EUR Research Information Portal

Instability of a (CGG)98 repeat in the Fmr1 promoter

Published in:
Human Molecular Genetics

Publication status and date:
Published: 01/01/2001

Document Version
Publisher's PDF, also known as Version of record

Citation for the published version (APA):
Bontekoe, CJM., Bakker, CE., Bakker, I., van der Linde, H., Lans, H., de Lange, D., Hirst, MC., & Oostra, B. (2001). Instability of a (CGG)98 repeat in the Fmr1 promoter. *Human Molecular Genetics*, *10*(16), 1693-1699. http://hdl.handle.net/1765/9699

Link to publication on the EUR Research Information Portal

Terms and Conditions of Use
Except as permitted by the applicable copyright law, you may not reproduce or make this material available to any third party without the prior written permission from the copyright holder(s). Copyright law allows the following uses of this material without prior permission:

- you may download, save and print a copy of this material for your personal use only;
- you may share the EUR portal link to this material.

In case the material is published with an open access license (e.g. a Creative Commons (CC) license), other uses may be allowed. Please check the terms and conditions of the specific license.

Take-down policy
If you believe that this material infringes your copyright and/or any other intellectual property rights, you may request its removal by contacting us at the following email address: openaccess.library@eur.nl. Please provide us with all the relevant information, including the reasons why you believe any of your rights have been infringed. In case of a legitimate complaint, we will make the material inaccessible and/or remove it from the website.

Erasmus University Rotterdam
Making Minds Matter

© 2001 Oxford University Press
Human Molecular Genetics, 2001, Vol. 10, No. 16 **1693–1699**

Instability of a $(CGG)_{98}$ repeat in the *Fmr1* promoter

C. J. M. Bontekoe, C. E. Bakker, I. M. Nieuwenhuizen, H. van der Linde, H. Lans, D. de Lange, M. C. Hirst[1] and B. A. Oostra*

CBG Department of Clinical Genetics, Erasmus University Rotterdam, PO Box 1738, 3000DR Rotterdam, The Netherlands and [1]Department of Biological Sciences, The Open University, Milton Keynes, UK

Received April 14, 2001; Revised and Accepted June 7, 2001

Fragile X syndrome is one of 14 trinucleotide repeat diseases. It arises due to expansion of a CGG repeat which is present in the 5′-untranslated region of the *FMR1* gene, disruption of which leads to mental retardation. The mechanisms involved in trinucleotide repeat expansion are poorly understood and to date, transgenic mouse models containing transgenic expanded CGG repeats have failed to reproduce the instability seen in humans. As both *cis*-acting factors and the genomic context of the CGG repeat are thought to play a role in expansion, we have now generated a knock-in mouse *Fmr1* gene in which the murine $(CGG)_8$ repeat has been exchanged with a human $(CGG)_{98}$ repeat. Unlike other CGG transgenic models, this model shows moderate CGG repeat instability upon both in maternal and paternal transmission. This model will now enable us to study the timing and the mechanism of repeat expansion in mice.

INTRODUCTION

The fragile X syndrome is one of over 14 human diseases associated with expanded trinucleotide repeats. Although much has been elucidated about the genetics of the trinucleotide repeat diseases, little is known about the mechanism(s) that cause the repeat instability. The highly polymorphic CGG trinucleotide repeat which is located in the 5′-untranslated region (5′-UTR) of the fragile X mental retardation gene (*FMR1*), is associated with the disease phenotype when the allele carries more than 200 triplets (1–3). Such alleles undergo methylation which extends across both the CpG island promoter region and the CGG repeat itself. This methylation blocks transcription of the gene which is normally translated into the fragile X mental retardation protein (FMRP) (4,5). This absence of FMRP results in the fragile X phenotype. The main characteristics of the fragile X syndrome are mental retardation and macroorchidism (6), with adult male mental retardation ranging from profound to borderline, with an average IQ in the moderate range. Macroorchidism is a common finding in post-pubescent affected males. With an incidence of 1:4000 males and 1:6000 females, this X-linked disorder is the most common form of inherited mental retardation (7,8).

FMR1 CGG repeat alleles can be divided into three groups: normal, premutation and full mutation alleles. Normal alleles range between 5 and 50 triplets. These alleles are stable upon transmission between generations. Premutations alleles, of between 50 to 200 triplets (9,10), are unstable upon transmission between generations with both expansions and contractions occurring. Since they allow FMRP expression (5,11) they do not result in the development of the fragile X phenotype, but they are prone to expand to full mutations of more than 200 triplets in the next generation. As described above, full mutation alleles with over 200 triplets undergo methylation and result in the absence of FMRP and thus the fragile X phenotype (5,11). These full mutations only arise upon transmission through the female germline, and males never transmit a full mutation to their daughters.

The exact timing of the repeat expansion is still under debate. Repeat expansions must occur during meiosis or early embryonic development (12,13). The most accepted model assumes that full mutations are already present in the oocyte and, thus, all cells in the resulting embryo will also have a full mutation. One, or several, mitotic contraction events to a premutation-length repeat could explain the mosaic pattern which is quite often observed in fragile X patients. The observation that oocytes of full mutation female fetuses also carry full mutation alleles appears to confirm this model, although it cannot be ruled out that the expansion from a premutation to a full mutation occurs during early embryogenesis (12,13). The repeat length(s) present in oocytes of premutation females is not known. In the male germline some kind of selection mechanism has to be assumed (14), as patients with a full mutation have only premutation alleles in their sperm. This mechanism appears, therefore, to protect the male germline against transmission of full mutations. The basis of such a selection mechanism is not known.

Examination of the mode of inheritance of premutations in fragile X families has shown that the risk of expansion to full mutation depends upon the size of the premutation. Small premutations give rise to both expansions and contractions, whereas premutations of over 90 triplets almost always expand to a full mutation in the next generation. The risk of this expansion to a full mutation increases with the length of the CGG repeat (15). This variation in risk accounts for the Sherman paradox (10,16). A more detailed insight into the repeat length and its behaviour upon transmission was gained by sequencing a large number normal and premutation alleles. It was found that most normal alleles are interspersed with

*To whom correspondence should be addressed. Tel: +31 10 4087198; Fax: +31 10 4089489; Email: oostra@kgen.fgg.eur.nl

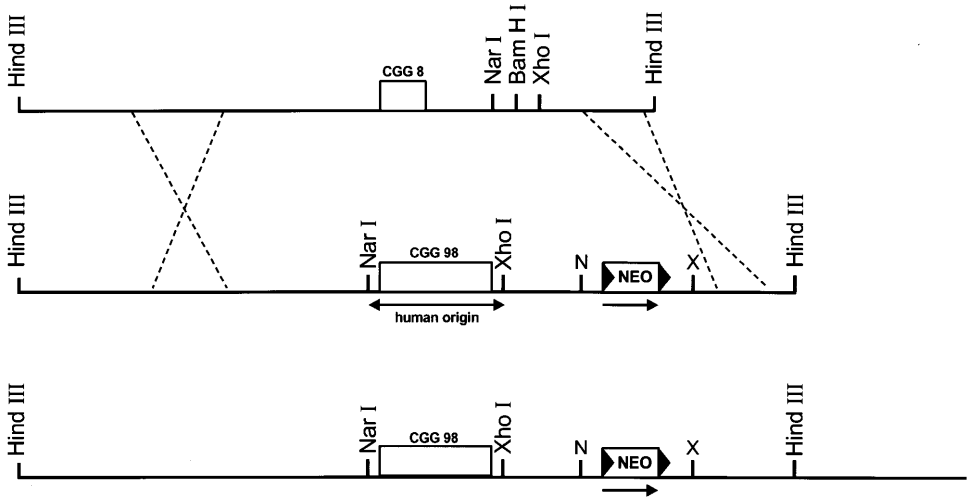

Figure 1. The targeting construct pCB66. The mouse endogenous *Fmr1* $(CGG)_8$ is exchanged for a $(CGG)_{98}$ repeat of human origin. X, removed *Xho*I site, N, removed *Nar*I site.

AGG triplets (17). The most common alleles of between 29 and 32 triplets are interspersed with two AGG interruptions. These AGG interruptions are normally found downstream of shorter tracts of 9 or 10 CGG triplets. In premutations, fewer AGG interruptions are present compared to normal alleles, variation in repeat length is polar and instability occurs always at the 3′ end of the repeat; the region where no or fewer AGG interruptions are present (17,18).

The most 3′ uninterrupted CGG tract appears to be the most important element of the CGG repeat; a pure CGG tract of more than 34–38 CGG triplets is enough to cause instability (19). Most premutation alleles contain only one or two AGG repeats, and the 3′ CGG tract is greater than 35 CGG triplets. Since the longest pure CGG tract is always found at the 3′ end of the CGG repeat and this is also the region where expansion occurs in fragile X families, this might give some insight into the mechanism of instability.

To study the timing and the mechanism of the CGG repeat expansion observed in the fragile X syndrome, it is, for obvious reasons, necessary to have an animal model. With such a model it would be possible to study the behaviour of the repeat through both the female and male germline and during (early) embryonic development. We have reported previously that a $[(CGG)_{11}AGG(CGG)_{60}CAG(CGG)_8]$ repeat was transmitted without any detectable length change through several generations of transgenic mice (20). Similar results were obtained in mice with a $[(CGG)_{22}TGG(CGG)_{43}TGG(CGG)_{21}]$ tract and with tracts having a pure 3′ CGG tract up to 97 triplets (21,22). Several hypotheses have been put forward to explain the stability of these repeats in mice. Firstly, the interruptions present in the first two transgenes discussed above may be acting as stabilizers. However, in the mice carrying a 97 repeat CGG tract at the 3′ end, other factors might be involved. Amongst these, the chromosomal or genomic context might be an important factor.

The *FMR1* gene is highly conserved among vertebrates. The murine homologue, *Fmr1,* is 97% identical in amino acid sequence to the human gene and exhibits an expression pattern very similar to that observed in humans (23,24).This homology extends to the repeat region and also across the promoter. We therefore generated a mouse model in which the endogenous mouse CGG repeat was replaced by a human CGG repeat carrying 98 CGG triplets. This was done using a homologous recombination strategy with a mouse promoter construct where the mouse endogenous $(CGG)_8$ repeat was exchanged for a $(CGG)_{98}$ repeat of human origin. We describe the first generation of such a 'knock-in' CGG triplet mouse and report the behaviour of the premutation allele in the *Fmr1* gene.

RESULTS

Construct *FMR1* promoter region

The mouse homologue of human *FMR1* was isolated from an E14 ES cell phage library (kindly provided by D.Meijer, Department of Genetics, Erasmus University, Rotterdam). A *neo* cassette, flanked by *lox*P sites, was cloned in the *Bam*HI site. The endogenous $(CGG)_8$ present in the mouse promoter was replaced by a $(CGG)_{98}$ of human origin. To allow cloning of the human CGG repeat minimal changes were made to the mouse promoter region (Fig. 1). Cloning of the CGG repeat was the last cloning step, because deletions were often found in plasmids containing the expanded CGG repeat.

Cloning of the mouse promoter region revealed that the mouse promoter region cloned into a vector and propagated in bacteria was, in itself, prone to deletions. Even in the absence of the CGG triplet repeat, deletions were often observed after a simple digestion, re-ligation and transformation into bacteria (different strains were used but in most cases we used the strain *E.coli* DH5α). The frequency of these deletions appeared to be dependent on the restriction enzyme being used or the localization of the restriction sites in the construct, as well as the ligation buffer used (data not shown). Since for most DNA constructs digestion, re-ligation and transformation is a very straightforward experiment, this could indicate that *cis*-acting factors important for instability might be present in the mouse promoter region. Comparison of the promoter sequence of *FMR1* and *Fmr1* showed that all the identified regulatory

elements were conserved. Whether or not all these elements are functional is not known.

Generation of the mice

Homologous recombinants were recognized by the absence of the endogenous mouse $(CGG)_8$ allele, and the presence of the expanded $(CGG)_{98}$ allele. Of the 1200 screened ES cell clones, 13 were identified as possible homologous recombinants. The length of the observed expanded CGG repeat differed between the clones. These different lengths most likely represent deletions of the repeat and/or flanking sequences in the plasmid DNA.

Clone 651 contained exactly the same length as the cloned $(CGG)_{98}$ repeat in the plasmid. This clone was therefore selected for blastocyst injection from which three chimeras with the same repeat length were identified. These animals were crossed onto an FVB background and the repeat length was determined in the next generation. The first heterozygous females were crossed either with a wild-type male or a male expressing *Cre*-recombinase. Expression of *Cre*-recombinase allows deletion of the *neo* cassette present between the *lox*P sites, minimizing the changes to the mouse genomic flanking sequence in the resultant transgenic animal. $(CGG)_{98}$ mice with [$(CGG)_{98}$/*neo*] and without [$(CGG)_{98}$/*neo*$^-$] the inserted *neo* marker were subsequently crossed and the offspring examined for repeat length.

In total, 155 $(CGG)_{98}$ transmissions were studied. $(CGG)_{98}$/*neo* was transmitted 34 times, 15 male transmissions and 19 female transmissions. $(CGG)_{98}$/*neo*$^-$ was transmitted 121 times, 80 maternal transmissions and 41 paternal transmissions. Offspring homozygous for the expanded repeat were included for paternal (13) and maternal (13) transmission. In total, 15 instabilities were found using a radioactive PCR assay to detect CGG length (Fig. 2). Confirmation and exact sizing of the repeat was carried out using the CGG expansion method of Perkin-Elmer. An example is shown in Figure 3 and a summary is presented in Table 1. In total, we have observed 15 alterations in length of the CGG repeat; two contractions and 13 expansions. Both contractions (of 11 triplets) were observed after (different) paternal transmission, but male transmission also led to six expansions. Five expansions were observed in transmission via the female germline. Two instabilities were observed in homozygous $(CGG)_{98}$ female offspring and in these cases the origin of the instability could be either paternal or maternal. The shortest CGG repeat present in the descendent mice is 87 CGGs and after passages through several generations the longest repeat is 108 CGGs. While the number of paternal changes (at least eight) outweigh maternal ones (at least five), the number of maternal transmissions (at least 86) appears to be greater than paternal ones (at least 43). Thus, the rate of change in paternal transmission of unstable repeats would appear to be 3-fold that of maternal transmission.

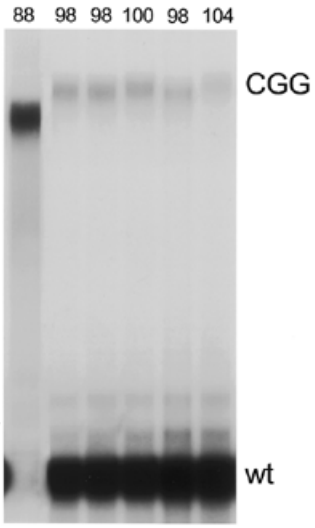

Figure 2. CGG repeat instabilities detected by radioactive PCR. On top of the lanes are the sizes as determined with the Fragile X polymorphism assay.

DISCUSSION

The mechanism underlying fragile X repeat instability in humans is still unknown, although many pathways and intermediates have been implicated. It has been postulated that during DNA replication, slippage synthesis can occur within the repeat due to the repetitive nature of the sequence. Through this and the formation of single-stranded breaks, both expansion and contraction can occur (25,26). It has also been suggested that secondary structures such as hairpins and tetraplex DNA might also play a role in repeat expansion, but the existence of these remains to be established *in vivo* (27). From *in vitro* studies, as well as studies in bacteria and yeast, it is also clear that repeat stability is dependent on the length of the repeat and the number of AGG interruptions, reminiscent of the situation in fragile X families (28,29). Furthermore, for yeast and bacteria it has been observed that host cell genotypes, the orientation of the repeat with respect to DNA replication and the position of the cloned repeat within the vector can all influence instability (30).

In humans carrying expanded CGG repeats, their length increases over generations. Once beyond a specific threshold the repeat becomes dramatically unstable, expanding rapidly up to a few thousand triplets. This phenomenon has given rise to the term 'dynamic mutation' (31). Although for some trinucleotide repeat mouse models small expansions or contractions have been observed, dynamic mutations such as those in human disease have never been observed. Before this study, transgenic mouse models have carried non-targeted autosomal CGG repeats and all have failed to show any evidence of instability (20–22). The absence of important *cis*-acting factors and the random integration of the transgene on one of the autosomes instead of the X chromosome are both possible explanations for the relative stability of the CGG repeats in the earlier mouse models (20).

To circumvent this, we generated a mouse with an expanded CGG repeat within the endogenous *Fmr1* promoter by using a homologous recombination technique. In this way, the behaviour of the expanded CGG repeat can be studied in its endogenous genomic localization. In total, 155 transmissions of the expanded $(CGG)_{98}$ allele were studied and 15 instabilities were observed; two contraction events and 13 expansions. As detection was performed using a radioactive PCR technique (where small changes are difficult to detect) this equates to a rate of instability of at least 10%. Both contractions occurred

Table 1. Total numbers of male and female transmissions for both $(CGG)_{98}$/*neo* and $(CGG)_{98}$/*neo*$^-$

	Female/male transmissions	Instabilities
$(CGG)_{98}$/*neo*	Female 19	0
$(CGG)_{98}$/*neo*	Male 15	4 (+2, +6, +4, +4)
$(CGG)_{98}$/*neo*$^-$	Female 67	5 (+2, +2, +1, +1, +1)
$(CGG)_{98}$/*neo*$^-$	Male 28	4 (–11, –11, +5, +3)
$(CGG)_{98}$/*neo*$^-$	Female/Male 13/13	2 (+2, +3)

Most instabilities are found upon male transmission. Only for a limited number the size of the instabilities is determined. Length changes are depicted in parentheses.

by paternal transmission, as did six of the expansions. Two expansions could be of either paternal or maternal origin. Given the total number of maternal and paternal transmissions, these results are surprising. Although the numbers are still limited, there appears to be a tendency of higher instability (both expansion and contraction) upon paternal transmission. In human fragile X premutations, small changes in repeat length of this magnitude are observed upon both male and female transmission. In contrast, dynamic mutations in humans are only observed upon female transmission. However, in this $(CGG)_{98}$ mouse model, no dynamic mutations were observed. In contrast to fragile X syndrome, most other trinucleotide repeat disorders show preferential instability upon male transmission. For one transgenic mouse model containing an expanded CAG repeat, it was found that the gender of the offspring determined the degree of instability (32). Whether the gender of the offspring in our CGG repeat mouse model plays a role in instability could not be determined.

The human and mouse FMR1 gene promoter regions are very homologous. With the human promoter, four 'footprints' have been identified reflecting positions of various DNA–protein interactions. These footprints correspond to consensus binding sites of various transcription factors and are absent in fragile X patients, indicating that they reflect functional regulatory elements (33). These regulatory elements are also present in the mouse promoter region. It is not known whether these sequences in the promoter region might be influencing the behaviour of the repeat instability. Despite the overall homology between the mouse and human promoters, it does not necessarily mean that DNA elements which influence repeat instability are conserved between mouse and human. Theoretically, there is still a possibility that the mouse promoter region does not contain the *cis*-acting elements involved in (large/dynamic) repeat instability which is seen in the human *FMR1* gene.

Interestingly, cloning of the mouse *Fmr1* promoter showed that this region was prone to deletions when maintained in plasmids in bacteria. These deletions occurred after digestion of the DNA, re-ligation and transformation into bacteria. Most, although not all, deletions occurred in the region directly down and upstream of the $(CGG)_8$ repeat. These results indicate that the *Fmr1* promoter region itself is unstable in bacteria. The reason for this instability might be the high GC content of the region flanking the CGG repeat. Although there is no direct evidence that a DNA region which is highly unstable in bacteria might influence instability in the mouse genome, this may well contribute to the instability of the CGG repeat, although we hypothesize that factors other than *cis*-acting factors alone play a role in determining instability.

Length variation found in fragile X alleles appears to be polar, instability always occurring at the 3′ end. This might be influenced by the direction of DNA replication through the repeat (29,30). The direction of replication is important in determining repeat instability in both *Escherichia coli* and yeast. The position of the origin of replication in artificial DNA constructs determines whether the 5′-CGG-3′ or the 5′-CCG-3′ strand is the leading or the lagging strand during DNA synthesis. For the mouse and human genomic *FMR1* loci the direction of replication is not known. It may well be that the direction of replication in the human situation might favour the occurrence of expansions, but that the situation might be different for the mouse *Fmr1* locus.

Flap endonuclease 1 (FEN1), a protein involved in DNA replication and long-patch base excision repair, is thought to play a role in trinucleotide repeat instability (34). In yeast, loss of flap endonuclease activity (*rad27* mutants) increases instabilities throughout the whole genome, including trinucleotide repeats (35–37). For normal and premutation size CGG repeats, a 10-fold elevated frequency of expansion in *rad27* yeast strain has been found (38). This suggests that FEN1 could play a role in CGG trinucleotide repeat instability.

The endogenous mouse *Fmr1* 5′-UTR contains a CGG repeat of between 8 and12 triplets. In the ES cells used in this study, the $(CGG)_8$ repeat was exchanged with a $(CGG)_{98}$ repeat in the expectation that this length of repeat exceeds a threshold of instability in the mouse. In humans, repeat instability occurs when the number of CGG triplets is greater than 50; the threshold for repeat instability in mice is as yet unknown. In mice containing an expanded Huntington CAG repeat, it has been found that the rate of instability is less than that observed for similar sized repeats in humans (39). Mice heterozygous for the CAG expansion show intergenerational repeat instability (+2 to –6) at a much higher frequency in maternal transmission than in paternal transmission. The majority of changes transmitted through the female germline were small contractions, as in humans, whereas small expansions occurred more frequently in paternal transmission. (40). The mouse Hdh gene with a knock-in of 90 and 109 units produced a graded increase in the mutation frequency to >70%, with instability being more evident in female transmissions. No large jumps in CAG length were detected in either male or female transmissions. Instead, size changes were modest increases and decreases, with expansions typically emanating from males and contractions from females. (41). For the Fmr1 mouse model described here, the rate of change in paternal transmission of unstable repeats would appear to be 3-fold that of maternal transmission. Extrapolation of the data obtained from the expanded trinucleotide mice suggests that, if there is a threshold for instability in mice, the threshold might be higher in mice than for humans.

Results of this CGG repeat expanded mouse model, together with the studies of other trinucleotide repeats, suggest that mouse might not be a perfect model in which to study repeat instability. This might be due to the fact that the mechanism involved in repeat instability in humans might be absent in mice, or perhaps that the environment leading to the instability

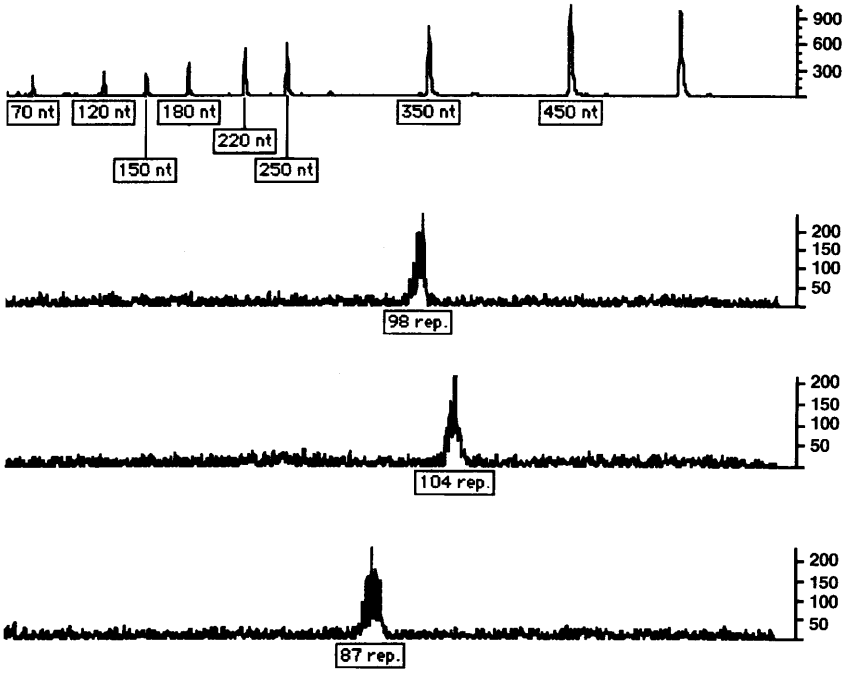

Figure 3. The Fragile X polymorphism assay is used to determine the exact repeat length of the observed instabilities.

in humans might be absent in mice. The mechanisms which have been proposed to play a role in repeat instability involve meiosis as well as DNA replication, recombination and repair. During all these processes, duplication of the DNA occurs, allowing an opportunity for DNA mutations and secondary structures like hairpins or tetraplex to form, errors which must be removed by proof-reading and repair processes. These processes, which aim to keep mistakes in the copying of DNA to a minimum, are balanced by processes in meiosis such as recombination which generate maximum genetic variation for a species. Although these processes are known to occur in humans as well as in mice, the balance between them might be different between the two species, giving rise to repeat instability in humans, but not in mice.

The involvement of DNA repair processes in repeat instability might be studied by crossing mice with expanded CGG repeats with mice deficient for different repair pathways. In yeast and *E.coli* it has been shown that the rate of instability changed in certain repair-deficient strains (38,42,43). Crossings between trinucleotide repeat mice and mice with certain repair deficiencies might give more insight into repeat instability in mice as well as in humans. Furthermore, it might be worthwhile to study the timing of the small repeat instabilities that occur as this might give more insight in the difference observed between male and female transmission.

Our transgenic $(CGG)_{98}$ repeat might also eventually also be used to study inactivation of FMRP expression. In the $(CGG)_{98}$ mice described in this study, FMRP expression was present, suggesting that the promoter region was not inactivated (data not shown). As small expansions are observed, it might be possible that expansion over several generations will generate a larger expanded allele which will inactivate the FMR1 gene and lead to loss of FMRP expression. Subsequently, both repeat expansion and FMRP expression could be studied in descendent generations. It has been reported recently that premutation alleles result in elevated mRNA levels (44) in human fragile X carriers. In contrast to this, FMRP expression in lymphocytes of premutation carriers with over 100 CGG triplets was reduced. This mouse model might, therefore, also be used to study the mRNA levels in premutation alleles, the translation of Fmr1 mRNA containing expanded CGG repeats, and maybe the mechanistic switch, which occurs when the *Fmr1* gene is silenced.

MATERIALS AND METHODS

Construction of pCB66 and ES cell electroporation

The construct pCB66 contained a 7 kb *Hin*dIII fragment of the mouse *Fmr1* promoter region inserted in a pBR322 derived vector (pBR322 – *Bam*HI 375–*Nar*I 1205). In this fragment, a *neo* cassette flanked by *lox*P sites was cloned in the *Bam*HI site present in intron 1 (Fig. 1). The mouse $(CGG)_8$ repeat was exchanged for a $(CGG)_{98}$ repeat of human origin. The human-derived $(CGG)_{98}$ repeat was isolated as an *Sfo*I–*Xho*I fragment from a cloned expanded CGG repeat. This cloned repeat was isolated from a yeast strain carrying a transgenic human premutation as described in references 28 and 38. To clone the expanded human CGG repeat in the murine *Fmr1* promoter, minimal changes were made to the flanking sequence. These changes involved a point mutation to abolish an *Nar*I site in intron 1 and the substitution of 4 bp (TCGA) to abolish an *Xho*I site in intron 1. The *Xho*I site flanking the CGG repeat in the human situation was not present in the mouse promoter. To facilitate the cloning of the expanded $(CGG)_{98}$ repeat an *Xho*I site was generated by an A→G transition based on the

homology between the human and the mouse promoter. The integrity of the cloned fragment was determined by sequencing and restriction enzyme digestion. Cloning was performed using standard procedures.

For electroporation of E14 ES cells to allow homologous recombination plasmid, pCB66 was linearized by an *Xba*I digest. After purification, linearized plasmid DNA was used to transfect ES cells. Electroporation was performed with 10^7 ES cells in 400 µl PBS using a Progenetor II Gene Pulser (1200 µF and 117 V during 10 ms). Using double selection with G418 (200 µg/ml) and Fiau (2 µM) the cells were cultured to allow colony forming. Colonies were picked and cultured separately. for DNA isolation.

DNA analysis

ES cell clones were grown to confluency in a 24-well plate. The medium was removed and 300 µl (10 mM Tris–HCl, 400 mM NaCl, 2 mM EDTA pH 7.3–7.4) and 30 µl of 10 mg/ml Prot K was added to lyse the cells during overnight incubation at 55°C. An aliquot of 150 µl of 6 M NaCl was added and the suspension was centrifuged. To the supernatant, 2 vol of 96% ethanol were added to precipitate the DNA. DNA was dissolved in 50 µl H_2O. For radioactive PCR, 1 µ1 DNA was used.

Radioactive PCR was performed to determine the repeat length in the ES clones. Primers C (5′-GCTCAGCTCCGTTT-CGGTTTCACTTCCGGT-3′) and F (5′-AGCCCCGCACTT-CCACCACCAGCTCCTCCA-3′) were used. PCR conditions were as described by Deelen *et al.* (45). PCR products were run in a 6 % denaturing polyacrylamide gel.

Generation of knockout mice

ES clone 651 was used for injection into C57/BL6J blastocysts. These blastocysts were transferred to pseudopregnant female mice. Three chimeras were generated and crossed with wild-type FVB and C57/Bl6J females. Female offspring of these chimeras was tested for the presence of the expanded $(CGG)_{98}$ repeat. Mice containing the expanded CGG repeat were crossed with wild-type mice as well as with $(CGG)_{98}$ littermates. Repeat instability upon transmission to offspring was tested by radioactive PCR.

The *neo* cassette inserted into the *Bam*HI was flanked by *lox*P sites. Because the presence of the *neo* cassette might disturb the natural environment of the CGG repeat, knock-in mice were crossed with mice expressing *Cre* recombinase. In this way the changes made to the *Fmr1* gene were kept minimal. These $(CGG)_{98}$/*neo*$^-$ mice were also crossed with wild-type mice, and with $(CGG)_{98}$/*neo*$^-$ littermates.

Fragile X size polymorphism assay

Radioactive PCR as described above is informative to determine whether instability occurs, but it does not give the exact length changes observed. Also, small changes ± 1 CGG triplet might be missed. The fragile X size polymorphism assay (Perkin Elmer Biosystems) allows us to determine the exact length of the CGG repeat. This test was used to determine the exact size changes. PCR conditions were as described by the manufacturer. PCR samples were analysed using an ABI377 sequencer (PE Biosystems).

ACKNOWLEDGEMENTS

We would like to thank T.de Wit for the generous gift of the mouse transgenic for *Cre*-recombinase, D.Meijer for the mouse ES cell phage library and R.van Haperen for blastocyst injections. We also thank Perkin Elmer for their gift of their fragile X size polymorphism assay. This study was supported by the Netherlands Organization for Scientific Research (N.W.O.) and the National Institute of Health (NIH 5 RO1 HD38038).

REFERENCES

1. Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F. and Mandel, J.L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, **252**, 1097–1102.
2. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P. *et al.* (1991) 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.
3. Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J.C., Warren, S.T., Schlessinger, D. *et al.* (1991) Fragile X genotype characterized by an unstable region of DNA. *Science*, **252**, 1179–1181.
4. Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T. and Nelson, D.L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, **66**, 817–822.
5. Verheij, C., Bakker, C.E., de Graaff, E., Keulemans, J., Willemsen, R., Verkerk, A.J., Galjaard, H., Reuser, A.J., Hoogeveen, A.T. and Oostra, B.A. (1993) Characterization and localization of the FMR-1 gene product associated with fragile X syndrome. *Nature*, **363**, 722–724.
6. Hagerman, R.J. (1996) Physical and behavioral phenotype. In Agri, J., Hagerman, R.J. and Silverman, A.C. (eds), *Fragile X Syndrome: Diagnosis, Treatment and Research*. The John Hopkins University Press, Baltimore and London, pp. 3–87.
7. Turner, G., Webb, T., Wake, S. and Robinson, H. (1996) The prevalence of the fragile X syndrome. *Am. J. Med. Genet.*, **64**, 196–197.
8. De Vries, B.B., van den Ouweland, A.M., Mohkamsing, S., Duivenvoorden, H.J., Mol, E., Gelsema, K., van Rijn, M., Halley, D.J., Sandkuijl, L.A., Oostra, B.A., Tibben, A. and Niermeijer, M.F. (1997) Screening and diagnosis for the fragile X syndrome among the mentally retarded: an epidemiological and psychological survey. Collaborative Fragile X Study Group. *Am. J. Hum. Genet.*, **61**, 660–667.
9. Rousseau, F., Heitz, D., Biancalana, V., Blumenfeld, S., Kretz, C., Boue, J., Tommerup, N., Van Der Hagen, C., DeLozier-Blanchet, C., Croquette, M.F. *et al.* (1991) Direct diagnosis by DNA analysis of the fragile X syndrome of mental retardation. *N. Engl. J. Med.*, **325**, 1673–1681.
10. Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.,Jr, Warren, S.T. *et al.* (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.
11. Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.P. and Mandel, J.L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat. Genet.*, **4**, 335–340.
12. Moutou, C., Vincent, M.C., Biancalana, V. and Mandel, J.L. (1997) Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic. *Hum. Mol. Genet.*, **6**, 971–979.
13. Malter, H.E., Iber, J.C., Willemsen, R., De Graaff, E., Tarleton, J.C., Leisti, J., Warren, S.T. and Oostra, B.A. (1997) Characterization of the full fragile X syndrome mutation in fetal gametes. *Nat. Genet.*, **15**, 165–169.
14. Reyniers, E., Vits, L., De Boulle, K., Van Roy, B., Van Velzen, D., de Graaff, E., Verkerk, A.J.M.H., Jorens, H.Z., Darby, J.K., Oostra, B.A. and Willems, P.J. (1993) The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. *Nat. Genet.*, **4**, 143–146.
15. Bat, O., Kimmel, M. and Axelrod, D.E. (1997) Computer simulation of expansions of DNA triplet repeats in the fragile X syndrome and Huntington's disease. *J. Theor. Biol.*, **188**, 53–67.
16. Sherman, S.L., Jacobs, P.A., Morton, N.E., Froster-Iskenius, U., Howard-Peebles, P.N., Nielsen, K.B., Partington, N.W., Sutherland, G.R., Turner, G. and Watson, M. (1985) Further segregation of the fragile X

syndrome with special reference to transmitting males. *Hum. Genet.*, **69**, 289–299.
17. Kunst, C.B. and Warren, S.T. (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell*, **77**, 853–861.
18. Eichler, E.E., Hammond, H.A., Macpherson, J.N., Ward, P.A. and Nelson, D.L. (1995) Population survey of the human FMR1 CGG repeat substructure suggests biased polarity for the loss of AGG interruption. *Hum. Mol. Genet.*, **4**, 2199–2208.
19. Eichler, E.E., Holden, J., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A. and Nelson, D.L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.*, **8**, 88–94.
20. Bontekoe, C.J.M., de Graaff, E., Nieuwenhuizen, I.M., Willemsen, R. and Oostra, B.A. (1997) FMR1 premutation allele is stable in mice. *Eur. J. Hum. Genet.*, **5**, 293–298.
21. Lavedan, C.N., Garrett, L. and Nussbaum, R.L. (1997) Trinucleotide repeats (CGG)22TGG(CGG)43TGG(CGG)21 from the fragile X gene remain stable in transgenic mice. *Hum. Genet.*, **100**, 407–414.
22. Lavedan, C., Grabczyk, E., Usdin, K. and Nussbaum, R.L. (1998) Long uninterrupted CGG repeats within the first exon of the human FMR1 gene are not intrinsically unstable in transgenic mice. *Genomics*, **50**, 229–240.
23. Ashley, C.T., Sutcliffe, J.S., Kunst, C.B., Leiner, H.A., Eichler, E.E., Nelson, D.L. and Warren, S.T. (1993) Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nat. Genet.*, **4**, 244–251.
24. Bakker, C.E., de Diego Otero, Y., Bontekoe, C., Raghoe, P., Luteijn, T., Hoogeveen, A.T., Oostra, B.A. and Willemsen, R. (2000) Immunocytochemical and biochemical characterization of FMRP, FXR1P, and FXR2P in the mouse. *Exp. Cell Res.*, **258**, 162–170.
25. Gacy, A.M., Goellner, G., Juranic, N., Macura, S. and McMurray, C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*. *Cell*, **81**, 533–540.
26. Nadel, Y., Weismanshomer, P. and Fry, M. (1995) The fragile X syndrome single strand d(CGG)(n) nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.*, **270**, 28970–28977.
27. Usdin, K. (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Res.*, **26**, 4078–4085.
28. Weisman-Shomer, P., Cohen, E. and Fry, M. (2000) Interruption of the fragile X syndrome expanded sequence d(CGG)(n) by interspersed d(AGG) trinucleotides diminishes the formation and stability of d(CGG)(n) tetrahelical structures. *Nucleic Acids Res.*, **28**, 1535–1541.
29. Shimizu, M., Gellibolian, R., Oostra, B.A. and Wells, R.D. (1996) Cloning and characterization, and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J. Mol. Biol.*, **258**, 614–626.
30. Hirst, M.C. and White, P.J. (1998) Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of *in vivo* lagging strand secondary structure. *Nucleic Acids Res.*, **26**, 2353–2358.
31. Richards, R.I. and Sutherland, G.R. (1992) Dynamic mutations: a new class of mutations causing human disease. *Cell*, **70**, 709–712.
32. Kovtun, I.V., Therneau, T.M. and McMurray, C.T. (2000) Gender of the embryo contributes to CAG instability in transgenic mice containing a Huntington's disease gene. *Hum. Mol. Genet.*, **9**, 2767–2775.
33. Schwemmle, S., de Graaff, E., Deissler, H., Glaser, D., Wohrle, D., Kennerknecht, I., Just, W., Oostra, B.A., Dorfler, W., Vogel, W. and Steinbach, P. (1997) Characterization of FMR1 promoter elements by *in vivo*-footprinting analysis. *Am. J. Hum. Genet.*, **60**, 1354–1362.
34. Tischkoff, D.X., Filosi, N., Gaida, G.M. and Kolodner, R.D. (1997) A novel mutation avoidance mechanism dependent on *S. cerevisiae* RAD27 is distinct from DNA mismatch repair. *Cell*, **88**, 5027–5031.
35. Freudenreich, C.H., Kantrow, S.M. and Zakian, V.A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science*, **279**, 853–856.
36. Schweitzer, J.K. and Livingston, D.M. (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum. Mol. Genet.*, **7**, 69–74.
37. Spiro, C., Pelletier, R., Rolfsmeier, M.L., Dixon, M.J., Lahue, R.S., Gupta, G., Park, M.S., Chen, X., Mariappan, S.V. and McMurray, C.T. (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell.*, **4**, 1079–1085.
38. White, P.J., Borts, R.H. and Hirst, M.C. (1999) Stability of the human fragile X (CGG)(n) triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism. *Mol. Cell Biol.*, **19**, 5675–5684.
39. Shelbourne, P.F., Killeen, N., Hevner, R.F., Johnston, H.M., Tecott, L., Lewandoski, M., Ennis, M., Ramirez, L., Li, Z., Iannicola, C., Littman, D.R. and Myers, R.M. (1999) A Huntington's disease CAG expansion at the Hdh locus is unstable and associated with behavioural abnormalities in mice. *Hum. Mol. Genet.*, **8**, 763–774.
40. Lorenzetti, D., Watase, K., Xu, B., Matzuk, M.M., Orr, H.T. and Zoghbi, H.Y. (2000) Repeat instability and motor incoordination in mice with a targeted expanded CAG repeat in the Sca1 locus. *Hum. Mol. Genet.*, **9**, 779–785.
41. Wheeler, V.C., Auerbach, W., White, J.K., Srinidhi, J., Auerbach, A., Ryan, A., Duyao, M.P., Vrbanac, V., Weaver, M., Gusella, J.F. *et al.* (1999) Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. *Hum. Mol. Genet.*, **8**, 115–122.
42. Iyer, R.R., Pluciennik, A., Rosche, W.A., Sinden, R.R. and Wells, R.D. (2000) DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in *Escherichia coli*. *J. Biol. Chem.*, **275**, 2174–2184.
43. Jakupciak, J.P. and Wells, R.D. (1999) Genetic instabilities in (CTG.CAG) repeats occur by recombination. *J. Biol. Chem.*, **274**, 23468–23479.
44. Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E. and Hagerman, P.J. (2000) Elevated levels of FMR1 mRNA in carrier males: A new mechanism of involvement in the Fragile-X syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
45. Deelen, W., Bakker, C., Halley, D. and Oostra, B.A. (1994) Conservation of CGG region in *FMR1* gene in mammals. *Am. J. Med. Genet.*, **51**, 513–516.