

# Trinucleotide repeats affect DNA replication *in vivo*

George M. Samadashwily, Gordana Raca & Sergei M. Mirkin

**(CGG)<sub>n</sub>·(CCG)<sub>n</sub> and (CTG)<sub>n</sub>·(CAG)<sub>n</sub> repeats of varying length were cloned into a bacterial plasmid, and the progression of the replication fork through these repeats was followed using electrophoretic analysis of replication intermediates. We observed stalling of the replication fork within repeated DNAs and found that this effect depends on repeat length, repeat orientation relative to the replication origin and the status of protein synthesis in a cell. Interruptions within repeated DNAs, similar to those observed in human genes, abolished the replication blockage. Our results suggest that the formation of unusual DNA structures by trinucleotide repeats in the lagging-strand template may account for the observed replication blockage and have relevance to repeat expansion in humans.**

More than a dozen human disorders are now attributed to the length polymorphism of simple DNA repeats within different human genes<sup>1–3</sup>. Trinucleotide repeats—specifically (CGG)<sub>n</sub>·(CCG)<sub>n</sub>, (CTG)<sub>n</sub>·(CAG)<sub>n</sub> and (GAA)<sub>n</sub>·(TTC)<sub>n</sub>—account for most of these cases. Trinucleotide repeats expand with a length-dependent probability. In normal individuals carrying five to 30 repeats, expansion is highly unlikely. Individuals with repeat numbers exceeding a threshold of approximately 30 can transmit expanded repeats to their progeny. In the following generations, expansions become more frequent, and each subsequent expansion has a higher probability than the previous one. This phenomenon is likely to account for the anticipation in the inheritance of these disorders<sup>4,5</sup>.

The dependence of expansion on repeat length suggests the involvement of unusual secondary DNA structures<sup>6</sup>. This notion is supported by several lines of experimental data. Double-stranded trinucleotide repeats differ from canonical B-DNA<sup>7,8</sup>. Single DNA strands of expandable repeats can fold into defined compact structures<sup>9–12</sup>. NMR analysis showed that these repeats fold into imperfect hairpins<sup>13–15</sup>, and it was hypothesized that the threshold length for expansion may reflect the threshold energy of hairpin formation<sup>14</sup>. Other structures, such as quadruplexes<sup>16,17</sup> and triplexes<sup>18</sup>, were also discussed.

The mechanisms of repeat expansion remain unknown, but most data implicate replication in this process. A good example is the polar variations observed by sequence analysis of fragile X families. (CGG)<sub>n</sub> repeats in the *FMR1* gene are usually interrupted by several dispersed AGG triplets<sup>19–21</sup>. Expansion of these repeats occurs exclusively at their 3' flanks, and the expanded part lacks any interruptions<sup>20,21</sup>. Similar polar variations were reported for (CAG)<sub>n</sub> repeat expansion<sup>22</sup>. This polarity can be explained by anomalous repeat replication from a single adjacent replication origin, assuming different fidelity of the leading- and lagging-strand synthesis<sup>23</sup>.

Further, *in vitro* DNA polymerization through trinucleotide repeats is abnormal. In both double-stranded<sup>24,25</sup> and single-stranded<sup>26</sup> DNA templates, polymerization was blocked within repeated segments, possibly because of their unusual secondary structures. This blockage can facilitate a misalignment between

the newly synthesized and the template DNA strand<sup>25</sup>, potentially leading to expansion.

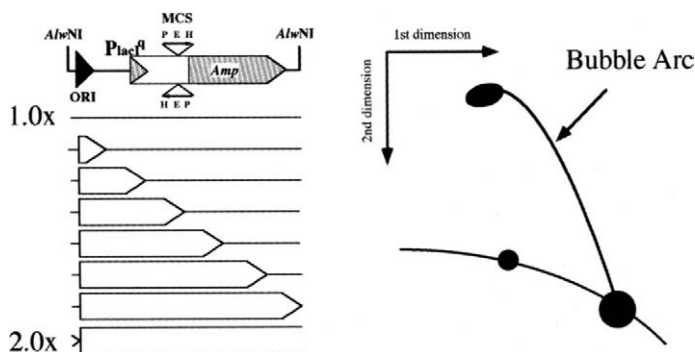
Studies of trinucleotide repeats in bacterial plasmids also support the replication hypothesis; whereas long (CTG)<sub>n</sub> or (CGG)<sub>n</sub> repeats are deleted from the lagging-strand template, but expanded in the leading-strand template<sup>27–31</sup>, suggesting that the formation of hairpins by these repeats in either the lagging-strand template or the newly synthesized lagging strand caused deletions or expansions, respectively.

Information on *in vivo* replication of trinucleotide repeats is scarce. The fragile-X mutation causes delayed replication<sup>32,33</sup>, although the latter occurs in a substantial part of the X chromosome and is unlikely to be due to the replication of the (CGG)<sub>n</sub>·(CCG)<sub>n</sub> tract itself<sup>34</sup>. To obtain direct data on trinucleotide repeats replication *in vivo*, we analysed the movement of the replication-fork through these repeats within bacterial plasmids by means of electrophoretic analysis of replication intermediates<sup>35</sup>. We found that (CGG)<sub>n</sub>·(CCG)<sub>n</sub> and (CTG)<sub>n</sub>·(CAG)<sub>n</sub> repeats blocked replication-fork progression, and the efficiency of blockage depended on the repeat length, its orientations relative to the replication origin and the presence of the protein synthesis inhibitor chloramphenicol. We believe that the structure of repeated DNA in the lagging-strand template is responsible for the replication blockage.

## Repeats arrest DNA replication *in vivo*

We analysed the effects of various trinucleotide repeats on the replication of bacterial ColE1-type plasmids *in vivo*. Because these plasmids replicate unidirectionally<sup>36</sup>, we were able to distinguish between the different strands of each repeat on leading- and lagging-strand DNA synthesis by cloning each repeat in two orientations relative to the replication origin.

We generated long (CGG)<sub>n</sub>·(CCG)<sub>n</sub> stretches by PCR using d(CGG)<sub>10</sub> and d(CCG)<sub>10</sub> oligonucleotides (see Methods) and cloned into a deletion-derivative of the pTrc99A plasmid<sup>37</sup> (Fig. 1). In this derivative, the *amp* gene is under the control of the *PlacI*<sup>q</sup>, and the polylinker is located between this promoter and the translational start of the *amp* gene. The two orientations of the polylinker allowed us to clone repeats in both orientations



**Fig. 1** Schematic representation of two dimensional N/N gel electrophoresis for the pTrc99 $\Delta$  plasmid. Plasmid structure is presented in the upper left corner. The restriction enzyme *AlwNI* cleaves this plasmid at a unique site located upstream of *ori*. Replicative intermediates digested by *AlwNI* are bubble-like, and the size of a bubble reflects the extent of replication (left panel). The right panel schematically shows the separation of bubble-like replicative intermediates by two dimensional gel electrophoresis.

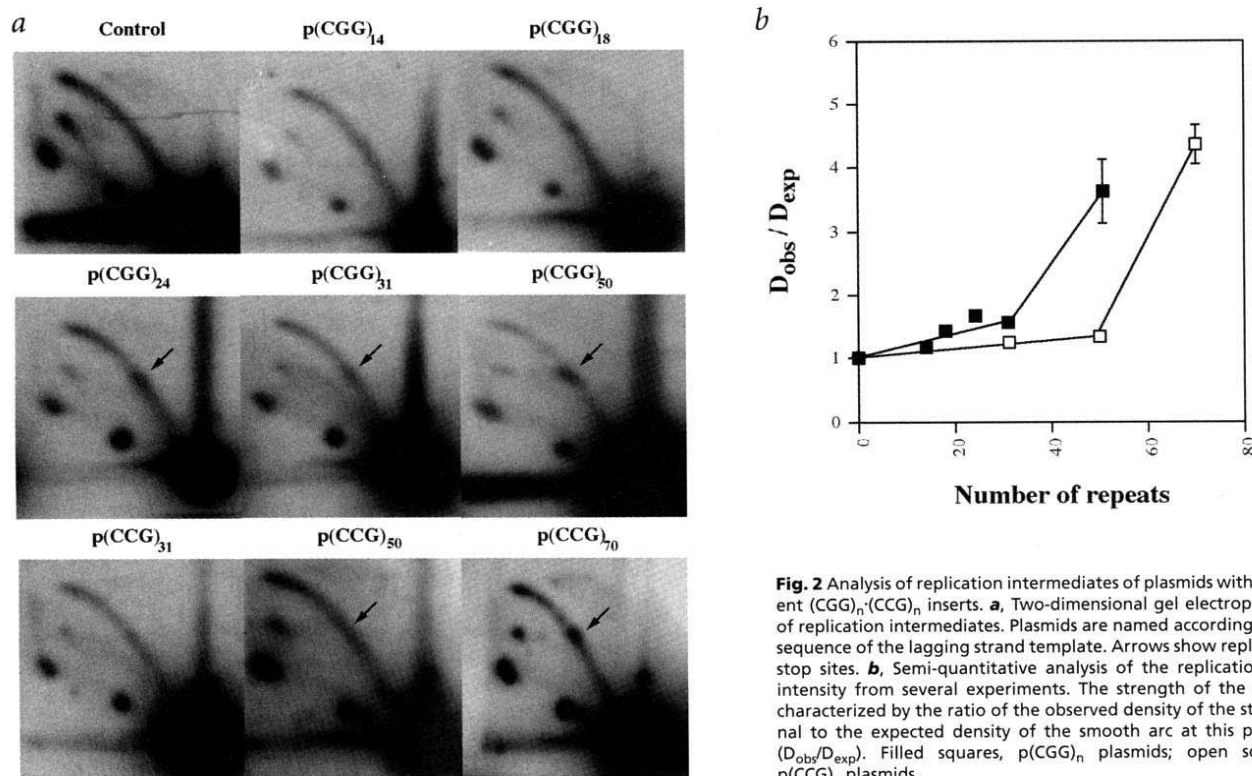
relative to the origin of replication (*ori*; Fig. 1). Cloned repeats were located approximately 1.4 kb downstream of *ori* and approximately 1 kb downstream from the *pasH* site—that is, in the area replicated by the DNA polymerase III.

To analyse the replication-fork movement, we used two-dimensional neutral/neutral electrophoresis of replication intermediates<sup>35</sup>. As our plasmids replicate unidirectionally, one might expect the appearance of bubble-like structures upon isolating the replicating DNA and cleaving it with a restriction enzyme upstream of *ori*<sup>38</sup>. Bubbles differ from non-replicated DNA in both their size and shape, allowing one to separate them by gel electrophoresis. As a result, one can observe a characteristic bubble arc (Fig. 1). Pausing of the replication fork at an insert would lead to the preferential accumulation of intermediates of a given length and shape—that is, the appearance of a bulge on an otherwise smooth bubble arc.

Owing to the directionality of plasmid replication, either (CGG)<sub>n</sub> or (CCG)<sub>n</sub> repeats appeared in the lagging-strand template; the plasmids are named accordingly. A control plasmid, pTrc99 $\Delta$ , has a smooth bubble arc (Fig. 2a), while plasmids containing long (CGG)<sub>n</sub>-(CCG)<sub>n</sub> repeats show evident replication stop signals in the bubble arc. Using semi-quantitative analysis (see Methods), we estimated the 'stop strength' as the ratio of the

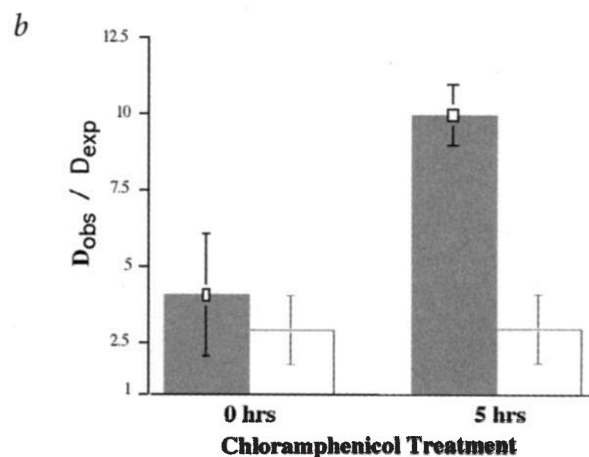
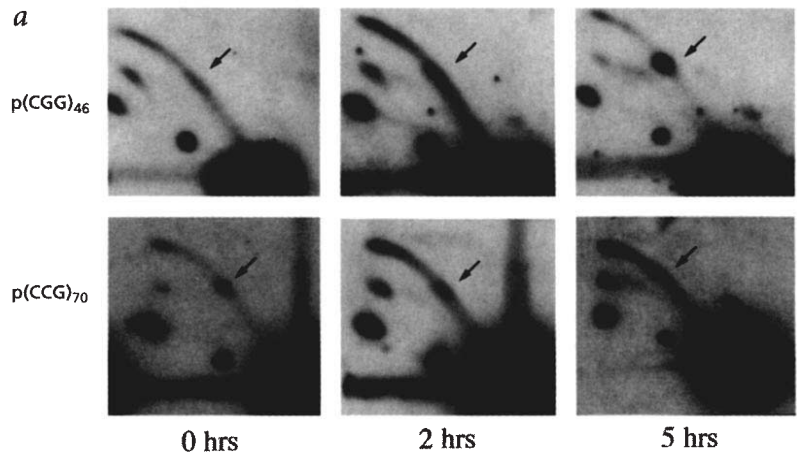
actual film density to the expected film density of a smooth replication arc at this point. The strength of the stops depended on both the length and the orientation of the repeat (Fig. 2b). For p(CGG)<sub>n</sub> plasmids, stops are evident but weak for 14 < n < 31, but become four times stronger for n = 50. For p(CCG)<sub>n</sub> plasmids, stops are barely distinguishable for n ≤ 50, but become strong at n = 70.

The length and orientation dependence of (CGG)<sub>n</sub>-(CCG)<sub>n</sub> repeats on replication suggests that the observed inhibition is due to their DNA structure. Protein-binding could not be ruled out as a mechanism, however. To address this concern, we analysed the replication of our plasmids in bacterial cells in the presence of chloramphenicol, which blocks the replication of the bacterial chromosome but not ColE1-type plasmids<sup>39</sup>. Consequently, plasmid DNA amplifies, while the protein content remains, at most, constant. If protein binding were essential, replication stops should diminish under chloramphenicol treatment. Prolonged chloramphenicol treatment (for five hours) does not abolish replication arrest for either p(CGG)<sub>46</sub> or p(CGG)<sub>70</sub> (Fig. 3), and even enhances the stop-intensity for the p(CGG)<sub>46</sub> plasmid. We believe, therefore, that protein binding is not responsible for the replication blockage.



**Fig. 2** Analysis of replication intermediates of plasmids with different (CGG)<sub>n</sub>-(CCG)<sub>n</sub> inserts. **a**, Two-dimensional gel electrophoresis of replication intermediates. Plasmids are named according to the sequence of the lagging strand template. Arrows show replication stop sites. **b**, Semi-quantitative analysis of the replication stop intensity from several experiments. The strength of the stop is characterized by the ratio of the observed density of the stop signal to the expected density of the smooth arc at this position ( $D_{obs}/D_{exp}$ ). Filled squares, p(CGG)<sub>n</sub> plasmids; open squares, p(CCG)<sub>n</sub> plasmids.

**Fig. 3** The effect of chloramphenicol on plasmid replication. **a**, Two-dimensional gel electrophoresis of replication intermediates. Plasmids are named as in Fig. 2a. Incubation time with chloramphenicol is indicated at the bottom of each column. Arrows show replication stop sites. **b**, Semi-quantitative analysis of replication stops from three experiments. Filled columns, p(CGG)<sub>46</sub>; open columns, p(CCG)<sub>70</sub>.



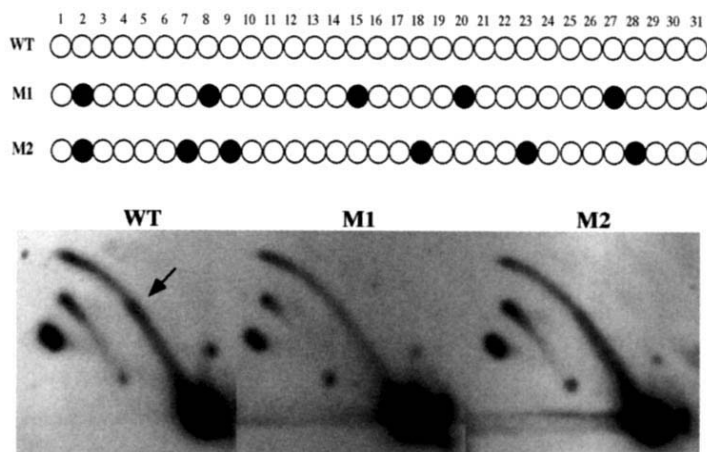
Previous analysis of fragile-X repeats from different human DNAs revealed that (CGG)<sub>n</sub> stretches are commonly interrupted by several AGG trinucleotides<sup>20</sup>. Twenty-four uninterrupted CGG blocks are believed to be prone to expansion<sup>20</sup>. To determine how AGG interruptions within the (CGG)<sub>n</sub> stretch would affect replication *in vivo*, we generated these repeats by PCR, using (CGG)<sub>5</sub>AGG(CGG)<sub>4</sub> and (CCG)<sub>5</sub>CCT(CCG)<sub>4</sub> primers to generate multiply interrupted (CGG)<sub>n</sub>·(CGG)<sub>n</sub> stretches. After cloning and sequencing, we picked two clones with five different AGG substitutions within the (CGG)<sub>31</sub> stretch (Fig. 4). Analysis of replication intermediates of such plasmids demonstrated that the AGG mutations abolish the replication arrest. Mutant 1, in which the longest consecutive (CGG)<sub>n</sub> block is six repeats, does not block replication, whereas mutant 2, with the longest consecutive (CGG)<sub>n</sub> block of eight repeats, shows traces of replication stops.

Because our inserts were located in the transcribed area, we were not certain whether their effects on replication were direct or mediated by transcription through these repeats. To address this question, we cloned two (CGG)<sub>n</sub>·(CCG)<sub>n</sub> repeats in both orientations into a mini-pBR plasmid (Fig. 5a). Our inserts were located upstream of the *amp* promoter P3 (ref. 40) in the non-transcribed area. The patterns of replication-fork movement were qualitatively similar to those observed for the pTrc derivatives (Fig. 5b). Replication arrest for (CGG)<sub>n</sub> repeats in the lagging strand was

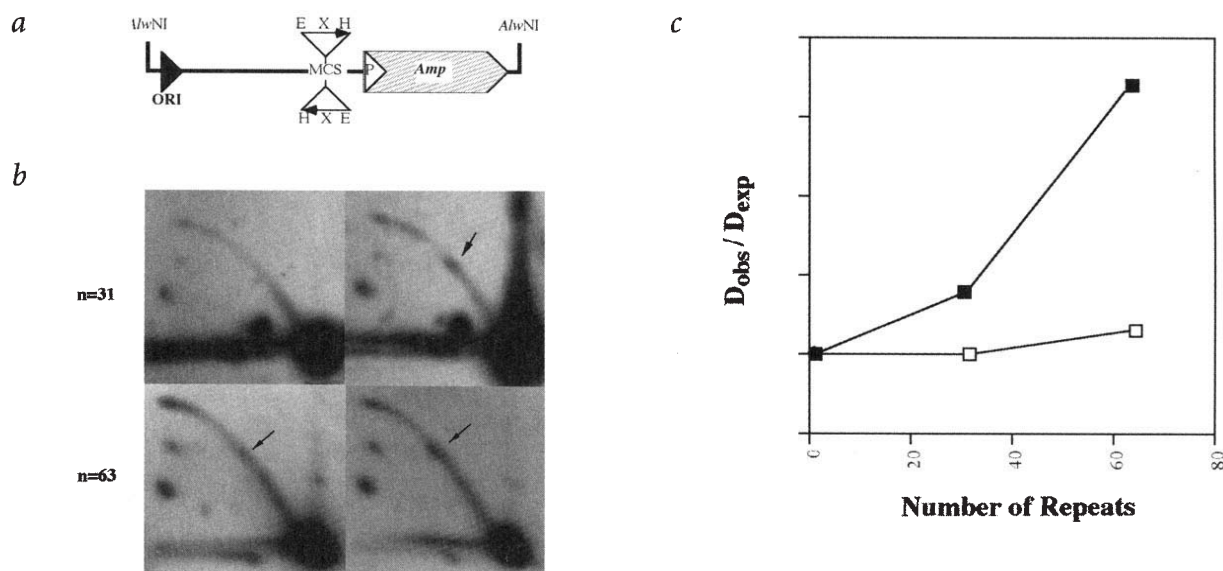
evident at n=31 and four times stronger at n=63 (Fig. 5c). In contrast, for (CGG)<sub>n</sub> repeats in the lagging-strand template, stops became evident only at n=63. We believe, therefore, that this replication blockage is not due to transcription.

To study the relation between the stop sites and the repeated inserts, we used a modified version of electrophoretic analysis of replication intermediates (Fig. 6a; ref. 41). After the first dimension of electrophoresis, replication intermediates were digested with a restriction enzyme in the gel. As a result, some bubble-shaped intermediates (marked by filled squares) converted into identical Y-shaped intermediates. In the second dimension of the electrophoresis, these intermediates migrate identically and are detected as a horizontal line upon hybridization with a probe adjacent to *ori*. If the replication fork is stalled within a repeat, restriction cleavage downstream of the repeat (relative to *ori*) would leave the stop on the bubble arc, while upstream cleavage would shift the stop to the horizontal line.

Cleavage of replication intermediates with *Eco*R1 (located downstream from the insert) leaves the replication stop on the bubble arc, while cleavage by upstream *Hind*III shifts the stop onto the horizontal line (Fig. 6b; spot 1). We conclude, therefore, that the replication fork is stalled within the (CGG)<sub>63</sub>·(CCG)<sub>63</sub> stretch. *Hind*III cleavage also results in the appearance of an additional spot (spot 2) which co-migrates with spot 1 in the first dimension but migrates more slowly in the second dimension. Although the nature of spot 2 is unknown, we speculate that it may reflect replica-



**Fig. 4** Effects of AGG interruptions in the (CGG)<sub>31</sub> stretch on replication blockage. The upper panel schematically presents the repeat composition, where open and solid circles correspond to CGG and AGG trinucleotides, respectively. The lower panel shows the 2D analysis of replication intermediates.



**Fig. 5** Analysis of replication intermediates of plasmids with  $(CGG)_n \cdot (CCG)_n$  inserts. **a**, Structure of the mini-pBR derivative. **b**, Two-dimensional analysis of replication intermediates. Plasmids are named according to the sequence of the lagging-strand template. The arrow shows replication stops. **c**, Semi-quantitative analysis of the replication stop strength. Filled squares,  $p(CGG)_n$ ; open squares,  $p(CCG)_n$ .

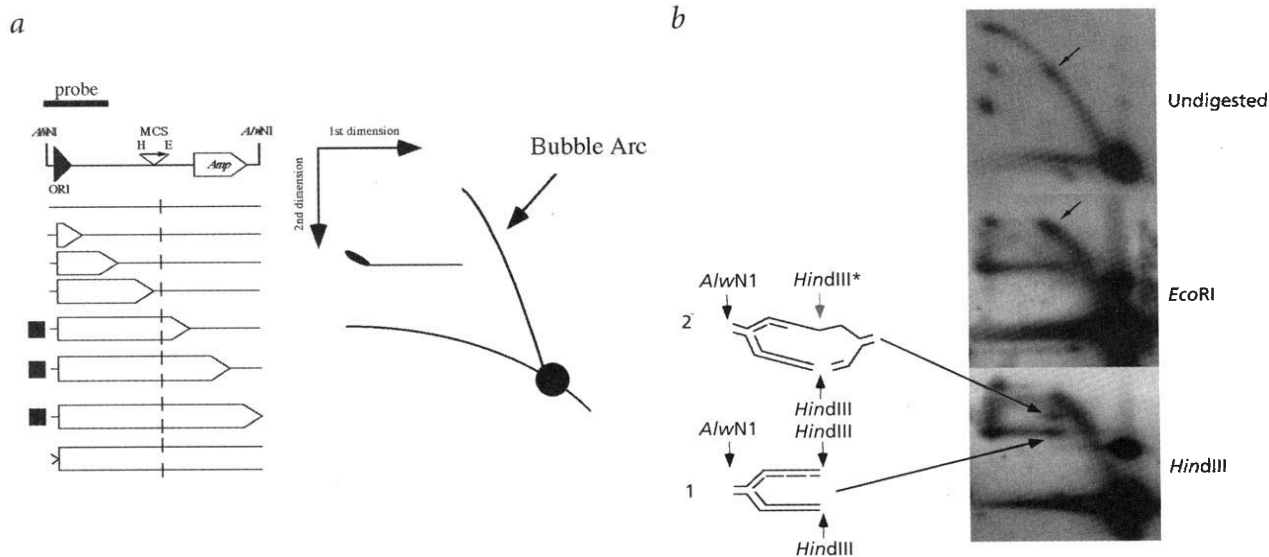
tion intermediates in which a portion of the lagging strand around the *HindIII* site was not synthesized. Partial *HindIII* digest of such intermediates would lead to the accumulation of butterfly-shaped DNA molecules (see diagram).

**Effects of repeats on plasmid replication**

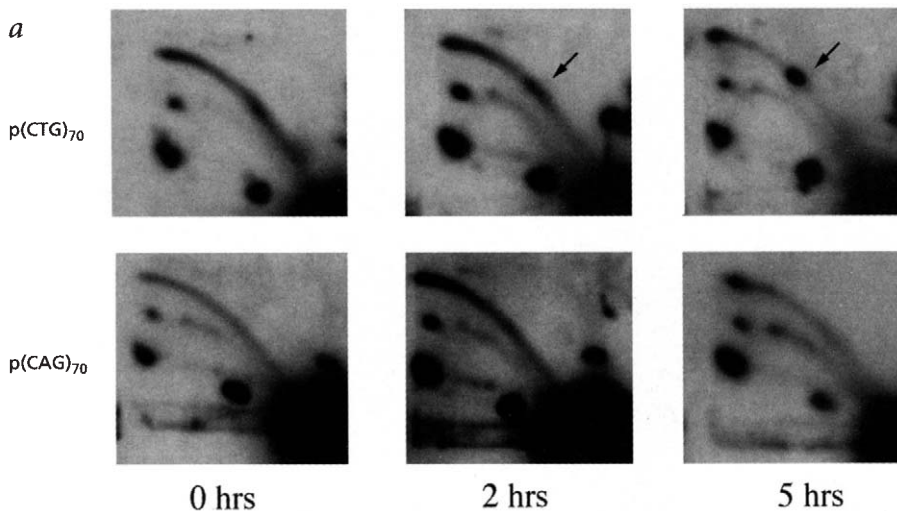
The most common trinucleotide repeat to undergo expansion in many human genes is  $(CTG)_n \cdot (CAG)_n$ . To determine whether these repeats would arrest replication in a manner similar to that found for the  $(CGG)_n \cdot (CCG)_n$  repeats, we generated  $(CTG)_n \cdot (CAG)_n$  stretches of varying length by PCR using  $d(CTG)_{10}$  and  $d(CAG)_{10}$  primers, and cloned them into the pTrc99Δ plasmid described above (Fig. 1) in two orientations relative to *ori*.

Analysis of the replication intermediates of these plasmids showed that only long inserts affected the replication-fork movement (data not shown). Moreover, inhibition of replication was evident only in the presence of the protein synthesis inhibitor chloramphenicol even for the longest inserts. The  $(CTG)_{70}$  repeat in the lagging-strand template causes very weak replication fork pausing; the stop signal is somewhat more evident after two hours with chloramphenicol, but becomes fivefold stronger after five hours of treatment (Fig. 7). In contrast, the  $(CAG)_{70}$  repeat in the lagging-strand template does not affect replication fork progression, regardless of chloramphenicol treatment.

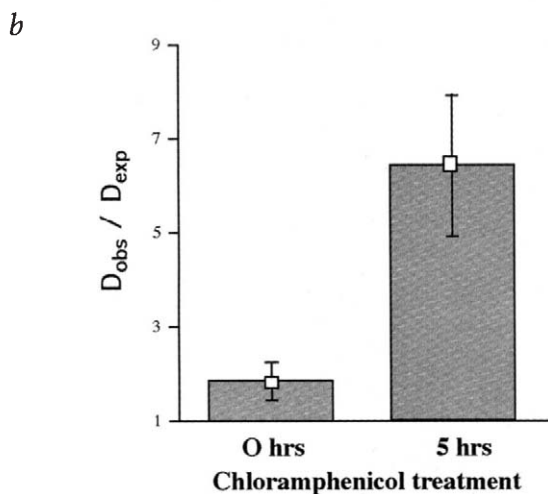
These experiments indicate that  $(CTG)_n \cdot (CAG)_n$  repeats cause a length- and orientation-dependent replication blockage, similar to that caused by  $(CGG)_n \cdot (CCG)_n$ . The inhibitory effects of



**Fig. 6** Fine mapping of the replication stop sites in the  $p(CGG)_{63}$  plasmid. **a**, Schematic representation of two dimensional gel electrophoresis upon restriction cleavage after the first dimension. The dotted vertical line shows the restriction cleavage. Filled squares mark identical Y-shaped intermediates resulting from cleavage. The right panel shows a horizontal line corresponding to the migration of the latter intermediates. **b**, Electrophoretic separation of the replication intermediates. Upper panel, no digestion; middle panel, *EcoRI* digestion; bottom panel, *HindIII* digestion. Small arrows show stop sites on the bubble arc. Long arrows show stop sites moving towards the horizontal line. Schematic representation of the structure of the intermediates in spots 1 and 2 are presented.



**Fig. 7** Analysis of replication intermediates in the presence of chloramphenicol. **a**, Two-dimensional gel electrophoresis of replication intermediates. Plasmids are named according to the sequence of the lagging-strand template. Incubation time with chloramphenicol is indicated at the bottom of each column. Arrows show replication stops. **b**, Semi-quantitative analysis of replication stop strength from four independent experiments for the p(CTG)<sub>70</sub> plasmid.



(CTG)<sub>n</sub>·(CAG)<sub>n</sub> repeats on replication, however, are weaker and become evident only upon protein synthesis blockage.

**Discussion**

Our results show that expandable (CGG)<sub>n</sub>·(CCG)<sub>n</sub> and (CTG)<sub>n</sub>·(CAG)<sub>n</sub> repeats block replication fork-progression *in vivo*, and this effect depends on the repeats' length, their orientation relative to the replication origin and the ratio of protein/DNA synthesis in bacterial cells. Fine mapping of the stop sites revealed that they are located within the repeated DNA sequences. The exact mechanisms of replication blockage are unknown, and several possibilities must be considered.

In principle, replication blockage could be due to the tight protein binding to repeated DNA. However, our data—which show that the protein synthesis inhibitor enhances, rather than abolishes, replication blockage—make this explanation highly unlikely. It is equally unlikely that the blockage is due to stalling of the RNA polymerase within long repeats<sup>42,43</sup>, because we observed virtually the same effects whether the repeats were in the transcribed or non-transcribed areas. The stability of trinucleotide repeats depends on the activity of the mismatch-repair system in bacterial and yeast cells<sup>44,45</sup>. Thus, it is possible that misalignments between the template and the novel DNA strands of a repeat during replication could impede the replication fork. This hypothesis, however, fails to explain the dependence of replication blockage on repeat orientation.

We believe, therefore, that the DNA structure of trinucleotide

repeats is directly responsible for the replication fork blockage. Individual strands of expandable trinucleotide repeats are able to form hairpin-like structures, whose stability is length dependent<sup>14</sup>. DNA hairpins significantly affect polymerization kinetics<sup>46,47</sup>. Our data on the threshold repeat length causing replication blockage (Fig. 2b) are in agreement with the data on the threshold of hairpin formation<sup>14</sup>. Thus, hairpin formation by expandable repeats could account for our observations.

The orientation dependence could be due to the differences in the mode of leading and lagging DNA strand synthesis<sup>36</sup>. Discontinuous synthesis of the lagging strand implies that a portion of the lagging-strand template must be transiently single-stranded, so that a trinucleotide repeat has a better chance to form a secondary structure when in the lagging-strand template. If this is correct, our data predict that the structural potential of different trinucleotide repeats *in vivo* is  $d(\text{CGG})_n > d(\text{CCG})_n > d(\text{CTG})_n > d(\text{CAG})_n$ . This prediction is in qualitative agreement with the data on hairpins formed by trinucleotide repeats *in vitro*<sup>10,13,14</sup>. The synthesis of both DNA strands during replication is believed to be co-ordinated<sup>36</sup>, so the arrest of the lagging-strand synthesis should stop the replication fork as a whole. This notion is additionally supported by our data on the mapping of replication stops (Fig. 6b), indicating the under-replication of the lagging strand.

Our hypothesis can also explain the increased blockage of replication in the presence of chloramphenicol. While an unusual DNA structure formed by repeated DNA might challenge DNA polymerase, the structure could still be resolved by accessory replication proteins, such as helicases or the single-strand binding protein<sup>36</sup>. We speculate that under chloramphenicol treatment, when plasmid DNA greatly amplifies while the protein level is at best stagnant, accessory proteins may be lacking, enhancing the negative effects of unusual secondary structures.

Are these results on replication in bacteria relevant to the repeat expansion in humans? Two of our observations may indicate that this is so. First, the threshold length of (CGG)<sub>n</sub>·(CCG)<sub>n</sub> repeats causing replication blockage ( $n \approx 30$ ) is very similar to the threshold length required for expansion *in vivo* ( $n \approx 25$ ; ref. 20). Second, it is believed that AGG interruptions within the (CGG)<sub>n</sub> repeat in the *FMR1* gene prevent expansion, and they alleviate replication blockage in our system.

The abnormalities of expandable repeat replication—and specifically, the imbalance between the leading and the lagging strand synthesis—is believed to contribute to repeat expansion. Our results provide the first direct proof that trinucleotide repeats affect replication and that the equilibrium between the leading- and lagging-strand synthesis is involved. Thus, our working model

is that the replication fork first stalls within the repeated DNA, which may eventually lead to its expansion/deletion.

It is obvious how the replication blockage can cause deletions, and our data are in agreement with the data on deletion formation in bacteria<sup>27,28,31</sup>. However, it is less obvious how replication stalling could lead to expansion. One explanation is that replication-fork blockage could occasionally lead to hairpin formation by the newly synthesized strand, resulting in expansion<sup>28</sup>. Another model is that the formation of a hairpin on the lagging-strand template leads to the priming of an extra Okazaki fragment, causing expansion<sup>2</sup>. Alternatively, the unprocessed 5' flaps of the Okazaki fragments within these repeats might lead to expansion<sup>48,49</sup>. Finally, the accumulation of trapped replication intermediates could stimulate recombination, provoking expansion. At present, we cannot distinguish between these possibilities.

## Methods

**Oligonucleotides.** Complementary oligodeoxyribonucleotides d(CGG)<sub>10</sub>/d(CCG)<sub>10</sub>, d(CGG)<sub>3</sub>AGG(CGG)<sub>4</sub>/d(CCG)<sub>5</sub>CCT(CCG)<sub>4</sub> and d(CTG)<sub>10</sub>/d(CAG)<sub>10</sub> were synthesized and purified as previously described<sup>50</sup>.

**Plasmids.** Plasmid pTrc99A was obtained by substitution of an *EcoRV*-*SspI* fragment of the pTrc99A<sup>37</sup> with an *SmaI*-*HincII* portion of the pBluescript SK(-) (Stratagene) multiple cloning site. A mini-pBR derivative was obtained by replacement of an *EcoRI*-*NruI* fragment of the pET15B (Novagen) with a synthetic polylinker.

**PCR.** Complementary oligonucleotides (500 nmol) were used in a standard 100- $\mu$ l PCR reaction. Vent (exo<sup>-</sup>) polymerase (4–6 U) was used for 15–20 rounds of amplification (1 min at 94 °C, 1 min at 37 °C and 2.5 min at 72 °C) in a Thermal Cycler 480 (Perkin Elmer). Because of the repeated nature of our oligonucleotides, primers could serve as templates and PCR products could serve as primers, resulting in a wide range of PCR products of different lengths. Long PCR products were cloned into the *EcoRV* site of pBluescript SK(-) and sequenced. Subsequently, the repeats were re-cloned into the pTrc99A plasmid.

**Bacteria.** Plasmids were maintained in the XL1-Blue *Escherichia coli* strain (Stratagene). Cells were grown in LB medium with 100  $\mu$ g/ml of ampicillin at 37 °C until the mid-logarithmic stage ( $A_{600} \approx 0.6$ ). Alternatively, cells were grown until early logarithmic stage ( $A_{600} \approx 0.2$ ), followed by the addition of 170  $\mu$ g/ml of chloramphenicol.

**Isolation of replication intermediates.** Cell cultures were rapidly cooled with ice-cold physiological solution and collected by centrifugation at 5,000 g. The cell pellet was washed in cold STE buffer, resuspended in 5 ml of cold 25% w/v sucrose, 0.25 M Tris-HCl pH 8.0, 10 mg/ml lysozyme, 0.1 mg/ml RNase A and incubated on ice for 5 min; 1 ml of 0.5 M EDTA was added, and the cell suspension was incubated on ice for 5 min. Cells were lysed by addition of 8 ml of 1% v/v Brij-58, 0.4% w/v sodium deoxycholate, 63 mM EDTA, 50 mM Tris-HCl pH 8.0. Lysates were incubated on

ice for 10–20 min and centrifuged at 36,000 g for 1 h at 4 °C. Plasmid DNA was precipitated from the supernatant by addition of 11 ml of 1.5 M NaCl, 25% PEG 8000. Proteins were hydrolyzed in 0.1% Sarcosyl, 100  $\mu$ g/ml proteinase E for 20 min at 37 °C, followed by phenol extraction. Plasmid DNA was then precipitated with ethanol.

Enrichment in replication intermediates was achieved by gradient centrifugation. The DNA sample was supplemented with CsCl (1 g/1 ml) and 200  $\mu$ g/ml of propidium iodide. Centrifugation was carried out at 400,000 g for 18 h. Replicative intermediates migrated between the supercoiled and linear DNA.

**Two dimensional gel electrophoresis.** Replication intermediates were digested by *AlwNI*, loaded on a 0.4% agarose gel in TBE and run at 0.7 V/cm for 36 h at room temperature with continuous buffer recirculation. The slice of the gel containing replication intermediates was cut out and implanted into a 1% agarose gel in TBE with 0.6  $\mu$ g/ml of ethidium bromide. The second direction was run perpendicular to the first one for 5–6 h at 4–5 V/cm at 4 °C. The gel was treated with 0.25M HCl for 10–15 min, followed by denaturation/renaturation, transfer onto a Zeta-Probe GT membrane (Bio-Rad) and hybridization with a plasmid.

**Mapping of the replication stop sites.** After the first dimension of the electrophoresis, the slice of 0.4% agarose gel (Sea Kem LE) containing replication intermediates was soaked twice in TE for 30 min and twice in  $\times 1$  restriction buffer for 30 min at room temperature. In-gel digestion was carried out in 5 ml of  $\times 1$  restriction buffer containing 3,000 U of an appropriate restriction enzyme for 15 h at 37 °C. Subsequently, the gel slice was washed twice with 15 ml of TE and embedded in 1.5% agarose for the second dimension of the electrophoresis.

**Semi-quantitative analysis of replication arcs.** Replication arcs developed on X-ray film (Fuji Medical) were digitally analysed with a Kohu KM-1 linear CCD camera and the NIH Image software. Calibration to standard amounts of radioactive label showed that the intensity of film exposure was proportional to the label over two orders of magnitude, including most of our readings from the 2D gels. To estimate the intensity of an arc at different positions, we sampled the film density along the arc and normalized the results to the background density of the film. The resultant values were plotted against their relative distances along the arc, generating a density profile of the replication arc, where local maxima corresponded to stops. The strength of the stop was measured as the ratio of its density to the expected density of a smooth arc at this position ( $D_{\text{obs}}/D_{\text{exp}}$ ).

## Acknowledgements

We thank S. Warren, C. McMurray, R. Wells, M. Mitas, R. Rothstein, J. Dahlberg and R. Davidson for discussions and R. Cox for critical reading of the manuscript. The work was supported by grants from the National Science Foundation (MCB-9405794) and from the Council for Tobacco Research (#4468) to S.M.M.; G.R. is a recipient of the Dorothea H. Fleming Student Fellowship.

Received 18 June; accepted 19 September 1997.

- Ashley, C., Jr. & Warren, S.T. Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.* **29**, 703–728 (1995).
- McMurray, C.T. Mechanisms of DNA expansion. *Chromosoma* **104**, 2–13 (1995).
- Wells, R.D. Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.* **271**, 2875–2878 (1996).
- Bates, G. & Lehrach, H. Trinucleotide repeat expansions and human genetic disease. *Bioessays* **16**, 277–284 (1994).
- Caskey, C.T., Pizzuti, A., Fu, Y.-H., Fenwick R.G., Jr. & Nelson, D.L. Triplet repeat mutations in human disease. *Science* **256**, 784–789 (1992).
- Cox, R. & Mirkin, S.M. Characteristic enrichment of DNA repeats in different genomes. *Proc. Natl. Acad. Sci. USA* **94**, 5237–5242 (1997).
- Kohwi, Y., Wang, H. & Kohwi-Shigematsu, T. A single trinucleotide, 5'AGC3'/5'GCT3', of the triplet-repeat disease genes confers metal ion-induced non-B DNA structure. *Nucleic Acids Res.* **21**, 5651–5655 (1993).
- Bacolla, A. et al. Flexible DNA: genetically unstable CTG-CAG and CGG-CCG from human hereditary neuromuscular disease genes. *J. Biol. Chem.* **272**, 16783–16792 (1997).
- Mitchell, J.E., Newbury, S.F. & McClellan, J.A. Compact structures of d(CNG)<sub>n</sub> oligonucleotides in solution and their possible relevance to fragile X and related human genetic diseases. *Nucleic Acids Res.* **23**, 1876–1881 (1995).
- Petruska, J., Arnheim, N. & Goodman, M.F. Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res.* **24**, 1992–1998 (1996).
- Yu, A. et al. The trinucleotide repeat sequence d(GTC)<sub>15</sub> adopts a hairpin conformation. *Nucleic Acids Res.* **23**, 2706–2714 (1995).
- Yu, A. et al. At physiological pH, d(CCG)<sub>15</sub> forms a hairpin containing protonated cytosines and a distorted helix. *Biochemistry* **36**, 3687–3699 (1997).
- Zheng, M., Huang, X., Smith, G.K., Yang, X. & Gao, X. Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding: an NMR and UV spectroscopic study. *J. Mol. Biol.* **264**, 323–336 (1996).
- Gacy, A.M., Goellner, G., Juranic, N., Macura, S. & McMurray, C.T. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**, 533–540 (1995).
- Chen, X. et al. Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* **92**, 5199–5203 (1995).
- Fry, M. & Loeb, L.A. The fragile X syndrome d(CGG)<sub>n</sub> nucleotide repeats form a stable tetrahedral structure. *Proc. Natl. Acad. Sci. USA* **91**, 4950–4954 (1994).
- Kettani, A., Kumar, R.A. & Patel, D.J. Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 638–656 (1995).

18. Kuryavyi, V.V. & Jovin, T.M. Triad-DNA: a model for trinucleotide repeats. *Nature Genet.* **9**, 339–341 (1995).
19. Reiss, A.L. et al. Frequency and stability of the fragile X premutation. *Hum. Mol. Genet.* **3**, 393–398 (1994).
20. Kunst, C.B. & Warren, S.T. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* **77**, 853–861 (1994).
21. Snow, K., Tester, D.J., Kruckeberg, K.E., Schaid, D.J. & Thibodeau, S.N. Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum. Mol. Genet.* **3**, 1543–1551 (1994).
22. Jodice, C. et al. Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia I. *Am. J. Hum. Genet.* **54**, 959–965 (1994).
23. Trinh, T.Q. & Sinden, R.R. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. *Nature* **352**, 544–547 (1991).
24. Kang, S., Ohshima, K., Shimizu, M., Amirhaeri, S. & Wells, R.D. Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes. *J. Biol. Chem.* **270**, 27014–27021 (1995).
25. Ohshima, K. & Wells, R.D. Hairpin formation during DNA synthesis primer realignment *in vitro* in triplet repeat sequences from human hereditary disease genes. *J. Biol. Chem.* **272**, 16798–16806 (1997).
26. Usdin, K. & Woodford, K.J. CGG repeats associated with DNA instability and chromosome fragility from structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.* **23**, 4202–4209 (1995).
27. Kang, S., Ohshima, K., Jaworski, A. & Wells, R.D. CTG triplet repeats from the myotonic dystrophy gene are expanded in *Escherichia coli* distal to the replication origin as a single large event. *J. Mol. Biol.* **258**, 543–547 (1996).
28. Kang, S., Jaworski, A., Ohshima, K. & Wells, R.D. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nature Genet.* **10**, 213–218 (1995).
29. Bowater, R.P., Rosche, W.A., Jaworski, A., Sinden, R.R. & Wells, R.D. Relationship between *Escherichia coli* growth and deletions of CTG/CAG triplet repeats in plasmids. *J. Mol. Biol.* **264**, 82–96 (1996).
30. Ohshima, K., Kang, S. & Wells, R.D. CTG triplet repeats from human hereditary diseases are dominant genetic expansion products in *Escherichia coli*. *J. Biol. Chem.* **271**, 1853–1856 (1996).
31. Shimizu, M., Gellibolian, R., Oostra, B.A. & Wells, R.D. Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. *J. Mol. Biol.* **258**, 614–626 (1996).
32. Webb, T. Delayed replication of Xq27 in individuals with the fragile X syndrome. *Am. J. Med. Genet.* **43**, 1057–1062 (1992).
33. Hansen, R.S., Canfield, T.K., Lamb, M.M., Gartler, S.M. & Laird, C.D. Association of fragile X syndrome with delayed replication of the FMR1 gene. *Cell* **73**, 1403–1409 (1993).
34. Subramanian, P.S., Nelson, D.L. & Chinault, A.C. Large domains of apparent delayed replication timing associated with triplet repeat expansion at FRAXA and FRAXE. *Am. J. Hum. Genet.* **59**, 407–416 (1996).
35. Brewer, B.J. & Fangman, W.L. The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* **51**, 463–471 (1987).
36. Kornberg, A. & Baker, T. *DNA Replication*, 2nd ed. (W.H. Freeman & Co., New York, 1992).
37. Amann, E., Ochs, B. & Abel, K.J. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene* **69**, 301–315 (1988).
38. Martin-Parras, L., Hernandez, P., Martinez-Robles, M. & Schwartzman, J.B. Unidirectional replication as visualised by two-dimensional agarose gel electrophoresis. *J. Mol. Biol.* **220**, 843–855 (1991).
39. Clewell, D.B. Nature of Col E1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. *J. Bacteriol.* **110**, 667–676 (1972).
40. Balbas, P. et al. Plasmid vector pBR322 and its special-purpose derivatives—a review. *Gene* **50**, 3–40 (1986).
41. Friedman, K.L. & Brewer, B.J. Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. *Methods Enzymol.* **262**, 613–627 (1995).
42. Reaban, M.E., Lebowitz, J. & Griffin, J.A. Transcription induces the formation of a stable RNA-DNA hybrid in the immunoglobulin  $\alpha$  switch region. *J. Biol. Chem.* **269**, 21850–21857 (1994).
43. Reaban, M.E. & Griffin, J.A. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature* **348**, 342–344 (1990).
44. Strand, M., Prolla, T.A., Liskay, R.M. & Petes, T.D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
45. Jaworski, A. et al. Mismatch repair in *Escherichia coli* enhances instability of (CTG) $n$  triplet repeats from human hereditary diseases. *Proc. Natl. Acad. Sci. USA* **92**, 11019–11023 (1995).
46. Hacker, K.J. & Albers, B.M. The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix: a model for polymerase release following the termination of each Okazaki fragment. *J. Biol. Chem.* **269**, 24221–24228 (1994).
47. Krasilnikov, A.S. et al. Mechanisms of triplex-caused polymerization arrest. *Nucleic Acids Res.* **25**, 1339–1346 (1997).
48. Gordenin, D.A., Kunkel, T.A. & Resnick, M.A. Repeat expansion—all in a flap? *Nature Genet.* **16**, 116–118 (1997).
49. Tishkoff, D.X., Filosi, N., Gaida, G.M. & Kolodner, R.D. A novel mutation avoidance mechanism dependent on *S. cerevisiae* RAD27 is distinct from DNA mismatch repair. *Cell* **88**, 253–263 (1997).
50. Samadashwily, G.M., Dayn, A. & Mirkin, S.M. Suicidal nucleotide sequences for DNA polymerization. *EMBO J.* **12**, 4975–4983 (1993).