Nucleic Acids Research, 1998, Vol. 26, No. 10 **2353–2358**

Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of *in vivo* lagging strand secondary structure

Mark C. Hirst* and Peter J. White

Institute of Molecular Medicine, The John Radcliffe Hospital, Headley Way, Headington, Oxford OX3 9DS, UK

Received February 2, 1998; Revised and Accepted March 23, 1998

ABSTRACT

The normal human *FMR1* gene contains a genetically stable $(CGG)_n$ trinucleotide repeat which usually carries interspersed AGG triplets. An increase in repeat number and the loss of interspersions results in array instability, predominantly expansion, leading to *FMR1* gene silencing. Instability is directly related to the length of the uninterrupted $(CGG)_n$ repeat and is widely assumed to be related to an increased propensity to form G-rich secondary structures which lead to expansion through replication slippage. In order to investigate this we have cloned human *FMR1* arrays with internal structures representing the normal, intermediate and unstable states. In one replicative orientation, arrays show a length-dependent instability, deletions occurring in a polar manner. With longer arrays these extend into the *FMR1* 5′-flanking DNA, terminating at either of two short CGG triplet arrays. The orientation-dependent instability suggests that secondary structure forms in the G-rich lagging strand template, resolution of which results in intra-array deletion. These data provide direct *in vivo* evidence for a G-rich lagging strand secondary structure which is believed to be involved in the process of triplet expansion in humans.

INTRODUCTION

The human fragile X syndrome is associated with dramatic expansion of a $(CGG)_n$ triplet repeat which lies within the *FMR1* promoter and is present in the 5′-untranslated region of its mRNA (1–3). In the normal population its length ranges from 6 to 52 repeats (4). In fragile X families it is expanded in length and is genetically unstable, with a high rate of length changes, which are almost exclusively expansions. Arrays longer than normal but <200 repeats, termed 'premutation' alleles, are found in non-penetrant carrier individuals. These arrays are somatically stable, but exhibit instability upon genetic transmission in a parental sex-specific manner, suggesting either an instability during germline formation and meiosis or during a window of very early embryonic development (5–9). Transmission from males is frequently associated with small decreases in array length, whilst female transmission usually results in expansion (10). Generally, when arrays exceed 200 repeats in length the *FMR1* promoter becomes extensively methylated (11,12), with concomitant loss of gene expression (13,14), although several exceptional cases have been identified (15).

The *FMR1* array in the normal size range has a periodic internal structure which results from regularly spaced AGG triplets positioned every 9 or 10 triplets (reviewed in 16). In contrast, unstable arrays consist of either uninterrupted $(CGG)_n$ or contain long portions of $(CGG)_n$ at their 3′-end (17–20). Expansion occurs exclusively within this uninterrupted portion of the array and transmission studies suggest that the degree of instability is directly related to its length (19,20). Therefore, the longer the uninterrupted $(CGG)_n$ portion the more unstable the array appears to be. Expanded triplet arrays have been identified in many other human disease genes, notably $(CAG)_n$ arrays within the genes causing Huntington's disease, spinocerebellar ataxias types I–VII and Kennedy's disease (see 21 and references therein); $(GAA)_n$ in the Friedrich's ataxia gene (22) and $(CTG)_n$ in the myotonic dystrophy gene (23). In addition, expanded $(CGG)_n$ arrays have also been found at the FRAXE (24), FRAXF (25), FRA16A (26) and FRA11B (27) fragile sites. In most of these, array instability is dependent upon array length and content, expansion occurring in arrays with >35–40 uninterrupted triplets (28).

This dependence upon array length suggests the involvement of length-dependent formation of non-B form DNA secondary structures in replicating DNA. Evidence for such structures has come from studies showing that single-stranded triplet DNA can assume unusual structures (29–32). They can also form hairpin-like structures (33,34) as well as quadraplexes (35,36) or triplexes (37). Primer extension studies through triplet arrays also provide indirect evidence for unusual structures which might cause a block to replication (38,39). The formation of a dramatically stable triplet-specific structure above a certain length threshold might thus be occurring. Whilst it is not known how formation of such unusual structures during replication would lead to expansion, replication stalling might allow a window of opportunity for slipped strand mispairing in the newly synthesized leading strand or perhaps through slippage of an Okazaki fragment or the newly synthesized lagging strand (for a review see 40). Recently, *in vivo* replication studies on $(CGG)_n$

*To whom correspondence should be addressed. Tel: +44 1865 222312; Fax: +44 1865 222500; Email: mhirst@worf.molbiol.ox.ac.uk

arrays have provided more direct evidence that unusual DNA structures in the lagging strand template lead to replication stalling (41).

As a first step toward a model system in which fragile X-associated triplet array stability can be studied we have established a range of cloned human *FMR1* arrays in bacteria. Unlike previous studies, these include *FMR1* flanking DNA and incorporate a wide range of interrupted repeat structures found in the human population, so that subsequent comparative instability studies can be performed.

MATERIALS AND METHODS

PCR amplification of *FMR1* arrays

PCR amplification was carried out on genomic DNA from selected individuals with known *FMR1* array structures as described previously (17). Amplification conditions were as follows: an initial denaturation at 98°C for 5 min was followed by 35 cycles of 98°C for 30 s and 70°C for 10 min. Each 20 µl reaction contained 20 mM Tris–HCl, pH 8.8, 10 mM KCl, 1.5 mM $MgCl_2$, 10 µM $(NH_4)_2SO_4$, 0.1% Triton, 100 µg/ml BSA, 0.5 µM each oligonucleotide (721 and 723), 320 µM each dNTP, 5% DMSO and 1 U native Pfu polymerase (Stratagene). Products were purified through 2% LMP agarose and isolated after Gelase solubilization (Epicenter Technologies) with Wizard PCR Prep reagents (Promega Corp.). For *FMR1* alleles longer than 50 repeats the concentration of DMSO was increased to 10%, which gives an increased yield of full-length product. All oligonucleotide sequences and their positions within the *FMR1* gene are taken from the HSFXDNA deposition in Genbank. Primers used were 721 (5′-AGCCCCGCACTTCCACCACCAGCTCCTCCA, complementary to 2617–2647) and 723 (5′-TTCACTTCCGGTGGAGGGCCGCCTCTGAGC, 2876–2838).

Cloning and insert analysis

Gel-purified PCR products were cloned into various vectors as described in Results. The vectors used were pBlueScript KS(–) (Stratagene), pBR322 (42) and pLEM8, a yeast integrating vector derived from pBR322 carrying a fragment of the *Saccharomyces cerevisiae* MAT gene (43). PCR products were phosphorylated using T4 PNK (NEB) and cloned either as blunt-ended fragments or after addition of *Eco*RI linkers. All plasmids were propagated in standard laboratory strain DH5α-MCR [RecA1, *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC); Life Technologies]. Recombinant plasmids were checked for integrity of the repeat array after each DNA preparation by restriction analysis, Southern blot hybridization and sequence analysis to confirm the structure of the repeat array. Orientation with respect to the plasmid origin of replication was determined by restriction analysis using the flanking *Nar*I and *Xho*I sites present in human *FMR1* exon 1 and plasmid restriction sites, notably the *Pvu*I site lying within the ampicillin resistance gene. For cloning of deletion arrays *Eco*RI-digested plasmid DNAs were electrophoresed in 2% agarose and deleted fragments purified and directly re-cloned into pBR322. Recombinant plasmids were selected by resistance to both ampicillin and tetracycline to ensure that original plasmid DNAs were not propagated. Deletion arrays, now re-cloned in the stable replicative orientation, were sequenced as described below. All plasmid DNA preparations were performed using standard alkaline lysis techniques with reagents supplied by Qiagen or Promega Corp. Southern blot transfer from agarose gels was performed with Hybond N^+ (Amersham International) using 0.4 M NaOH and filters were hybridized to the radiolabelled 721–723 PCR product which had been random prime radiolabelled by standard techniques.

Sequencing

Primers used for sequencing were either standard M13 or pBR322 vector primers designed to flank the cloning sites used or the *FMR1*-specific primers 170 and 172 described below. Sequencing was performed with ^{32}P-end-labelled primers using the exonuclease-deficient Pfu polymerase cyclist kit (Stratagene) as described in Hirst *et al.* (17). Products were resolved on 5% denaturing polyacrylamide sequencing gels (USB) and visualized by autoradiography. *FMR1*-specific sequencing primers used were 170 (5′-GGCGGTGACGGAGGCGCC, complementary to 2678–2695) and 171 (5′-CCTGCTAGCGCCGGGAGC, 2807–2824).

RESULTS

Cloning of *FMR1* arrays

Human genomic DNA from individuals with previously characterized *FMR1* array structures were used in order to establish a panel of cloned arrays reflecting those occurring *in vivo* (Fig. 1a). The amplified region includes most of *FMR1* exon 1 and extends to cover 90 bp distal and 120 bp proximal of the $(CGG)_n$ array. Fragments were amplified using a protocol established for direct sequence analysis of *FMR1* arrays which yields sufficient product for cloning into various vectors (17). This protocol has been used extensively for *FMR1* triplet array amplification and has a high degree of fidelity (17,44–46). This ensured that the arrays were amplified intact and could be analysed in their naturally occurring DNA context. Inclusion of *FMR1* exon 1 DNA also facilitated further manipulation with unique flanking restriction sites. We had previously observed that *Escherichia coli* strain DH5α-MCR permitted longer repeat arrays to be propagated than several other common laboratory strains tested and this strain was therefore used in all subsequent studies.

Initial experiments with the high copy number plasmid pBS (250–300 copies/cell) revealed that long *FMR1* arrays were highly unstable in this vector. In order to assess the degree of instability we assayed the length of the cloned insert by restriction analysis using the *Eco*RI (the cloning site) and *Xho*I and *Nar*I sites present in the co-amplified human flanking DNA. Southern blot hybridization with a radiolabelled $(CGG)_n$ probe was used to detect deletion arrays and then direct sequence analysis was used to check for array integrity. Using these stringent criteria we found evidence for some orientation-dependent deletion in arrays with as few as 13 uninterrupted triplets, even when they also contained AGGs. In the (+) orientation the arrays $(CGG)_{13}$, $(CGG)_{27}$, 9A23, 9A9A19 and 10A9A7A9 [where numbers represent the length of $(CGG)_n$ and an A represents an interspersed AGG repeat] exhibit a smear of deletion products upon Southern blot hybridization analysis (Fig. 1b). We occasionally encountered plasmids which appeared to contain stable arrays but upon closer examination by sequence analysis these were found to have acquired interruptions in the array. Obviously, the presence of interrupting repeats and deletions

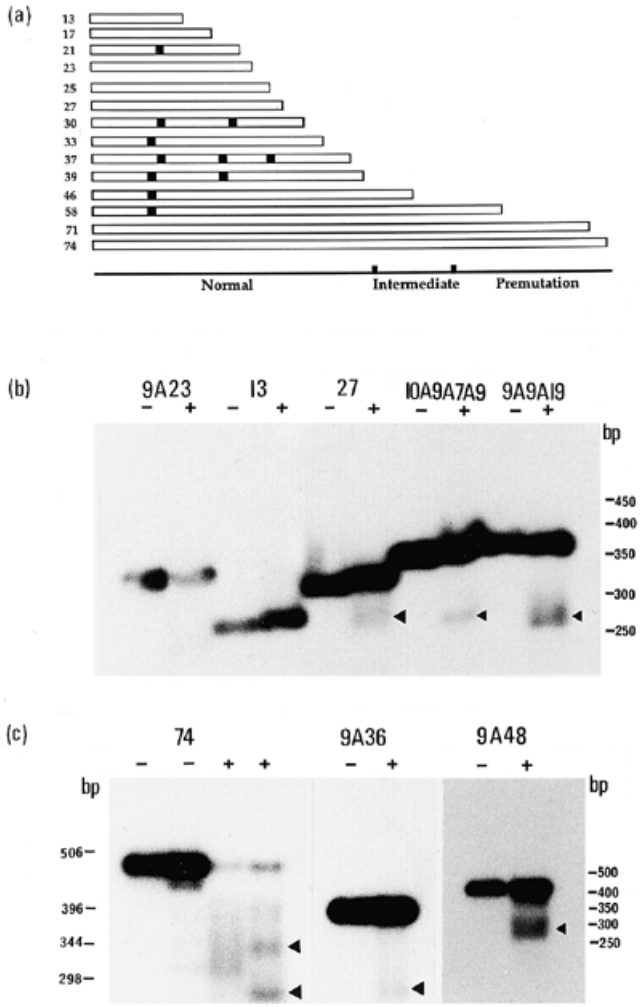

Figure 1. Cloned *FMR1* triplet arrays. (**a**) The cloned human *FMR1* arrays ranging from 13 to 74 triplets in length are shown pictorially to highlight important features of $(CGG)_n$ length (open boxes) and AGG (solid boxes) interspersion pattern. The overall length of each array is shown in total triplet number and the scale represents the status of these arrays with regard to fragile X instability and mutation class. From the top the cloned arrays are $(CGG)_{13}$, $(CGG)_{17}$, 10A11, $(CGG)_{23}$, $(CGG)_{25}$, $(CGG)_{27}$, $(CGG)_{30}$, 10A9A9, 9A23, 10A9A6A9, 9A10A18, 9A36, 9A48, $(CGG)_{71}$ and $(CGG)_{74}$, with the structure of the array being abbreviated such that numbers represent the length of $(CGG)_n$ and an A represents an interspersed AGG repeat. (**b**) *FMR1* array instability of normal length triplet arrays with respect to the origin of replication in pBS. The two replicative orientations of the triplet array were nominally termed + or –. Cloned inserts were released by *Eco*RI digestion, resolved on 2.5% agarose gels and detected by hybridization after Southern transfer. The DNA molecular weight ladder is in base pairs. Deletion products are highlighted by arrows. (**c**) *FMR1* array instability of longer intermediate and premutation length triplet arrays in both orientations with respect to the origin of replication in plasmid pLEM8, which carries a pBR322 origin of replication. Details as in (b) above.

would have invalidated any subsequent experiments assessing instability. Plasmids carrying the same arrays in the (–) orientation showed no evidence of deletion. When the array contained >36 uninterrupted repeats in pBS it proved impossible to obtain plasmid DNA with arrays in either orientation which did not carry a substantial proportion of deleted material. We therefore looked for a more stable system in which to isolate these longer arrays.

In order to isolate recombinant premutation arrays, vectors carrying a lower copy number ColE1 unidirectional origin of replication were employed. These vectors, with a copy number of only15–20 plasmids/cell, were found to stabilize longer repeat arrays considerably. Several vectors, including pBR322 and pLEM8, were used successfully to propagate these longer arrays, suggesting that the critical factor is most likely the origin of replication or copy number, rather than any other additional DNA element present. The orientation effect we had observed previously was even more pronounced with these longer arrays, with the proportion of full-length array present diminishing with increased length. It proved impossible to obtain DNA preparations with full-length cloned arrays in the (+) orientation (Fig. 1c). For example, with the $(CGG)_{74}$ array, DNA preparations for the (+) orientation plasmids contain very little full-length array (Fig. 1c, tracks 3 and 4) compared with the opposite (–) orientation (Fig. 1c, tracks 1 and 2). The same is true for the 9A48 (Fig. 1c, tracks 7 and 8) and 9A36 arrays (Fig. 1c, tracks 5 and 6). Using the combination of low copy number vectors and the (–) orientation we successfully cloned stable full-length 9A39, 9A48, $(CGG)_{71}$ and $(CGG)_{74}$ arrays. Sequence analysis of these arrays confirmed their integrity. Attempts to rescue longer arrays from the gel using the approach of Shimizu *et al.* (47) were unsuccessful. We did not detect any significant expansion events as judged by this technique or by extensive Southern blot analysis of plasmid DNA preparations in either orientation. In contrast to the problems encountered with *FMR1* arrays, arrays of $(CAG)_{65}$ and $(CGT)_{75}$ were stably propagated in pBS (data not shown). This highlights the unstable nature of the *FMR1* $(CGG)_n$ arrays.

The structure of orientation-specific deletions in premutation arrays

For the 9A36 and 9A48 arrays, which represent intermediate and small premutation *FMR1* alleles, deletions appeared to be intra-array events, as judged by retention of the flanking *Nar*I and *Xho*I restriction sites. In contrast, deletions in the 71 and 74 repeat arrays frequently removed the 5′ *Nar*I site (data not shown). In order to examine this further, deleted arrays from plasmids carrying $(CGG)_{74}$ and 9A48 in the (+) orientation were isolated after *Eco*RI digestion and re-cloned into pBR322. Plasmids in which the deletion array had been re-cloned in the stable (–) orientation were then subjected to sequence analysis. The data obtained from this are summarized in Figure 2. In both cases deletions appear to occur in a polar manner. For the 9A48 array, where the proximal AGG serves as an internal landmark, deletions extend from the 3′-end of the array toward the 5′-end, frequently removing the interspersion (14 out of 22 cases; Fig. 2a), but extended into the flanking DNA. Deletion events frequently gave products where the longest uninterrupted repeat was between 8 and 12 triplets, with typically 40 and 50 repeats being removed (16 out of 22). In the case of the $(CGG)_{74}$ repeat array the deletion product size range is similar, suggesting a common mechanism or end-point to the process (Fig. 2b). However, in contrast to 9A48, deletions were found to extend into the 5′-region of *FMR1* exon 1 (3 out of 7 cases; Fig. 2c) with 5′ end-points at positions 2649 and 2674 in the flanking *FMR1* exon 1. At these points lie short arrays of $(CGG)_2$ and $(CGG)_3$ and the deletion products appear as $(CGG)_{12}$ and $(CGG)_{13}$ fusion arrays. This suggests that they have been generated in a homology-dependent manner. Interestingly, these deletion end-points 5′ of

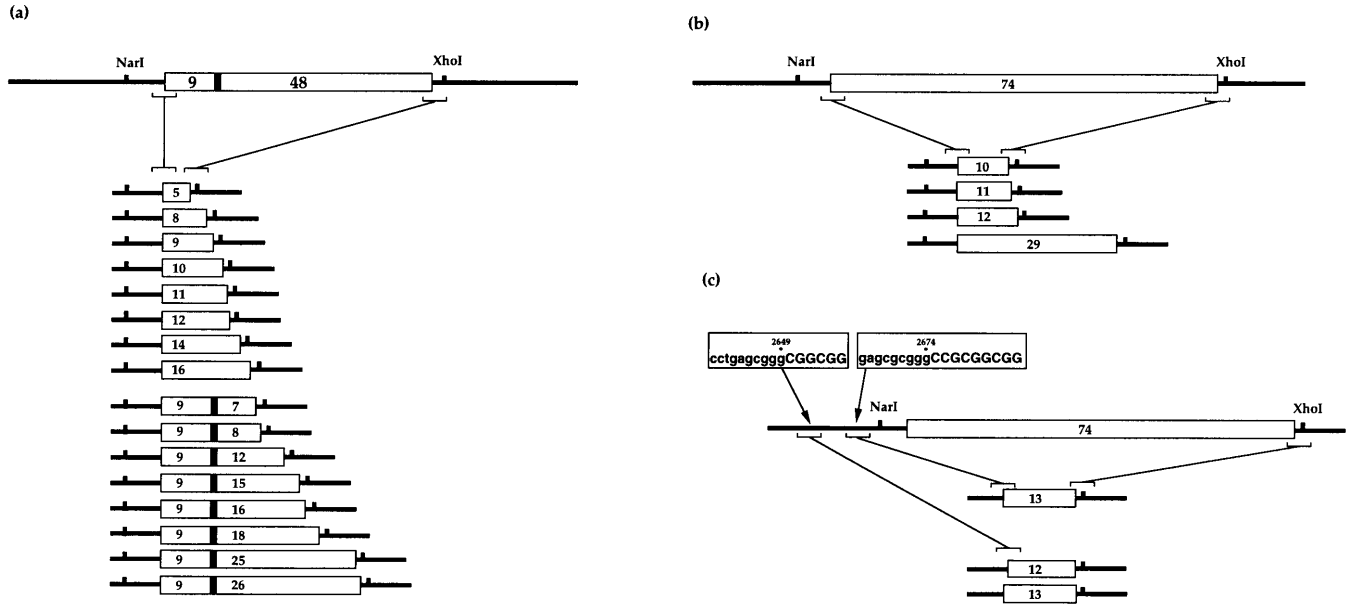

Figure 2. Structure of deleted *FMR1* arrays. The structure of deleted arrays from the unstable (+) orientation are shown from various original plasmids. The original array length is shown and below are the sequenced deleted arrays derived from each and the number of independent clones analysed. Deletions originating from the 9A48 array are shown in (**a**), whilst those in (**b**) are from the $(CGG)_{74}$ array. In (**c**) deletion of the $(CGG)_{74}$ array into the flanking 5′-region is shown. Nucleotide numbering refers to the human *FMR1* sequence HSFXDNA deposited in GenBank.

the array are close to a deletion hot-spot found to occur in mosaic fragile X males (48).

DISCUSSION

We have successfully cloned a representative series of normal, intermediate and small premutation human *FMR1* triplet arrays. We found that in certain configurations they demonstrate a pronounced length-, orientation- and interspersion-dependent instability, frequently giving rise to deletions. Whilst most arrays tested show some degree of instability, those with uninterrupted lengths greater than $(CGG)_{36}$ appear to be particularly unstable. These observations confirm previous studies of both fragile X and $(CTG)_n$ arrays, which demonstrated a preferential stability in one replicative direction (47,49). These data suggest that an unusual structure forms in the lagging strand template which interferes with the normal process of replication and results in a plasmid population carrying intra-array deletions. The polar manner of these deletions suggests that this structure can form from as few as 13 CGG triplets and is nucleated within the triplet array. Furthermore, with longer arrays, flanking DNA regions appear to contribute to the stability of these structures.

Plasmid cloning experiments found that longer *FMR1* arrays are more stable in lower copy number plasmids. This is presumably related to the number of replication events that an array undergoes during the multiple bacterial divisions required for a standard DNA preparation, which for a high copy number vector such as pBS is 10–20 times greater than that in a low copy vector. In the simplest case, if the likelihood of a deletion event occurring in a full-length array is the same for each round of replication, then the frequency of deletion would be decreased in lower copy number vectors. Several other studies have reported the use of low copy number vectors for cloning of *FMR1* arrays. The study of Shimizu *et al.* (47) used a concatenation process to generate arrays of up to 240 repeats in overall length, but these contain multiple interruptions and carry only $(CGG)_{60}$ as their longest stretch of uninterrupted triplet repeat. Whilst these artificial arrays did exhibit the same orientation effect as reported in this study, their degree of instability is less pronounced. This is most likely due to the interrupted nature of these arrays, which obviously limits the value of any subsequent instability studies using them, as instability is known to be dependent upon the length of the uninterrupted repeat. A more recent report by Sandberg and Schalling (50) described the use of a $(CGG)_{200}$ array isolated by PCR from a fragile X male, although no data were included to establish the integrity of either the PCR product or the cloned array. The apparent stability of this array is almost certainly an indication that it has undergone a stabilizing event, most likely acquisition of interruptions (perhaps during PCR), although we cannot rule out additional stabilizing bacterial host cell factors. Determining the integrity of cloned *FMR1* arrays longer than 70–80 triplets is not a simple task. Sequence analysis is difficult because repetitive G/C-rich arrays cause extensive slippage during both template preparation and sequencing reactions, which makes detection of interruptions extremely difficult.

A major factor which determines cloned array stability is its orientation with respect to the direction of replication. This indicates that the G/C-rich strands behave differently as leading and lagging strand templates. For $(CGG)_n$ this is most pronounced where replication proceeds through the triplet array from its 5′-end. In this orientation the G-rich strand is the lagging strand template and is thus transiently single-stranded during replication (Fig. 3a). Several studies have suggested that single-stranded

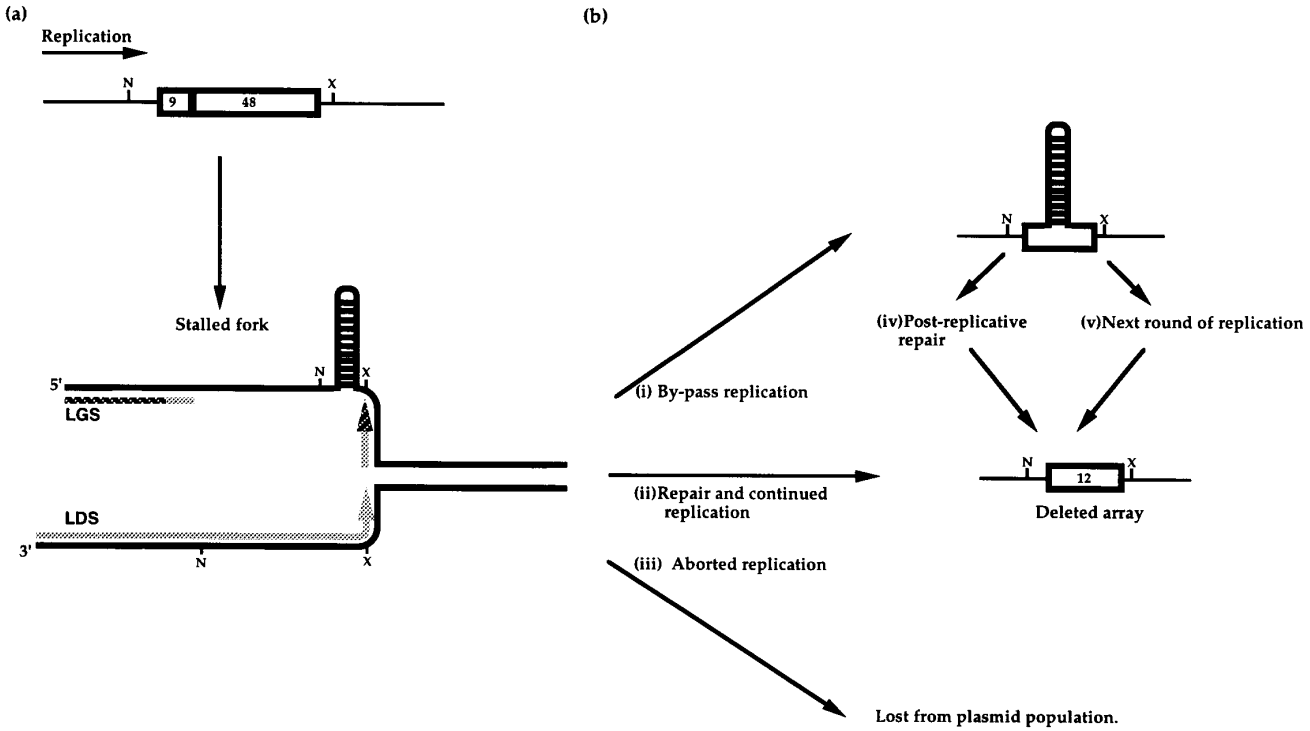

Figure 3. Model of lagging strand secondary structure-induced instability in cloned *FMR1* triplet arrays and the possible deletion mechanisms in the unstable (+) orientation of the 9A48 *FMR1* array. (**a**) A stalled replication fork due to G-rich secondary strand structure can be resolved by (**b**): (i) by-pass replication; (ii) repair; (iii) abortion of replication. After continued by-pass synthesis of the lagging strand (i), a deleted array could arise either by post-replicative repair (iv) or in the absence of repair in the next round of replication (v). Aborted replication would most likely result in loss of the plasmid from the population. LDS, leading strand (grey line); LGS, lagging strand (hatched line); open blocks represent $(CGG)_n$ and a solid box represents an AGG triplet. N, *Nar*I site; X, *Xho*I site.

$(CGG)_n$ arrays have a propensity to form unusual stable secondary structures (29–39). These were suggested to be causing the stalled replication forks observed by Samadashwily *et al.* (41), but how might these lead to deletions in cloned arrays? In our model, replication through the *FMR1* array from its 5′-end leads to a stalled replication fork-induced lagging strand template secondary structure (Fig. 3a and b). If this cannot be resolved, it must either be (i) by-passed by the newly synthesized lagging strand or (ii) removed by DNA repair, both cases allowing replication to continue. In the latter case, generation of a deleted array requires removal of DNA from the template strand, which is normally marked by hemi-methylation. This suggests that either the deletion event is independent of this discriminating marking of the template strand or that it occurs after *de novo* methylation of the newly synthesized strand. In the alternative case of by-passed structures these might be either subsequently repaired or will result in deleted arrays after the next round of replication. A third possible fate of stalled templates which are not by-passed or repaired might be termination of replication and loss from the plasmid population. The DNA yield from a standard plasmid DNA preparation represents a population of closed circular plasmids present after 20–25 bacterial divisions. $(CGG)_n$ arrays are known to be difficult templates to replicate and the longer the array the greater that difficulty. Under constant antibiotic selection pressure replication of plasmids carrying truncated arrays, with reduced replicative difficulty, would be favoured and propagation of expanded arrays strongly disfavoured.

In the unstable (+) orientation, deletion of the *FMR1* array occurs in a polar manner. This is most likely a result of the direction of replication in which they arise. At a stalled replication fork the lagging strand template is transiently single-stranded in the region between Okazaki fragments, providing an opportunity for secondary structure to form (Fig. 3a). In the case of the (+) orientation this single-stranded region will often extend through the array and into the 5′-flanking DNA, allowing any secondary structure to extend in this direction. As we observed frequent deletions of the $(CGG)_{74}$ array in homologous regions 5′ of the array, this suggests that this secondary structure might become more extensive with longer repeats, probably also involving flanking sequences. Additionally, this suggests that the resolution event which removed the secondary structure in these cases did so in a homology-dependent manner.

The behaviour of cloned *FMR1* triplet arrays is very different from what occurs in humans. However, these studies do tell us a great deal about the basis of triplet instability. There is now overwhelming evidence that replication through $(CGG)_n$ triplet arrays is problematical. Evidence presented in this study suggests that at least part of this is due to lagging strand template structure. How might this be involved in the process of expansion? A favoured model to explain triplet expansion invokes replication stalling due to G-rich secondary structure, which allows an opportunity for slippage of the lagging strand (40). It has been shown in other studies that the C-rich strand is most likely to expand by slippage (38). This might result from either simple

slipped strand re-annealing or with large arrays it might additionally involve slippage of an un-tethered Okazaki fragment which could give rise to larger expansions (51). The loss of plasmids carrying stalled replication forks might also explain why we see no expansion of our cloned triplet arrays. If plasmid templates with stalled replication forks, which are the potentially active templates on which lagging strand slippage would occur, are aborted and lost from the plasmid population, then no opportunity for expansion would exist in our plasmids. Rare expansions might occur, but the resulting plasmids would be at a replicative disadvantage and would hence be likely to subsequently be lost.

Resolution of slippage events as expanded arrays in humans is probably due to differences in the factors processing replicative intermediates. This might simply represent a difference between prokaryotic and eukaryotic replication and repair systems. Additionally, in order to account for differential instabilities between mitotic, meiotic and embryonic stages, expression of these factors must vary within human cell lineages. Our panel of cloned arrays will be useful in addressing these issues in model replication systems such as *S.cerevisiae*, where repair pathways have been extensively studied, and in mouse transgenesis. Indeed, several of these arrays have already been established in these systems and studies on their instability are currently underway. The use of these cloned arrays in such studies should provide new insights into both the origins of unstable *FMR1* arrays and the pathway of their expansion to full fragile X mutations.

ACKNOWLEDGEMENTS

This work was funded by a fellowship awarded to M.H. by the Wellcome Trust. We would like to thank Prof. Sir David Weatherall and Prof. Kay E.Davies for their support and Prof. Charles Laird and members of his laboratory at the FHCRC (Seattle, WA) for discussion and additional comments.

REFERENCES

1 Verkerk,A.J.M.H. *et al*. (1991) *Cell*, **65**, 904–914.
2 Yu,S. *et al.* (1991) *Science*, **252**, 1179–1181.
3 Oberle,I., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boue,J., Bertheas,M.F. and Mandel,J.L. (1991) *Science*, **252**, 1079–1102.
4 Fu,Y.-H. *et al*. (1991) *Cell*, **67**, 1047–1058.
5 Worhle,D., Hirst,M., Davies,K. and Steinbach,P. (1992) *Hum. Genet*., **89**, 114–116.
6 Reyniers,E. *et al*. (1993) *Nature Genet*., **4**, 143–146.
7 Worhle,D., Hennig,I., Vogel,W. and Steinbach,P. (1993) *Nature Genet*., **4**, 140–142.
8 Devys,D. *et al*. (1992) *Am. J. Med. Genet*., **43**, 208–213.
9 Malter,H.E., Iber,J.C., Willemsen,R., de Graff,E., Tarleton,J.C., Leisti,J., Warren,S.T. and Oostra,B.A. (1997) *Nature Genet*., **15**, 165–169.
10 Fisch,G.S. *et al.* (1995) *Am. J. Hum. Genet*., **56**, 1147–1155.
11 Bell,M.V. *et al.* (1991) *Cell*, **64**, 861–866.
12 Hansen,R.S., Gartler,S.M., Scott,C.R., Chen,S.-H. and Laird,C.D. (1992) *Hum. Mol. Genet*., **1**, 571–578.
13 Pierreti,M., Zhang,F., Fu,Y.-H., Warren,S.T., Oostra,B.A., Caskey,C.T. and Nelson,D.L. (1991) *Cell*, **66**, 817–822.
14 Sutcliffe,J.S., Nelson,D.L., Zhang,F., Pieretti,M., Caskey,C.T., Saxe,D. and Warren,S.T. (1992) *Hum. Mol. Genet*., **1**, 397–400.
15 Smeets,H.J., Smits,A.P., Verheij,C.E., Theelen,J.P., Willemsen,R., van de Burgt,I., Hoogeveen,A.T., Oosterwijk,J.C. and Oostra,B.A. (1995) *Hum. Mol. Genet*., **4**, 2103–2108.
16 Hirst,M.C. (1995) *J. Med. Genet*., **32**, 761–763.
17 Hirst,M.C., Grewal,P. and Davies,K.E. (1994) *Hum. Mol. Genet*., **3**, 1553–1560.
18 Snow,K., Tester,D.J., Kuckeberg,K.E., Scaid,D.J. and Thibodeau,S.N. (1994) *Hum. Mol. Genet*., **3**, 1543–1551.
19 Eichler,E.E., Holden,J.J., Popovich,B.W., Reiss,A.L., Snow,K., Thibodeau,S.N., Richards,C.S., Ward,P.A. and Nelson,D.L. (1994) *Nature Genet*., **8**, 88–94.
20 Kunst,C.B. and Warren,S.T. (1994) *Cell*, **77**, 853–861.
21 Warren,S.T. (1996) *Science*, **271**, 1374–1375.
22 Campuzano,V. *et al*. (1996) *Science*, **271**, 1423–1437.
23 Brook,J.D. *et al.* (1992) *Cell*, **68**, 799–808.
24 Knight,S.J.L. *et al.* (1993) *Cell*, **74**, 127–134.
25 Ritchie,R.J., Knight,S.J., Hirst,M.C., Grewal,P.K., Bobrow,M., Cross,G.S. and Davies,K.E. (1994) *Hum. Mol. Genet*., **3**, 2115–2121.
26 Nancarrow,J.K., Holman,K., Mangelsdorf,M., Hori,T., Denton,M., Sutherland,G.R. and Richards,R.I. (1995) *Hum. Mol. Genet*., **4**, 367–372.
27 Jones,C., Slijepcevic,P., Marsh,S., Baker,E., Langdon,W.Y., Richards,R.I. and Tunnacliffe,A. (1994) *Hum. Mol. Genet*., **3**, 2123–2130.
28 Mandel,J.L. (1997) *Nature*, **386**, 767–769.
29 Mitchell,J.E., Newbury,S.F. and McClellan,J.A. (1995) *Nucleic Acids Res*., **23**, 1876–1881.
30 Petruska,J., Arnheim,N. and Goodman,M.F. (1996) *Nucleic Acids Res*., **24**, 1992–1998.
31 Yu,A., Dill,J., Wirth,S.S., Huang,G., Lee,V.H., Haworth,I.S. and Mitas,M. (1995) *Nucleic Acids Res*., **23**, 2706–2714.
32 Yu,A., Barron,M.D., Romero,R.M., Christy,M., Gold,B., Dai,J., Gray,D.M., Haworth,I.S. and Mitas,M. (1997) *Biochemistry*, **36**, 368736–368799.
33 Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) *Cell*, **81**, 533–540.
34 Chen,X., Mariappan,S.V., Catasti,P., Ratliff,R., Moyzis,R.K., Laayoun,A., Smith,S.S., Bradbury,E.M. and Gupta,G. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 5199–5203.
35 Fry,M. and Loeb,L.A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4950–4954.
36 Kettani,A., Kumar,R.A. and Patel,D.J. (1995) *J. Mol. Biol*., **254**, 638–656.
37 Kuryavyi,V.V. and Jovin,T.M. (1995) *Nature Genet*., **9**, 339–341.
38 Ji,J., Clegg,N.J., Peterson,K.R., Jackson,A.L., Laird,C.D. and Loeb,L.A. (1996) *Nucleic Acids Res*., **24**, 2835–2840.
39 Usdin,K. and Woodford,K.J. (1995) *Nucleic Acids Res*., **23**, 4202–4209.
40 McMurray,C.T. (1995) *Chromosoma*, **104**, 2–13.
41 Samadashwily,G.M., Raca,G. and Mirkin,S.M. (1997) *Nature Genet*., **17**, 298–304.
42 Bolivar,F., Rodriguez,R.L., Greene,P.J., Betlach,M.C., Heyneker,H.L. and Boyer,H.W. (1977) *Gene*, **2**, 95–113.
43 Hunter,N., Chambers,S.R., Louis,E.J. and Borts,R.H. (1996) *EMBO J*., **15**, 1726–1733.
44 Macpherson,J.N., Curtis,G., Crolla,J.A., Dennis,N., Migeon,B., Grewal,P.K., Hirst,M.C., Davies,K.E. and Jacobs,P.A. (1995) *J. Med. Genet*., **32**, 236–239.
45 Hirst,M.C., Arinami,T. and Laird,C.D. (1997) *Hum. Genet*., **101**, 214–218.
46 Mornet,E., Chateau,C., Hirst,M.C., Thepot,F., Taillandier,A., Cibois,O. and Serre,J.L. (1996) *Hum. Mol. Genet*., **5**, 821–825.
47 Shimizu,M., Gellibolian,R., Oostra,B.A. and Wells,R.D. (1996) *J. Mol. Biol*., **258**, 614–626.
48 de Graaff,E., Rouillard,P., Willems,P.J., Smits,A.P., Rousseau,F. and Oostra,B.A. (1995) *Hum. Mol. Genet*., **4**, 45–49.
49 Kang,S., Jaworski,A., Oshima,K. and Wells,R.D. (1995) *Nature Genet*., **10**, 213–218.
50 Sandberg,G. and Schalling,M. (1997) *Nucleic Acids Res*., **25**, 2883–2887.
51 Sutherland,G.R. and Richards,R.I. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3636–3641.