Am. J. Hum. Genet. 45:458–464, 1989

High-Resolution Analysis of a Hypervariable Region in the Human Apolipoprotein B Gene

Erwin H. Ludwig,* Walter Friedl,† and Brian J. McCarthy*

*Gladstone Foundation Laboratories for Cardiovascular Disease, Cardiovascular Research Institute, University of California, San Francisco; and †First Department of Medicine, Landeskrankenanstalten Salzburg, Salzburg

Summary

A hypervariable region occurs immediately 3′ of the human apolipoprotein B gene. Several allelic variants of this tandemly repeated sequence can be resolved by genomic blotting. Higher resolution among size variants may be obtained by polymerase-chain-reaction amplification of this region followed by electrophoresis in a denaturing acrylamide gel. Fourteen different alleles containing 25–52 repeats of the basic 15-bp unit were distinguished in a population study of 318 unrelated individuals. This approach should be applicable to pedigree and linkage analysis with the apolipoprotein B gene or other tandemly repeated sequence elements.

Introduction

The human genome contains a class of tandemly repeated sequence elements, usually 10–100 bp in length. In general, they are scattered throughout the genome and may be classified into families of related sequence (Nakamura et al. 1987). In many cases, the number of tandem repeats is highly variable, so that multiple alleles are detectable as length polymorphisms by Southern blot analysis (Ali and Wallace 1988). The degree of heterogeneity within various minisatellites differs, but in extreme cases the frequency of heterozygosity may approach 100% (Jeffreys et al. 1985, 1988*a*). Several of these minisatellites occur in association with various genes of clinical interest. The hyperallelism resulting from the existence of many different hypervariable regions (HVRs) at a single locus within the population has been exploited for linkage studies with the genes for human globin (Jarman et al. 1986), insulin (Bell et al. 1982), collagen type II (Stoker et al. 1985), and apolipoprotein (apo) B (Knott et al. 1986; Huang and Breslow 1987; Jenner et al. 1988), among others. Five different minisatellites occur in the vicinity of the globin gene cluster (Jarman and Higgs 1988).

Apolipoprotein B-100 is the major protein constituent of low-density lipoproteins (LDL) and acts as the ligand for LDL receptor–mediated uptake of LDL. Levels of LDL and apo B-100 in the plasma correlate with the incidence of premature heart disease. In a pig model of atherosclerosis, particular apo B alleles detectable with alloantisera are associated with susceptibility and elevated plasma cholesterol (Rapacz et al. 1986). In the human population, at least 14 common apo B alleles are distinguishable through the use of alloantisera from multiply transfused patients (Lusis 1988). Similarly, a number of apo B gene RFLPs have been reported, some of which are associated with plasma LDL level (Berg 1986). One particular apo B gene point mutation displays a high degree of correlation with hypercholesterolemia (Weisgraber et al. 1988; Soria et al. 1989).

The human apo B gene also contains a minisatellite sequence less than 100 bp 3′ of its second canonical polyadenylation signal (Knott et al. 1986; Huang and Breslow 1987; Jenner et al. 1988). Alleles differ in length by virtue of variable numbers of 15-bp hypervariable elements (HVEs). Larger alleles of this HVR are reported to be more common in myocardial infarction patients (Hegele et al. 1986). In the present paper, we present a rapid method for the analysis of the apo B HVR, a method that provides a higher degree of reso-

Received February 14, 1989; revision received May 10, 1989.

Address for correspondence and reprints: B. J. McCarthy, Ph.D., Gladstone Foundation Laboratories, P.O. Box 40608, San Francisco, CA 94140–0608.

0002-9297/89/4503-0015$02.00

lution among alleles of different sizes than does that offered by Southern blot analysis and that thereby increases the apparent frequency of heterozygosity.

Material and Methods

Plasmids

Plasmid pBB120 contains a 13.5-kb *SalI* fragment from a λ clone isolated from a human genomic library (Blackhart et al. 1986) containing part of the human apo B gene inserted into a Bluescribe™ (Stratagene) vector. Plasmid p12D1Eco contains a 4.8-kb *Eco*RI fragment of the above insert in the same vector.

Oligonucleotides

All oligonucleotides were synthesized using an Applied Biosystems 380B synthesizer. One of them (PCR3) was end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP to 10^6 cpm/ng.

The three oligonucleotide primers used had the following sequences:

1. PCR1—5′ CACAGCAAAACCTCTAGAACA 3′. The 5′ end of this sequence lies 369 bp 3′ of the codon specifying the carboxy-terminal leucine and 121 bp 5′ of the first 15-bp HVE.
2. PCR2—5′ GTTCCTCAGGATCAAAGTATGTAC 3′. The 5′ end lies 125 bp 3′ of the last 15-bp HVE.
3. PCR3—5′ GGAGAAATTATGGAGGGAAAT 3′. This 5′ end lies 427 bp 3′ of the codon specifying the carboxy-terminal leucine and 63 bp 5′ of the first 15-bp HVE.

Polymerase Chain Reaction (PCR)

PCRs were carried out using an automated thermal cycler (Perkin-Elmer/Cetus) and *Thermus aquaticus* DNA polymerase obtained from Cetus. Reactions contained 500 ng genomic or 5 ng plasmid DNA, 100 ng PCR1, 200 ng PCR2, and 200 μM of each dNTP in the buffer recommended by Saiki et al. (1988). Initial denaturation was at 97°C for 10 min. Subsequently, 25 cycles were used, each consisting of 94°C for 1 min for denaturation, 1 min annealing at 55°C, and 5 min extension at 72°C. To label the amplified product, 1 ng ^{32}P-labeled PCR3 was added to each reaction mixture, and five more cycles of PCR amplification were performed under the same conditions.

Electrophoresis

Electrophoresis was conducted in a vertical DNA sequencing apparatus in 4% acrylamide gels in a buffer containing 7 M urea at 45–50°C for 3.5–4.0 h. One-tenth of each PCR reaction was dried down, resuspended in formamide dye mix, and heated to 72°C for 3 min prior to loading. Gels were dried onto paper and autoradiographed.

DNA Sequencing

Sequencing was performed with double-stranded plasmid DNA by the method of Kraft et al. (1988) and with double-stranded linear-amplified DNA by using ^{32}P-labeled primers and Sequenase.™

Genomic DNA

DNA was prepared from white blood cells of 318 Austrian subjects (Paulweber et al., in press).

Results

Sequence Comparison of HVR Alleles

The 3′ end of the human apo B gene is depicted in figure 1. The HVR begins 73 bp 3′ of the second AATAA polyadenylation signal (fig. 1). Different alleles vary in length over a range of about 400 bp as a result of the numbers of two related 15-bp HVEs. One of these HVRs was cloned and sequenced from human genomic DNA in order to compare it with sequences published previously (Knott et al. 1986; Huang and Breslow 1987; Jenner et al. 1988). Rather than being presented in terms of its actual DNA sequence, this HVR is expressed in terms of the convention used by Huang and Breslow (1987) (fig. 2). In this scheme, the first 15-bp sequence (ATAATTAAATATTTT) is designated x, and the second one (ATAATTAAAATATTT) is designated as y; these two sequences differ only at positions 10–12. These repeats alternate and are followed by other repeats of slightly different sequences, designated x′ and y′, for example. Such variants contain single-base substitutions or, in one case y‴, a 4-bp deletion. Figure 2 compares, according to this convention, the six published sequences and the new sequence. It is clear that most of the variation in length of the seven HVR alleles is attributable to the number of x and y HVEs. Furthermore, all seven sequences differ from one another.

Our intention was to develop a high-resolution method for distinguishing HVR alleles differing in length. For that reason, we have used a nomenclature that describes the number of HVEs. Since the sequence is composed of tandem repeats of 15-bp variants, we term these different-length alleles as HVE x, where x is the number of 15-bp repeats. Thus, in figure 2, al-

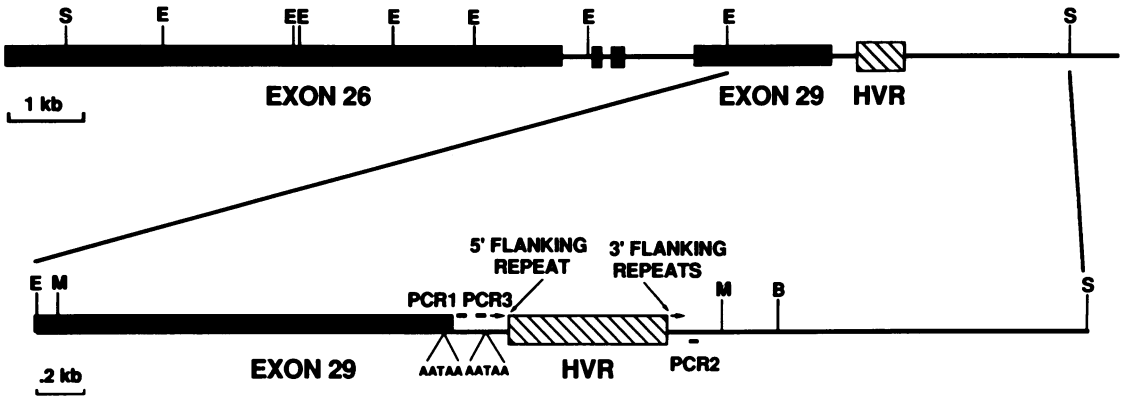

Figure 1 Schematic diagram illustrating the structure of the 3′ end of the human apo B gene (Ludwig et al. 1987). The upper scale shows a 13.5-kb *Sal*I (S) fragment together with seven *Eco*RI (E) sites. The lower scale delineates a subcloned fragment including part of exon 29, the HVR, and its flanking regions. The loci of the two transcriptional stop signals (AATAA) and the three oligonucleotides used as primers for amplification (PCR1, 2, and 3) are indicated. Also shown are *Msp*I (M) and *Bam*HI (B) sites within this subclone. Exons are depicted as solid bars, and the HVR is depicted as a hatched region.

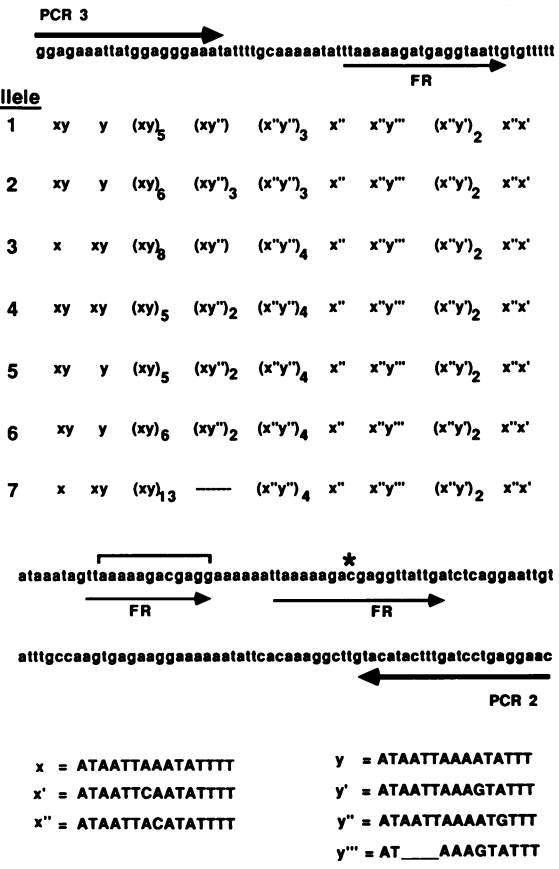

Figure 2 Comparison of seven sequences in the polymorphic AT-rich region at the 3′ end of the human apo B gene. This figure compares six previously published sequences (alleles 2–7) and the plasmid that we sequenced (allele 1) in a schematic mode devised by Huang and Breslow (1987). The core 15-base repeat HVEs designated x (ATAATTAAATATTTT) and y (ATAATTAAAATATTT) differ only at the bases in boldface. The sequences of several modified HVEs (x′, x″, y′, y″, and y‴) are listed at the bottom of the figure. Allele 1 is the sequence in plasmid p12D1Eco described in the present paper, allele 2 is the sequence from Knott et al. (1986), allele 3 is from Jenner et al. (1988), and alleles 4–7 are those published by Huang and Breslow (1987). The nucleotide sequences at the 5′ and 3′ ends are the nonrepetitive regions between the PCR primers PCR3 (arrows, top) and PCR2 (arrows, bottom). The bracket over the 3′ flanking repeat (FR) represents a deletion reported by Jenner et al., and the asterisk (*) over the C in the most 3′ FR is reported by Jenner et al. (1988) to be a T. With these two exceptions, the sequences outside the HVRs reported by the four groups are identical.

leles 1–7 are designated HVE 30, HVE 36, HVE 38, HVE 35, HVE 34, HVE 36, and HVE 46, respectively. However, although this terminology does not distinguish alleles 2 and 6, the data of figure 2 show that these two alleles differ as a result of sequence microheterogeneity.

Amplification of the HVR Region

To examine the length heterogeneity in the range of HVE 30 and HVE 46, we developed a high-resolution method based on localized DNA amplification by the PCR (Saiki et al. 1988). Two primers, PCR1 and PCR2, were used to amplify the HVR from genomic DNA (see fig. 1). After 25 cycles, a third, internal ^{32}P-labeled oligonucleotide, PCR3, was added for five more cycles. The ^{32}P-labeled product was then electrophoresed on a denaturing acrylamide gel. High resolution was obtained that could distinguish alleles differing only by 15 bases in length, i.e., by only one HVE. Figure 3 il-

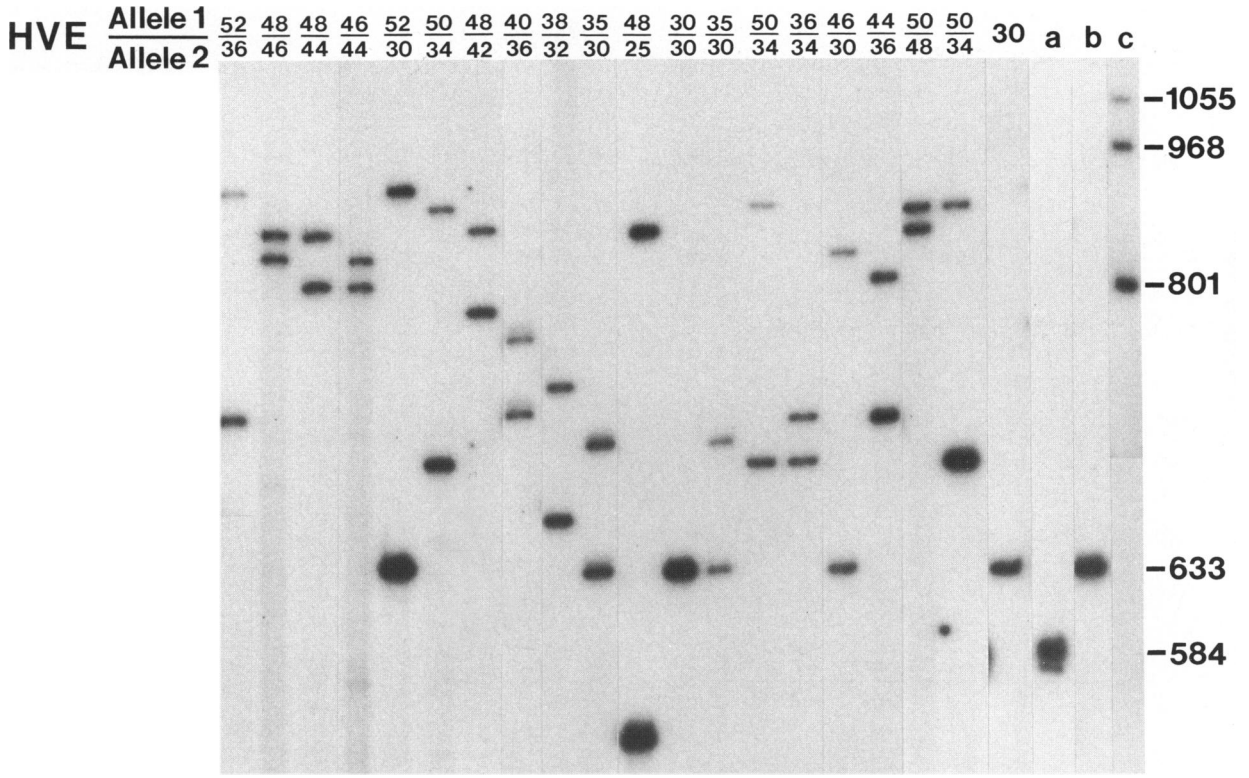

Figure 3 Autoradiograph of an acrylamide gel representing all 14 HVR alleles occurring within the group of subjects examined. The lanes on the left show the two HVR alleles, of 19 subjects, amplified from total DNA by using oligonucleotides PCR3 and PCR2. Numbers in the top scale indicate the number of repeats in each allele. The lane marked 30 is the plasmid 12D1Eco, which was sequenced and known to contain 30 HVEs. Lanes a–c are markers created by amplifying a segment of the apo B gene of known sequence (from 183 bases 5′ of exon 29 to base 13087 of the cDNA map) with only one ^{32}P-labeled primer. Lane a, Digestion with *Eco*RV to yield a fragment of 584 bases; lane b, digestion with *Pst*I to yield a fragment of 633 bases; lane c, partial digestion with *Sau*3A to yield fragments of 801 and 968 bases, together with the undigested 1,055-base fragment. The number of repeats represented in each allele was computed from the plot of migration vs. apparent size in fig. 4. The number of hypervariable repeats is equal to the size of the band in bases minus the sum of 5′ and 3′ nonrepetitive flanking sequences divided by 15, the repeat length; e.g., HVE 52 is 970 − 188, divided by 15 = 52.1.

lustrates some representative data; each lane shows the HVR amplified from genomic DNA of a different individual. Fourteen different-size fragments are apparent, and their sizes can be computed by comparison of migration distance relative to markers in lanes at the right. These estimates were made by plotting the number of bases against the distance migrated (fig. 4). Knowing the total length and the length of 5′- and 3′-flanking sequences both between PCR3 and the first HVE and between PCR2 and the first HVE, we calculated the number of HVEs in the repeat. The genotype of each individual computed in this manner is recorded above each lane in figure 3.

Overamplification as a Means of Measuring HVR Length

For routine purposes, we used 25 cycles of amplification, followed by five more after the addition of ^{32}P-labeled PCR3. However, in the process of optimizing the conditions, we discovered that, with a higher degree of amplification, a very diverse population of amplified DNA fragments was generated. When these were electrophoresed on a sequencing gel, a ladder was produced. Examples are shown in figure 5; panel *A* shows three overamplified samples. Bands occur at each 1-base (b) interval, as in a sequencing gel; those at 15- and 30-b intervals are more intense. Higher up on the gel, the single-base ladder is no longer evident, but 15- and 30-b periodicities persist. A longer run of one of these same samples is shown in panel *B*, together with normally amplified markers representing a variety of alleles. The scale on the left of figure 5*B* shows the number of HVE 15-b repeats read from the sequencing

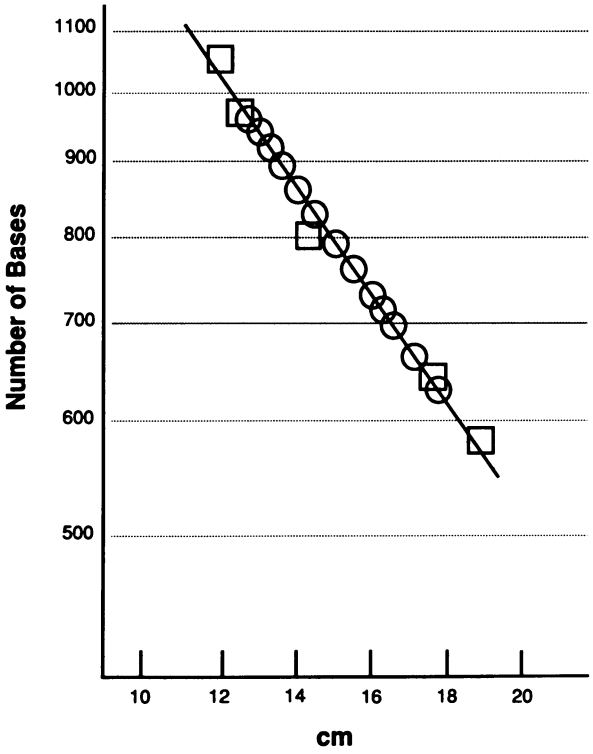

Figure 4 Relationship between the size of HVR alleles in bases and migration in an acrylamide gel. The migrations of the various fragments and markers seen in fig. 3 are plotted on a semilogarithmic scale vs. distance migrated. The molecular-weight standards are plotted as squares, and the individual HVR alleles are circles. All HVR variants are shown, except the smallest, which migrated beyond the smallest marker. The line drawn is a best fit to the marker points.

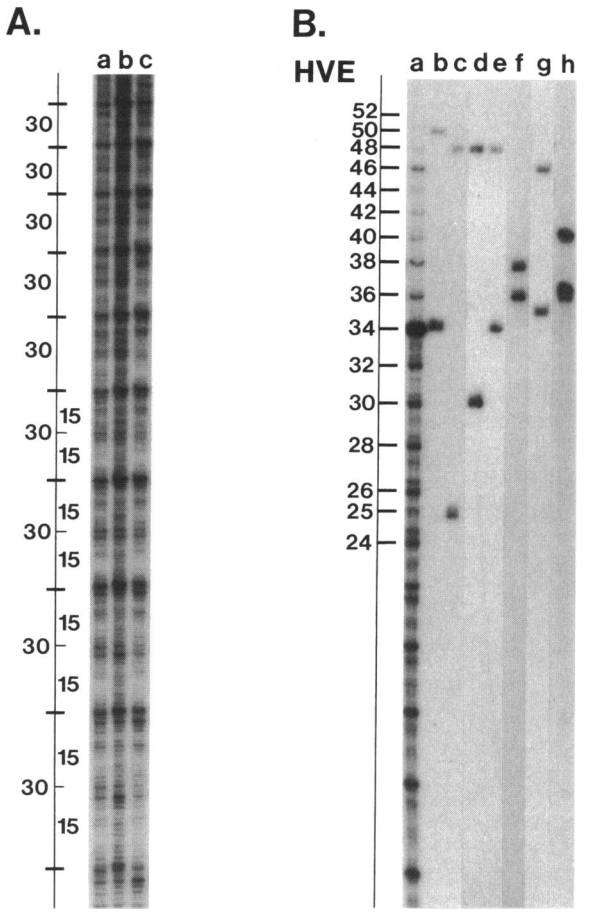

Figure 5 Sequence ladders obtained by overamplification of HVRs. *A*, Lanes a–c contain genomic DNA amplified for 35 cycles with primers PCR1 and PCR2 followed by 10 cycles using ^{32}P-labeled PCR3 and unlabeled PCR2. Note the single-base ladder, with every 15th and 30th base having a higher intensity. Higher on the gel, the single-base ladder is no longer evident, but the 15-base and 30-base periodicities remain. Also shown is a longer run of one of these same samples (panel *B*, lane a). Lanes b–h in panel *B* contain various samples with alleles of known size acting as markers. For example, the HVE 25 allele lies between the 24 and 26 repeat bands on the ladder.

ladder. Thus, the size of the various alleles may be deduced by comparison of migration with this sequence ladder. In this way, the number of repeats present in each of several alleles was verified. Excellent agreement was obtained between estimates of the numbers of repeats made from migration alone and those made on the basis of comparison with sequence ladders from overamplified DNA.

Distribution of HVR Alleles within the Population

To determine the frequency distribution of size alleles within the population, HVR regions were amplified from 318 subjects from an Austrian population previously studied in terms of apo B RFLPs (Paulweber et al., in press). The size of both alleles of each subject was measured by the methods described above, and the results are plotted in figure 6. HVEs containing 25–52 repeats were observed. Except for HVE 25 and HVE 35, the alleles contained even numbers of repeats. HVE 36 was the most prevalent allele, a result in agreement with the fact that two of seven sequenced alleles (fig. 2) have this repeat number. The index of heterozygosity in this group of subjects was 78%. Mendelian inheritance of HVR alleles was observed in two families, in agreement with earlier studies by Huang and Breslow (1987) and Jenner et al. (1988).

Discussion

Tandemly repeated minisatellite sequences are scattered throughout the human genome and also occur closely associated with specific genes such as the hemoglobin cluster (Jarman and Higgs 1988). The high degree of allelic variability makes these informative genetic markers useful for pedigree and linkage analysis. For such applications, it is important to develop methods of analysis that can distinguish the maximum number of alleles. The HVR of the human apolipoprotein B gene is an example of a minisatellite region associated with a gene of clinical interest. This region consists of approximately 25–52 tandem repeats of two related 15-bp sequences and minor variations of them. The hyperallelism of this region has been exploited as a marker for family and population studies; seven of these different alleles have been characterized by sequence analysis (Knott et al. 1986; Huang and Breslow 1987; Jenner et al. 1988; also see fig. 2).

The use of PCR amplification–radiolabeled primers has made it possible to resolve, by means of electrophoresis on denaturing acrylamide gels, alleles that differ by only a single 15-bp repeat. In a study of 318 unrelated individuals, 14 different alleles were distinguishable. This is obviously a minimal estimate of the heterogeneity in the population, because only half of the possible number of alleles varying in repeat number by 25–52 were actually documented. Nevertheless, even with 14 alleles, the average heterozygosity within this group amounted to 78%, a number that compares favorably with that for several other HVRs, which were surveyed by Nakamura et al. (1987).

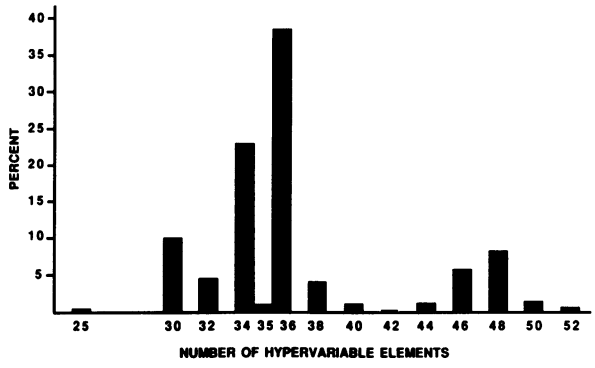

Figure 6 Frequency distribution of various HVR size alleles in a population. The figure illustrates the distribution of alleles of different lengths in a population of 318 subjects.

The method for amplifying the HVR employs a standard PCR protocol, with one important proviso. Although sharp bands are obtained with limited numbers of amplification cycles, this clarity is obscured if excessive amplification is undertaken (fig. 5). With overamplification—e.g., 35 cycles—the full-length product is still observed, but it is accompanied by a sequence ladder of prematurely terminated products. Although we have not examined the mechanism in detail, we attribute this effect to annealing, during chain extension, brought about by the high concentration of product. Because the HVR is a tandem repeat, this annealing may be out of register and may lead to chain termination predominantly at the end of successive 15-bp repeats. The occurrence of a sequence ladder in which strong bands at 15- and 30-bp intervals are superimposed on a pattern of terminated products representing single-base intervals is consistent with this view. Jeffreys et al. (1988*b*) have also stressed the importance of using limited numbers of cycles when amplifying minisatellite DNA. It should be emphasized, however, that this effect of excessive numbers of cycles is not a limitation of the approach; because radiolabeled primers are used, only a small number of cycles are necessary to obtain sufficient product for analysis. In fact, the generation of sequence ladders by overamplification can be used to some advantage. Thus, it is possible, by analysis of the sequence ladder, to obtain an accurate measure of the number of repeated elements constituting a given HVR allele (fig. 5).

While the final version of the manuscript of this paper was being prepared, an article by Boerwinkle et al. (1989) appeared that discussed the hyperallelism of the apo B HVR. These authors also employed the PCR technique but used nondenaturing agarose gels to resolve length differences among the amplified products. They were able to distinguish 12 alleles in a population of French origin. One size allele was absent in the population studied by us; therefore, the combined number of different alleles resolved in the two studies is 15.

The frequency distribution of size alleles reported by Boerwinkle et al. (1989) was remarkably similar to that shown in figure 6. One difference, in terminology, between the two studies should be pointed out. Although most of the HVEs are 15 bp in length and differ at only one or two bases (fig. 2), other sequences exist in the HVR that are more divergent. Thus, the actual number of HVEs assigned to a given HVR depends on the criterion used to define repeat elements. For example, an 11-bp sequence at the 3′ boundary of the HVR

was counted as a repeat by Boerwinkle et al. (1989) but not by us. Accordingly, they designated the most frequent allele 3′β37, whereas we have referred to exactly the same allele as HVE 36. Nomenclature applied to all other size alleles distinguished in the two studies also differs by one unit. A further difference between the two studies is the ability to resolve alleles differing, in our denaturing gel system, by one 15-bp repeat. The alleles reported by Boerwinkle et al. differed only by an even number of 15-bp repeats.

Finally, it is probable that this same approach will be applicable to analysis of other minisatellite sequences and to simultaneous amplification of sets of minisatellites, as described by Jeffreys et al. (1988*b*). If this proves to be the case, increased resolution would become available for a variety of genetic purposes.

Acknowledgments

This research was supported by National Institutes of Health grant HL 38781. The authors would like to thank Al Averbach and Sally Gullatt Seehafer for editorial assistance, Kate Sholly and Tony Gridley for typing the manuscript, and Charles Benedict and Tom Rolain for graphics.

References

Ali S, Wallace RB (1988) Intrinsic polymorphism of variable number tandem repeat loci in the human genome. Nucleic Acids Res 16:8487–8495

Bell GI, Selby MJ, Rutter WJ (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature 295:31–35

Berg K (1986) DNA polymorphism at the apolipoprotein B locus is associated with lipoprotein level. Clin Genet 30:515–520

Blackhart BD, Ludwig EH, Pierotti VR, Caiati L, Onasch MA, Wallis SC, Powell L, et al (1986) Structure of the human apolipoprotein B gene. J Biol Chem 261:15364–15367

Boerwinkle E, Xiong W, Fourest E, Chan L (1989) Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3′ hypervariable region. Proc Natl Acad Sci USA 86: 212–216

Hegele RA, Huang L-S, Herbert PN, Blum CB, Buring JE, Hennekens CH, Breslow JL (1986) Apolipoprotein B-gene DNA polymorphisms associated with myocardial infarction. N Engl J Med 315:1509–1515

Huang L-S, Breslow JL (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

Jarman AP, Higgs DR (1988) A new hypervariable marker for the human α-globin gene cluster. Am J Hum Genet 43:249–256

Jarman AP, Nicholls RD, Weatherall DJ, Clegg JB, Higgs DR (1986) Molecular characterization of a hypervariable region downstream of the human α-globin gene cluster. EMBO J 5:1857–1863

Jeffreys AJ, Royle NJ, Wilson V, Wong Z (1988*a*) Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA. Nature 332:278–281

Jeffreys AJ, Wilson V, Neuman R, Keyte J (1988*b*) Amplification of human mini-satellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Res 16:10953–10971

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Jenner K, Sidoli A, Ball M, Rodriguez JR, Pagani F, Giudici G, Vergani C, et al (1988) Characterization of genetic markers in the 3′ end of the apo B gene and their use in family and population studies. Atherosclerosis 69:39–49

Knott TJ, Wallis SC, Pease RJ, Powell LM, Scott J (1986) A hypervariable region 3′ to the human apolipoprotein B gene. Nucleic Acids Res 14:9215–9216

Kraft R, Tardiff J, Krauter KS, Leinward LA (1988) Using mini-prep plasmid DNA for sequencing double-stranded templates with Sequenase. Biotechniques 6:544–547

Ludwig EH, Blackhart BD, Pierotti VR, Caiati L, Fortier C, Knott T, Scott J, et al (1987) DNA sequence of the human apolipoprotein B gene. DNA 6:363–372

Lusis AJ (1988) Genetic factors affecting blood lipoproteins: the candidate gene approach. J Lipid Res 29:397–429

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, et al (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616–1622

Paulweber B, Friedl W, Krempler F, Humphries SE, Sandhofer F. Arteriosclerosis (in press)

Rapacz J, Hasler-Rapacz J, Taylor KM, Checovich WJ, Attie AD (1986) Lipoprotein mutations in pigs are associated with elevated plasma cholesterol and atherosclerosis. Science 234:1573–1577

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Soria LF, Ludwig EH, Clarke HRG, Vega GL, Grundy SM, McCarthy BJ (1989) Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. Proc Natl Acad Sci USA 86:587–591

Stoker NG, Cheah KSE, Griffin JR, Pope FM, Solomon E (1985) A highly polymorphic region 3′ to the human type II collagen gene. Nucleic Acids Res 13:4613–4622

Weisgraber KH, Innerarity TL, Newhouse YM, Young SG, Arnold KS, Krauss RM, Vega GL, et al (1988) Familial defective apolipoprotein B-100: enhanced binding of monoclonal antibody MB47 to abnormal low density lipoproteins. Proc Natl Acad Sci USA 85:9758–9762