The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing

Catherine B Kunst, Esther P Leeflang, Jane C Iber, Norman Arnheim, Stephen T Warren

Abstract
Fragile X syndrome results from the unstable expansion of a CGG repeat within the FMR1 gene. Three classes of FMR1 alleles have been identified, normal alleles with 6–60 repeats, premutations with 60–200 repeats, and full mutations with >230 repeats. Premutations are exquisitely unstable upon transmission. Normal alleles, while generally stable upon transmission, are thought to have different intrinsic mutation frequencies, such that some normal alleles may be predisposed towards expansion while others may be more resistant to such change. One variable that may account for this difference is the occurrence of one or more AGG triplets punctuating the normal CGG repeat. The AGG interruptions lead to alleles that have equivalent overall length but different lengths of perfect repeats. To test the influence of the length of perfect repeats on stability, we examined the CGG repeat of single sorted sperm from two males, each with 39 total repeats, but distinct AGG interruption patterns. Sorted sperm of each donor showed ~15% variation in repeat length, consistent with previous studies of sorted sperm at other triplet repeat loci. However, when discounting the majority variation of ±1 repeat, the male with 29 perfect repeats showed 3% expansion changes while the donor with only 19 perfect repeats had none (<0.9%). Moreover, >90% of all variant sperm, including all those observed with expansions, showed expansion of the FMR1 repeat of the 3' end of the repeat array. These data are consistent with the hypothesis that perfect repeat tracts influence the repeat stability and that changes of the FMR1 repeat exhibit polarity.

Keywords: FMR1; CGG repeat; sperm typing

Fragile X syndrome, an X linked dominant disorder with reduced penetrance, is a frequent cause of mental retardation. The molecular basis of fragile X syndrome has been determined to be the loss of function of the FMR1 gene located at Xq27.3. The loss of function is unusual as it is the consequence of a CGG repeat expansion in the 5' untranslated portion of FMR1. Penetrant males with fragile X syndrome have an average of ~800 repeats which are associated with aberrant methylation of the FMR1 gene and transcriptional silencing. The absence of the encoded nucleocytoplasmic RNA binding protein, FMRP, is responsible for the phenotype.

Non-penetrant carriers of fragile X syndrome carry an unmethylated and expressed FMR1 gene with between ~60 and 200 repeats that is referred to as a premutation. Premutations, when transmitted from either sex, are unstable leading to offspring with repeat lengths distinct from the transmitting parent. Although male transmission of a premutation leads to only premutation offspring, female transmission can result in either premutation offspring or children with full mutations, carrying an excess of 230 repeats. The probability of a child having a full mutation rather than one bearing a premutation is directly related to the maternal premutation repeat length. Therefore, it has been proposed that longer tracts of perfect CGG repeats result in a proportionately higher likelihood of DNA slippage or other undefined molecular event expanding the repeat allele.

Studies of the normal FMR1 repeat are consistent with the concept that perfect repeat arrays promote length variation. Normal FMR1 alleles are polymorphic in length as well as content. The overall repeat array varies from 6 to ~60 units with a mean of 30 repeats and can include one to three AGG interruptions. These interruptions are commonly situated at every 10th repeat from the 5' end of the array. Therefore, tracts of >10 continuous CGG repeats are most often found towards the 3' end of the array, which is the same region that expands in premutations. Studies of nearby flanking polymorphisms have defined haplotypes which exhibit linkage disequilibrium among normal and fragile X chromosomes, suggesting that some normal haplotypes may be predisposed towards instability.

Indeed, these haplotypes are those in which the normal FMR1 alleles contain fewer interruptions and therefore longer perfect repeat runs. Conversely, haplotypes displaying negative linkage disequilibrium are associated with normal alleles containing two or more interruptions and consequently shorter perfect repeats. Although there is evidence for cis or trans acting elements affecting repeat stability, the AGG interruptions and resulting length variation of perfect CGG repeat tracts is the most clearly defined influence to date.

Direct measurement of the effect of the purity of CGG repeat tracts on stability has been difficult and is limited to population or family studies. Here we test the stability of the

Howard Hughes Medical Institute and Departments of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA
C B Kunst*
J C Iber
S T Warren

Molecular Biology Program, University of Southern California, Los Angeles, CA 90089, USA
E P Leeflang
N Arnheim

Correspondence to:
Dr Warren, Howard Hughes Medical Institute, Emory University School of Medicine, Room 4035
Rollins Research Center, 1510 Clifton Road, Atlanta, Georgia 30322, USA.
*Present address: Eleanor Roosevelt Institute, Denver, CO 80206, USA.

Received 19 February 1996
Revised version accepted for publication 26 March 1997
FMRI repeat by examining the sorted sperm from two normal men. The donors have identical overall repeat lengths but differ in their number of AGG interruptions and therefore the lengths of their perfect CGG repeats.

Materials and methods
Individual sperm from the chosen donors were sorted into 96 well dishes by published methods. Sperm were lysed and neutralised in these dishes and were used without further purification for PCR amplification. After briefly centrifuging the dishes to remove condensation, 3 μl of additional lysis solution (200 mmol/l KOH, 50 mmol/l DTT) and 3 μl of additional neutralisation solution (900 mmol/l Tris, pH 8.3, 300 mmol/l KCl, 200 mmol/l HCl) were added to each well. These samples were transferred to UV irradiated PCR tubes which contained UV irradiated oil. To prevent contamination, all solutions (except primers) were UV irradiated for these experiments. The initial PCR amplification was carried out as previously described in a total reaction volume of 30 μl except that a Taq polymerase/Phu polymerase blend (16:1) was used for the PCR amplifications to minimise repeat slippage and glycerol was added to a final concentration of 10%. Because Tris and KCl are in the lysis and neutralisation buffers, no additional PCR buffer was used in the primary round of amplification, although MgCl2 was added to a final concentration of 1.25 mmol/l. As described previously, 25 1 μl of the primary amplification product was used for the second round of PCR amplification and samples were separated on a 6% denaturing polyacrylamide gel containing 32% denatured formamide. Repeats were sized against size markers and FMRI repeats of known sequence. All variant PCR products were directly sequenced by automated fluorescent methods as previously described.

Results
Examination of 155 anonymous semen donors identified two males each with 39 repeats at FMRI but with distinct AGG interruption patterns (at a frequency similar to previous population screens13-16). Donor 727 had a single AGG located at triplet position 10 resulting in a tract of 29 perfect CGG repeats. In contrast, donor 791 had two AGG interruptions at triplet positions 10 and 20 resulting in only 19 perfect CGG repeats. Based upon previous population studies, it has been suggested that FMRI alleles of various lengths, each containing >24 perfect CGG repeats, may be predisposed towards expansion and comprise a pool of ancestral fragile X syndrome alleles. Thus, donor 727 would harbour such a predisposed allele while donor 791 would not.

Sperm of donors 727 and 791 were sorted into 96 well dishes, each containing eight empty wells as internal negative controls. The individual sperm were lysed and neutralised as previously described24 and used without further purification. Using two sets of primers flanking the FMRI repeat, a nested PCR amplification was carried out on each well with a3-P-dCTP added during the final set of amplifications.25 The PCR products were then resolved by polyacrylamide gel electrophoresis and detected by autoradiography. As shown in fig 1, variation in repeat length of individual sperm products could be easily discerned. Since no signals were ever observed in the empty control wells, the observed bands more than likely emanated from sperm DNA amplification. From subsequent sequence analysis (below), as little as ±1 triplet could be detected. Although some slippage during PCR is evident by the smearing of the signal, as is typical in microsatellite amplifications, at optimal exposures the predominant product was evident and scored. In selected cases of variation from the expected repeat length, reamplification of the initial PCR reactions with the nested primer set resulted in identical product lengths with the first round. It has already been shown that slippage during PCR amplification does not make a significant contribution to the detection of new allele sizes. Although slippage during PCR cannot be totally excluded, the overall trend of the data based on many sperm should reflect the actual germline mutation process.

From 510 wells into which sperm of donor 727 was sorted, 214 wells resulted in PCR products. This 42% well amplification agrees well with the theoretical maximum of 50% X-bearing sperm. Amplification of sorted sperm of donor 791 was less efficient resulting in 119 products from the amplification of 531 wells (22% total or 44% of the X chromosomes). The level of PCR efficiency for sperm from donor 791 is within the range seen in other studies and could reflect difficulties in sperm sorting, sperm lysis, or other technical factors.24 26 27

As shown in table 1, amplified sperm of both donors showed ~15% variation in length from the constitutional 39 repeats. In donor 791 nearly all the observed variation (94%) resulted from ±1 repeat change while only 63% of the variation for donor 727 resulted from ±1 repeat. Direct sequence analysis of all the variant PCR products confirmed the repeat number and showed that >90% of all variant sperm from either donor maintained the expected AGG interruption pattern. Thus, nearly all variation observed occurred in the 3' portion of the repeat array (fig 2). In donor 727 a loss or gain of a single CGG repeat in the first nine repeats (8+29 and 10+10, table 1) accounted for all changes in the AGG pattern and in donor 791 a similar single repeat loss or
The effect of FMR1 CGG repeat interruptions on mutation frequency

Table 1  Variation of the FMR1 repeat in sorted sperm of two normal males differing in repeat sequence but not length*

<table>
<thead>
<tr>
<th>Change in repeat length</th>
<th>Donor 727 (39 repeats; 9+29)</th>
<th>Repeat sequence</th>
<th>Donor 791 (39 repeats; 9+9+19)</th>
<th>Repeat sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>+6</td>
<td>1 (0.5%)</td>
<td>45; 9+35</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>+2</td>
<td>6 (2.8%)</td>
<td>41; 9+31</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>+1</td>
<td>11 (5.1%)</td>
<td>40; 9+30</td>
<td>8 (6.7%)</td>
<td>40; 9+9+20</td>
</tr>
<tr>
<td>0</td>
<td>182 (85%)</td>
<td>39; 9+29</td>
<td>99 (83%)</td>
<td>39; 9+9+19</td>
</tr>
<tr>
<td>−1</td>
<td>9 (4.2%)</td>
<td>38; 9+28</td>
<td>8 (6.7%)</td>
<td>38; 9+9+18</td>
</tr>
<tr>
<td>−2</td>
<td>1 (0.5%)</td>
<td>38; 9+29</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>−3</td>
<td>1 (0.5%)</td>
<td>36; 9+26</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>−4</td>
<td>1 (0.5%)</td>
<td>35; 9+25</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>−18</td>
<td>1 (0.5%)</td>
<td>21; 10+10</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>−19</td>
<td>0</td>
<td>—</td>
<td>1 (0.8%)</td>
<td>1 (0.8%)</td>
</tr>
</tbody>
</table>

* Repeat sequence designated by total triplet number followed by nomenclature where a + sign represents the position of an AGG within pure tracts of CGG repeat, the length of which is designated by numbers. Thus (CGG)AGG(CGG)_{n} is 39; 9+29.

Figure 2  Sequence analysis of the FMR1 repeat from donor 727 sorted sperm. Panel A shows the representative sequence of the predominant allele with 39 repeats and an AGG interruption at position 10. Panel B shows an expanded allele with 45 repeats with retention of the AGG at position 10. Blue dots indicate CGG triplets while red dots indicate AGG interruptions; 5' to 3' polarity is left to right according to the FMR1 reading frame.

Table 2  Sorted sperm results of changes of at least two repeat units*

<table>
<thead>
<tr>
<th>Size of expansions</th>
<th>Donor 727 (9+29) (n=214)</th>
<th>Donor 791 (9+9+19) (n=117)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Size of contractions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

* Data taken from table 1 considering changes of +1 to −1 repeat units as no change.

Discussion

Instability of trinucleotide repeats leading to disease causing repeat expansions remains a poorly understood process. Currently, 11 disorders resulting from trinucleotide repeat expansion have been identified in humans. 28, 29 Fragile X syndrome, one of the first such mutations discovered, is among the best understood of these disorders although few clues exist to explain the mechanisms behind the expansion process. One clue relates to the number and position of AGG interruptions which define tracts of perfect CGG repeats. Kunz and Warren 30 found that the most prevalent haplotypes among fragile X chromosomes are those which are commonly found on normal chromosomes with fewer interrupted FMR1 repeats and therefore longer perfect CGG tracts. Moreover, the long perfect repeats were found at the 3' portion of the array indicating that variation at the FMR1 repeat was highly polar. It was suggested that perfect CGG tracts in excess of 24 triplets may be predisposed towards expansion. Indeed, Eichler et al. 28 found that when the tract of perfect repeats exceeds 34 triplets, instability is frequent enough to observe in small nuclear families. Thus, AGG interruptions, particularly two or more, are believed to stabilise the FMR1 repeat. Loss of one or more AGG triplets, while lengthening the 3' tract of perfect CGG repeats, may enhance instability and result in alleles which may be predisposed towards fragile X syndrome. 18-21 The notion that triplet interruption confers stability has been independently observed with the ataxin gene; the
unstable expansion of a perfect CAG trinucleotide repeat leads to spinocerebellar ataxia type 1 but the normal, stable alleles are interrupted by CAT triplets.26

Sperm analysis has begun to provide insight into trinucleotide repeat instability. Zhang et al.27 examined the CAG repeat of the androgen receptor (AR) gene in sorted sperm and found a fourfold greater mutation frequency for sperm from donors with 28-31 perfect repeats compared to those donors with only 20-22 repeats. Monckton et al.28 showed, by small pool PCR, that sperm from myotonic dystrophy males with >80 perfect CTG repeats showed a very high mutation frequency with expansions as well as contractions. Similarly, studies with sorted sperm from a patient with Kennedy's disease, with 47 perfect CAG repeats at the AR locus, showed striking instability with contractions as well as more frequent expansions of the repeat tract.29 Similarly, sorted sperm analysis of the CAG repeat of the Huntington's disease (HD) gene showed a clear increase in instability in sperm of donors with increasing repeat length.30 While alleles with only 15-18 perfect repeats had very low sperm mutation frequencies (~0.6%), a male with 30 perfect repeats had a sperm mutation frequency of 11% and a male with 36 repeats had a mutation frequency of 53%. These studies all point to the importance of tracts of lengthy perfect repeats, greater than approximately 24 triplets, in the instability process. Indeed, for fragile X syndrome, Mornet et al.31 found an increased rate of instability in sperm DNA of a male with 55 repeats (longest perfect tract of 36 triplets) compared to sperm DNA of a male with 29 repeats (longest perfect tract of nine triplets).32 Although these data support the trend of increased instability with longer perfect repeats, the mutation frequencies were considerably less than those using sorted sperm since the method of small pool PCR cannot reliably distinguish changes of <4 repeats.

The data reported above, while consistent with those of Mornet et al.,33 are distinct on several levels. Sorted sperm were used, which allowed small changes of repeat units to be distinguished. This, in turn, leads to mutation frequencies higher than those of Mornet et al.33 but more similar to those of Leeffang et al.34 using sorted sperm analysis at the HD locus (11% change of 30 perfect HD repeats compared here with 15% change of 29 perfect FMR1 repeats). Variant sperm in our study were subjected to DNA sequencing to verify the exact repeat length and structural variation incurred. The choice of two sperm donors with identical overall repeat lengths but distinct in terms of their longest perfect repeat tract allowed direct examination of the influence of AGG interspersions on FMR1 stability.

Examination of a total of 1041 wells within which individual sperm were sorted resulted in PCR products of the FMR1 repeat in 214 wells from donor 727 and 117 wells from donor 791. Both donors, each with 39 total repeat arrays, showed an overall mutation frequency of ~15%. However, donor 727, who carried 29 perfect CGG repeats owing to a single AGG interspersion, had more expansions and contractions than donor 791 with only 19 perfect repeats owing to two AGG interspersions. This was more evident when changes of ≥2 repeat units were considered. Moreover, when analysed in this way, only donor 727 showed mutations of the repeat array. Although simple slippage mechanisms could account for the observed loss of single 5’ CGG triplets from both 727 and 791 and the gain or loss of CGG repeat units from the 3’ uninterrupted portion of their repeats, it is difficult to model the mechanism by which the 10+10 and 10+9 alleles were generated. However, such complex contractions of the CGG repeat, although rare (<2% of the sperm analysed), could contribute to the mode of CGG repeat alleles observed ≥20 repeats. Greater than 90% of all variations observed in both donors’ sperm involved the lengthening or contraction of the 3’ portion of the perfect repeat array without alteration of AGG localisation. These data support the idea that AGG interspersions promote stability of the FMR1 repeat and that variation of the normal repeat involves polarity. It is also possible that variation among people in repeat instability owing to trans factors could account for this difference. However, in conjunction with previous studies using both sorted sperm and population data at a number of loci involving trinucleotide repeat expansion, these data point to the common importance of perfect trinucleotide repeat tracts of greater than approximately 24 triplets influencing repeat instability.

STW is an investigator and CBK is an associate of the Howard Hughes Medical Institute. EPL is a postdoctoral fellow of the Hereditary Disease Foundation. This work was supported in part by NIH grants R37 GM36745 (NA) and R37 HD20521 (STW).

The effect of FMR1 CGG repeat interruptions on mutation frequency

17 Richards RJ, Sutherland GR. Simple repeat DNA is not replicated simply. Nat Genet 1994;6:114-16.