ApoB gene MspI RFLP in exon 26 changes amino acid 3611 from Arg to Gln

Li-Shin Huang, Jacqueline de Graaf, and Jan L. Breslow

Rockefeller University, New York, NY 10021

Abstract An apolipoprotein B gene MspI RFLP was identified by the use of a probe to a portion of the 3' end of the gene. By Southern blotting analysis after digestion with MspI, this probe detected either a 9 kb or a 2.6 kb fragment. Family studies showed that these corresponded to alleles that segregated in a simple Mendelian fashion. The minor allele (9.0 kb) had a frequency of approximately 12% in an unrelated Caucasian population. Restriction mapping showed that the minor allele was due to the loss of an MspI site in exon 26. Sequencing of both alleles in the region containing the polymorphic MspI site revealed a single-base pair alteration which abolished the MspI site at codon 3611 of the mature apoB protein. In the major allele, this codon is CGG, which specifies Arg; whereas in the minor allele, it was CAG, which codes for Gln.—**Huang, L-S., J. de Graaf, and J. L Breslow.** ApoB gene MspI RFLP in exon 26 changes amino acid 3611 from Arg to Aln. *J. Lipid Res.* 1988. **29:** 63–67.

Supplementary key words genetic variants • restriction fragment length polymorphism

Apolipoprotein B-100 (apoB-100) is the major protein of low density lipoprotein (LDL). It is the ligand for the LDL receptor, which mediates high affinity cellular LDL uptake (1). Recent studies show that apoB-100 mRNA is 14,121 bases in length (2, 3) and codes for a 4563-amino acid long protein containing a 27-amino acid signal peptide (4, 5). The apoB gene maps to chromosome 2 p24-pter, is 43 kb in length, and contains 28 introns (6–8).

In epidemiological studies, LDL-cholesterol and apoB levels are positively correlated with coronary heart disease incidence (9). Because of the role of apoB in lipoprotein and cholesterol metabolism, genetic variations in apoB may play a role in atherosclerosis susceptibility. Genetic variants of apoB in pigs, detected by alloantisera, have been shown to be associated with the development of raised lesions in the aortic wall in response to a high fat diet (10). In the case of human apoB, we have used apoB cDNA and genomic probes to identify several restriction fragment length polymorphisms (RFLPs). Some of these RFLP markers have been associated with myocardial infarction (11).

In the current report, a probe to a portion of the 3' end of the apoB gene was used to demonstrate an MspI RFLP. This polymorphism is due to a loss of an MspI site resulting from a base substitution in codon 3611 of the mature B-100 protein. The base substitution causes an amino acid change from arginine to glutamine.

MATERIALS AND METHODS

Probe preparation

The apoB probes used were isolated from plasmid clones pB27 and pB23 (see Ref. 11 and Fig. 1). Plasmid DNAs from both clones were digested with EcoRI to release the inserts and subjected to low melting temperature agarose gel electrophoresis. These inserts were then isolated and nick-translated (12) for use as probes either in Southern blotting analysis or plasmid library screening.

Southern blotting analysis

DNAs from unrelated individuals or members of families were isolated from peripheral lymphocytes or lymphoblast cell lines as previously described (7). Lymphoblast cell lines were obtained from the National Institute of General Medical Science's Human Genetic Mutant Cell Repository (Camden, NJ). Southern blots were prepared following complete digestion of DNAs with MspI and electrophoresis through 0.7% agarose gels (12). Compared to other MspI sites, this particular one tended to be digested partially. To ensure complete cleavage of genomic DNA, 80 units of MspI (New England Biolabs) was added to 10 μg of DNA and digestion was carried out at 37°C for 3 hr. Subsequently, an additional aliquot of enzyme was added, and incubation was continued overnight. The blots were hybridized with a nick-translated

Abbreviations: apo, apolipoprotein; LDL, low density lipoprotein; RFLP, restriction fragment length polymorphism.

pB27 or pB23 probe, washed with 0.1 × SSC, 0.1% SDS (1 × SSC = 150 mM NaCl and 15 mM sodium citrate) at 65°C, and exposed to X-ray film.

Restriction mapping of the 3' end of the human apoB gene

Restriction mapping was carried out with genomic DNAs from two individuals homozygous for the different alleles. The probes used were pB23 and pB27. The relationship of intron and exon junctions is taken from the map described by Blackhart et al. (8).

Molecular cloning of the MspI polymorphic alleles

Clone pB23 is a subclone of lambda genomic clone λB1 (11), which has an EcoRI fragment (**Fig. 1**) containing the polymorphic MspI site. To obtain the allele that lacks the MspI site, DNA from an individual homozygous for that allele (lane 4 of Fig. 2) was digested to completion with EcoRI. The digested DNA was electrophoresed in a 4% polyacrylamide gel, and DNA between 1.6 to 1.8 kb in length was selected and eluted from the gel. The size-selected fraction was ligated to EcoRI-digested pUC 18 vector and used to transform *E. coli.* DH5 competent cells (BRL) (13). The size-selected genomic library was screened with the pB23 probe and clone pB72 was isolated.

DNA sequencing analysis

Two oligonucleotide primers were synthesized by phosphoramidite chemistry on a DNA synthesizer (Applied Biosystems, Model 381 A). Both primers are 15 bases long. Primer A (GAATGCTAACACTAA) is 30 bases 5' to the polymorphic MspI site (CCGG). Primer B (TGGTCATTGGAAAGC) is complementary to the sense strand and is 30 bases 3' to the polymorphic MspI site. These primers were used to sequence both strands from clones pB23 and pB72 with the dideoxy method of Sanger, Nicklen, and Coulson (14).

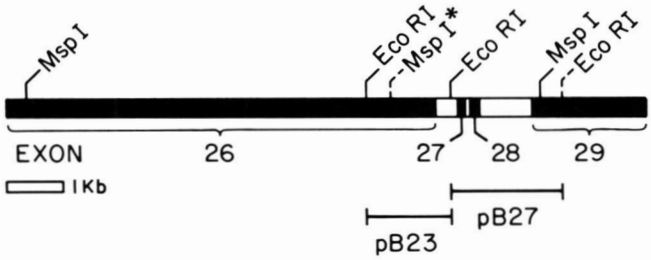

Fig. 1. Restriction map of the 3' end of the human apoB gene. Genomic DNAs from individuals with different alleles were digested with the indicated restriction enzymes. Dash lines indicate polymorphic sites. The asterisk indicates the polymorphic MspI site described in the text. Solid bars are exons, and open bars are introns. The relationships of introns and exons are cited from Blackhart et al. (18). The scale is shown below the map. Two plasmid clones, pB23 and pB27, containing different EcoRI fragments of the apoB gene are indicated below the map. The scale of the map is shown in kb.

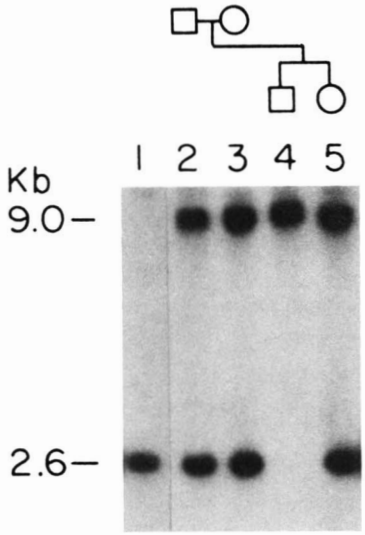

Fig. 2. An apoB MspI RFLP and its inheritance. Human DNAs isolated from peripheral blood lymphocytes of a normal individual (lane 1) and members from a family (lanes 2 to 5) were digested with MspI and subjected to Southern blotting analysis. The membrane was hybridized to ^{32}P-labeled pB27 probe and exposed to X-ray film overnight. Two fragments that hybridized to the probe are 9.0 and 2.6 kb, respectively. The segregation of the two alleles that give 9.0 or 2.6 kb fragments is shown in a family (lanes 2–5). The pedigree of this family is shown directly above the autoradiogram.

RESULTS

ApoB gene MspI RFLP and its inheritance

DNAs from members of a representative family were digested with MspI, Southern blotted, and hybridized to probe pB27. **Fig. 2** (lanes 2 to 5) shows that there are two alleles that give fragments with sizes of 2.6 and 9.0 kb, respectively. These two alleles segregated in a simple Mendelian fashion. Studies in three other families (data not shown) confirmed this pattern. No homozygotes of the 2.6 kb allele were found in this family. Included in Fig. 2 (lane 1) is DNA from an unrelated normal individual who was a homozygote for the 2.6 kb band. In order to determine the frequency of these two alleles, DNAs from 28 unrelated individuals were subjected to the same Southern blotting analysis. The results show that the 9.0 kb allele is the minor allele and has a frequency of approximately 12% in a normal Caucasian population (data not shown).

MspI RFLP is due to a single base substitution

The position of the polymorphic MspI site was determined by restriction mapping of genomic DNAs from individuals with two different alleles (Fig. 1). When pB23,

which begins in exon 26 and ends in intron 26, was used to hybridize to the MspI genomic blot, an extra 6.4 kb fragment was seen (data not shown) in addition to the 9.0 and 2.6 kb fragments seen with the pB27 probe. This result indicated that the allele giving rise to the 9 kb fragment is due to the disappearance of an MspI site, and the polymorphic MspI site is within the 1.7 kb EcoRI fragment included in pB23. The polymorphic and invariant MspI sites in exon 26 (Fig. 1) were confirmed by the sequence of the apoB cDNA previously reported (2, 3).

Restriction mapping and DNA sequencing of clone pB23 showed that this clone represents the major allele **(Fig. 3,** right panel) which has the polymorphic MspI site (CCGG) at a position corresponding to codons 3610 and 3611. Codon 3611 (CGG) codes for the amino acid arginine. Restriction mapping and DNA sequencing of clone pB72, obtained from an individual homozygous for the absence of the MspI site, shows a single base substitution of G to A in this DNA region. Codon 3611 has been changed to CAG which codes for the amino acid glutamine (Fig. 3, left panel). This abolishes the recognition sequence for the MspI enzyme.

DISCUSSION

An apoB gene MspI RFLP at the 3' end of the gene is reported. The polymorphism is due to a loss of an MspI site resulting from a base substitution in codon 3611 of the mature protein. The codon 3611 was changed from CGG to CAG which changes the amino acid from arginine to glutamine. The minor allele, which has the Gln in codon 3611, has a frequency of approximately 12% in a Caucasian population.

ApoB binds to the LDL receptor as does apoE. Structural variants of apoE have been identified where arginine has been replaced with cysteine which resulted in the loss of receptor binding (15). The MspI RFLP involves loss of an arginine at residue 3611, but one cannot be sure this is in a region that affects apoB function. It has been proposed that the basic regions of apoB between amino acids 3147 and 3157 and 3359 to 3367 may be involved in LDL receptor binding (4).

The MspI restriction enzyme recognition site contains the dimer sequence CpG. The cytosine in CpG sequences is known to be frequently methylated in mammals, and the deamination of 5-methyl cytosine results in the formation of a thymine. This is thought to result in a hot spot for C to T transition (16). In the CpG to CpA transition reported here, the deamination has presumably occurred first in the C on the antisense strand according to this proposed theory. In human factor VIII gene, two point mutations occurred frequently in the CpG dinucleotides in exon 18 and 22, respectively (17). They have been shown to occur independently in several unrelated families that have hemophilia A and, therefore, support the view that CpG dinucleotides are mutation hotspots (17).

Genetic variants in the structural or regulatory regions of the apoB gene could cause abnormalities in lipoprotein metabolism and even premature atherosclerosis. Previous studies have demonstrated the presence of structural variants of apoB detected with allotypic antisera (18). For example, antigenic variants of human LDL have been found and designated the Ag system. This system has five pairs of antigenic variants: Ag(al/d), Ag(c/g), Ag(h/i), Ag(t/z), and Ag(x/y). A recent study has assigned these variants to the apoB protein, and the Ag alleles likely reflect apoB protein polymorphisms (19). It has been shown that serum cholesterol and triglyceride levels are higher in Ag(x –) than in Ag(x +) individuals (20). Since the Ag(x) epitope is associated with altered plasma lipid levels, it might be in a domain that is important for lipoprotein metabolism. The Ag(x) epitope has also been shown to be in strong linkage disequilibrium with an apoB XbaI RFLP (19). The variant XbaI site is due to a silent cytosine to thymine substitution in the third base of the threonine codon at residue 2488 in the mature protein (19). Therefore, the XbaI RFLP cannot represent the Ag(x) epitope. In our studies, the apoB MspI RFLP is in linkage equilibrium with the XbaI RFLP (data not

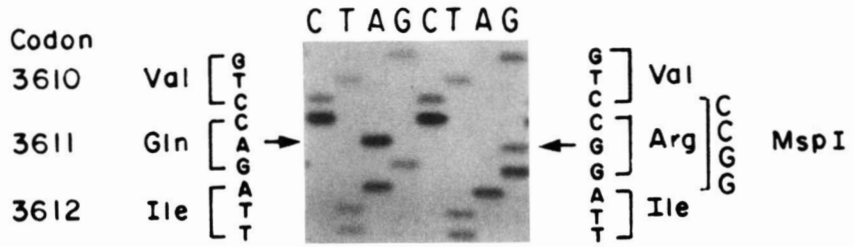

Fig. 3. DNA sequence analysis of clone pB72 (left) and pB 23 (right). Both clones were sequenced as described in Materials and Methods. A portion of the autoradiogram from an 8% polyacrylamide/7 M urea gel is shown. Arrows indicate the one-base difference in two clones. The major allele contains the polymorphic MspI site (CCGG), whereas the minor allele contains the sequence CCAG. The amino acid codon numbers are indicated.

shown) and probably not the site of the Ag(x/y) alleles. The Ag(t/z) alleles have been shown to be the site of the apoB EcoRI RFLP (21). The EcoRI site polymorphism alters the apoB protein sequence by substituting lysine (AAA) for glutamic acid (GAA) at codon 4154 in the mature protein (19). The amino acid difference for different EcoRI alleles is thought to be the cause for the Ag(t/z) epitopes (21). A monoclonal antibody, MB19, has been shown to detect two alleles of the apoB protein (22, 23). One allele codes for apoB that has high affinity for the antibody, whereas the other allele codes for apoB that has low affinity for the antibody. Heterozygotes for the alleles have an intermediate affinity for the antibody. The high affinity allele has been shown to correspond to the Ag(c) epitope and the low affinity allele to the Ag(g) allele (24). Since the MB19 antibody reacts with the B-26 portion of B-100 protein, which is in the amino terminal region (25) of the protein, and the MspI RFLP reported here is caused by codon 3611 variation, the MspI RFLP could not be the basis for MB19 reactivity. Due to the unavailability of the Ag antisera, we have been unable to determine whether the MspI RFLP corresponds to one of the two remaining pairs of Ag alleles, Ag(al/d) or Ag(h/i).

In addition to the EcoRI and XbaI RFLPs described above, an insertion-deletion (ID) type length polymorphism has been shown just 3' to the end of the apoB gene. The ID polymorphism is due to variable numbers of tandem repeats of AT-rich core sequences (26). Alleles of all three RFLPs, EcoRI, XbaI and ID, have been associated in a case-control study with myocardial infarction but not with elevated levels of LDL-cholesterol or apoB (11). The structural variant of apoB reported in this paper will be useful as a new marker for clinical studies.

This research was supported in part by grants from the National Institutes of Health (HL33714, HL32435, HL36461, AG04727, CA29502) and a General Clinical Research Center grant (RR00102), as well as general support from the Pew Trusts. Dr. Jan L. Breslow is an Established Investigator of the American Heart Association. We wish to thank Ms. Mona S. Spector for the synthesis of oligonucleotides and Ms. Lori Cersosimo for her expertise in preparing the manuscript.

Manuscript received 6 August 1987.

REFERENCES

1. Herbert, P. N., G. Assmann, A. M. Gotto, Jr., and D. S. Fredrickson. 1983. Familial lipoprotein deficiency: abetalipoproteinemia, hypobetalipoproteinemia, and Tangier disease. *In* The Metabolic Basis of Inherited Disease. J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson, editors. McGraw-Hill Publications, New York, NY. 589-621.
2. Chen, S-H., C-Y. Yang, P-F. Chen, D. Setzer, M. Tanimura, W-H. Li, A, M., Gotto, Jr., and L. Chan. 1986. The complete cDNA and amino acid sequence of human apolipoprotein B-100. *J. Biol. Chem.* **261:** 12918-12921.
3. Knott, T. J., S. C. Wallis, L. M. Powell, R. J. Pease, A. J. Lusis, B. Blackhart, B. J. McCarthy, R. W. Mahley, B. Levy-Wilson, and J. Scott. 1986. Complete cDNA and derived protein sequence of human apolipoprotein B-100. *Nucleic Acids Res.* **14:** 7501-7503.
4. Knott, T. J., R. J. Pease, L. M. Powell, S. C. Wallis, S. C. Rall, Jr., T. L. Innerarity, B. Blackhart, W. H. Taylor, Y. Marcel, R. Milne, D. Johnson, M. Fuller, A. J. Lusis, B. J. McCarthy, R. W. Mahley, B. Levy-Wilson, and J. Scott. 1986. Complete protein sequence and identification of structural domains of human apolipoprotein B. *Nature.* **323:** 734-738.
5. Yang, C-Y., S-H. Chen, S. H. Gianturco, W. A. Bradley, J. T. Sparrow, M. Tanimura, W-H. Li, D. A. Sparrow, H. DeLoof, M. Rosseneu, F-S. Lee, Z-W. Gu, A. M. Gotto, Jr., and L. Chan. 1986. Sequence, structure, receptor-binding domains and internal repeats of human apolipoprotein B-100. *Nature.* **323:** 738-742.
6. Knott, T. J., S. C. Rall, Jr., T. L. Innerarity, S. F. Jacobson, M. S. Urdea, B. Levy-Wilson, L. M. Powell, R. J. Pease, R. Eddy, H. Nakai, M. Byers, L. M. Priestley, E. Robertson, L. B. Rall, C. Bertsholtz, T. B. Shows, R. W. Mahley, and J. Scott. 1985. Human apolipoprotein B: structure of carboxyl-terminal domains, sites of gene expression and chromosomal localization. *Science.* **230:** 37-43.
7. Huang, L-S., D. A. Miller, G. A. P. Bruns, and J. L. Breslow. 1986. Mapping of the human apoB gene to chromosome 2p and demonstration of a two-allele restriction fragment length polymorphism. *Proc. Natl. Acad. Sci. USA.* **83:** 644-648.
8. Blackhart, B. D., E. M. Ludwig, V. R. Pierotti, L. Caiati, M. A. Onasch, S. C. Wallis, L. Powell, R. Pease, T. J. Knott, M-L. Chu, R. W. Mahley, J. Scott, B. J. McCarthy, and B. Levy-Wilson. 1986. Structure of the human apolipoprotein B gene. *J. Biol. Chem.* **261:** 15364-15367.
9. Heiss, G., and H. A. Tyroler. 1982. Are apolipoproteins useful for evaluating ischemic heart disease? A brief overview of the literature. *In* Proceedings of the Workshop on Apolipoprotein Quantification. National Institutes of Health, Bethesda, MD, NIH Publication 83-1266: 7-24.
10. Rapacz, J., C. E. Elson, and J. J. Lalich. 1977. Correlation of an immunogenetically defined lipoprotein type with aortic intimal lipidosis in swine. *Exp. Mol. Pathol.* **27:** 249-261.
11. Hegele, R. A., L-S. Huang, P. N. Herbert, C. B. Blum, J. E. Buring, C. H. Hennekens, and J. L. Breslow. 1986. Apolipoprotein B-gene DNA polymorphisms associated with myocardial infarction. *N. Engl. J. Med.* **315:** 1509-1515.
12. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
13. Nicholls, R. D., A. V. S. Hill, J. B. Glegg, and D. R. Higgs. 1985. Direct cloning of specific genomic DNA sequences in plasmid libraries following fragment enrichment. *Nucleic Acids Res.* **13:** 7569-7578.
14. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74:** 5463-5467.
15. Rall, S. C., Jr., K. H. Weisgraber, T. L. Innerarity, and R. W. Mahley. 1982. Structural basis for receptor binding heterogeneity of apolipoprotein E from type III hyperlipoproteinemic subjects. *Proc. Natl. Acad. Sci. USA.* **79:** 4696-4700.
16. Barker, D., M. Schafer, and R. White. 1984. Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell.* **36:** 131-138.

17. Youssoufian, H., H. H. Kazazian, Jr., D. G. Phillips, S. Aronis, G. Tsiftis, V. A. Brown, and S. E. Antonarakis. 1986. Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. *Nature.* **324:** 380–382.
18. Butler R., E. Brunner, and G. Morganti. 1974. Contribution to the inheritance of the Ag groups. *Vox Sang.* **26:** 485–496.
19. Berg, K., L. M. Powell, S. C. Wallis, R. Pease, T. J. Knott, and J. Scott. 1986. Genetic linkage between the antigenic group (Ag) variation and the apolipoprotein B gene: assignment of the Ag locus. *Proc. Natl. Acad. Sci. USA.* **83:** 7367–7370.
20. Berg, K., C. Hames, G. Dahlen, M. H. Frick, and I. Krishan. 1976. Genetic variation in serum low density lipoproteins and lipid levels in man. *Proc. Natl. Acad. Sci. USA.* **73:** 937–940.
21. Ma, Y., V. N. Schumaker, R. Butler, and R. S. Sparkes. 1987. Two DNA restriction fragment length polymorphisms associated Ag(t/z) and Ag(g/c) antigenic sites of human apolipoprotein B. *Arteriosclerosis.* **7:** 301–305.
22. Schumaker, V. N., M. T. Robinson, L. K. Curtiss, R. Butler, and R. S. Sparkes. 1983. Anti-apoprotein B monoclonal antibodies detect human low density lipoprotein polymorphism. *J. Biol. Chem.* **259:** 6423–6430.
23. Young, S. G., S. J. Bertics, L. K. Curtiss, D. C. Casal, and J. L. Witztum. 1986. Monoclonal antibody MB19 detects genetic polymorphism in human apolipoprotein B. *Proc. Natl. Acad. Sci. USA.* **83:** 1101–1105.
24. Robinson, M. T., V. N. Schumaker, R. Butler, K. Berg, and L. K. Curtiss. 1986. Ag(c): recognition by a monoclonal antibody. *Arteriosclerosis.* **6:** 341–344.
25. Curtiss, L. K., and T. S. Edgington. 1982. Immunochemical heterogeneity of human plasma apolipoprotein B. I. Apolipoprotein B binding of mouse hybridoma antibodies. *J. Biol. Chem.* **257:** 15213–15221.
26. Huang, L-S., and J. L. Breslow. 1987. A unique AT-rich hypervariable minisatellite 3' to the apoB gene defines a high information restriction fragment length polymorphism. *J. Biol. Chem.* **262:** 8952–8955.