# A High Frequency of Length Polymorphisms in Repeated Sequences Adjacent to Alu Sequences

Giovanni Zuliani and Helen H. Hobbs

Departments of Molecular Genetics and Internal Medicine, University of Texas Southwestern Medical Center, Dallas

## Summary

We describe a new class of DNA length polymorphism that is due to a variation in the number of tandem repeats associated with Alu sequences (Alu sequence–related polymorphisms). The polymerase chain reaction was used to selectively amplify a $(TTA)_n$ repeat identified in the 3-hydroxy-3-methylglutaryl coenzyme A (HMG CoA) reductase gene from genomic DNA of 41 human subjects, and the size of the amplified products was determined by gel electrophoresis. Seven alleles were found that differed in size by integrals of three nucleotides. The allele frequencies ranged from 1.5% to 52%, and the overall heterozygosity index was 62%. The polymorphic TTA repeat was located adjacent to a repetitive sequence of the Alu family. A homology search of human genomic DNA sequences for the trinucleotide TTA (at least five members in length) revealed tandem repeats in six other genes. Three of the six $(TTA)_n$ repeats were located adjacent to Alu sequences, and two of the three (in the genes for β-tubulin and interleukin-1α) were found to be polymorphic in length. Tandemly repetitive sequences found in association with Alu sequences may be frequent sites of length polymorphism that can be used as genetic markers for gene mapping or linkage analysis.

## Introduction

The human genome contains many interspersed repetitive DNA sequences that are polymorphic in length because of variability in the number of repeat units in a tandemly repeated sequence. One class of repetitive sequences that is highly polymorphic is the hypervariable region or variable number of tandem repeats (VNTRs). These sequences differ in the number of a tandemly repeated sequence that ranges in length from nine to 40 nucleotides (Jeffreys et al. 1985; Nakamura et al. 1987). The different VNTR alleles can be distinguished by Southern blot analysis or the polymerase chain reaction (PCR) and can be used as genetic markers in linkage analysis or in gene mapping (Jeffreys et al. 1985; Donis-Keller et al. 1987; Nakamura et al. 1987) Owing to their very high heterozygosity index (approaching 90%; Wong et al. 1987), VNTRs can be

very informative genetic markers (Jeffreys et al. 1985). Their usefulness, however, is limited by the fact that they are not randomly distributed in the human genome but instead are preferentially located in the telomeric regions of chromosomes (Royle et al. 1987; Nakamura et al. 1988).

Microsatellites are another class of tandemly repeated sequences that are more evenly distributed in the human genome. Each microsatellite consists of a variable number of the dinucleotide repeat $(dC-dA)_n$, where n ranges from 10 to 60. There are an estimated 50,000–100,000 copies of these sequences in the human genome (Miesfeld et al. 1981; Hamada et al. 1984; Litt and Luty 1989). Oligonucleotides homologous to unique sequences flanking a microsatellite can be used to selectively amplify the repeated sequence by PCR. The alleles (which can differ in length by as few as 2 bp) can be differentiated by fractionating the amplified DNA product on a high-resolution gel (Litt and Luty 1989). The heterozygosity indices of microsatellites range between 34% and 74%, and thus they are somewhat less informative than VNTRs (Weber and May 1988).

In the present paper we describe a new class of tan-

demly repetitive sequences which are located adjacent to Alu sequences (Schmid and Jelinek 1982) and are highly polymorphic in length. We have called these length polymorphisms "Alu sequence–related polymorphisms."

## Material and Methods

### Material

Poly (dA-dC) · (poly (dT-dG) and bacteriophage M13 replicative form were obtained from Pharmacia LKB Biotechnology (Piscataway, NJ). Genomic DNA was extracted and purified from blood leukocytes by using a model 340A DNA extractor (Applied Biosystems, Foster City, CA). The thermal cycler used for PCR was obtained from Perkin Elmer Cetus (Norwalk, CT). [$\gamma$-$^{32}$P] ATP ($\approx$ 7,000 Ci/mmol) and Biotrans nylon membranes were purchased from ICN Radiochemicals (Costa Mesa, CA). [$\alpha$-$^{32}$P]dATP ($\approx$3,000 Ci/mmol) was obtained from Dupont New England Nuclear (Boston). Geneclean™ was obtained from BIO 101 (La Jolla, CA). A model 380A DNA synthesizer was used to synthesize the oligonucleotides used for PCR, and a model 370A DNA sequencer was used for sequencing; both machines were purchased from Applied Biosystems (Foster City, CA). Homology searches were performed using Microgenie™ (updated 3/89) from Beckman (Palo Alto, CA). DNA probes were radiolabeled by hexamer priming using a random-primer DNA-labeling kit from Boehringer-Mannheim Biochemicals (Indianapolis). A Charon 4A λ bacteriophage (λHRed-1) containing a 14-kb insert from the human 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase) gene (Luskey and Stevens 1985) was obtained from Dr. Ken Luskey, University of Texas Southwestern Medical Center, Dallas.

### Identification and Sequencing of (TTA)$_n$ Repeat in the HMG CoA Reductase Gene

Two micrograms of λ HRed-1 were digested with HindIII and EcoRI in the buffers suggested by the supplier. The DNA was size-fractionated on a 0.8% agarose gel in 1 × TBE (0.05 M Tris, 0.05 M boric acid, 1 mM EDTA, at pH 8), transferred to a nylon membrane, and then hybridized with radiolabeled poly (dA-dC) · poly (dT-dG). The hybridization and washing conditions were exactly as described elsewhere (Litt and Luty 1989). A single 1.5-kb fragment that hybridized strongly to the probe was isolated by extraction from a low-melting-temperature gel by using Geneclean™. The fragment was subcloned into bacteriophage M13 and sequenced with

an Applied Biosystems DNA Sequencer using the universal M-13 primer.

### PCR-directed Amplification of (TTA)$_n$ Repeat

Two 20-base oligonucleotides, end-labeled GZ-1 and unlabeled GZ-2, complimentary to sequences flanking the (TTA)$_n$ repeat in the HMG-CoA reductase gene (fig. 2A), were used to selectively amplify the repeated sequence from genomic DNA by PCR according to a method described elsewhere (Saiki et al. 1988), with the following modifications; (1) the annealing and elongation reactions were performed at 60°C for 2 min and (2) the denaturation was performed at 95°C for 1 min. One-tenth of the amplification product was size-fractionated by electrophoresis on an 8% denaturing polyacrylamide gel in 2 × TBE for 4 h at 1,200 V and 50 mA. Radiolabeled MspI-digested pBR22 DNA was used as a size standard. The gel was exposed to Kodak XAR-5 film for 30 min at −20°C.

A homology search of human DNA sequences was performed to identify TTA (or TAA) repeats of at least five copies in length; (TTA)$_n$ repeats were found in association with the following genes: enkephalin (Comb et al. 1983), factor IX (Yoshitake et al. 1985), fibrinogen (Kant et al. 1983), α-globin (Hess et al. 1983) β-tubulin (Lee et al. 1984), and interleukin-1α (Furutani et al. 1986). Oligonucleotides corresponding to sequences flanking these repeats were synthesized and used in PCR-directed amplification of genomic DNA from six unrelated individuals as described above.

## Results

In an effort to find a polymorphic microsatellite in the HMG-CoA reductase gene, we sequenced a fragment that hybridized on Southern blot with a poly (dA-dC)·(dG-dT) probe. We did not identify a (dA-dC) repeat but did find 10 tandem copies of the triplet TTA at the end of an Alu sequence located 10 kb 3' of exon 2 (fig. 1). Two 20-base oligonucleotides complementary to sequences flanking the repeat (GZ-1 and GZ-2 in fig. 2A) were used to amplify, by PCR, the intervening sequences from genomic DNA of eight unrelated individuals. The amplification products were size-fractionated on a denaturing polyacrylamide gel, and the gel was subjected to autoradiography (fig. 3A). In each lane we observed one or two intense bands and additional minor bands. All together a total of seven different-sized major bands were found. The number of TTA repeats in the smallest and largest fragment was determined by direct sequencing of an end-labeled am-
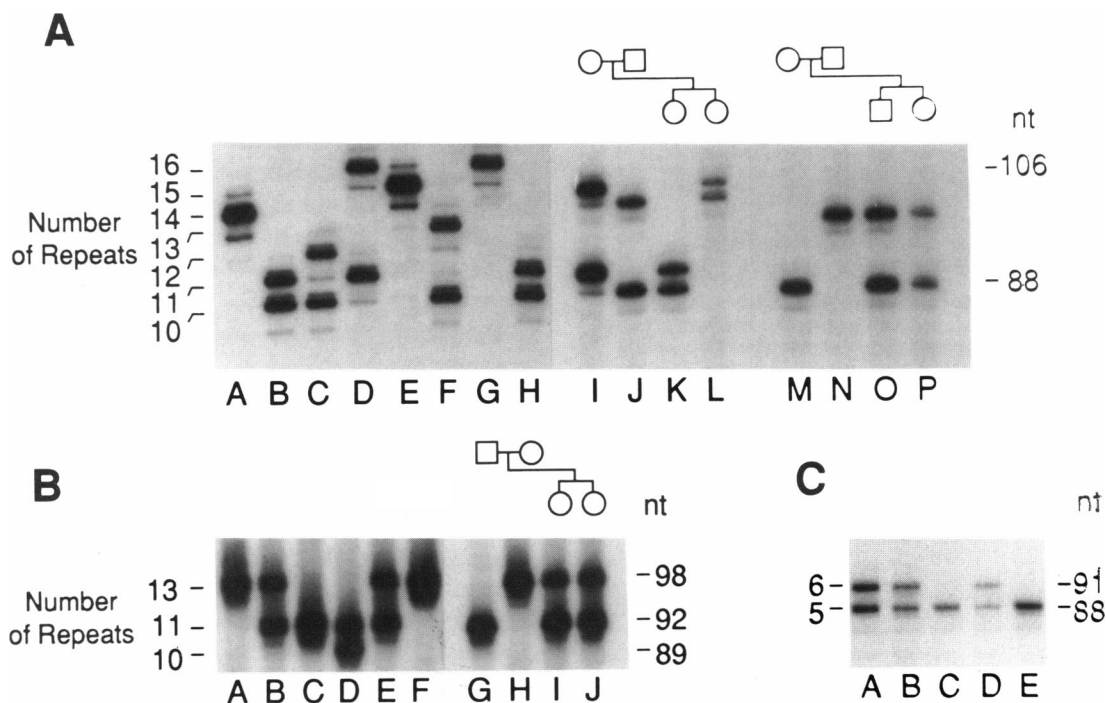
## A  HMG-CoA-REDUCTASE

CATGACCATATGAATCTACTCATGCTTTCATGCAAATAATCATGTTCCATCCATGTCTGCTTAATATGGTTCTTTCTTTTAAATATATGTTAATACGTTT
　　　　　　　GZ-1 ✱├──────────────→
ATTGGTAAAATGCAATTTTTTGCCAGCTTTATTGAGGTATACTTGACAAAATTTTATT[TTATTATTATTATTATTATTATTATTA]TGTTTTGGAGA
──────────┤GZ-2　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　↑282
CAGAGTGTCACTCTGTCACCCAGGCTGGAGTACAGTGGCACCATCTTGGCTCACAGCAACCTCCACCTCCCAGGTTCAAGCGATTCTTCTCTATCAGCCT
CCCGAGTAGCTGGGATTACAGGCCATCCGCCAACCCATGTCTGGGCTAATTTTTTGTGTTTTTAGCAGAGATGGGATTTCACTATGTTGGCCAGGCTGGT
CTCAAACTCCTGGCCTCAAGTGGTCTGCCTGCCAAGGCCTCCCAAAGTGCTGGGATTACAGGTATGAGCCACTCGATGCCGGCCTGCATTCTTCAATCAT
ATGCTTTTAGCTAACTTAGTCAG

## B  β-TUBULIN

AGGCAGGGGGCGGGGGAGTGTTTTGTGTGTATGTGTGTTTTTGAGATGGAGTCTCACTCCGTCACCCAGGCTGGAATGCAGTGGCTCAATCTTGGCTCCC
TGCAGCCTCCTCCTCCCAGGTTCAAGCAATTCTCTTGCCTCAGCCTCCCGAGTAGCCAGGGACTACGATCGCTCACCAGCACACTGGCTATAATTGTTAT
　　　　　　　　　　　　　　　　　　　　　　　　　　　GZ-3✱├───────────────→
TA[TTATTATTATTATTATTATTATTATTA]TTTGAGACGGAGCTCGCTCTGTTGCCCAGGCTGCAGTGCAGTGGCACAATGTCGGCTCACTGCCAA
　　　　　　　　　　　　　　　　　├──────────GZ-4
　　　　　　↑282
GCTTCCGCCTCCTGGGTTCACGCCATTCTCCTGCCTAGCCTCCCGAGTAGCTAGGACTACAGGCGCCCGCACCACTCCCGGCTAATTTTTTTTGTAATTT
TAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGTTGGTCTCGATCTCCTGACCTCGTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGTG
TGAGCCACCGCACCCGGCTATTTGTTATTATTTTTGATAGTTACTA
　　　　　　　↑1

## C  INTERLEUKIN-1α

ACTATCA[TTATTATTATTA]TTTATTTTTTTTTTTTTGAGATGGAGTCTGGCTCTGTCACCCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCACTGCAA
　　　　　　　　　　　　↑282
GCTCTGCCTCCTGGGTTCATGCCATTCTCCTGCCTCAGCCTCCCGAGTAAGCTGGGAATACAGGCACCCGCCACTGTTCCCGGCTCCTGGGTTCATGCC
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　GZ-5├─
ATTCTCCTGCCTAATTTTTTGTATTTTTAGTAGAGACGGAGTTTCACCGTGGTCTCCATCTCCTCGTGATCCACCCACCTTGGCCTCCCAAAGTGCCGG
──────────────────→　　　　　　　　　←────────────────　✱ GZ-6
GATTACAGGCGTGAGCCACCGCGCCCGGCCTA[TTATTATTATTATTA]CTACTACTACTACCTATATGAATACTACCAGCAATACTAATTTATTAATGAC
　　　　　　　　　　　　　　　　　↑1
TGGATTATGTCTAATCTCATTTTACATAAAAGGAAACTAAGCTCATTGAGATAGGTAAACTGCCCAATGGCATACATCTGTAAGTGGGAGAGCCTCAAA
TCTAATTCAGTTCTA

**Figure 1**    Partial sequences of the human HMG-CoA reductase, β-tubulin, and interleukin-1α genes. The HMG-CoA reductase (A), β-tubulin (B), and interleukin-1α (C) genes contain 10, 10, and 5 copies of the (TTA)n repeats, respectively. The (TTA)n repeated sequences are boxed. Alu sequences are underlined and numbered according to the consensus sequence. The direct repeats flanking the Alu sequences are double underlined. The arrows show the location of the oligonucleotides used to amplify the three blocks of (TTA)n repeats by PCR, and the asterisk denotes the end-labeled oligonucleotide.

plified fragment from individuals homozygous for the smallest (10 repeats) and largest (16 repeats) bands (Maxam and Gilbert 1980). The number of repeats in the other five bands were deduced by reference to the size standards and to the two sequenced alleles. The seven different alleles differed in size by integrals of three nucleotides, and the alleles segregated in two families in a codominant fashion (fig. 3A, lanes I–L and M–P). When the end-labeled fragments were sequenced, only a single sequence was visualized (data not shown), suggesting that the minor "shadow" bands were generated as artifacts of the PCR reaction. Similar PCR artifacts have been observed when a (dC·dA)n repeat was analyzed in an analogous fashion (Litt and Luty 1989).

The distribution of the seven alleles was determined in 41 unrelated Caucasian Americans (82 alleles), and the results are shown in figure 4A. The most frequent allele was the one with the fewest repeats, and this accounted for 52% of all alleles. The heterozygosity index was 62%; this is close to the calculated frequency of 66% that would be expected if all seven alleles were in Hardy-Weinberg equilibrium.

In order to determine whether the (TTA)n size polymorphism was present at other genetic loci, we searched the GenBank Depository (updated 3/89) for human genomic sequences that contained five or more copies of the TTA trimer or its complement, TAA. Tandem repeats of the trimer were identified in or near the genes for β-tubulin (10 repeats), enkephalin (five repeats), factor IX (five repeats), α-globin (five repeats), interleukin-1α (five repeats), and fibrinogen (five repeats). For each of these loci, oligonucleotides flanking the tandemly repeated trimer were used to amplify the sequence, and the number of repeats for six individuals was determined by fractionating the PCR-amplified product on a denaturing gel. For the enkephalin, factor IX, fibrinogen, and α-globin genes, no length polymorphisms were identified (data not shown). The TTA repeats in each of these four genes are composed of five copies of the sequence TTA. With the exception of the α-globin gene, none of these repeats was located next to an Alu sequence. The TTA repeats found in the β-tubulin and interleukin-1α genes were polymorphic in length among these six individuals. In both of these genes the TTA repeats were located within or adjacent to one of the direct repeats flanking an Alu sequence (figs. 2B and 2C). The results of PCR amplification of the β-tubulin repeat when oligonucleotides GZ-3 and GZ-4 were used are shown in figure 3B. On the basis of the published β-tubulin sequence, a single 89-bp amplification product was expected. However, bands that were three or six nucleotides longer than expected were generated from some individuals, owing to the presence of additional copies of the trinucleotide repeat. Codominant inheritance was demonstrated in one family (fig. 3B,

### H Red - 1

EXON 2

kb

1.5 kb

(TAA)n

**Figure 2**    Location of (TTA)n repeat in λHRed-1. λHRed-1 is a Charon 4A bacteriophage containing a 14-kb insert from the human HMG-CoA reductase gene. The location of exon 2 and restriction sites for EcoRI (E) and HindIII (H) are shown.

**Figure 3** Electrophoresis of PCR-amplified products containing the (TTA)$_n$ repeats associated with the human HMG-CoA reductase, β-tubulin, and interleukin-1α genes. Panel A, End-labeled oligonucleotide GZ-1 (5'-TTTATTGAGGTATACTTGAC-3') and oligonucleotide GZ-2 (5'-CAGAGTGACACTCTGTCTCC-3') were used to amplify the intervening sequences from 1 µg of genomic DNA of 41 Caucasian American individuals. The amplified products were fractionated on an 8% denaturing polyacrylamide gel by using MspI-digested pBR322 DNA as a size standard. The gel was subjected to autoradiography for 1 h at −20°C. The number of (TTA)$_n$ repeats was determined by sequencing PCR-amplified DNA from an individual homozygous for the shortest allele (with 10 repeats) and from individuals E and G, who had 15 and 16 repeats, respectively. The sizes of the other alleles were determined by reference to size standards and to the sizes of the alleles with the fewest and most (TTA)$_n$ repeats. Lanes I–L and M–P show the segregation of these alleles in two families whose pedigrees are indicated. Similar analyses were done for the (TTA)$_n$ repeats associated with the β-tubulin gene (panel B) by using end-labeled GZ-3 (5'-GATCGCTCACCAGCACACTGGCTAT-3') and GZ-4 (5'-CTGGGCAACAGAGCGAGCTCCGTCT-3') and for the interleukin-1α gene (panel C) by using GZ-5 (5'-GGGATTACAGGCGTGAGCCACCGCG-3') and end-labeled GZ-6 (5'-TTAGTATTGCTGGTAGTATTCATAT-3').

lanes G–J). The frequencies of the three different alleles were determined in 35 unrelated individuals (70 alleles), and a summary of the results is shown in figure 4B. The heterozygosity index for this polymorphism was 57%.

The TTA repeat associated with the interleukin-1α gene was also polymorphic in length, but only two alleles were identified, and these differed in size by one repeat (fig. 3C). The DNA from 20 individuals was analyzed. The smaller allele, which has five repeats, accounted for 77% of the sample. Of the individuals tested, 45% were heterozygous, although a frequency of 35% is predicted if the alleles are in Hardy-Weinberg equilibrium. The difference between the expected and observed frequencies may be due to the small sample size.

The polymorphic TTA repeat in the β-tubulin gene

is located at the 3' end of an Alu sequence, whereas the polymorphic repeat at the interleukin-1α locus was found at the 5' end. Both sequences were part of, or were directly adjacent to, one of the direct repeats flanking an Alu sequence (fig. 2). For each locus, the other direct repeat flanking the Alu sequence was analyzed in a similar fashion by PCR, and no length polymorphisms were found (data not shown). In the HMG-CoA reductase gene the TTA repeat is located adjacent to the 3' poly-A tail of the Alu sequence, but there are no direct repeats flanking this element.

## Conclusions

We describe a new class of DNA length polymorphisms that is attributable to a variation in the number of trimers of a repeated trinucleotide, TTA. The heter-
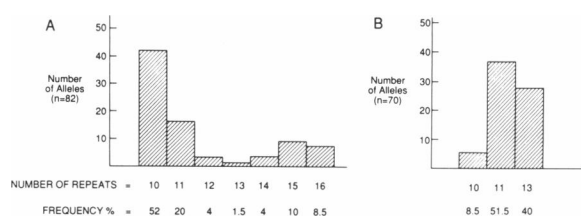
**Figure 4** Frequency distribution of alleles of the human HMG-CoA reductase and β-tubulin genes in a sample of unrelated individuals. Genomic DNA from 41 and 35 Caucasian Americans was used to selectively amplify the $(TTA)_n$ repeats of the HMG-CoA reductase gene ($A$) and the β-tubulin gene ($B$), as described in Methods. A total of seven and three different alleles were detected in the HMG-CoA reductase and β-tubulin genes, respectively.

ozygosity index of these repeated sequences ranges from 37% to 62%, a range which is similar to that seen with microsatellites (Weber and May 1988) and which thus makes these repeated sequences potentially useful as genetic markers for linkage studies. In this study, polymorphic TTA repeats were identified in the human HMG-CoA reductase, β-tubulin, and interleukin-1α genes.

The TTA repeat in the HMG-CoA reductase gene is located immediately 3' of an Alu sequence, and in the other two genes the repeats are part of, or are immediately adjacent to, one of the direct repeats flanking an Alu sequence. Direct repeats are found immediately 3' of the poly-A tail and at a variable distance 5' of most Alu repeats. The sequences of the direct repeats are not part of the Alu consensus sequence and are thought to be generated at the site of Alu integration into the genome. Alu sequences seem to preferentially insert into A-rich regions of the genome (Bains 1986), which may account for the A-T-rich composition of these polymorphic repeats. At both the β-tubulin and interleukin-1α loci, only one of the two TTA repeats flanking the Alu sequence was found to be polymorphic in length, suggesting that the size polymorphism was generated subsequent to the insertion of the Alu sequence.

In some cases the tandem repeat probably evolved by slipped-strand mispairing (Fresco and Alberts 1960; Streisinger et al. 1966) after an adenosine-to-thymidine transition in the poly-A tail of the Alu sequence. In vitro studies of slipped-strand mispairing of $(dC \cdot dA)_n$ repeats have shown that the degree of length polymorphism is directly related to the number of repeats and that deletions are more frequent than insertions (Levinson and Gutman 1987). These studies predict that the most frequent allele would be the one with the fewest

repeats needed to maintain the stability of the misaligned DNA intermediate. At the HMG-CoA reductase and interleukin-1α loci, the allele with the fewest number of TTA repeats was the most frequent, but at the β-tubulin locus the allele with an intermediate number of repeats predominated.

Four of the TTA repeats identified on the homology search were associated with Alu sequences, and three of these were found to be polymorphic in length, whereas no size polymorphism was identified in the repeats not associated with Alu sequences. This difference in the degree of the polymorphism cannot be attributed to differences in the number of repeat units, since the number of TTA repeats found in the factor IX, enkephalin, and fibrinogen genes (which were not associated with Alu sequences) was the same as the number of TTA repeats in the α-globin and interleukin-1α genes (which were associated with an Alu sequence). Yet, only the TTA repeat found in the interleukin-1α gene was polymorphic in length. Too few loci have been examined to determine whether there is a causal relationship between the frequency of length polymorphism and the physical proximity to Alu sequences, but these studies suggest such an association.

Sequences in close proximity to Alu repeats might be more susceptible to mispairing events if Alu sequences were transcriptionally active and/or preferentially involved in recombinational events. Alu sequences (which constitute 9% of the human genome) have been shown to be concentrated in the R (reverse) bands by in situ hybridization studies of human chromosomes (Kornberg and Rykowski 1988). These regions are very GC rich and have been found to be more transcriptionally active. There is good evidence in vitro, and conflicting evidence in vivo, that Alu sequences are transcribed by RNA polymerase III (Duncan et al. 1981; Elder et al. 1981; Fuhrman et al. 1981; Johnson and Jelinek 1986; Paulson and Schmid 1986). Both the close physical association of Alu sequences with transcribed genes and the possibility that some Alu sequences are transcriptionally active may make sequences adjacent to Alu repeats more prone to slipped-strand mispairing events.

Recombination seems to occur more frequently in the R bands, and this may be related to the concentration of Alu sequences within these regions. Alu sequences have been implicated as sites of recombinational events responsible for the duplication and evolution of new genes in the human genome (Barsh et al. 1983; Kudo and Fukada 1989). Alu sequences have also been implicated as hot spots for gene rearrangement in both the low-density-lipoprotein (LDL)

and β-globin genes (for review, see Lehrman et al. 1987), as well as in numerous other gene loci (Myerowitz and Hogikyan 1987; Nicholls et al. 1987; Rouyer et al. 1987; Markert et al. 1988; Vnencak-Jones et al. 1988; Huang et al. 1989). Sequences in close proximity to Alu repeats may be more susceptible to slipped-strand mispairing events, owing to a higher frequency of recombinational events associated with Alu sequences.

On the basis of these findings it would be expected that other tandemly repeated sequences adjacent to Alu repeats would be polymorphic in length. To test this hypothesis we examined two sets of tandem repeats adjacent to an Alu sequence: a $(TTTA)_n$ repeat located in the apolipoprotein B gene (Huang et al. 1989) and a $(TTTC)_n$ repeat in the third intron of the Apo C-III gene (Protter et al. 1984). Both repeats were found to be highly polymorphic in length, with heterozygosity indices of 66% and 83%, respectively (Zuliani and Hobbs, in press-a, in press-b). Also, Economou et al. (1989) have reported a length polymorphism in the β-globin gene cluster. Therefore, the association found in the present study between length variation in a repeated sequence and physical proximity to an Alu repeat is a more generalized finding.

*Note added in proof.*—Since this paper was submitted we have identified an additional Alu sequence-related polymorphism. There is a $(TTTA)_n$ repeat located at the 3' end of an Alu sequence in intron 6 of the lipoprotein lipase gene that is polymorphic in length. There are three different alleles, and the heterozygosity index is 54%.

## Acknowledgment

## References

Bains W (1986) The multiple origins of human Alu sequences. J Mol Evol 23:189–199

Barsh GS, Seeburg PH, Gelinas RE (1983) The human growth hormone gene family: structure and evolution of the chromosomal locus. Nucleic Acids Res 11:3939–3958

Comb M, Rosen H, Seeburg P, Adelman J, Herbert E (1983) Primary structure of the human proenkephalin gene. DNA 2:213

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, et al (1987) A genetic linkage map of the human genome. Cell 51:319–337

Duncan CH, Jagadeeswaran P, Wang RRC, Weissman SM (1981) Structural analysis of templates and RNA polymerase III transcripts of Alu family sequences interspersed among the human β-like globin genes. Gene 13:185–196

Economou EP, Bergen AW, Antonarakis SE (1989) Novel DNA polymorphic system: variable poly A tract 3' to Alu I repetitive elements. Am J Hum Genet 45 [Suppl]: A138

Elder JT, Pan J, Duncan CH, Weissman SM (1981) Transcriptional analysis of interspersed repetitive polymerase III transcription units in human DNA. Nucleic Acids Res 9: 1171–1189

Fresco JR, Alberts BM (1960) The association of noncomplimentary bases in helical polyribonucleotides and deoxyribonucleic acids. Proc Natl Acad Sci USA 46:311–321

Furhman SA, Deininger PL, LaPorte P, Friedmann T, Geiduschek EP (1981) Analysis of transcription of the human Alu family ubiquitous repeating element by eukaryotic RNA polymerase III. Nucleic Acids Res 9:6439–6457

Furutani Y, Notake M, Fukui T, Ohue M, Nomura H, Yamada M, Nakamura S (1986) Complete nucleotide sequence of the gene for human interleukin 1 alpha. Nucleic Acids Res 14:3167–3179

Hamada H, Petrino M, Kakunaga T (1984) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomics. Proc Natl Acad Sci USA 79:6465–6469

Hess JF, Fox M, Schmid C, Schen C-KJ (1983) Molecular evolution of the human adult α-globin-like gene region: insertion and deletion of Alu family repeats and non-Alu DNA sequences. Proc Natl Acad Sci USA 80:5970–5974

Huang L-S, Ripps ME, Korman SH, Deckelbaum RJ, Breslow JL (1989) Hypobetalipoproteinemia due to an apolipoprotein B gene exon 21 deletion derived by alu-alu recombination. J Biol Chem 264:11394–11400

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

Johnson EM, Jelinek WR (1986) Replication of a plasmid bearing a human Alu-family repeat in monkey COS-7 cells. Proc Natl Acad Sci USA 83:4660–4664

Kant JA, Lord ST, Crabtree GR (1983) Partial mRNA sequences for human Aα, Bβ, and fibrinogen chains: evolutionary and functional implications. Proc Natl Acad Sci USA 80:3953–3957

Kornberg JR, Rykowski MC (1988) Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell 53:391–400

Kudo S, Fukuda M (1989) Structural organization of glycophorin A and B genes: glycophorin B gene evolved by homologous recombination at Alu repeat sequences. Proc Natl Acad Sci USA 86:4619–4623

Lee MG-S, Loomis C, Cowan NJ (1984) Sequence of an expressed human β-tubulin gene containing ten Alu family members. Nucleic Acids Res 12:5823–5836

Lehrman MA, Goldstein JL, Russell DW, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48:827–835

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Luskey KL, Stevens B (1985) Human 3-hydroxy-3-methylglutanyl coenzyme A reductase: conserved domains responsible for catalytic activity and sterol regulated degrulation. J Biol Chem 280:10271–10277

Markert ML, Hutton JJ, Wiginton DA, States JC, Kaufman RE (1988) Adenosine deaminase (ADA) deficiency due to deletion of the ADA gene promoter and first exon by homologous recombination between two Alu elements. J Clin Invest 81:1323–1327

Maxam AM, Gilbert W (1980) Sequencing end labeled DNA with basic-specific chemical cleavages. Methods Enzymol 65:499–560

Miesfeld R, Krystal M, Arnheim N (1981) A member of a new repeated sequence family which is conserved throughout eukaryotic evolution is found between the human γ and β globin genes. Nucleic Acids Res 9:5931–5947

Myerowitz R, Hogikyan ND (1987) A deletion involving Alu sequences in the β-hexosaminidase α-chain gene of French Canadians with Tay-Sachs Disease. J Biol Chem 262: 15396–15399

Nakamura Y, Lathrop M, O'Connell P, Leppert M, Barker D, Wright E, Skolnick M, et al (1988) A mapped set of DNA markers for human chromosome 17. Genomics 2:302–309

Nakamura Y, Leppert M, O'Connell P, Wolfe R, Holm T, Culver M, Martin C, et al (1987) Variable number of tandem repeats (VNTR) markers for human gene mapping. Nature 235:2616

Nicholls RD, Fischel-Ghodsian N, Higgs DR (1987) Recombination at the human α-globin gene cluster: sequence features and topological constraints. Cell 49:369–378

Paulson KE, Schmid CW (1986) Transcriptional inactivity of Alu repeats in HeLa cells. Nucleic Acids Res 14:6145–6158

Protter A, Levy-Wilson B, Miller J, Bencen G, White T, Seilhamer J (1984) Isolation and sequence analysis of the human apolipoprotein CIII gene and the intergenic region between the Apo AI and Apo CIII genes. DNA 3:449–456

Rouyer F, Simmler MC, Page DC, Weissenbach J (1987) A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. Cell 51:417–425

Royle NJ, Clarkson R, Wong Z, Jeffreys AJ (1987) Preferential localization of hypervariable minisatellites near human telomeres. Human Gene Mapping 9. Cytogenet Cell Genet 46:685

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 230:487–491

Schmid CW, Jelinek RW (1982) The Alu family of dispersed repetitive sequences. Science 216:1065–1070

Streisinger G, Orada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M (1966) Frameshift mutations and the genetic code. Cold Spring Harbor Symp Quant Biol 31: 77–84

Vnencak-Jones CL, Phillips JA III, Chen EY, Seeburg PH (1988) Molecular basis of human growth hormone gene deletions. Proc Natl Acad Sci USA 85:5615–5619

Weber JL, May PE (1988) An abundant new class of human DNA polymorphisms. Am J Hum Genet (Suppl) 43:A161

Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann Hum Genet 51:269–288

Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K (1985) Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). Biochemistry 24:3736–3750

Zuliani G, Hobbs HH. Tetranucleotide repeat polymorphism in the Apo C-III gene. Nucleic Acids Res (in press-a)

——— Tetranucleotide repeat polymorphism in the apolipoprotein B gene. Nucleic Acids Res (in press-b).