© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

Length of uninterrupted CGG repeats determines instability in the *FMR1* gene

Evan E. Eichler[1], Jeanette J.A. Holden[2], Bradley W. Popovich[3], Allan L. Reiss[4], Karen Snow[5], Stephen N. Thibodeau[5], C. Sue Richards[1], Patricia A.Ward[1] & David L. Nelson[1]

Analysis of 84 human X chromosomes for the presence of interrupting AGG trinucleotides within the CGG repeat tract of the *FMR1* gene revealed that most alleles possess two interspersed AGGs and that the longest tract of uninterrupted CGG repeats is usually found at the 3' end. Variation in the length of the repeat appears polar. Alleles containing between 34 and 55 repeats, with documented unstable transmissions, were shown to have lost one or both AGG interruptions. These comparisons define an instability threshold of 34–38 uninterrupted CGG repeats. Analysis of premutation alleles in Fragile X syndrome carriers reveals that 70% of these alleles contain a single AGG interruption. These data suggest that the loss of an AGG is an important mutational event in the generation of unstable alleles predisposed to the Fragile X syndrome.

[1]*Department of Molecular & Human Genetics, Human Genome Center, Baylor College of Medicine, Houston, Texas 77030, USA*
[2]*Department of Psychiatry, Queen's University, and Cytogenetics and DNA Research Laboratory, Ongwanada Resource Center, Kingston, Ontario K7L 3N6, Canada*
[3]*Department of Medical and Molecular Genetics, Oregon Health Sciences University, Portland, Oregon 97201-3098, USA*
[4]*Kennedy Krieger Institute and Departments of Psychiatry and Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA*
[5]*Laboratory Genetics, Mayo Clinic and Foundation, Rochester, Minnesota 55905, USA*

Correspondence should be addressed to D.L.N.

Trinucleotide instability has been found in several neurological disorders exhibiting anticipatory non-mendelian inheritance[1]. Fragile X syndrome was the first human genetic disease in which unstable trinucleotide repeats were described[2–5]. In the Fragile X syndrome, CGG triplet repeat expansion occurs in the 5' untranslated region of the first exon of the *FMR1* gene, leading to hypermethylation of the promoter region, extinction of gene expression and mental impairment[6–8]. An individual's degree of risk in a Fragile X lineage is dependent on his position in a pedigree with penetrance increasing in subsequent generations[9]. A molecular basis for the phenomenon was established with the discovery that the length of the CGG repeat in premutations correlated with a propensity for amplification and disease[6,10,11].

A difficulty with the Fragile X model for triplet repeat expansion, and one with considerable clinical significance, has been the inability to define precise thresholds for prediction of instability and risk of hyperexpansion. A "grey zone" has been defined in which there is considerable overlap in repeat length between alleles at the high-end of the range found in the general population, and the low-end of premutation alleles in Fragile X families[1,6,12]. Several groups have identified alleles of similar size which differ markedly in their intergenerational stability[12–14]. This observation indicates that repeat length is not the sole factor accounting for the instability of Fragile X alleles. Linkage disequilibrium of Fragile X chromosomes to specific *FRAXAC1–DXS548–FRAXAC2* haplotypes, especially in Northern European populations[14–18], suggests that the predisposing instability event is a cis-acting factor located near or within the repeat. The observation of AGG trinucleotides interrupting the *FMR1* CGG repeat[2–4] has led to the suggestion that these provide stability to the repeat and that instability might result from loss of AGG triplets[6,18,19]. Precedence for this model was established for spinocerebellar ataxia type 1 (*SCA1*), in which the loss of CAT interruptions within the CAG repeat was shown to have occurred in all disease alleles[20].

To test this model we have developed a simple indirect method to determine the presence or absence of internal AGG trinucleotides within the *FMR1* CGG repeat tract. This technique combines the robust ability of the *Pfu (Pyrococcus furiosus)* polymerase to amplify this region[21] and digestion with *MnlI* restriction endonuclease whose recognition sequence (GAGG) identifies internal AGG trinucleotides within a CGG repeat tract. This analysis suggests that alleles with 33 or fewer uninterrupted CGGs will be inherited with stability, while those with 39 or more CGGs will show marked instability characteristic of premutations.

AGG interspersion analysis

MnlI digests of *FMR1* CGG repeat-containing PCR products, amplified using primers A and 571R (ref. 21), generate a total of ten fragments (Fig. 1). Based on the arrangement of interspersed AGGs in cDNA clone bc72 $(CGG)_{10}AGG(CGG)_9AGG(CGG)_9$ (ref. 3) three of these fragments should contain CGG repeats. Two of these, the 54 bp fragment and a 61 bp fragment, can be visualized on a 3% Metaphor agarose gel while the third (a 30 bp fragment) comigrates with fragments of a similar size (see Fig. 2*b*, lane 30 repeats). In order to detect fragments of interest, digests were transferred and probed with a 30 mer $(CGG)_{10}$ end-labelled oligonucleotide. Hybridization revealed that the three CGG repeat-containing fragments

© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

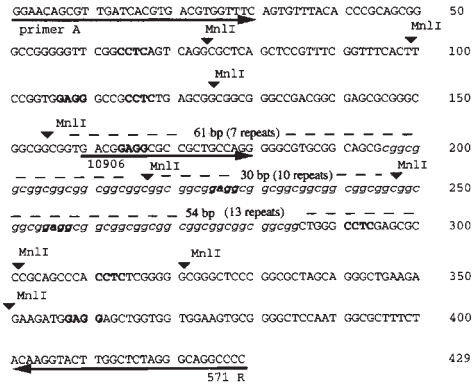

Fig. 1 Sequence of the primers A/571R CGG repeat-containing PCR product. The positions of the primers are indicated by arrows below the sequence. Primer 10906 was used as a hybridization probe to detect fragments containing the most-5' portion of the repeat. *Mnl*I recognition sequence (CCTC/GAGG) is depicted in bold and sites of *Mnl*I cleavage are indicated by arrowheads. The CGG repeat is presented in italics. The arrangement of interspersed AGG trinucleotides is based on *FMR1* cDNA clone bc72. The *Mnl*I restriction fragment sizes from 5' to 3' for this product are 74, 25, 26, 61,30, 54, 20, 29 and 79 bp.

hybridized with differential intensity. *Mnl*I cleaves seven basepairs 3' to its recognition sequence ($CCTCN_7$). The net effect is that the most 5' CGG repeat-containing fragment from this region is shortened by three CGG repeats, whereas the most 3' fragment acquires an additional three CGG repeats (Fig. 1) relative to the position of the AGG interspersion. This results in more efficient hybridization of CGG oligonucleotide probes to the 3'-most 54 bp fragment than to the 61 bp fragment (Fig. 2*b*, lane 30). Despite *Mnl*I over digestion and the use of various PCR protocols, partial digestion of PCR fragments was always observed. However, comparison of ethidium-stained gels with the autoradiograph clearly identified those fragments which were the result of partial digestion (Fig. 2*a*,*b*). To determine which fragments were changing with increasing CGG repeat number, the position of the 5'-most fragment was assessed by hybridization with a second oligonucleotide, 10906 (ref. 22). In nearly every case, the variation in allele length was due to changes in length of the 3'-most *Mnl*I fragment (data not shown).

Classification of *FMR1* CGG repeat alleles

We examined a total of 84 male *FMR1* CGG repeat alleles for the frequency and position of interspersed AGG trinucleotides within the *FMR1* CGG repeat tract (see Methodology for sample selection). The majority of alleles (52.3% or 44/84) had two interspersed AGG trinucleotides (Table 1 and lane 30, Fig. 2*b*). If we consider only alleles ranging in size from 27–45 repeats, the frequency of two AGG's increases to 85% (34/40 alleles) (Table 1). Alleles devoid of AGG's or having three AGG interspersions are extremely rare within this range, accounting for only 5% (2/40) of alleles (Table 1). For repeat sizes ranging from 18–26, 10/13 or 77% have a single AGG interspersion. All alleles less than 18 repeats were found to be completely devoid of AGG trinucleotides (Table 1).

Polarized variability within the CGG repeat

The general formula for the *FMR1* CGG repeat is $(CGG)_{8-15}AGG(CGG)_{9-13}AGG(CGG)_x$, where x equals the remaining number of repeats (see Table 1). If we consider the position of the AGG within the repeat, *FMR1* CGG repeat alleles may be placed into one of 16 classes. The most frequent class of alleles is 9+9+ (see Table 1) followed by 10+9+, representing 20% and 19% of the alleles respectively. Although the position of the first AGG interruption can vary from 8–15 starting with the first CGG repeat, in 81% of alleles which have one or more AGGs it punctuates at the tenth or eleventh position within the repeat. Similarly for alleles with two or more AGG interruptions, the distance between the first and second AGG is usually nine or ten CGG repeat units (40/45 or 88%). Since the positions of the first two AGGs are largely invariant, this would necessitate that for larger

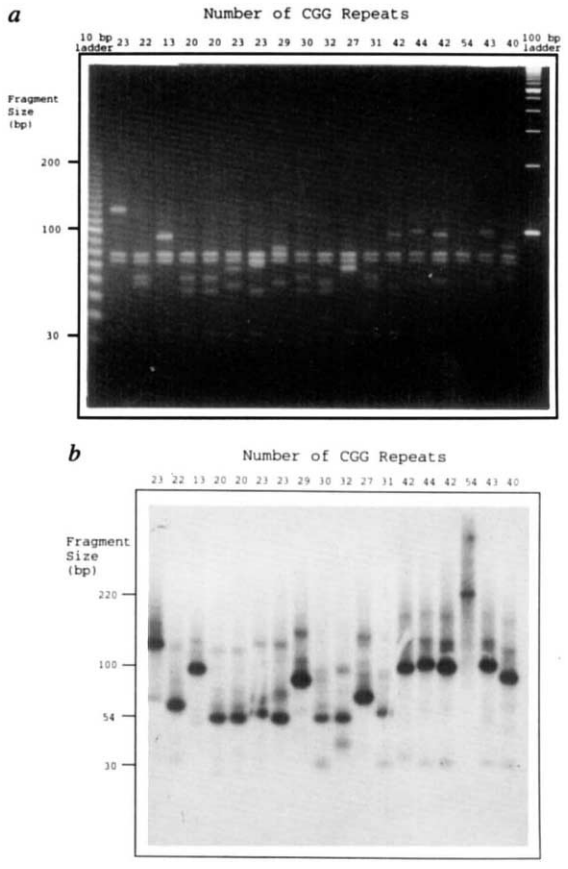

Fig. 2 AGG interspersion analysis. *a*, *Mnl*I digest of *FMR1* CGG repeat PCR products. PCR products were amplified using primers A and 571R (ref. 21), purified and digested with 4 U of *Mnl*I. Repeat length, shown above each lane, was determined using standard protocols[6] and corrected by addition of 1 repeat unit[41]. *b*, Hybridization of *Mnl*I Digests. *Mnl*I digests were transferred and probed with an end-labelled $(CGG)_{10}$ oligonucleotide. CGG repeat-containing fragments hybridize with differential intensity (see lane with repeat length of 30; fragment sizes 61, 54 and 30 bp) based on the number of CGG repeats in each fragment (Fig. 1). Note the presence of partial digest products above 100 bp. The sum of the CGG repeat-containing *Mnl*I fragments predicted a repeat number greater by 3 bp than had been measured by PCR estimates from denaturing polyacrylamide gels[6], confirming previous observations[1,41].

© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

Table 1 Classes of *FMR1* CGG repeat alleles

Allele class	Position of AGG interspersion within CGG Repeat (x=remaining number of repeats)	Repeat length <18	18–26	27–35	36–45	46–55	>55	Total number of alleles
0	$(CGG)_x$	4	2	1		2	1	10
9+	$(CGG)_9$**AGG**$(CGG)_x$			2	1	4	7	14
10+	$(CGG)_{10}$**AGG**$(CGG)_x$		5				1	6
11+	$(CGG)_{11}$**AGG**$(CGG)_x$					1		1
12+	$(CGG)_{12}$**AGG**$(CGG)_x$		2			1	1	4
13+	$(CGG)_{13}$**AGG**$(CGG)_x$		3					3
x-10+	$(CGG)_x$**AGG**$(CGG)_9$			1				1
8+10+	$(CGG)_8$**AGG**$(CGG)_{10}$**AGG**$(CGG)_x$				1		1	2
9+9+	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_x$			7	4	4	2	17
9+10+	$(CGG)_9$**AGG**$(CGG)_{10}$**AGG**$(CGG)_x$			2				2
9+11+	$(CGG)_9$**AGG**$(CGG)_{11}$**AGG**$(CGG)_x$			2				2
9+12+	$(CGG)_9$**AGG**$(CGG)_{12}$**AGG**$(CGG)_x$			1				1
10+9+	$(CGG)_{10}$**AGG**$(CGG)_9$**AGG**$(CGG)_x$			14	1	1		16
12+9+	$(CGG)_{12}$**AGG**$(CGG)_9$**AGG**$(CGG)_x$				1			1
12+13+	$(CGG)_{12}$**AGG**$(CGG)_{13}$**AGG**		1					1
13+9	$(CGG)_{13}$**AGG**$(CGG)_9$**AGG**$(CGG)_x$					1		1
15+9+	$(CGG)_{15}$**AGG**$(CGG)_9$**AGG**$(CGG)_x$				1			1
10+9+10+9	$(CGG)_{10}$**AGG**$(CGG)_9$**AGG**$(CGG)_{10}$**AGG**$(CGG)_9$				1			1
	Total number of alleles	4	13	30	10	14	13	84

Classification of CGG repeat alleles is based on the positions of AGG interspersions within the *FMR1* CGG repeat. The designation within each class is indicated in the far left-hand column, where a + sign represents an AGG, the number denotes the uninterrupted CGG repeats and x is the remaining number of repeats. The number of each allele class within the different size categories is indicated in the right-hand columns.

alleles the vast majority of repeat length variability has occurred at the 3' end of the repeat. For alleles greater than 36 repeats this is clearly the case (Fig. 2*b*; Table 1). Furthermore, we have demonstrated in pedigrees which have demonstrated instability (see below) that changes in repeat length occur in the CGG repeat tract 3' to the last interrupting AGG trinucleotide, suggesting polarity of repeat variation.

Unstable and stable alleles

When unstable alleles were compared to stable alleles of similar size by AGG interspersion analysis, a characteristic *Mnl*I restriction difference was observed (Fig. 3*a*,*b*). In every unstable allele examined, one or both of the smaller CGG-containing *Mnl*I fragments were absent, corresponding to the loss of interspersed AGG's and a concomitant increase in pure CGG repeat length consistent with the measured length of the repeat (Table 2). If the length of pure CGG repeat is compared among unstable and stable transmissions, a pattern emerges (Fig. 4). Alleles which possess between 34–37 pure CGG repeats have a relatively high probability of transmitting an allele unstably. Although the numbers are quite small ~30% or 3/9 of transmissions from alleles within this range were unstable. For alleles greater than 37 pure CGG repeats the probability of an unstable transmission approaches unity (Table 2), as it does for premutations, while no instability was identified in alleles with 33 or fewer uninterrupted repeats.

Premutation analysis

We have analysed a total of 13 premutation-sized alleles

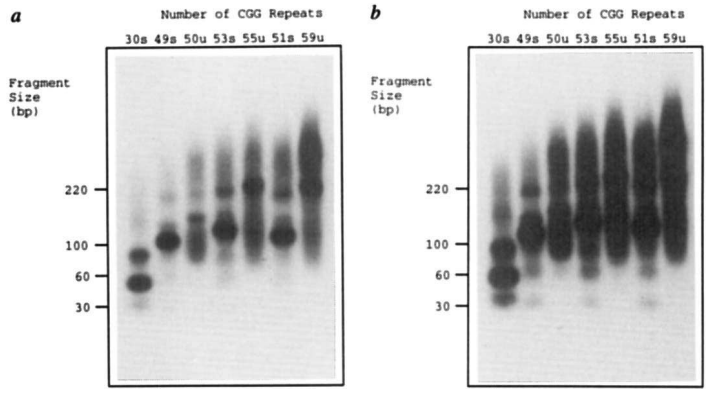

Fig. 3 *Mnl*I digests of stable and unstable alleles. *FMR1* PCR product was digested, transferred and hybridized as in Fig. 2*b*. The predicted number of CGG repeats is indicated above each lane. S, alleles where only stable transmissions have been observed. U, alleles in which unstable transmissions have been detected. (Note that the number of transmissions per allele is small.) *a*, A short exposure (4 h) allows visualization of larger *Mnl*1 fragments (comparison with ethidium bromide gels allows determination of partial digest products). *b*, An overnight exposure provides detection of smaller 61 bp and 30 bp fragments. Interpretation of the internal composition is presented in Tables 2 and 3 where 49s, 50u, 53s, 55u, 51s and 59u correspond to samples IX-1, XII-3, VII-1, 1408-08, 10822 and 7544, respectively. The smeared background of the longer exposure is due to a PCR artefact[21].

© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

Table 2 Unstable and stable alleles

Individual	Total number of repeats	CGG_{10} oligo hybridizing *Mnl*1 fragments (bp)	Predicted internal composition of repeat	Stable transmissions	Unstable transmissions	Longest tract of pure CGGs	Number of AGGs
IX-1	49	110, 58, 30	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_{29}$	2	0	29	2
8125	50	115, 58, 30	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_{30}$	2	0	30	2
10822	51	120, 58, 30	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_{31}$	5	0	31	2
XIX-1	53	125, 58, 30	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_{33}$	1	0	33	2
C-1	34	160	$(CGG)_{34}$	1	1	34	0
A-17	44	130, 58	$(CGG)_9$**AGG**$(CGG)_{34}$	2	1	34	1
10608	47	140, 58	$(CGG)_9$**AGG**$(CGG)_{37}$	2	0	37	1
XXI-1	47	140, 58	$(CGG)_9$**AGG**$(CGG)_{37}$	0	1	37	1
IV-1	48	140, 60	$(CGG)_{10}$**AGG**$(CGG)_{37}$	1	0	37	1
B-1	51	145, 65	$(CGG)_{11}$**AGG**$(CGG)_{39}$	0	2	39	1
XIII-3	50	150, 58	$(CGG)_9$**AGG**$(CGG)_{40}$	0	1	40	1
D-3	54	155, 68	$(CGG)_{12}$**AGG**$(CGG)_{41}$	0	1	41	1
1408-08	55	220	$(CGG)_{55}$	0	6	55	0

A comparison of the internal composition of the CGG repeat. Samples alleles representing each pedigree studied are arranged in ascending order of longest tract of pure CGG repeats. The number of stable and unstable transmissions observed for each allele in the pedigree is compared with the number of interrupting AGGs (right hand column). Stable alleles are defined as those alleles which have shown only stable transmissions, although the total number of transmissions observed is small.

(>55 repeats) for interrupting AGG trinucleotides. There is evidence for expansion to a full mutation in the majority of the pedigrees from which these alleles are derived (Table 3). Nine of 13 of these alleles contained a single AGG interruption located at the 5' end of the repeat (Table 3). A single premutation was identified which was completely devoid of AGGs. This last allele was the smallest premutation characterized (59 repeats) and expanded to a full mutation within two generations (Table 3, individual 7544). The remaining three alleles (23%) contained two interspersed AGGs, the smallest of which (individual 15,135, Table 3) demonstrated two stable transmissions. Analysis of three premutations from a single kindred (Table 3, individuals 75-787, 75-828 and 75-830) showed that all variation in repeat number occurred in the longest tract of pure CGG repeats located at the 3' end of the repeat. Finally, a consideration of the size of premutation alleles (and inferred pure CGG repeat number) which expanded to a full mutation after passage through a female germline, indicated that pure CGG repeat sizes ranging from 56–74 were sufficient to expand to a full mutation.

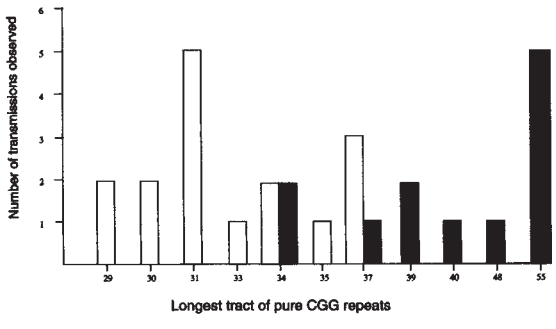

Fig. 4 Instability threshold. The longest tract of uninterrupted CGG repeats (in repeat number) is compared against the number of stable and unstable transmissions observed for each length. For unstable alleles (filled bars), the internal content of the parental allele which demonstrated instability upon transmission is considered. These data are derived from 13 pedigrees in which 12 unstable and 16 stable transmissions have been observed. White bars denote numbers of stable transmissions.

Discussion

Stability threshold of pure CGG repeats. Since the first detailed description of the repeat variation in normal and Fragile X chromosomes[6], stable alleles have been found with more than 50 repeats[12], while a number of unstable alleles have been identified with as few as 34 repeats, in both Fragile X and "normal" families (Table 2)[13]. This has led to the concept of a grey zone ranging, from 35–55 repeats, in which alleles can be either stable or unstable[23]. The presence of stable and unstable alleles of similar sizes[13] suggested that a feature other than length, but intrinsic to the allele itself was involved in stability. In this study, we have determined that the length of uninterrupted CGG triplet can be used to predict stability, with marked instability beginning when alleles exceed 33 pure CGGs. Alleles smaller than this also show instability[14] in families (and historically from the extensive levels of polymorphism). However, these events are expected to be at frequencies similar to other microsatellites.

It is significant that the length conferring instability in Fragile X alleles (34–37 pure CGGs) is in agreement with the lengths at the high end of the range found in the general population for other loci involved in triplet repeat expansion (37,34,36,36 and 34 triplets in myotonic dystrophy, Kennedy's disease, Huntington disease, spinocerebellar ataxia type 1 (*SCA1*) and dentatorubral pallidoluysian atrophy, respectively)[1,24]. With the exception of *SCA1*, an uninterrupted repeat is found at each of these loci. In *SCA1*, conversion from the normal interrupted repeat to a pure repeat is found in disease alleles[20]. This common threshold suggests a common mechanism for instability and predisposition to expansion in each disorder.

In this study, one allele which would be termed a premutation based on length (66 repeats), was found to transmit unchanged from a mother to two of her offspring (Table 3; Individual 15,135). This allele is remarkable in that it is stably transmitted despite the presence of 46 uninterrupted CGGs, and represents an exception to the

© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

Table 3 AGG interspersion in premutation alleles

Individual	Total number of repeats	Predicted sequence	Number of pure CGG repeats	Transition size to full mutation and number of CGG repeats (pure)
15,135	66	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_{46}$	46	stable
92002	61	$(CGG)_{12}$**AGG**$(CGG)_{48}$	48	ND
75–830	66	$(CGG)_9$**AGG**$(CGG)_{56}$	56	80 (70)
68–650	68	$(CGG)_9$**AGG**$(CGG)_{58}$	58	83 (73)
7544	59	$(CGG)_{59}$	59	68 (68)
75–787	72	$(CGG)_9$**AGG**$(CGG)_{62}$	62	80 (70)
80–813	82	$(CGG)_9$**AGG**$(CGG)_9$**AGG**$(CGG)_{62}$	62	ND
21,579	74	$(CGG)_{10}$**AGG**$(CGG)_{63}$	63	ND
75–828	76	$(CGG)_9$**AGG**$(CGG)_{66}$	66	80 (70)
16,415	92	$(CGG)_9$**AGG**$(CGG)_{10}$**AGG**$(CGG)_{71}$	71	95 (74)
A–4	83	$(CGG)_9$**AGG**$(CGG)_{73}$	73	84 (74)
15–522	104	$(CGG)_9$**AGG**$(CGG)_{94}$	94	66 (56)
66–691	110	$(CGG)_9$**AGG**$(CGG)_{100}$	100	80 (70)

Samples are arranged in ascending order of longest tract of pure CGG repeats. Transition size is the size of the *FMR1* CGG repeat measured in a female in each pedigree whose offspring obtained a full mutation Fragile X allele. Number of pure repeats in these females is inferred from the AGG interspersion pattern found in the male allele and is shown in brackets.
ND signifies status of female carrier not determined.

threshold proposed above. Several possibilities exist to explain the behaviour of this allele. First, only two transmissions have been observed, and each of these preserved the CGG repeat number. Further analysis of this allele may reveal instability when additional meioses are studied. Alternatively, this allele may be stabilized by a different interruption present within the 45 CGGs at the 3' end of this repeat, since *Mnl*1 digestions will only detect an AGG interruption. While this is the sole triplet other than CGG found by sequencing over 70 human alleles (E.E.E. and K.S., unpublished results), sequence analysis of this rare allele would help clarify this possibility. The presence of two AGG interruptions suggests a third explanation for the stability of this allele. The 34–37 repeat threshold is based on 12 unstable alleles, each of which contains no more than one AGG. It is possible that the presence of two AGGs has a stabilizing effect at these lengths.

Polarized variability. Mutational polarity in VNTR minisatellites has been reported for several mouse and human loci[25,26] Similarly, the majority of changes in the *FMR1* repeat appear to be occurring in the most 3' tract of CGG trinucleotides (relative to the orientation of transcription). We have demonstrated that the position of the AGGs remains invariant in expansions (Table 3). Two explanations, which are not mutually exclusive, for polarized variability within repeat sequences have been suggested. Sequence differences flanking the triplet repeat may influence directional change in the repeat[25] or the polarity may be the result of different mutation potentials for the lagging and leading strands during replication[27]. Although at this point it is impossible to distinguish which is more likely for *FMR1*, it may be noteworthy that deletions of palindromic repeats in *E. coli* have been found to occur much more frequently on the lagging than the leading strand[28].

Mechanism of mutation. Several mutational models were originally proposed to account for the generation of unstable and hyperexpanded alleles in the Fragile X Syndrome[2,6,15]. Polarized variability within a continuous tract of CGG repeats and the fact that changes involve differences of multiples of 3 bp clearly favour slipped-strand mispairing as the most likely mechanism of mutation[6,29]. Recently, unequal sister chromatid exchange and gene conversion have been suggested as a mechanism for variability of VNTR length[25]. If this were the dominant mode of expansion for Fragile X repeats, one might expect larger alleles to contain three or four AGGs as they are constructed in a cassette-like arrangement by conversion or unequal cross-over. Such alleles have been observed (Table 1), but appear to be rare. In contrast the majority of unstable larger alleles tend to have fewer, rather than more, interspersed AGGs. This would suggest that one of

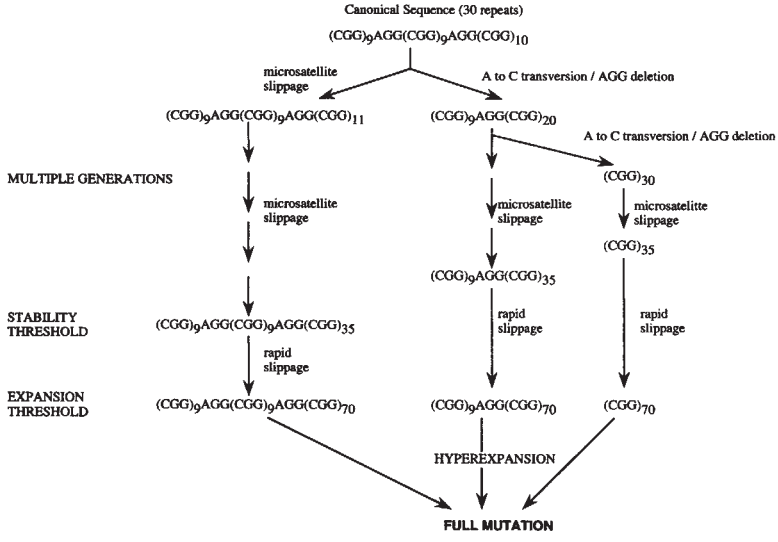

Fig. 5 Mechanism of mutation. A hypothetical depiction of multiple pathways for the generation of a Fragile X allele. Increases in pure CGG repeat length may be acquired either by slipped strand mispairing, deletion of AGG triplets or A to C transversions. The latter two would likely reach instability thresholds more rapidly when compared to the multiple generations required by slippage alone.

© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

the main predisposing events in Fragile X lineages has been the loss of an interrupting AGG. Such an event could be the result of an A to C transversion or a deletion, as has been proposed for other triplet repeat conversions[20,30]. We propose a model depicting several ways in which unstable *FMR1* CGG repeat alleles could be generated (Fig. 5). The recurrent loss of AGG trinucleotides may explain the occurrence of Fragile X alleles on various haplotypes[14].

Okazaki fragment slippage: a model for instability. Polarized variability, as observed for the *FMR1* CGG triplet repeat, may result from differential mutation potentials for the leading and lagging strands[27]. A recent study has suggested that Okazaki fragments are more likely to incur slipped-structures, resulting in a greater propensity for mutation[30]. Polarized variability observed for the *FMR1* CGG triplet repeat may therefore be the result of differential potential of the CCG or CGG strand to form slipped structures relative to the origin of replication. Studies of Fragile X kindreds indicate that the likelihood of expansion to full mutation for maternally transmitted alleles with greater than 90 repeats is close to 100%[6,12]. If such premutation alleles have a maximum of two interrupting AGGs, this number would suggest a minimum of 70 pure CGG repeats. Indeed, our analysis of premutation alleles indicates that pure CGG repeats ranging from 56–74 are sufficient for hyperexpansion upon female transmission in a single generation (Table 3).

The average length of an Okazaki fragment is 150–200 bp[31]. For alleles, with ~70 pure CGG repeats (210 bp), there is a increased likelihood that Okazaki fragment synthesis begins and ends within a continuous tract of pure CGG repeats. Such fragments, without single copy anchor points, may be prone to formation of slipped structures[32] (Fig. 6). Eukaryotic mismatch repair enzymes, which have been implicated in repeat instability[33–36], may distinguish parental and newly synthesized DNA, similarly to bacterial systems, by virtue of the fact that DNA is hemimethylated. Failure to repair such structures might result in exponential increases in repeat length characterizing the transition to the Fragile X full mutation. Unique microenvironmental conditions within the developing zygote, such as depletion of ATP pools[37,38], which results in a shortening of average Okazaki fragment length[39] or delay in activation of embryonic methylases may restrict such expansion events to initial stages of embryological development as has been suggested for the Fragile X syndrome[40].

In summary, both sequence length and sequence context are important parameters in the assessment of stability of *FMR1* CGG repeat alleles. We have demonstrated that the absence of interspersed AGG trinucleotides correlates with instability of normal-sized alleles. Such unstable repeats are likely precursors to fully-expanded disease alleles. Comparisons of internal content between stable and unstable alleles suggests a stability threshold between 34–37 pure CGG repeats. AGG interspersion analysis of 84 human chromosomes revealed that the majority of changes in repeat length have occurred in the 3' end. Based on the observations of polarized variability and that 56–74 pure CGG repeats is sufficient for hyperexpansion, we propose Okazaki fragment slippage as a possible model for rapid expansions of repeat length associated with the Fragile X syndrome.

Methodology

DNA samples. A total of 84 individuals were analysed for the internal composition of the *FMR1* CGG repeat. 55 of these were selected from unrelated male patients who had been submitted to the Kleberg BCM DNA Diagnostic Laboratory for Fragile X testing, but were found to have repeats within the normal range (13–54 repeats). Allele sizes were selected to reflect the somewhat trimodal distribution (20,30,40) in the normal population[6]. The remaining 29 subjects were obtained from pedigrees in which high-end normal stable or unstable transmissions of alleles had been observed. Of these, 13 alleles could be classified as premutations based on size (> 55 repeats), and of these, nine demonstrated expansion to full mutation

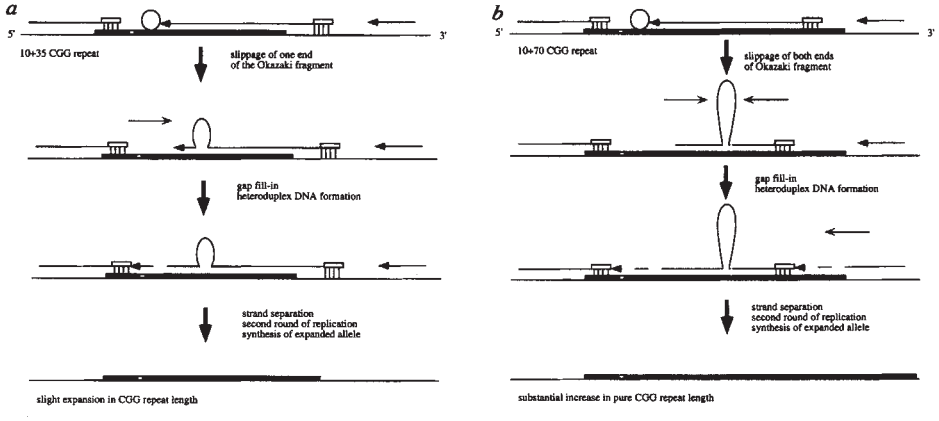

Fig. 6 Models for instability and hyperexpansion involving Okazaki fragment slippage. CGG repeat on the lagging strand is depicted by a blackened horizontal bar and the AGG interspersion is represented as a white square within the bar. Direction of recently synthesized DNA is shown with a horizontal arrow. The circle depicts DNA polymerase, while the white box with vertical lines depicts a DNA-RNA primer duplex. Fishook arrows show the direction of slippage. *a*, Instability slippage of a repeat consisting of 35 pure CGG repeats. Synthesis of DNA ends within the CGG repeat tract resulting in the formation of a small hair-pin structure. Resolution of heteroduplex DNA by second round synthesis results in moderate increase in CGG repeat length. *b*, Hyperexpansion slippage of a repeat consisting of 70 pure CGG repeats. Synthesis begins and ends within a continuous tract of pure CGGs. Okazaki fragment slippage occurs at both free ends. Multiple rounds of replication result in hyperexpansion of the triplet repeat.

© 1994 Nature Publishing Group http://www.nature.com/naturegenetics

by pedigree analysis. Size estimates were made on a 5% denaturing sequencing gel relative to a M13 sequencing ladder as previously described[6].

Amplification of the *FMR1* CGG repeat region. Amplification of the *FMR1* CGG repeat region was performed using a previously described method[21] with the exception that amplification of premutations was performed with native *Pfu* and native *Pfu* buffer, due to the decreased intensity of smear products when compared to Exo (–) *Pfu*. Two 50 µl reactions were amplified for each sample. These were combined and 15 µl was electrophoresed on a 1% agarose gel to ensure that sufficient amplification had occurred.

AGG interspersion analysis. The remaining 85 µl of each amplification were purified using a QuiaQuick purification kit (Quiagen) to eliminate excess salt and DMSO. DNA was eluted in a final volume of 50 µl TE. 10 µl (~200 ng) of each sample was digested overnight in a total reaction volume of 30 µl with 4 U *Mnl*1 (2 U µl^{-1}, New England Biolabs), 10 µg µl^{-1} BSA and 1× NEB Buffer II. Samples were electrophoresed at 10 V cm^{-1} for 7 h on a 3.0% Metaphor™ Agarose gel (FMC Bioproducts) at 4 °C. Gels were bidirectionally transferred to GeneScreen plus nylon membrane (Dupont), U.V. crosslinked and prehybridized for at least 1 h at 65 °C with 0.25 M Na PO_4, O.25 M NaCl, 5% SDS, 10% PEG, 1 mM EDTA and 100 µg ml^{-1} herring sperm DNA. A 30 mer oligonucleotide, $(CGG)_{10}$, was end-labelled with γ-^{32}P-ATP and 1 U Klenow fragment (NEB) for 45 min at 37 °C . The end-labelled oligonucleotide was purified through G-25 Sephadex columns and hybridized to Southern filters overnight at 65 °C in a rotisserie oven. Blots were washed three times for 30 min each at 65 °C in 0.05 M $NaPO_4$, 0.5% SDS and 1 mM EDTA solution and exposed to autoradiographic film. Hybridization with primer 10906 (see Fig. 1) was similar except that hybridization and washes were performed at 45 °C.

Acknowledgements

We would like to thank A. McAughan and M. Eichler for technical assistance, K. Kohlemainen and S. Chong for useful discussion and R. Sinden for valuable comments in the preparation of this manuscript. This work was supported in part by NIH grants (HD29256) to D.L.N. and MH50047 and MH01142 to A.L.R, as well as grant 1005-88-92 from the Ontario Mental Health Foundation to J.J.A.H. and B.N. White.

Received 20 May; accepted 23 June 1994.

1. Warren, S.T. & Nelson, D.L. Advances in molecular analysis of fragile X syndrome. *JAMA* **271**, 536–542 (1994).
2. Oberlé, I. *et al.* Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**, 1097–1102 (1991).
3. Verkerk, A.J.M.H. *et al.* Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
4. Yu, S. *et al.* Fragile X genotype characterized by an unstable region of DNA. *Science* **252**, 1179–1181 (1991).
5. Kremer, E.J. *et al* . Isolation of a human DNA sequence which spans the fragile X. *Am. J. hum. Genet.* **49**, 656–661 (1991).
6. Fu, Y.H. *et al.* Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047–1058 (1991).
7. Pieretti, M. *et al.* Absence of expression of the *FMR-1* gene in fragile X syndrome. *Cell* **66**, 817–822 (1991).
8. Ashley, C.T. *et al.* Human and murine FMR-1: evidence for alternative splicing and translational initiation downstream of the CGG-repeat. *Nature Genet.* **4**, 244–251 (1993).
9. Sherman, S.L. *et al.* Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum. Genet.* **69**, 289–299 (1985).
10. Yu, S. *et al.* Fragile X syndrome: unique genetics of the heritable unstable element. *Am. J. hum. Genet.* **50**, 968–980 (1992).
11. Heitz, D., Devys, D., Imbert, G., Kretz, C. & Mandel, J.L. Inheritance of the fragile X syndrome: size of the fragile X premutation is a major determinant of the transition to full mutation. *J. med. Genet.* **29**, 794–801 (1992).
12. Snow, K. *et al.* Analysis of a CGG sequence at the FMR-1 locus in the fragile X families and in the general population. *Am. J. hum. Genet.* **53**, 1217–1228 (1993).
13. Reiss, A.L. *et al.* Frequency and stability of the fragile X premutation. *Hum. molec. Genet.* **3**, 393–398 (1994).
14. Macpherson, J.N., Bullman, H., Youings, S.A. & Jacobs, P.A. Insert size and flanking haplotype in fragile X and normal populations: possible multiple origins for the fragile X mutation. *Hum. molec. Genet.* **3**, 399–405 (1994).
15. Richards, R.I. *et al.* Evidence of founder chromosomes in fragile X syndrome. *Nature Genet.* **1**, 257–260 (1992).
16. Hirst, M.C. *et al.* Origins of the fragile X syndrome mutation. *J. med. Genet.* **30**, 647–650 (1993).
17. Buyle, S. *et al.* Founder effect in a Belgian-Dutch fragile X population. *Hum. Genet.* **92**, 269–272 (1993).
18. Oudet, C. *et al.* Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that Fragile-X chromosomes are derived from a small number of founder chromosomes. *Am. J. hum. Genet.* **52**, 297–304 (1993).
19. Richards, R.I. *et al.* Fragile X syndrome: genetic localisation by linkage mapping of two microsatellite repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. *J. med. Genet.* **28**, 818–23 (1991).
20. Chung, M.-Y. *et al.* Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genet.* **5**, 254–258 (1993).
21. Chong, S.S., Eichler, E.E., Hughes, M.R. & Nelson, D.L. Robust amplification of the fragile X syndrome CGG repeat using *Pfu* polymerase: ethidium bromide detection of normal and premutation alleles. *Am. J. med. Genet.* **51**, 522–526 (1994).
22. Pergolizzi, R.G., Erster, S.H., Goonewarda, P. & Brown, W.T. Detection of full fragile X mutation. *Lancet* **339**, 271–272 (1992).
23. Nelson, D.L. *et al.* Characterization of the FRAX(A) locus in man. *Am. J. hum. Genet.* **51**, A186 (1992).
24. Nelson, D.L. & Warren, S.T. Trinucleotide repeat instability: when andwhere? *Nature Genet.* **4**, 107–108 (1993).
25. Jeffreys, A.J. *et al.* Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* **6**, 136–145 (1994).
26. Armour, J.A.L., Harris, P.C. & Jeffreys, A.J. Allelic diversity at minisatellite MS205 (*D16S309*): evidence for polarized variability. *Hum. molec. Genet.* **2**, 1137–1145 (1993).
27. Richards, R.I. & Sutherland, G.R. Simple repeat DNA is not replicated simply. *Nature Genet.* **6**, 114–116 (1994).
28. Trinh, T.Q. & Sinden, R.R. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**, 544–548 (1991).
29. Levinson, G. & Gutman, G.A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *A. Rev. Biochem.* 203–221 (1986).
30. Gostout, B., Liu, Q. & Sommer, S.S. Cryptic repeating triplets of purines and pyrimidines (cRRY(i)) are frequent and polymorphic: analysis of coding cRRY(i) in the proopiomelanocortin (POMC) and TATA-binding protein (TBP) genes. *Am. J. hum. Genet.* **52**, 1182–1190 (1993).
31. Thömmes, P. & Hübscher, U. Eukaryotic DNA replication. *European J. Biochem.* **194**, 699–712 (1990).
32. Wells, R.D. & Sinden, R.R. in *Genome Anlaysis: Genome Rearrangement and stability* (eds Davies, K.E. & Warren, S.T.) 107–138 (Cold Spring Harbor Laboratory Press, New York, 1993).
33. Wooster, R. *et al.* Instability of short tandem repeats (microsatellites) in human cancers. *Nature Genet.* **6**, 152–156 (1994).
34. Strand, M., Prolla, T.A., Liskay, R.M. & Petes, T.D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
35. Thibodeau, S.N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819 (1993).
36. Parsons, R. *et al.* Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227–1236 (1993).
37. Alexiou, M. & Leese, H.J. Purine utilisation, de novo synthesis and degradation in mouse preimplantation embryos. *Development* **114**, 185–192 (1992).
38. Gibbs, M., Collick, A., Kelly, R.G. & Jeffreys, A.J. A tetranucleotide repeat mouse minisatellite displaying substantial somatic instability during early preimplantation development. *Genomics* **17**, 121–128 (1993).
39. Nethanel, T., Zlotkin, T. & Kaufmann, G. Assembly of simian virus 40 Okazaki pieces from DNA primers is reversibly arrested by ATP depletion. *J. Virol.* **66**, 6634–6640 (1992).
40. Reyniers, E. *et al.* The full mutation in the *FMR-1* gene of male fragile X patients is absent in their sperm. *Nature Genet.* **4**, 143–146 (1993).
41. Brown, W.T. *et al.* Rapid fragile X carrier screening and prenatal diagnosis using a nonradioactive PCR test. *JAMA* **270**, 1569–1575 (1993).