acid/20 mM EDTA, pH 7.8).

Oligonucleotides were prepared with an Applied Biosystems' synthesizer. Genomic DNAs were amplified by employing primers designed from the genomic sequences for apo(a); for amplification and sequencing to detect polymorphisms in the 5'-flanking region, 5' side-CTTGAATTCCCAAAGTGCTGGGATTACAGAG (A2-53; underlined nucleotides are identical to those of the apo(a) gene) & 3' side-TAAGGATCCGGCATATGTATTTTACTACATTGTGGGAG (A5FL-3).

RESULTS AND DISCUSSION

Two homologous genes which may code for apo(a) [Apo(a)I & II in Figure 4A in ref. 5] were identified in the previous study. It is, however, very likely that only one of these genes is the true apo(a) gene which expresses the protein, since no individual has more than two apo(a)

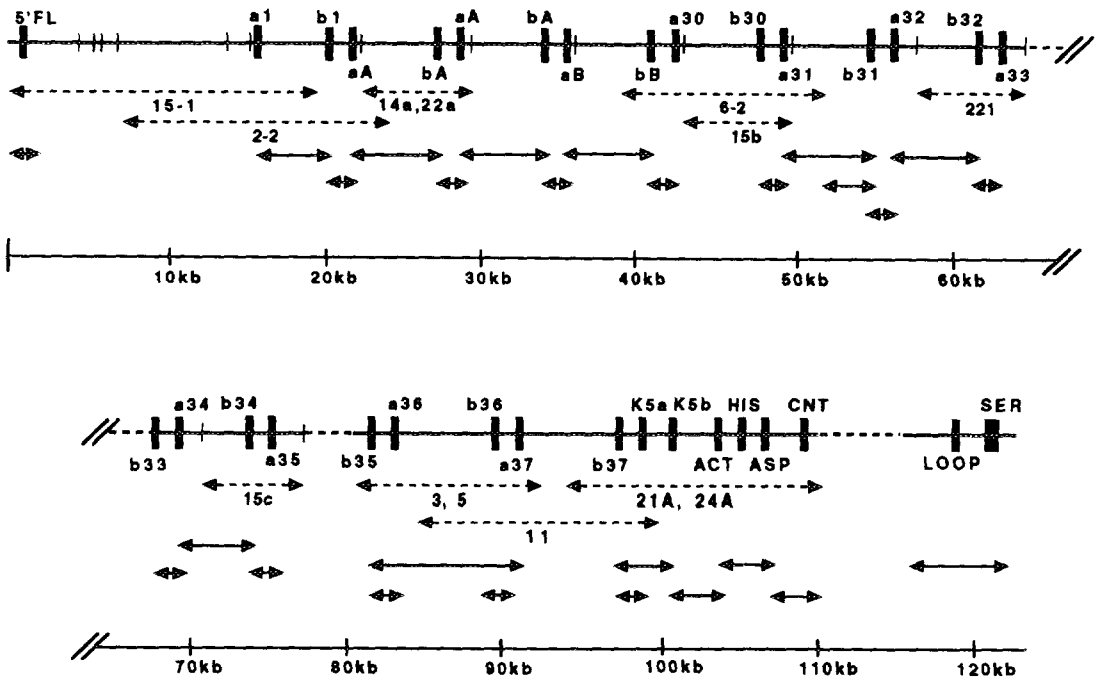

Figure 1. Organization of the gene for apo(a). The gene model is based on isolated genomic phage clones (broken arrows) except for two 3'-exons. In addition, all exons and their boundaries including the two 3'-exons of apo(a) were obtained by *in vitro* amplification of genomic DNAs (solid arrows). Exons are shown by wide vertical bars with the names of regions as defined for the plasminogen gene in ref. 6. K4a stands for an exon coding for a first half and K4b for a second half of a kringle 4 repeat.

isoforms, each of which is inherited from either parent. The true apo(a) gene is most likely apo(a)I because it has typical regulatory elements for gene expression, an exon encoding a signal peptide, and exons coding for kringle 4 repeats which correspond exactly to the cDNA for apo(a) (4), while Apo(a)II does not. It is confirmed that Apo(a)I contains exons coding for kringles 1-33 of apo(a) (13); thus, Apo(a)I is designated as the apo(a) gene hereafter. The gene for apo(a) spans more than 120 kb (Figure 1), assuming that it has one each of 11 types of kringle 4 (1, A, B, 30-37 according to ref. 4). This is consistent with a report that a yeast artificial chromosome (YAC) clone contained the apo(a) gene of about 120 kb, although its 3'-portion was truncated (13). The total length of the apo(a) genes depends upon the number of kringle 4 repeats in each individual (14).

The organization of the human apo(a) gene, in terms of the locations of introns (Figure 1) and the types of splice junctions, was identical to the gene coding for plasminogen (Table 1). All splice donors and acceptor

Table 1: Sequence and type of Intron/Exon boundaries of the apo(a) gene

Region	Size (bp)	Exon	Boundary sequence	Intron	Boundary sequence	Exon	type
5'-NC	(95)	I	..CTGAAATCAG GTAAGA...	A	...TTTTAG CAGCACCTGA..	II	I
K4a1	160	II-1	..ACCCAAATGC GTATGT...	B1	...TTCCAG TGGCTTGATC..	III1	II
K4b1	182	III-1	..TCCGAACAAG GTAAGG...	C1	...TTTCAG CACCGACTGA..	IV1	I
K4aA	160	II-2	..ACCCAAATGC GTATGT...	B2	...TTCCAG TGGCTTGATC..	III2	II
K4bA	182	III-2	..TCCGAACAAG GTAAGG...	C2	...TTTCAG CACCGACTGA..	IV2	I
K4aB	160	II-24	..ACCCAAATGC GTATGT...	B24	...TTCCAG TGGCTTGATC..	III24	II
K4bB	182	III-24	..TCCGAACAAG GTAAGG...	C24	...TTTCAG CACCGACTGA..	IV24	I
K4a30	160	II-30	..ACCCAAATGC GTATGT...	B30	...TTCCAG TGGCTTGATC..	III30	II
K4b30	182	III-30	..TCTGAACAAG GTAAGG...	C30	...TTTCAG CACCAACTGA..	IV30	I
K4a31	160	II-31	..ACCCAAATGC GTATGT...	B31	...TTCCAG TGGCTTGATC..	III31	II
K4b31	182	III-31	..TTTGAACAAG GTAAGA...	C31	...TTTCAG CACTGACTGA..	IV31	I
K4a32	160	II-32	..ACCCAAATGC GTCTGT...	B32	...TTGCAG TGGCCTGACC..	III32	II
K4b32	182	III-32	..TCTGAAGAAG GTAGGA...	C32	...TTTCAG CACCAACGGA..	IV32	I
K4a33	160	II-33	..ACCCAAATGG GTACAA...	B33	...TTGCAG TGGCCTGACC..	III33	II
K4b33	182	III-33	..TCTGAACAAG GTAAGA...	C33	...TTTCAG CACCAACGGA..	IV33	I
K4a34	160	II-34	..ACCCAAATGG GTATGT...	B34	...TTGCAG TGGCCTGACC..	III34	II
K4b34	182	III-34	..TCTGAAGAAG GTAAGA...	C34	...TTTCAG CACCAACTGA..	IV34	I
K4a35	160	II-35	..ATCCAAATGC GTATGT...	B35	...TTGCAG TGGCCTGACC..	III35	II
K4b35	182	III-35	..TCTGAACAAG GTAAGA...	C35	...TTTCAG CACCACCTGA..	IV35	I
K4a36	160	II-36	..ACCCAAATGC GTATGT...	B36	...TTGCAG TGGCCTGACC..	III36	II
K4b36	182	III-36	..TCTGAAGCAG GTAAGA...	C36	...TTTCAG CACCAACTGA..	IV36	I
K4a37	160	II-37	..ACCCAAATGA GTATGG...	B37	...TTCCAG TGGCCTGACA..	III37	II
K4b37	182	III-37	..TCTGAACAAG GTAAGA...	C37	...CTACAG ACTGTATGTT..	IV37	I
K5a	149	IV	..GGAAAAAAAT GTAAGC...	D	...TTTCAG TACTGCCGTA..	V	0
K5b	94	V	..CCTCTCTGTG GTAAGT...	E	...CCACAG CATCCTCTTC..	VI	I
ACT	121	VI	..TTAGGACAAG GTAAGA...	F	...TTCTAG GTTTGGAAAG..	VII	II
HIS	75	VII	..GCTTGAAGAA GTACGT...	G	...TTCTAG GTCCTCAAGG..	VIII	II
ASP	141	VIII	..AGCTAAGCAG GTACTC...	H	...TTACAG GCCTGCCGTC..	IX	II
CNT	107	IX	..GAAACCCAAG GTGAGA...	I	...ACACAG GTACCTTTGG..	X	I
LOOP	119	X	..CAGTTGCCAG GTAAGA...	J	...GTTTAG GGTGACAGTG..	XI	0
SER	403	XI					
+3'NC			A		TT T		
		CONSENSUS	AG GTGAGT-		-CCNCAG G		

sequences are consistent with the consensus sequence (15). All 11 types of kringle 4 repeats in apo(a) were coded by two exons, and introns were inserted at positions precisely identical to the plasminogen gene——in the middle and at both ends of each kringle 4. The splice junctions of the exons coding for the (pseudo)serine protease domain of apo(a) were also located in exactly the same positions as those of plasminogen (6). These findings support the hypothesis that these genes were derived from a common ancestor through duplication and exon shuffling.

As shown in Figure 2, the apo(a) and plasminogen genes share more than 95% sequence identity up to about 710 nucleotides upstream of the initiator Met, while apo(a)II has a high degree of sequence similarity to these two genes about 160 nucleotides upstream. There are typical regulatory sequences, including TATAA and CAAT, in front of the proposed transcription site 141 nucleotides upstream of the initiator Met (16). One mismatch at nucleotide -21 of the apo(a) gene was observed when compared with the apo(a) cDNA (4); however, this anomaly turned out to be one of the sequence polymorphisms among individuals as described below.

Figure 2. Nucleotide sequence of exon I and its flanking regions in the apo(a) and homologous genes. Nucleotide numbers are shown in the right margin. The top (P), middle (A), and bottom (X) lines represent the sequences of the genes for plasminogen, Apo(a) [Apo(a)I], and Apo(a)II, respectively. Asterisks show nucleotides identical to those of the plasminogen gene. Sequences used in designing amplification primers are underlined.

Since both the plasma concentration and apparent molecular size of apo(a) are inherited in an autosomal co-dominant manner (17), there could be differences in its genomic sequence among individuals. The DNA-size polymorphism, that is, the number of kringle 4 repeats, explains in part the variation in plasma Lp(a) levels, that is, apo(a) levels (18). However, plasma Lp(a) concentrations vary widely even within the same apo(a) isoform (19,20). It is also suggested that there are *cis*-acting sequences at the apo(a) locus, which would account for the variation in plasma Lp(a) levels among individuals (18). Thus, it is very likely that the 5'-flanking region of the apo(a) gene is important in determining Lp(a) levels through transcriptional regulation and that there are sequence polymorphisms in that region.

As expected, nucleotide sequence analysis of the 5'-flanking region in the apo(a) gene among individuals revealed that there are differences in four nucleotides at positions -914, -103, -49, and -21, respectively (21). In contrast, no difference in nucleotide sequence was found in the Apo(a)II gene among these individuals. Some selective pressure may have kept the Apo(a)II gene unchanged because of its providing an important——though as yet unknown——function(s).

The apo(a) genomic sequence shown in Figure 2 also has several nucleotides which are different from those reported by Wade et al. (16). Some of these may be additional genomic polymorphisms and the others may be sequencing artifacts.

ACKNOWLEDGMENTS

This paper was presented in part at the XIIth ICTH meeting in Amsterdam (1991) and the XIth ICF meeting in Copenhagen (1992). The author thanks Prof. E. W. Davie for his support, Prof. G. Brown for providing blood samples, E. S. Espling for his excellent assistance, J. Harris for synthesis of oligonucleotides in the initial part of this study, and L. Boba for her help in the preparation of the manuscript. This work was supported by research grants from the NIH USA (HL 16919), Yamagata University, the Ministry of Education, Science and Culture, Japan (05454327), Yamagata Prefecture, Ono Medical Research Foundation, CIBA-GEIGY Foundation for the Promotion of Science, and T. Nanba Memorial Health Care Foundation.

REFERENCES

1. Berg, K. (1963) Acta Pathol. Microbiol. Scand. 59, 369-382.
2. Utermann, G. (1989) Science 246, 904-910.
3. Scanu, A. M., and Fless, G. M. (1990) J. Clin. Invest. 85, 1709-1715.

4. McLean, J. W., Tomlinson, J. E., Kuang, W. -J., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M., and Lawn, R. M. (1987) Nature 330, 132-137.
5. Ichinose, A. (1992) Biochemistry 31, 3113-3118.
6. Petersen, T. E., Martzen, M. R., Ichinose, A., and Davie, E. W. (1990) J. Biol. Chem. 265, 6104-6111.
7. Rhoads, G. G., Dahlen, G., Berg., K., Morton, N. E., and Dannenberg, A. L. (1986) Am. J. Med. Assoc. 256, 2540-2544.
8. Murai, A., Miyahara, T., Fujimoto, N., Matsuda, M., and Kameyama, M. (1986) Atheroscler. 59, 199-204.
9. Aoki, N., Moroi, M., Sakata, Y., Yoshida, N., and Matsuda, M. (1978) J. Clin. Invest. 61, 1186-1195.
10. Ichinose, A., Espling, E. S., Takamatsu, J., Saito, H., Shinmyozu, K., Maruyama, I., Petersen, T. E., and Davie, E. W. (1991) Proc. Natl. Acad. Sci. USA 88, 115-119.
11. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A 74, 5463-5467.
12. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) Science 239, 487-491.
13. Malgaretti, N., Acquati, F., Magnaghi, P., Bruno, L., Pontoglio, M., Rocchi, M., Sccone, S., Valle, G. D., D'Urso, M., LePaslier, D., Ottolenghi, S., and Taramelli, R. (1992) Proc. Natl. Acad. Sci. USA 89, 11584-11588.
14. Lackner, C., Boerwinkle, E., Leffert, C. C., Rahmig, T., and Hobbs, H. H. (1991) J. Clin. Invest. 87, 2153-2161.
15. Senapathy, P., Shapiro, B. M., and Harris, N. L. (1990) Method. Enz. 183, 252-278.
16. Wade, D. P., Clarke, J. G., Lindahl, G. E., Liu, A. C., Zysow, B. R., Meer, K., Schwartz, K., and Lawn, R. M. (1993) Proc. Natl. Acad. Sci. USA 90, 1369-1373.
17. Utermann, G., Menzel, H. J., Kraft, H. G., Duba, H. C., Kemmler, H. G., and Seitz, C. (1987) J. Clin. Invest. 80, 458-465.
18. Boerwinkle, E., Leffert, C. C., Lin, J., Lackner, C., Chiesa, G., and Hobbs, H. H. (1992) J. Clin. Invest. 90, 52-60.
19. Perombelon, Y. F. N., Soutar, A. K., and Knight, B. L. (1994) J. Clin. Invest. 93, 1481-1492.
20. Rader, D. J., Cain, W., Ikewaki, K., Tally, G., Zech, L. A., Usher, D., and Brewer, H. B. (1994) J. Clin. Invest. 93, 2758-2763.
21. Ichinose, A., and Kuriyama, M. (1995) Biochem. Biophys. Res. Commun. 209, 372-378.