Hotspot for deletions in the CGG repeat region of FMR1 in fragile X patients

Esther de Graaff, Patricia Rouillard1, Patrick J.Willems2, Arie P.T.Smits3, Francois Rousseau1 and Ben A.Oostra*

Department of Clinical Genetics, Erasmus University, PO Box 1738, 3000 DR Rotterdam, The Netherlands, 1Unité de Recherche en Génétique Humaine et Moléculaire, Hôpital St-Francois-d'Assise and Department of Biochemistry Laval University, Quebec, Canada, 2University of Antwerp, Antwerp, Belgium and 3University of Nijmegen, Nijmegen, The Netherlands

Received September 15, 1994; Revised and Accepted October 11, 1994

The fragile X syndrome is the most frequent cause of inherited mental retardation. The molecular mechanism of the disorder is based on the expansion of a CGG repeat in the 5' UTR of the FMR1 gene in the majority of fragile X patients. The instability of this CGG repeat containing region is not restricted to the CGG repeat itself but expands to the flanking region as well. We describe four unrelated fragile X patients that are mosaic for both a full mutation and a small deletion in the CGG repeat containing region. Sequence analysis of the regions surrounding the deletions showed that both the (CGG)n repeat and some flanking sequences were missing in all four patients. The 5' breakpoints of the deletions were found to be located between 75-53 bp proximal to the CGG repeat. This suggests the presence of a hot spot region for deletions in the CGG repeat region of the FMR1 gene and emphasizes the instability of this region in the presence of an expanded CGG repeat.

INTRODUCTION

The fragile X syndrome is the most common inherited form of mental retardation (1) and the disorder is associated with the expression of a fragile site at Xq27.3. Besides mental impairment, the patients are characterized by typical morphological features such as large, prominent ears, elongated face with mandibular prognathism and macroorchidism (2). The incidence of fragile X syndrome is estimated to be approximately 1 in 1250 males and 1 in 2500 females (for review see 3).

In the majority of the patients, the molecular mechanism of the disorder is based on an unstable (CGG)n trinucleotide repeat in the 5' UTR of the FMR1 gene (4–7). In the normal population this trinucleotide repeat is polymorphic, varying between 6 and 52 repeat units, and is stable upon transmission (7). Premutation alleles, ranging from 50–200 repeats, exhibit instability, usually resulting in increases in repeat number in the offspring when compared to the parents. Premutations are not associated with a clinical phenotype and are found in female carriers and normal transmitting males. During or after passage through female meiosis, a premutation can expand to a full mutation of over 200 repeats resulting in methylation of the CGG repeat and the CpG island 250 bp proximal to the CGG repeat (5,8–10). This methylation is associated with repression of FMR1 transcription, thereby resulting in the severe reduction of the level of FMR1 protein leading to the fragile X phenotype (11,12). Although the function of the FMR1 protein is not yet fully understood, evidence is accumulating that the protein plays a role in RNA metabolism. Two motifs have been identified that are present in RNA binding proteins: a RGG box and a KH domain (13,14).

Transmission of an expanded CGG repeat mostly results in increase in repeat length, although occasionally, regression of the CGG repeat has been reported as well. Reduction of a premutation to a smaller premutation has been detected (6,7,15–17) as well as reduction of a premutation to a repeat length within the normal range (18). Decrease from a full mutation to a premutation has been reported only once (15), whereas the reverse mutation, from full mutation to a normal allele has not yet been described. This is in contrast to the CTG repeat in myotonic dystrophy, in which a number of independent reverse mutations have already been described (19–21).

Several deletions in the FMR1 gene resulting in the fragile X syndrome are now known. Two large deletions of 2.5 Mb (22) and 3 Mb (23) have been described including the entire FMR1 gene plus flanking sequences. In a third patient a 250 kb deletion resulted in the absence of the FMR1 gene associated CpG island and the first five exons of the gene (24), whereas in a fourth patient ~100 kb, including the first eight exons and upstream sequences, were missing (25). Recently a small deletion of 1.6 kb immediately 5' of the CGG repeat was detected, removing regulatory sequences of the gene. This deletion resulted in four affected males within a fragile X family (26).

The instability of the CGG repeat is not only seen between generations, but it can also be seen within one individual. Patients with a full mutation display a variety of repeat lengths,
visible as a smear in Southern blot analysis. In addition, ~20% of the affected males are mosaic for both a full mutation and a premutation allele (5,7,11,15).

In this paper we describe four unrelated fragile X patients that are mosaic for a full mutation in the majority of their lymphocytes and a deletion of the CGG repeat and flanking sequences in ~5-10% of their blood lymphocytes. Sequence analysis of the deletion junctions revealed that all 5' breakpoints are located within 30 bp of each other. This emphasizes the instability of the CGG repeat region of \textit{FMR1}.

**RESULTS**

We studied the length of the CGG repeat and the methylation pattern of the CpG island proximal of the \textit{FMR1} gene using Southern blot analysis. In normal males and females digestion of DNA with \textit{EcoRI} results in a 5.2 kb fragment. The majority of fragile X males exhibit a smear of over 5.8 kb, due to expansion of the CGG repeat to over 200 repeats. In four male patients, digestion with \textit{EcoRI} not only resulted in this smear but in addition a fragment slightly smaller than the normal 5.2 kb band was found (data not shown). These novel fragments were estimated to be derived from ~5-10% of the cells in patients 1, 2 and 4 and 15% in patient 3. Double digestion of DNA of normal males with \textit{EcoRI} and the methylation sensitive enzyme \textit{EagI} leads to a 2.8 kb fragment (Fig. 1, lanes 6 and 9). In females both the 2.8 kb fragment and the methylated 5.2 kb fragment are detected, due to normal X chromosome inactivation (Fig. 1, lanes 7 and 8). In fragile X males the smear of over 5.8 kb remains due to methylation of the \textit{EagI} site. However, in the four male patients we detected both this smear as well as a smaller fragment (Fig. 1, lanes 1, 3, 5, 10). These fragments varied in size between 2.7 kb in patient 1 (lane 1) to 2.2 kb in patient 3 (lane 5), again indicating the presence of deletions that varied from 150-600 bp, all falling within the \textit{EcoRI-EagI} 2.8 kb fragment. The \textit{EagI} restriction site could clearly be digested, which implies the absence of hypermethylation of the novel fragment. DNA derived only from the mothers of patient 1 and 2 was also subjected to Southern blot analysis (Fig. 1, lanes 2 and 4, respectively).

In order to determine the exact junction of the four deletions we PCR amplified the region surrounding the CGG repeat. The deletion in the \textit{EagI-EcoRI} fragment was only present in a low percentage of their cells. Therefore, in order to get a sufficient amount of PCR product, we performed two cycles

![Figure 1. Southern blot analysis of the four patients. DNA was digested with \textit{EcoRI-EagI} and following electrophoresis, fragments were blotted on filters and hybridized with \textit{pP2}. Lanes 1, 3, 5 and 10 represent patients 1, 2, 3 and 4 respectively. Lane 2: mother of patient 1; lane 4: mother of patient 2; lanes 6 and 9: control male; lanes 7 and 8: control female. The normal length of the fragments are indicated.](image)

![Figure 2. Schematic representation, not drawn to scale, of the deletions found thus far in the CGG region. The position of the primers in the pE5.1 sequence are indicated (32) as triangles. WT: wild type sequence; ATG: startcodon. Patients (Pat.) 1-4 are described in this paper; XTY26 and Pat. 24 were described by Kremer et al. (27) and Meijer et al. (26), respectively.](image)

<table>
<thead>
<tr>
<th>Patient</th>
<th>Deletion breakpoint*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>CCGACCGGGAGCGGCGGGG-------ttttctcttgTCTCTCCGG</td>
</tr>
<tr>
<td>Patient 2</td>
<td>CCGGGCGGGAGCGGCTCTAG-------ccggccggcccACCTCTCCGG</td>
</tr>
<tr>
<td>Patient 3</td>
<td>GCCGCCTGGGACCGGGCGGG-------ggggcggaaacGCCGTAAGT</td>
</tr>
<tr>
<td>Patient 4</td>
<td>CCGGGCGGGAGCGGCGGGG-------ggggcggaaacGCCGTAAGT</td>
</tr>
<tr>
<td>Pat. 24</td>
<td>TTAACAGGTGGGGATAT-------ggggggggggggGGCGGGCGG</td>
</tr>
<tr>
<td>XTY26 5'</td>
<td>CCGGGCGGGAGCGGCGGGG-------ggggcggaaacGCCGTAAGT</td>
</tr>
<tr>
<td>XTY26 3'</td>
<td>CCGGGCGGGAGCGGCGGGG-------ggggcggaaacGCCGTAAGT</td>
</tr>
</tbody>
</table>

*The wild type sequence is shown. The deleted sequences are indicated by lower case letters. Short regions of homology at the deletion termini are underlined. The asterisks denote complementary regions. The Chi-like element is indicated in bold italics. Pat. 24 and XTY26 were described by Meijer et al. (26) and Kremer et al. (27) respectively.
of amplification standardly. After a Centricon-100 purification, the PCR products were sequenced using the same primers as were used for amplification. A schematic representation of the deletion breakpoints of the four patients described are shown in Figure 2. This figure also includes the deletion described by Meijer et al. (26; pat. 24) and Kremer et al. (32, XTY26). The exact sequences of the junctions of all deletions are shown in Table 1.

**DISCUSSION**

The instability of the expanded CGG repeat in the *FMRI* gene is a well known phenomenon. Transmission of a premutation to the offspring can result in increase and decrease of the CGG repeat, whereas increase to a full mutation only occurs after transmission by a female (6,7,28). The variation of repeat length is not only found between generations but is also present within individual fragile X patients and is called somatic instability or mosaicism. Two different types of mosaicism have been observed. Firstly, there is the wide range of repeat lengths of the full mutation, visible as a smear on Southern blot analysis (7,29,30). Secondly, 20% of the fragile X patients have a premutation allele in addition to the full mutation (7,11,15). Mosaicism is established in early fetal life. Identical patterns of mosaicism were demonstrated in different fetal tissues of monozygotic twins (30) and a 13 week old fetus (31). In vitro studies on proliferation of single fibroblasts from adults resulted in several clones, each carrying only one repeat length (29). This indicates that once mosaicism is established early in development it remains stable during life.

In this paper we describe a third kind of mosaicism; in four unrelated fragile X males we detected both a full mutation and a deletion of the CGG repeat and flasking sequences. The deletions in patients 1 and 3 encompasses the first ATG of *FMRI*. Therefore these patients will not produce a functional protein. The other two patients may have *FMRI* mRNA and protein expression but this could not be tested. However, this may be irrelevant, for it has already been shown that mosaic males, with both a full mutation and a premutation, have some mRNA expression but do not have a significantly different clinical phenotype than patients with a full mutation only (11,15).

The deletions reported in this paper are not the first deletions known in the region flanking the CGG repeat. In 1991, Kremer et al. (27) sequenced the 1 kb *PstI* fragment, that was found to contain the unstable CGG repeat. This sequence was obtained from a YAC (XTY26) that was subcloned from a hybrid cell line (X.3000.1) expressing a fragile X chromosome. Southern blot analysis had shown that this cell line contained ~350 triplets. However, only 43 CGGs were detected in the YAC, indicating that a large part of the repeat had been deleted during cloning. Sequence analysis of this region in DNA of both fragile X patients and unaffected individuals performed independently by Fu et al. (7) revealed that in the YAC some of the CGG flanking sequences were lacking as well. Remarkably, the 5′ endpoint of the deletion in the YAC, at position 2671 of the pE5.1 fragment, is also located within the 35 bp interval in which the 5′ endpoints of the other four deletions are found (Fig. 2; 32).

Recently, a deletion in the CGG region has been described in a fragile X family (26). The 3′ endpoint of this 1.6 kb deletion was located within the CGG repeat and the deletion was assumed to be caused by an expanded, unstable CGG repeat. Previously, four large deletions, spanning 2.5 Mb, 3 Mb, 250 kb and 100 kb respectively, in the fragile X gene have been described (22–25). The CGG repeats of the mothers of these patients were in the normal range suggesting that these deletions were not caused by an unstable CGG repeat. The four deletions described in this paper appear to be directly derived from an expanded, unstable CGG repeat; all four patients expressed a full mutation smear on Southern blot analysis besides the deletion fragment (Fig. 1).

An intriguing question remains: when did the deletion take place? In 1993, Reyniers et al. (33) showed the presence of a premutation and not a full mutation in sperm of patients with a full mutation in their blood cells. From this they postulated two hypotheses on the possible timing of the CGG repeat amplification. The first model assumed expansion during meiosis of the oocyte, resulting in the presence of a full mutation in all somatic cells. In sperm cells this full mutation regressed to a premutation and due to a selection mechanism only cells with a premutation will proliferate. In the second model a premutation expands to a full mutation during postzygotic proliferation of the somatic cells, after the separation of the germ line (day 5–6 post conception). Recent breeding studies with fragile X knock out mice demonstrated that Fmr1 gene expression is not required for gametogenesis (34). In addition, Meijer et al. (26) showed that the 1.6 kb deletion resulted in the absence of *FMRI* mRNA in the affected males of the family and that this deletion was originally derived from a male with a progeny of five children. This confirms that the expression of *FMRI* is not required for spermatogenesis, thus excluding the presence of a selection mechanism in sperm. The second model therefore seems more likely.

The deletions of the four patients described in this paper could originate from either a premutation or a full mutation. In the first case the premutation allele expands, resulting in a full mutation in the offspring. However, due to a certain error, in a few cells instead of expansion a deletion takes place. In the latter case, the premutation allele of the mother expands to a full mutation. This full mutation, which is somatically much more unstable than premutations, later, during subsequent proliferation and development of the zygote, results in a deletion in a low percentage of the cells. It is possible that all...
premutations found in mosaic patients are the result of regression of a full mutation indicating that reduction of the CGG repeat is a more frequent event than expected. Female carriers of a full mutation can have children with either both a full and a premutation or a full mutation alone. In these mosaics the premutation may well be caused by a deletion of a large number of CGGs in the full mutation of the mother. A similar deletion mechanism could have resulted in a patient described by Van den Ouweland et al. (35). This patient had, in addition to a full mutation, two alleles within the normal range. The repeat length of these normal alleles, 21 and 33 respectively, did not correspond to the mother’s normal allele of 29 repeats. The sequences flanking the repeat were found to be normal suggesting that regression within the mother’s expanded CGG repeat resulted in the two normal alleles present in this patient. In most cases the deletion that causes the reduction from full mutation to premutation will be restricted to the CGG repeat meaning that on Southern blot analysis either a normal or a premutated fragment will be detected. These deletions will not be noticed in diagnostic screening, resulting in an underestimation of the total percentage of deletion events. Occasionally, the deletion will also involve sequences flanking the CGG repeat, resulting in the four deletions described in this paper, now estimated to be below 1%. It should be noted that these latter deletions will not be detected if the diagnostic analysis is performed by PCR using primers immediately flanking the repeat.

Regression of a full mutation to a premutation would involve two mechanisms: the reduction in number of CGG repeats and demethylation. Both the premutations found in mosaic patients and the deletions described in this paper are unmethylated fragments (Fig. 1). Demethylation may be explained by the fact that methylation of the CpG island is an active process, maintained by the presence of a full mutation. Disappearance of the full mutation would then result in passive demethylation. Methylation being an active mechanism might explain the occurrence of unmethylated full mutations (36). The possibility that demethylation is an active process in itself, causing the deletion by, for instance, nicking, can not be excluded.

Several studies on the mechanism of deletions have shown that ~40% of the large deletions in human disorders are flanked by very short direct repeats of 2–6 bp (37–39). It is proposed that these short repeats slip and mispair during replication, resulting in the formation of a loop between the two repeats (37,38). Subsequent excision of this loop removes both the sequences between the repeats as well as one of the repeats. Of the seven deletions in the CGG region of FMR1 now known five have such a short region of homology (Table 1). Patient 3 and the deletion described by Meijer et al. (26) have a GC flanking the deletion. The deletion of XTY26 at the 3′ site of the repeat is flanked by a CG and in patient 2 the deletion is flanked by a CC. It should be noted however, that the CGG region of FMR1 is very C-G rich and that both CG, GC and CC are present rather often in this region. It can therefore not be excluded that the presence of the 2 bp repeats represents a chance occurrence. The deletion at the 5′ site of the CGG repeat of XTY26 is a better example of this slippage mispairing mechanism. This deletion is flanked by a sequence repeat of 11 bp (Table 1).

Another mechanism may account for the deletion in patient 4. This deletion is flanked by a short inverted repeat; 9 out of 10 bp are complementary (Table 1). These sequences can align together in a quasi-palindromic manner, resulting in the formation and subsequent excision of the loop, as previously described (40,41; Fig. 3).

Neither of the two mechanisms can explain the close vicinity of the 5′ endpoint of the different deletions. Recently Jeffrey et al. (42) demonstrated that both gain and loss of repeat units in a minisatellite mainly occurred at one end of the repeat. They postulated the presence of a specific mutation initiator element adjacent to the repeat that would cause the instability by inducing a double strand break. Such an element could be a Chi element, a short sequence that is often located in the vicinity of breakpoints. Chi or Chi-like elements are thought to play a major part in recombination processes (43) and may also be involved in replication slippage (44). At position 2626, immediately 5′ of the breakpoint of patient 2, a Chi-like sequence is present (GTGCGAGG; Table 1) and could cause a DNA break. Subsequent misrepair of this break, possibly caused by the presence of a large CGG repeat, would then result in a deletion.

Our results that only the 5′ and not the 3′ breakpoints of the deletions are located within a hotspot, correlate well with the previous findings of Kunst and Warren (45). They demonstrated that expansion of the CGG repeat occurs at the 3′ end of the repeat. Both studies thus indicate that variability of the CGG repeat is mainly present at the 3′ site, suggesting that the replication of this region goes in one direction.

In conclusion we have described a hotspot of deletions ~70 bp upstream of the (CGG)n repeat. These findings may help in future unravelling of the mechanism of the CGG repeat expansion.

**MATERIALS AND METHODS**

**Patients**

Patient 1 has a Prader—Willi phenotype and was previously described by de Vries et al. (46). The other three patients have the classical Martin—Bell phenotype.

**Southern blot analysis**

Genomic DNA was isolated from blood leukocytes according to standard procedures (47). Seven μg DNA was digested to completion with EcoRI and EcoQ, electrophoresed on a 0.7% agarose gel and transferred to Hybond N+ blotting membrane. Probes StB123.5 (5) or p2, a 1 kb Pvu II fragment derived from pES1 which identifies the (CGG)n repeat and the preceding CpG island (48), were used after labelling by the random oligonucleotide priming method (49). After overnight hybridization the filters were washed to 0.1×SSC, 0.1% SDS at 65°C prior to exposure to X-ray film.

**PCR and sequencing analysis**

Amplification was performed using 200 ng of genomic DNA in a total volume of 50 μl in the presence of 0.2 mM each of dATP, dCTP and dTTP, 0.05 mM dGTP, 0.15 mM 7-deaza-dGTP, 10 mM Tris—Cl pH 8.3, 50 mM KCl, 10% DMSO, 1 mM MgCl2, 2.5 U Taq polymerase (BRL) and 1 μM of each primer. The reactions were first denatured for 5 min at 95°C, followed by 30 cycles of 2 min at 95°C, 2 min at 61°C and 2 min at 72°C. A final extension of 10 min was carried out at 72°C. In order to get enough product to use for sequencing, a sample of 2.5 μl of the PCR product was additionally amplified another 30 cycles using the same conditions. Products were analysed on a 1% agarose gel, stained with ethidium bromide. After a Centricon-100 purification (Amicon), the PCR products of the second amplification were directly sequenced using the BRL cycle sequencing kit. The primers used for the PCR reactions and the sequencing protocols as well as their position in the pES1 sequence are indicated (between brackets) (52):

- 5′-CCGCGTCTGTTCTTTGCAGCC (2409–2429)
- 5′-TGAAGGCTCTAACAGCGCC (3073–3093)
- 5′-CTTCACAGGGCCTCCGCTCAAACAG (3408–3430)
- 5′-CCGCTACGCTCTTGTTCCTTGTC (2997–2619, ref.27)
- 5′-CTCCATCTTCTCTTGCCCGCTCTAG (2816–2842, ref.27)
ACKNOWLEDGEMENTS

We thank Drs L.B.A.de Vries, B.A.van Oost, D.Chityat and V.M.Der Kalouslian for referring fragile X cases. We thank Wouter Katje (Department of Cytochemistry, Lübeck) and Bert Eussen (Department of Clinical Genetics, Rotterdam) for providing the photograph of the X chromosome with the fragile site (FRA-X) as shown by Atomic Force Microscopy. P.Rouillard holds a studentship from the Medical Research Council of Canada. This study was supported by a grant of the MOCZWN to B.A.Oostra and the foundation of Clinical Genetics to E.de Graaff.

REFERENCES