Mutation rate variation in the hypervariable VNTR g3 (D7S22) is associated with a flanking DNA sequence polymorphism near the repeat array.

Rune Andreassen and Bjørnar Olaisen
Institute of Forensic Medicine, University of Oslo.
0027 Oslo, Norway.

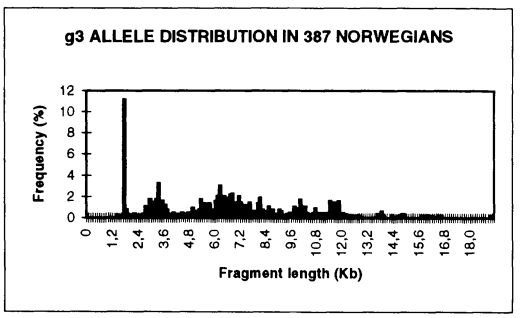
INTRODUCTION

The human minisatellite D7S22 (probe g3) is extremely polymorphic (heteroz. freq. 97%) with a paternal mutation rate estimated at 1.4% (Stenersen et al.1994).The size distribution of g3 alleles (fig. 1) reveals some size clustering of alleles. The smallest alleles (0-2 kb) show comparably reduced allelic diversity with one common small allele (14 repeats, freq 12.5%, Andreassen and Olaisen, 1994). No mutated 14 repeat alleles were found in a material with 2808 father-child observations, indicating that this allele has a reduced mutation rate.

A recent study of MS32 (locus D1S8, Moncton et al. 1994) has indicated that minisatellite stability is regulated by elements in the flanking DNA. A polymorphism in such a sequence could affect its abilities to act as mutagenic initiator. VNTR alleles with a suppression variant of such a flanking polymorphism would be more stable and be allowed to drift to relatively high population frequencies.

The reduced allelic diversity among small alleles in g3 could possibly reflect such a mechanism.To test this hypothesis, we have nucleotide sequenced the flanking DNA 3' to the tandem repeat array in 40 small family groups with de novo germline mutants as well as in 50 small alleles, in an attempt to reveal if there is any association between flanking polymorphism(s) and mutation rate.

Fig. 1.



MATERIAL AND METHODS

Sothern blots consisting of 6388 consecutive parent-child triplets were screened for mutations in the VNTR g3. 40 de novo germline mutants were identified using 0.5 mm. as the largest difference in band migration accepted for a band match. The progenitor band was chosen as the parental band most similar in size to the mutant allele (Wolff et al. 1989).The paternity (or maternity) was confirmed by typing in six other VNTR loci.

All mutations were found to be changes in the repeat array rather than a change in the HinfI restriction site when digested with PstI. Thus, a total of 101 individuals were further analysed, aiming at an unambigous haplotyping of all 40 progenitor and mutated alleles as well as the 80 accompanying ones.

12 homozygous and 10 heterozygous individuals with the 14 repeat allele as well as 16 samples with rare small alleles (1.5-2.0 Kb on HinfI blots) were also chosen for further studies.

PCR-amplification:

The primers A (5'TGTAAAACGACGGCCAGTGGAACAGACATTGCTGTAAG3') with a -21M13 tail and B (5'TCTGTGAGACGCTGCGTATC 3') with a biotinylated 5' end were choosen

for amplification of the flanking sequence 3' to the repeat array (236 bp).

If necessary, allele specific PCR was performed in individuals with small alleles (<2Kb) using primer C (5'AGGCTGCCTGCAGATTGCCT 3') located on the other side of the repeat array and primer B, followed by a semi-nested PCR reaction (primer A and B) to amplify the flanking sequence.

When haplotyping longer alleles, samples were HinfI digested and separated on a 0.7% agarose gel. The area containing the alleles was excised from the gel and the DNA was then recovered from the gel slice using the Geneclean II kit (BIO 101 Inc.) and then used as template in a PCR amplification using primer A and B as described.

DNA sequencing

An improved DNA sequencing method using streptavidin coated magnetic beads (Dynabeads M280, Dynal AS) was used to generate single stranded, purified and consentrated template (Leren et al. 1993). DNA sequencing was performed using Sequenase® version 2.0, -21M13 sequencing primers and sequenase dye primer sequencing kit (Applied Biosystems inc.) as described by the manufacturer. The DNA sequence analyses were performed using the ABI 373 sequencing system software version 1. 2. 0. (Applied Biosystem inc.) followed by visual examination of the chromatograms.

The haplotypes were determined by family typing. When both parental samples were heterozygous, an allele specific PCR was performed followed by a DNA sequencing of this haplotype.

RESULTS AND DISCUSSION

Two base substitution polymorphisms were found: a C/G transversion and a A/G transition 54 and 173 bp upstream of the repeat array, respectively. The frequencies of 54G or C alleles were 36% and 64%, and those of 173A or G alleles were 71% and 29%, respectively, when estimated from the accompanying alleles (nonmutated alleles in offspring and non progenitor alleles in progenitor). All combinations of haplotypes were found.

All the 50 selected small alleles were 54G and 173A, indicating a common origin of the 14 repeat alleles.

Using the Chi-square test with the accompanying alleles as "matched controls", there was a significant association (x^2= 9.6, p= 0.002) between mutation rate and the alleles at the 54C/G polymorphic site. Haplotypes with 54C tend to mutate more frequently than those with 54G. Most of the mutations in g3 were of paternal origin (85%) and they tend to be gain mutations. However, among the four mutated 54G alleles, two were of maternal origin and three represent losses. Variants of the 173A/G polymorphism did not reveal any significant association to mutation rate.

There was a higher mutation rate in the 4-9 Kb allele size cluster than those with smaller or larger size (fig. 2). Thus , the mutation rate is probably not directly related to allele size.

The 54G variant, associated with a relatively low mutation rate and present in all small alleles, could be responsible for the low diversity in this allele size group. The 173A variant, also found in most of the mutating alleles, could not. These findings migth indicate that the mutation rate variation is primarily associated with the 54C/G polymorphic site.

The 54C/G polymorphism is located in one of three basepairs that interrupts a sequence (ATGCACAC/GAATAA) that otherwise would have been a perfect match with the octamer motif (ATGCAAAT, Parslow et al. 1984). Previous studies have shown that there excist both degenerated variants (Baumruker T. et al. 1988) and variants extending the sequence by having basepairs that interrupt the motif (ATGCATAAAT, Castrillo et al.1989).

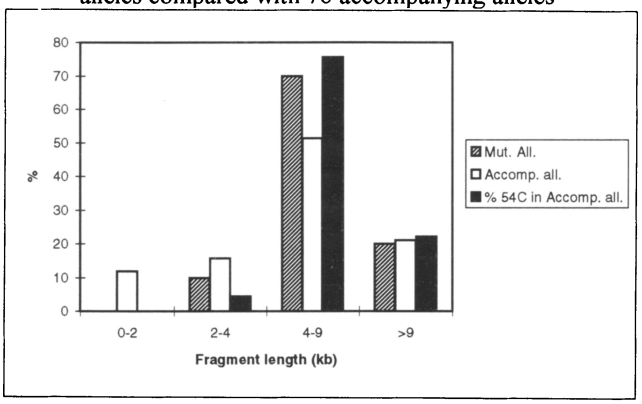
One variant of the octamer motif (ATGATAAT) located in the adenovirus 2 inverted terminal repeat is involved in adenovirus DNA replication (Pruijn et al. 1986). Another study (Iguchi-Ariga and Ogawa 1993) suggests that the octamer sequence (ATGCAAATNA) might

serve as part of a replication origin in mammalian cells. Furthermore, using the GCG computer software, a 100% fit with the octamer like sequence (ATGCACACAAT) was found within a 983 bp human sequence with autonomously replicating activity (Wu et al. 1993), suggesting that this sequence contains a replication origin.

One could speculate that the octamer like sequence found in this study could act as a regulatory element binding a factor essential for a possible origin of replication. The variation in stability might then be related to the ability of the two variants of the octamer like sequence to act as a component in an origin of DNA replication. The main mutational mechanism responsible for the repeat array instability could then be replication slippage during DNA replication. This would be consistent with an excess of small size changes (Levison et al. 1987), and could also account for the difference in mutation rate between the sexes found in this VNTR.

In the MS32 study the mutation rate associated polymorphism was found in a 16 bp palindrome sequence.The octamer motif like sequence is located almost in the same position relative to the repeat array, but it shows no sequence similarity with this palindrome sequence. Nevertheless, our findings support the hypothesis that there exist elements in the flanking DNA affecting the repeat array stability.

Fig.2. Fragment size distribution of 40 mutated g3 alleles compared with 76 accompanying alleles



REFERENCES

Andreassen R, Olaisen B (1994) Study of short alleles in the VNTR locus D7S22 (λg3). Paper presented at the twelwth nordic meeting of forensic medicine, Lillehammer, Norway August 10-13.

Baumruker T, Sturm R, Herr W (1988) OBP100 binds remarkably degenerate octamermotifs through spesific interaction with flanking sequences. Genes Dev. 2: 1400-1413.

Castrillo JL, Bodner M, Karin M (1989) Purification of growth hormone-spesific transcription factor GHF-1 containing homeobox. Science 243:814-817

Igushi-Ariga SM, Ogawa N (1993) Identification of the initiation region of DNA replication in the murine immunoglobulin heavy chain gene and possible function of the octamer motif as a putative DNA replication origin in mammalian cells. Biochimica et biophysica Acta 1172(1-2):73-81.

Leren TP, Rødingen OK, Røsby O, Solberg K, Berg K (1993) Screening for point mutations by semi-automated DNA sequencing using Sequenase and magnetic beads. Bio techniques 14:618-623.

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

Moncton DG, Neumann R, Guram T, Fretwell N, Tamaki K, Macleod A, Jeffreys AJ (1994) minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. Nature genet. 8: 162-170.

Parslow TG, Blair DL, Murphy WJ, Granner DK (1984) Strukture of the 5' ends of immunoglobulin genes: a novel conserved sequence. Proceed. of the Nat. Acad of Science, USA 81: 2650-2654.

Pruijn GJM, van Driel W, van der Vliet PC (1986) Nuclear factor III, a novel sequence- spesific DNA-binding protein from HeLa cells stimulating adenovirus DNA replication. Nature 322: 656-659.

Stenersen M, Hoff-Olsen P, Olaisen B (1994) Paternity investigations experiences using single locus probes. Paper presented at the twelwth nordic meeting of forensic medicine, Lillehammer, Norway August 10-13.

Wolff RK, Plaetke R, Jeffreys AJ, White R (1989) Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics 5: 382-384.

Wu C, Friedlander PA, Lamoreux C, Zannis-Hadjopoulus M, Price GB. cDNA clones contain autonomous replication activity. AC L08438 EM.Unpublished.