

Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells

John D. Cleary^{1,2}, Kerrie Nichol¹, Yuh-Hwa Wang³ & Christopher E. Pearson^{1,2}

Published online: 22 April 2002, DOI: 10.1038/ng870

The mechanism of disease-associated trinucleotide repeat instability involves *cis*-acting factors (*cis*-elements) in the vicinity of the repeat, but the nature of these elements is unknown. One *cis*-element may be the location of the replication origin relative to the repeat. We have used an SV40 DNA replication system to investigate the effect of the location of replication initiation on (CTG)_n•(CAG)_n stability in primate cells. Depending on the distance between the SV40 replication origin and the repeat tract, templates with 79 repeats yield predominantly expansions or predominantly deletions or remain intact. All templates with 17 repeats are stable. Thus, *cis*-elements that affect the sites of Okazaki fragment initiation relative to the repeat are crucial determinants of instability. This model system recapitulates the bias for expansions observed in many of the diseases associated with trinucleotide repeats. Our results might explain the variable amounts of CTG/CAG instability that are observed in different chromosomal contexts.

Introduction

The expansion of certain gene-specific trinucleotide repeats (TNRs) is a form of mutation that is responsible for at least 15 human diseases. Twelve of these, including myotonic dystrophy type 1 (DM1), spinobulbar muscular atrophy and Huntington disease^{1,2}, are caused by unstable (CTG)_n•(CAG)_n repeats. In the nonaffected population, the length of the repeat tracts is generally polymorphic, ranging from 4 to 24 repeat units, and these lengths are transmitted stably. Lengths with more than 34 repeats are genetically unstable. Depending on the gene, expansion to 25–35 or 34–90 repeats can lead to disease or be asymptomatic. Asymptomatic expansions are known as intermediate, pre- or protomutation alleles³.

In DM1, premutation lengths can expand rapidly to disease-associated lengths that contain between a hundred and thousands of repeats. The genetic instability of the repeats is sensitive to the length of the repeat tract, as the products of an expansion mutation are more likely to undergo a subsequent mutation than is the original substrate. These genetic alterations are therefore known as dynamic mutations^{4,5}. In individuals with DM1, the expanded (CTG)_n tract shows an extraordinarily large amount of somatic mosaicism, owing to high somatic instability both within and between tissues. Somatic instability becomes apparent during fetal development, is evident during the life of affected individuals^{6,7} and can be detected in cultured DM1 cells^{8,9}. Although both deletions and expansions are observed somatically^{6,7} and during transmission¹⁰, there is a bias towards expansion. The mechanism of (CTG)_n instability for myotonic dystrophy and other TNR-associated diseases is unknown.

In addition to the repeat itself, the sequences in the vicinity of the repeat (*cis*-elements) contribute to the mechanism of instability for several disease loci^{2,11}. Complete allelic association between the DM1 expansion and several insertion/deletion polymorphisms flanking the repeat^{12,13} show that a specific chromosomal background is associated with (CTG)_n instability, which suggests that chromosomal context and *cis*-elements may be required for expansion. The variable stability of similar CAG tract lengths at different disease loci provides further support for a role of flanking sequences^{2,11,14,15}. Transgenic mouse models of TNR instability also indicate that *cis*-elements that flank the repeat tract may 'drive' repeat instability^{16,17}. The nature of such *cis*-elements is unknown and has remained elusive for over a decade^{2,11,18}.

Bacterial (*Escherichia coli*)^{19,20} and yeast (*Saccharomyces cerevisiae*)^{21–24} models support a role for DNA replication in TNR instability in that the direction of replication through the TNR tract affects its stability. In bacterial and yeast systems, repeat deletions predominate regardless of the direction of replication, but the frequency of deletions is higher when the CTG strand, rather than the CAG strand, is the template for lagging-strand synthesis¹⁹. Unlike in humans, in which the repeats tend to expand, the tendency for repeats to be deleted in bacteria may be caused by differences in the replication fork dynamics between single-cell organisms and primates. For example, bacterial Okazaki fragments are considerably longer (approximately 2,000 nt) than those formed in mammalian cells (135–145 nt)^{25–27}. Analysis of TNR instability in a primate DNA replication system is lacking.

¹Program of Genetics & Genomic Biology, The Hospital for Sick Children, 555 University Avenue, Elm Wing 11-135, Toronto, Ontario M5G 1X8, Canada.

²Department of Molecular & Medical Genetics, University of Toronto. ³Department of Biochemistry, University of Medicine & Dentistry of New Jersey Robert Wood Johnson Medical School, Piscataway, New Jersey, USA. Correspondence should be addressed to C.E.P. (e-mail: cepearson@genet.sickkids.on.ca).



To understand TNR instability in a primate system, we established an SV40 DNA replication assay. A detailed understanding of primate replication fork dynamics has been gained through the SV40 viral DNA replication system^{28,29}. Each replication fork comprises a continuous leading strand and a discontinuous lagging strand, which consists of several Okazaki fragments each of roughly 135–145 nt (Fig. 1a)²⁵. Unlike synthesis at the leading strand, which is maintained mostly as a duplex, a portion of the lagging-strand template must be rendered single-stranded to permit the priming of an Okazaki fragment. For SV40, this region of single-stranded template DNA of roughly 290 nt is defined as the Okazaki initiation zone (OIZ; Fig. 1a)^{25,30}. Each Okazaki fragment is initiated by RNA priming, extended by polymerase- α to an RNA–DNA primer, which is extended further by polymerase- δ . Synthesis is complete on reaching the initiator RNA of the previous Okazaki fragment. The processing of Okazaki fragments involves removal of the RNA primer, gap filling and ligation^{25,29}. An alternative, but not mutually exclusive model for the production of Okazaki fragments is the ‘nested discontinuity model’ (Fig. 1a)³¹, whereby several initiations of short primer RNA–DNAs (approximately 34 nt) are processed and ligated. Specific template sequences might determine whether full-length or nested Okazaki fragments are produced.

Current models of somatic (CTG)_n•(CAG)_n instability involve aberrant processing of Okazaki fragments that initiates within the repeat tract^{32–34}. Because Okazaki processing occurs only at sites of Okazaki initiation, we thought that *cis*-elements that affect the frequency of Okazaki initiations in the repeat tract might affect repeat stability. Thus, *cis*-elements that modulate replication fork dynamics with respect to the TNR tract would be determinants of repeat instability.

To test our hypothesis, we used the SV40 DNA replication system to determine the effect of three *cis*-elements on (CTG)_n•(CAG)_n instability: first, the sequence (CTG or CAG) of the Okazaki template, which is determined by the direction of replication; second, repeat tract length and third, location of the SV40 replication origin relative to the repeat tract. The CTG or CAG template might experience sequence-specific preferences for Okazaki-specific events. Longer repeat tracts would occupy a larger proportion of the single-stranded OIZ and might experience increased Okazaki initiation events. Altering the distance between the replication origin and the repeat will alter the location of the repeat within the single-stranded OIZ. Placing the repeat tract at the upstream or downstream regions of the single-stranded OIZ might increase or decrease the frequency of initiation events within the repeat.

We found that all three variables affect repeat stability. Unexpectedly, changing only the location of replication initiation by roughly 130 nt relative to the repeat tract, while maintaining the direction of replication, resulted in profound differences in the type of mutation observed (expansions only versus deletions only). Because DNA templates were replicated in cells that are competent in DNA replication and repair activities, our results provide strong evidence for the participation of *cis*-elements in TNR instability.

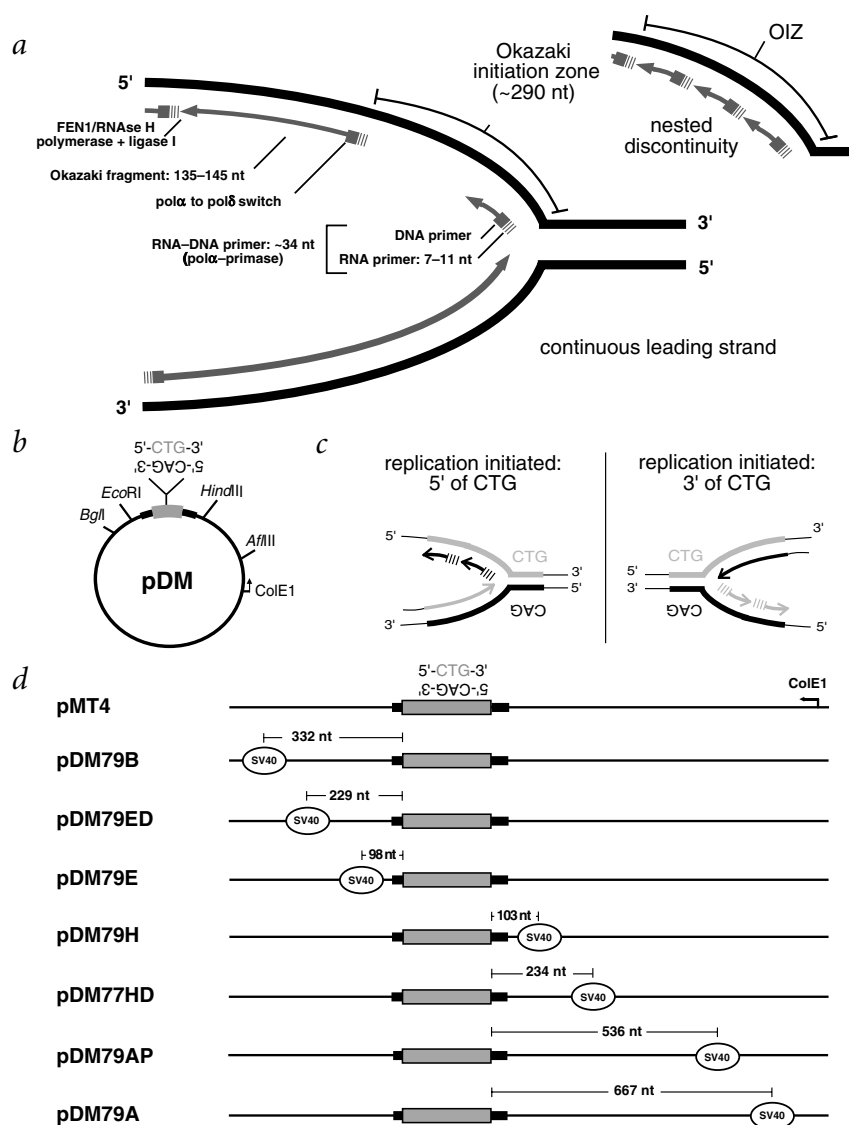


Fig. 1 Replication fork and replication templates. **a**, Each replication fork comprises a continuous and a discontinuous lagging strand. A portion of the lagging-strand template must be rendered single-stranded to permit Okazaki initiation; this is the OIZ^{28,30}. Shown are two different but not mutually exclusive mechanisms for producing Okazaki fragments: one produces a fragment with an average length of 135–145 nt; the other produces a fragment through several initiations of short primer RNA–DNAs (upper right)³¹. **b**, The templates contained human DM1 genomic (CTG)_n and (CAG)_n repeats (*n*=17 or 79), with human nonrepeating sequences (bold lines) flanking the repeat (sites 417–436 and 451–494 from GenBank accession number S86455). Repeats are in the stable orientation relative to the bacterial unidirectional ColE1 origin of replication. Each clone contained one copy of the SV40-ori. The SV40-ori fragment inserted at B (*Bgl*I), E (*Eco*RI), H (*Hind*III) or A (*Afl*III) contains the OBR, the unique nucleotide (SV40 site 5,210/5,211) at which replication initiation occurs³⁰. Templates are drawn to scale. **c**, Location of the SV40-ori relative to the repeat defines the direction of replication and which repeat strand (CTG or CAG) will serve as the leading- or lagging-strand template. **d**, Distance between the SV40-OBR and the repeat tract. The location and cloning orientation of the SV40-ori fragment determines the distance between the SV40-OBR (enclosed circle) and the repeat tract. The distance to the repeat tract was measured from the SV40-OBR to the nucleotide preceding the 5' or 3' end of the CTG repeat tract. Shown are premutation templates (79 repeats); distances are drawn to scale.

Results

A primate replication assay for stability

To investigate the role of primate DNA replication on TNR stability, we used the SV40 replication system that, in addition to being a model of the primate replication fork^{28,29}, has been used to study replication fidelity³⁵, DNA mutagenesis and DNA repair. Replication of SV40 plasmids in transfected cells uses both T-antigen and the replication machinery of the host cell. Replication initiates in the SV40 origin of replication (SV40-*ori*) and begins at a unique nucleotide position—the origin of bidirectional replication (OBR)³⁰. Replication proceeds bidirectionally, with both forks progressing at similar rates.

Inserting the SV40-*ori* closer to one side of the repeat tract (Fig. 1b) determines the direction of replication fork progression through the repeats and thus which strand (CTG or CAG) will serve as the lagging-strand template (Fig. 1c). To analyze the effect that location of replication initiation might have on repeat stability, we inserted the SV40-*ori* at various locations on either side of the repeat tract, producing a series of SV40-DM1 shuttle vectors (Fig. 1d). Three of the replication templates used the CTG strand as the lagging-strand template and four templates used the CAG strand as the lagging-strand template.

To study expansions from normal (5–30 repeats) to premutation lengths (34–90 repeats) and expansions of premutation to disease-associated lengths (approximately 100 repeats), we constructed genomic DM1 clones^{36–38} with either 17 or 79 (CTG)_n repeats (Fig. 1d). After their transfection into T-antigen-expressing primate cells (COS-1)³⁹, the DM1/SV40 templates were replicated by primate proteins and assayed for repeat stability.

Mutation analysis—the STRIP assay

We developed an assay to determine the role of primate DNA replication on the stability of trinucleotide repeats by individual product (STRIP) analysis (Fig. 2). We isolated the products of replication from transfected COS-1 cells⁴⁰ and digested them with *DpnI* to eliminate unreplicated parental DNAs. *DpnI* digests only DNAs that are methylated on both strands on adenine residues at GATC sites, a modification that occurs in *dam*⁺ bacteria but not in primate cells. Only *DpnI*-resistant material—the products of primate replication—is capable of transforming bacteria⁴¹. Individual bacterial colonies therefore represent individual products of primate replication.

Plasmids isolated from the individual bacterial colonies were digested with restriction enzymes to release the fragment containing the repeat (Fig. 3a), separated on high-resolution polyacrylamide gels and scored for expansion or deletion events (a representative example is shown in Fig. 3b). Expansion events are observed as distinct bands that migrate more slowly than the starting (CTG)₇₉ band, whereas deletion events are observed as bands that migrate faster. To confirm that these events were caused by changes in the repeat tract, we electrotransferred and probed the samples with ³²P-labeled (CTG)₁₅ oligonucleotide (Fig. 3c). We also sequenced individual mutation products, which confirmed that the observed changes were limited to increases or decreases in the integral number of repeat units.

To focus on mutation events mediated by primate replication, we needed to correct for both the low amount of repeat length heterogeneity that is present in the template preparation³⁸ and the limited heterogeneity that is produced during amplification of miniprep plasmids (see Methods). In both human and model systems, TNR tract lengths are best described as a distribution. The background levels for each template were determined by direct bacterial transformation of the same starting template DNA used for primate cell transfection. We analyzed individual molecules (colonies) for repeat length (Fig. 3) and classified them into one of three categories: 'less than 79 repeats', '79 repeats' and 'greater than 79 repeats' (Fig. 4). The molecules in each of these categories could experience primate replication-mediated expansion or deletion events. After primate replication, the length distribution of the replicated material for each template was determined (Fig. 4).

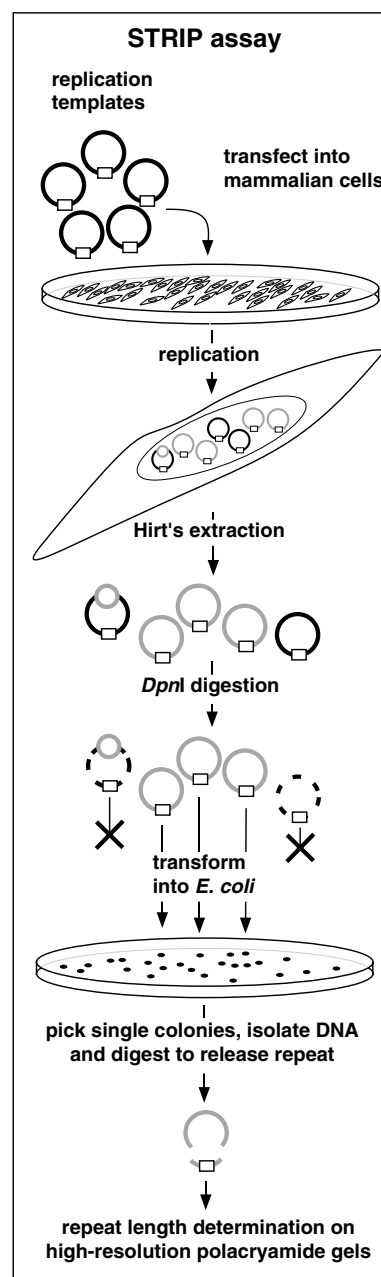


Fig. 2 Experimental strategy. Templates replicated in primate cells were analyzed by STRIP assay. COS-1 cells were transfected with replication templates. After a 48-h replication period, material was collected by Hirt's lysis and digested with *DpnI*. The *DpnI*-resistant material (gray molecules) represents products of complete replication by primate cell proteins. *DpnI* digestion eliminated unreplicated (shown in bold) or partially replicated material (shown in gray and bold). Reactions were transformed into *E. coli*. Only the *DpnI*-resistant material yielded colonies, from which DNAs were isolated and the repeat-containing fragment released and analyzed for length on high-resolution polyacrylamide gels.

The frequencies of primate replication-mediated expansion and deletion (Table 1) were corrected by subtracting the background frequencies. Primate replication products that did not show a significant change (χ^2 , $P>0.05$) in the number of molecules in either of the length categories relative to the parental material are classified as stable or having '0' primate replication-induced events (Table 1). Templates that showed a statistically significant difference (χ^2 , $P<0.05$) in only the number of primate cell-replicated molecules with 'greater than 79 repeats' or 'less than 79 repeats' are classified as having a bias for expansions or deletions, respectively.

An increased percentage of molecules in one category is necessarily coupled with a decrease in the percentage of molecules in at least one other category (no increased events after primate replication or '0'; Table 1). A template that showed a significantly increased number of molecules in both the 'less than 79' and the 'greater than 79 repeats' categories experienced both deletions and expansions, which was necessarily coupled with a decreased proportion of molecules with 79 repeats. This rigorous approach indicated that the observed spectrum of expansion or contraction products is the result of primate cell-mediated mutations and not a function of the distribution of repeat arrays in the starting material.

Expansion of the premutation template pDM79H

For pDM79H, replication initiates 103 nt to the 3' side of the CTG tract (Fig. 1d), and the CAG strand serves as the lagging-strand template. Replication of pDM79H in primate cells resulted in mutations that were predominantly expansions at a frequency of 9% (Table 1). Comparison of only primate replication products with 'greater than 79 repeats' with background

molecules having 'greater than 29 repeats' confirmed a significant bias for expansions ($P=0.016$, $\chi^2=5.8$) after primate replication. The largest expansions after primate replication for pDM79H were increases of +28 repeats (mean +8), which were far greater than the largest molecule of +8 repeats (mean +5) observed in the same parental template preparation. The primate cell-replicated expansion products were also significantly larger (Wilcoxon two-sided test, $P=0.02$) than the molecules contained in the starting material. Although bacteria can stably propagate plasmids with as many as 255 (CTG) \bullet (CAG) repeats³⁶, expansions in bacteria are relatively rare¹⁹. Thus, both the frequency of expansions and the magnitudes of repeats gained are statistically significant, which is consistent with a bias for expansions generated by primate replication of this template.

Direct analysis of replication products

We also analyzed the pDM79H primate replication products by three independent direct methods. Owing to its inherent lack of sensitivity, direct Southern-blot analysis of the Hirt's replicated material did not detect changes in repeat length (data not shown). We analyzed individual products of replication by colony PCR, which eliminates bacterial-induced instability during culture (Fig. 5a). The PCR results showed a bias for expansions (11% frequency, $\chi^2=7.09$, $P=0.029$), which is consistent with the 9% expansion frequency detected by STRIP analysis (Table 1).

We analyzed replicated DNAs by electron microscopy (Fig. 5b–d), which for small DNA fragments can detect length differences of as few as 30–60 bp (10–20 repeats)³⁸. The replicated material contained longer, but not shorter, molecules ranging from 65.6 nm to 159.6 nm (mean 117.5 ± 25.9 nm), whereas the parental material ranged from 65.8 nm to 152.02 nm (mean 100 ± 13.4 nm). The replicated material also contained a greater proportion of longer molecules relative to the parental material (Fig. 5d); thus, the electron microscopy results support qualitatively both the STRIP and PCR results. The length distribution difference observed by electron microscopy and the mutation spectra generated by both the PCR and STRIP analyses are therefore consistent with a bias for expansions for primate replication of pDM79H.

DNA replication actively contributed to (CTG) \bullet (CAG) instability. To show that DNA replication is essential for the observed repeat instabilities, we carried out control experiments in which replication was not permitted. Plasmid replication in transfected cells requires both the presence of the SV40 replication origin and the expression of SV40 T-antigen. We therefore transfected plasmid pMT4, which contains 79 CTG and CAG repeats but no SV40-ori (Fig. 1d), into COS-1 cells, and the SV40 replication template pDM79H into CV-1 cells (precursor cells of COS-1), which do not express SV40 T-antigen³⁹. In both cases, the transfected DNAs were passed through the primate cells without replicating, and STRIP analysis of the recovered episomal DNAs (without *DpnI* digestion) did not show any significant repeat expansions or deletions (Table 1). Thus, the observed repeat instability is mediated by replication in primate cells and does not arise through repair or recombination events that are independent of replication forks.

Replication direction

To test the effect of Okazaki template sequence, which is defined by replication direction, we constructed plasmid pDM79E by inserting the SV40-ori on the opposite side of the repeat tract, as in pDM79H (Fig. 1d). For pDM79E, replication initiated approximately 100 nt from the repeat tract, similar to pDM79H, but pDM79E used the lagging-strand template instead of the CAG strand used by pDM79H.

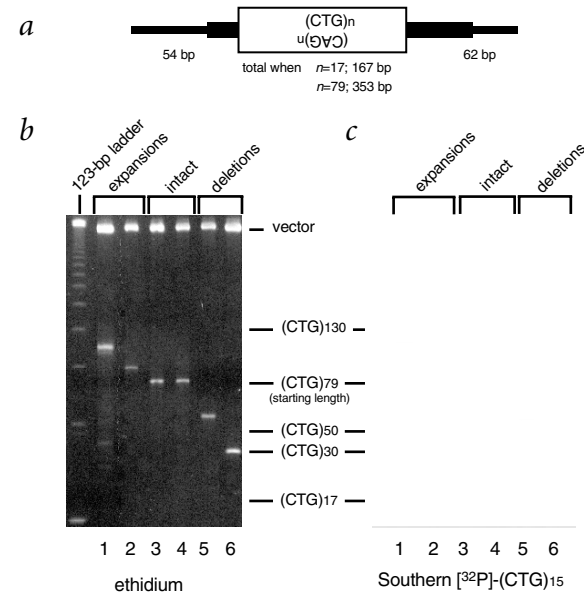


Fig. 3 Representative example of STRIP analysis. **a**, Map of the repeat-containing fragment from the DM1 plasmid replication templates. Human nonrepetitive flanking sequences are shown as thick lines; plasmid vector sequence is shown as a thin line. **b**, Primate replication products were isolated and processed as shown in Fig. 2. Individual bacterial colonies represent individual products of primate replication. Plasmids isolated from individual bacterial colonies were digested with restriction enzymes to release the repeat-containing fragment, resolved on high-resolution 4% polyacrylamide gels, and scored for expansion (lanes 1 and 2), intact (lanes 3 and 4) and deletion (lanes 5 and 6) events. Replication products are shown for pDM79 replication templates. **c**, Southern blot of gel shown in **a** hybridized with ³²P-labeled (CTG)₁₅. Scored results were corrected for background as shown in Fig. 4 and are given in Table 1.

Replication of pDM79E resulted in mutations that were predominantly deletions (13%, Table 1). The magnitude of repeats deleted ranged from –3 to –69 repeats (mean –24 repeats). The shift from a bias for expansions in pDM79H to a bias for deletions in pDM79E indicates that the direction of primate replication has a marked effect on repeat instability.

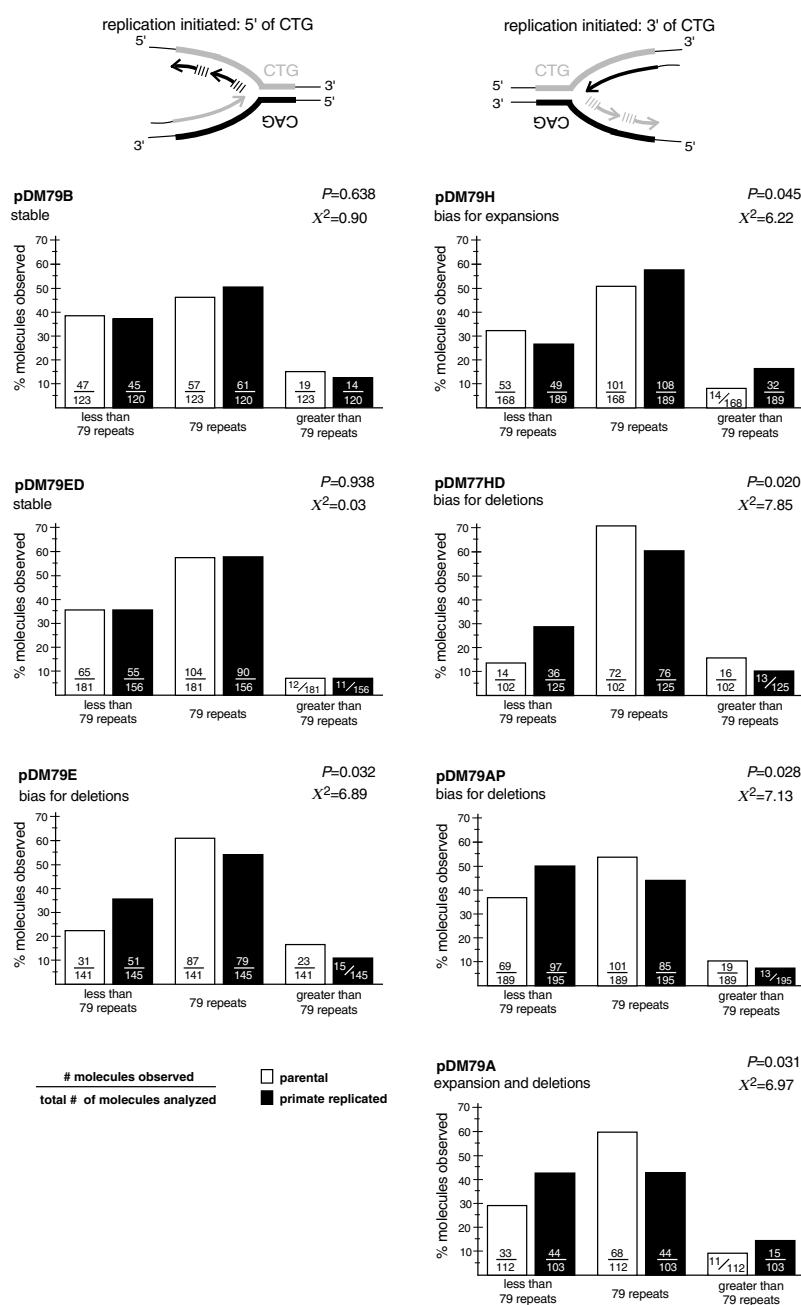
Location of replication initiation

Altering the location of the SV40-ori relative to the repeat tract may affect the frequency of Okazaki fragment initiations that occur in the repeat tract, which in turn may affect instability. To test the effect of replication initiation site on repeat stability, we placed the SV40-ori at varying distances from the repeat tract (Fig. 1d). For pDM79A (Fig. 1d), replication proceeded through the repeat in the same direction as in pDM79H but initiated further from the repeat. Replication of pDM79A resulted in a

greater overall frequency of mutation compared with replication of pDM79H (a total of 18% versus 9%) and showed both expansions (5%) and deletions (13%, Table 1). For pDM79B (Fig. 1d), replication proceeded in the same direction as pDM79E but initiated further from the repeat tract. Replication of pDM79B did not result in significant amounts of expansion or deletion, in contrast to the deletions (13%) obtained with replication of pDM79E. These results indicate that the position of the SV40-ori and therefore the location of replication initiation relative to the repeat tract may contribute to TNR instability.

Small adjustments in the distance between the location of replication initiation (SV40-OBR) and the repeat resulted in profound differences in stability (Table 1). In pDM77HD (Fig. 1d), replication initiated roughly 130 nt further from the repeat tract than in pDM79H. Replication of pDM77HD yielded mutations that were predominantly deletions (15%), in marked contrast to the bias for

Fig. 4 Mutation analysis. The bacterial preparation of the parental template molecules contained a distribution of repeat tract lengths, each of which was determined by direct bacterial transformation. Lengths of individual molecules were then assessed (Fig. 3) and classified into three categories: 'less than 79 repeats', '79 repeats' and 'greater than 79 repeats' (open bars). After primate replication, the length distribution of the replicated material for each template was similarly determined (filled bars). To facilitate the comparison of the relative proportions of the length distributions between parental and primate cell-replicated DNAs, the bar height (y axis) reflects the percentage of molecules within each length category. For each length category, the number of molecules observed out of the total number of molecules analyzed is indicated inside the bars. After primate replication, templates for which there was no significant change (χ^2 -test, $P > 0.05$) in the length distribution were classed as stable. Templates with statistically significant (χ^2 -test, $P < 0.05$) increased numbers of only molecules with 'greater than 79 repeats' were classed as having a bias for expansions, whereas templates with significant increased numbers of only molecules with 'less than 79 repeats' were classed as having a bias for deletions. The primate replication-mediated expansion and deletion frequencies reported in Table 1 were corrected by subtracting the background for the same template preparation used for primate cell transfection. After primate replication of pDM79A, for example, the frequency of expansion events, 'greater than 79 repeats', was calculated as $5\% = [(15/103) - (11/112) / (15/103 + 33/112)] \times 100$; the frequency of deletion events, 'less than 79 repeats', was calculated as $13\% = [(44/103) - (33/112) / (44/103 + 33/112)] \times 100$. Indicated in the figure are the χ^2 statistics and P values for each template (2 degrees of freedom). By using this rigorous experimental approach and treatment of results, we have focused on the mutations mediated by primate cells.



expansions (9%) derived from pDM79H (Table 1). Because this shift in mutation bias resulted from a change in the location of replication initiation (Fig. 1d) and not in the direction of replication, the location of initiation must be a major determinant of instability. Similarly, pDM79ED (Fig. 1d) was stable (Table 1), in contrast to the bias for deletions (13%) derived from pDM79E, which initiated replication roughly 130 nt closer to the repeat tract (Fig. 1d). We also found an effect of initiation location on instability for initiation sites further from the repeat tract. Replication of pDM79AP (Fig. 1d) produced mutations that were predominantly deletions (13%), in contrast to the mixture of expansions (5%) and deletions (13%) observed from replication of pDM79A, which initiated replication approximately 130 nt from the repeat tract (Fig. 1d). The differences in mutation bias caused by minor changes in the distance between replication initiation and the repeat, without changing the direction of replication, underscores the contribution of *cis*-elements to repeat instability.

Repeat length

To determine the effect of repeat length on instability, an identical series of pDM replication templates containing 17 (CTG)_n•(CAG)_n repeats were replicated in COS-1 cells. All of the pDM17 templates were completely stable after primate DNA replication (Table 1). Similarly, a template with 30 repeats—a length that is at the high end of the normal and stable range—yielded no significant mutations. But many of the templates containing 79 (CTG)_n•(CAG)_n repeats showed some degree of instability. For each of the unstable pre-mutation templates, the

mutation frequency and the magnitude of repeat units gained or lost was always greater than that observed by bacterial replication. The expansions and deletions derived from many of the templates containing 79 repeats, and the lack of instability in any of the 17 or 30 repeat templates, indicated that there is an association between repeat length and instability; that is, longer repeat tracts are more prone to mutation.

Discussion

We have developed a rapid and efficient assay to study the effect of primate DNA replication on the stability of trinucleotide repeat tracts in living cells. Instability required replication mediated by primate cells. We used this system to analyze the effect of *cis*-elements on the stability of (CTG)_n•(CAG)_n repeats. The factors that we have examined include the length of the repeat tract, the direction of replication and the location of replication initiation. We found that these factors influence both the type of mutation (expansions versus deletions) and the frequency of mutations.

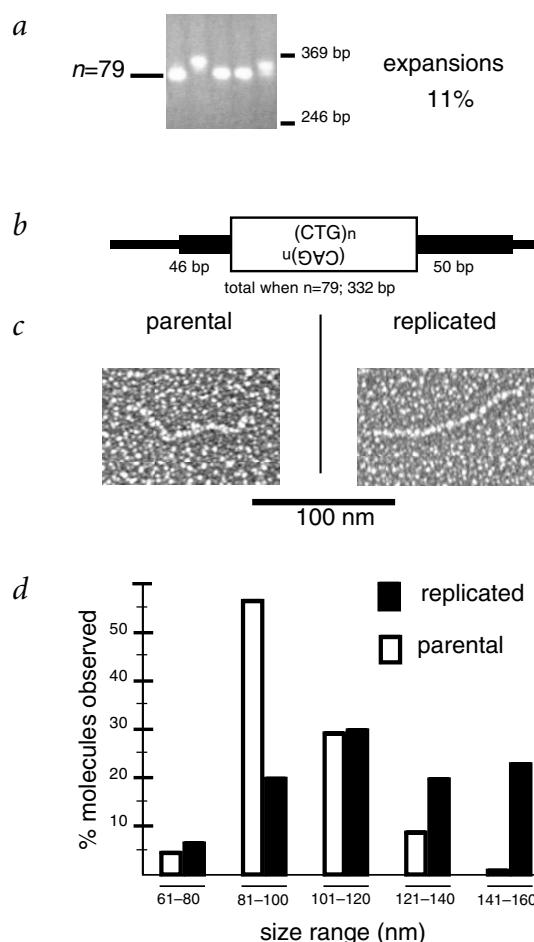
Notably, for a specific pre-mutation template, we observed a high frequency of predominantly expansion mutations, the detection of which did not require a sensitive selection-based assay. In bacterial and yeast models of TNR instability, the predominant mutations are invariably deletions, with only rare instances of expansion^{19,21}. Along with primate cell-specific differences, the major reason we observe a bias for expansion mutations for one of seven pre-mutation templates was the location of the replication origin relative to the repeat tract—a variable that has not been examined previously.

Table 1 • Corrected frequencies following primate replication

Templates	Repeat length	Distance from SV40-ori (nt)	Total colonies	Frequency (%)		
				expansions	deletions	
Replication initiated 5' to CTG						
pDM79B	79	332	120	0	0	
pDM79ED	79	229	156	0	0	
pDM79E	79	98	145	0 ^a	13	
Replication initiated 3' to CTG						
pDM79H	79	103	133	9	0 ^a	
pDM77HD	77	234	125	0 ^a	15	
pDM79AP	79	536	195	0 ^a	13	
pDM79A	79	667	103	5	13	
pDM17 E, H, A and B	17	see above	> 40 each	0	0] replication not permitted
pDM79H in CV-1	79	103	52	0	0	
pMT4	79	NA	64	0	0	

Corrected mutation frequencies were calculated only for repeat length distributions that were significantly different from the parental template background (by χ^2 -test or Fisher's exact test; see Fig. 4). Repeat length distributions that were not significantly different from the parental background ($P>0.05$) were classed as stable or having no primate cell-induced events ('0'). Expansion and deletion frequencies mediated by primate cell replication were corrected for background by subtracting the frequencies of 'greater than 79 repeats' or 'less than 79 repeats' events for the parental template molecules from the frequencies of the same category for replicated products. For those events that increased after replication, the positive value obtained is reported, whereas those events that decreased after replication are reported as '0^a' (no increased events).

Fig. 5 Independent analyses of primate cell-replicated material. **a**, PCR analysis of individual pDM79H replication products (see Methods). Shown are some of the electrophoretically resolved products and the corrected frequency of mutations. Of a total of 106 primate cell-replicated molecules analyzed, 16 had 'greater than 79 repeats' and 25 had 'less than 79 repeats'. When corrected for parental template background (of a total of 109 molecules analyzed, 5 had 'greater than 29 repeats' and 33 had 'less than 79 repeats'), a significant bias for expansions (11% frequency, $\chi^2=7.09$, $P=0.029$) was observed. **b–d**, Electron microscopic analysis of pDM79H primate replication products. **b**, Hirt's extracted replication products were linearized by *SacI* digestion, the unreplicated material was digested by *DpnI*, and the *DpnI*-resistant *SacI* material was resolved by agarose gel electrophoresis and gel purified. This material was then digested with *PstI* to release the short repeat-containing fragment. To account for the limited amounts of contaminating genomic DNA⁴⁰, Hirt's lysates were prepared from mock-transfected cells and purified as above. With the exclusion of *DpnI* digestion, the bacterial parental template was treated identically. **c**, Reverse contrast electron micrographs of DNAs. Scale bar, 100 nm. **d**, Determination of the contour lengths of individual repeat-containing fragments (see Methods). Individual parental ($n=124$, open bars) and primate cell-replicated molecules ($n=30$, filled bars) were batched according to size and scaled according to the percentage of molecules with that size. The length distribution of the replicated material is significantly different (longer) than that of the parental material (Wilcoxon two-sided test, $P=0.0005$).



Effect of tract length

In individuals affected by TNR-associated disease, both the unstable transmission and the somatic instability (expansions and deletions) of CTG/CAG repeats are sensitive to the length of the repeat tract, which is a hallmark of dynamic mutations^{1,2,5}. We observed a similar association between the length of the repeat tract and repeat instability. All of the templates containing 17 (CTG)•(CAG) repeats were stable, whereas five out of seven templates containing 79 (CTG)•(CAG) repeats showed instability.

Most notably, the pDM79H template showed a tendency for expansion, which is, to our knowledge, the first experimental observation of a predominance for expansions. The stability of a template containing 30 repeats—a length that has been observed rarely to be genetically unstable in humans—places the stability threshold between 30 and 79 repeats. Thus, our primate replication system recapitulates both the length-dependent instability and the bias for expansions seen in affected humans.

Effect of replication direction

We observed that the direction of replication affected the stability of (CTG)•(CAG) repeats. Bacterial and yeast models of TNR instability tend to delete repeat tracts, with only rare instances of expansion^{19,21,22,24}. In these systems, deletions predominate regardless of the direction of replication; however, the frequency of deletions is greater when the CTG strand is the template for lagging-strand synthesis^{19,21,22,24}. Unlike bacterial and yeast models^{19,21,22}, the effect of replication direction in our primate system manifested as differences in mutation type (expansions versus deletions). Changing the direction of replication while maintaining the distance of initiation to the repeat resulted in marked shifts from biases for expansions to biases for deletions (compare pDM79H with pDM79E) or shifts from biases for deletions to stable intact repeats (compare pDM77HD with pDM79ED; Table 1 and Fig. 6). Thus, changing only the direction of replication while maintaining the location of initiations has a marked effect on TNR instability.

Effect of location of replication initiation

Systematic analysis of the effect of location of replication initiation on TNR stability showed that profound differences in repeat stability result from changing the location of replication initiation. Of seven pre-mutation templates, one showed predominantly expansions, three showed predominantly deletions, one showed both expansions and deletions and two were stable. The

location of initiation was as strong a determinant of repeat stability as was replication direction. Changing the location of replication initiation while maintaining the direction of replication resulted in marked shifts from biases for expansions to biases for deletions (compare pDM79H with pDM77HD) or shifts from biases for deletions to stable intact repeats (compare pDM79E with pDM79ED; Table 1 and Fig. 6).

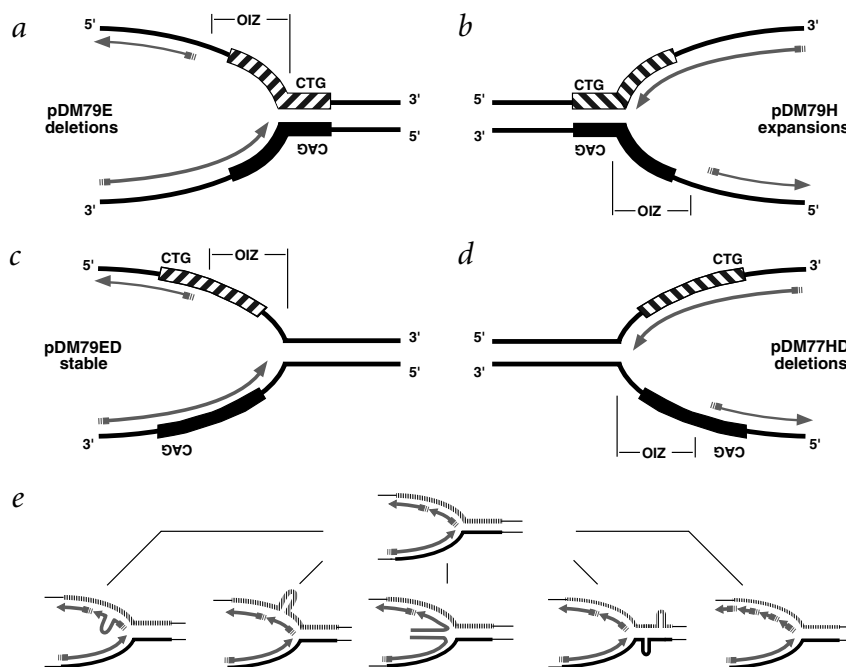
Only long tracts were unstable, and expansions were detected primarily for a template that used CAG as the lagging-strand template, which provides support for the hairpin model^{19,42} that is based on the increased thermodynamic stability of the CTG hairpin over the CAG hairpin. However, deletions and not expansions were detected for a different template (pDM77HD) that used the CAG as the lagging-strand template (but that was located in a different portion of the OIZ). This shows that the location of the repeat with respect to the OIZ is a more crucial determinant of repeat instability than is the differential ability to form hairpins.

Why were types of mutation events altered by a minor change (approximately 130 nt) in the location of replication initiation? Several factors might contribute to this location-dependent effect, and understanding the processes that occur at a primate replication fork may provide some insight.

Replication fork dynamics and dynamic mutations

Repeat instability might require the initiation of Okazaki fragments within particular locations of the repeat tract, which would result in an Okazaki fragment with repeats at its 3' or 5' termini. Various factors contribute to the selection of sites used for Okazaki initiation and affect whether initiation occurs within

Fig. 6 Replication fork dynamics and instability of CTG and CAG repeats. The location and the sequence of the repeat in the OIZ^{25,30} determines both instability and mutation type (expansion versus deletion). The OIZ is the region of roughly 290 nt of single-stranded, lagging-strand template DNA at the replication fork^{25,30}. The locations of the repeat tracts within the indicated OIZ were determined from the location of the SV40-OBR³⁰ where replication is initiated. **a**, For pDM79E, the OBR is 98 nt from the repeat, the last two-thirds of the OIZ consists of CTG repeats, and replication yields predominantly deletions. **b**, For pDM79H, the OBR is 103 nt from the repeat, the last two-thirds of the OIZ consists of CAG repeats, and replication results in a bias for expansions. **c**, For pDM79ED, the OBR is 229 nt from the repeat, the first two-thirds of the OIZ consists of CTG repeats, and repeat tracts were stably replicated. **d**, For pDM79HD, the OBR is 234 nt from the repeat tract, the first two-thirds of the OIZ consists of CAG repeats, and replication yields predominantly deletions. **e**, The location of Okazaki fragment initiation may determine the propensity to form specific mutagenic intermediates, the aberrant metabolism of which may result in repeat instabilities. Longer, rather than shorter (rightmost), nascent repeat tracts at Okazaki termini may more readily form these structures.



the repeat or not. Human and SV40 Okazaki fragments have an average length of 135–145 nt (refs 25,27,30), and their priming is coordinated with leading-strand synthesis^{28,29} (Fig. 1a). Priming occurs within an OIZ²⁸, which is a region of single-stranded template DNA of approximately 290 nt (Fig. 1a).

The boundaries of the OIZ are defined at the downstream end by the initiation site of the previous Okazaki fragment and at the upstream end by the progressing fork. Short repeat tracts would occupy a small proportion of the OIZ, whereas premutation repeat tracts (roughly 80 repeats or 240 nt) or longer tracts could span the OIZ and might experience increased initiation events. Changing the location of the OBR³⁰ will alter the boundaries of the OIZ, which in turn will alter the sites available for Okazaki priming and processing. Altering the distance of the replication origin from the repeat will probably affect the location of the repeat in the single-stranded OIZ (Fig. 6). This may determine where Okazaki priming will occur within the repeat. Priming at the beginning of the repeat tract will result in an Okazaki fragment with a long stretch of repeats at its termini, whereas priming at the end of the tract will result in a short stretch of repeats.

In our system, moving the location of the SV40-OBR relative to the pre-mutation repeat tract had marked effects on the stability of the repeat. For pDM79H, in which the last two-thirds of the OIZ consists of CAG repeats (Fig. 6b), we found a bias for expansions. In contrast, for pDM77HD, in which the first two-thirds of the OIZ comprises CAG repeats (Fig. 6d), we observed predominantly deletions. Similarly, we saw predominantly deletions for pDM79E, in which the last two-thirds of the OIZ consists of CTG repeats (Fig. 6a). The repeat was stable for pDM79ED, in which the first two-thirds of the OIZ comprises CTG repeats (Fig. 6c). Clearly, marked alterations in stability result from moving the site of initiation by increments of an Okazaki fragment (approximately 130 nt) relative to the repeat tract (Fig. 1d). Expansion events seem to require priming within a particular location of the template CAG repeat tract, which results in an Okazaki fragment that terminates in a long tract of CTG repeats.

How can the location of the repeat tract within the OIZ determine repeat instability? An Okazaki fragment that terminates in a long tract of repeats might facilitate the formation of muta-

genic intermediates. Such intermediates (Fig. 6e) include hairpins⁴² or slipped-strand structures^{36–38} that form between nascent and template strands; between the two nascent strands, which may occur through replication fork reversal^{43–45}; or between the two template strands, which may block or pause fork progression²⁰. The sequence (CTG or CAG) and the number of repeats at the terminus might affect both the type and the propensity of structure formation. Longer rather than shorter nascent repeat tracts at Okazaki termini might form these structures more readily. The metabolism of the structural intermediates may lead to efficient or error-prone processing by replication^{22,24,32–34}, repair^{37,46,47} or recombination proteins^{43–45,48}, all of which could lead to instability. Although other explanations are possible, it is clear that striking differences in (CTG)•(CAG) stability result from the location of replication initiation relative to the repeat tract.

Implications

Changing the location of the SV40-ori relative to the repeat tract results in profound differences in mutation. A similar location effect may explain DM1 founder chromosomes and their possible predisposition to CTG expansion^{12,13}. The location of the human chromosomal replication origin that replicates the DM1 region is not known, and it may differ between individuals affected with DM1 and nonaffected individuals. The distance between the chromosomal replication origin and the CTG repeat may be affected by the 1-kb *Alu* insertion found on all chromosomes with expanded DM1 repeats¹³. (The *Alu* insertion is 5 kb telomeric of the CTG repeat.) A similar distancing effect may be caused by other DM1 chromosome-specific insertion/deletion polymorphisms of 0.5 kb that are either telomeric or centromeric of the CTG repeat¹³. In addition to differential replication origin use, the location of the OIZ relative to the repeat tract might vary between different tissues—a factor that may explain the tissue-specific instabilities⁸.

Factors other than the location of replication origins can affect the location of priming initiation of Okazaki fragments relative to the repeat tract. Sequence variations in the proximal nonrepeating regions that flank the repeats may alter replication fork

dynamics at the repeats: differences in the density of 3'-PurineT-5' priming 'hotspots'^{30,49} on either side of the repeat tract may affect the location of Okazaki priming before, within or after the repeat tract. Sequence- or tissue-specific nucleosome positioning in the flanks may affect the priming sites of Okazaki fragments²⁸. In this manner, the different chromosomal contexts might explain the variable amounts of instability of similar lengths of CTG/CAG tracts observed at the DM1 locus compared with those observed at the locus associated with spinobulbar muscular atrophy^{2,11,14,15}.

We have presented a model system that recapitulates both the length effect and the bias for expansions observed in individuals affected with TNR-associated diseases. Because the bias for expansions occurs on a specific replication template within primate cells that are competent for DNA replication and repair processes, these results are conclusive evidence for the participation of *cis*-elements in TNR instability.

Methods

Construction of SV40 replication templates. Genomic clones of DM1 (CTG)_n have been described^{36,38}. We subcloned the *EcoRI*-*HindIII* (CTG)_n-containing fragments into pBLUESCRIPT KSII+ (Stratagene). All clones used in this study had the repeat in the stable orientation relative to the unidirectional bacterial ColE1 origin of replication (CAG strand is the lagging-strand template). We confirmed the length and purity of all (CTG)_n repeat tracts by sequencing. The *SphI*-*HindIII* fragment (viral nucleotide positions 128 and 5,171, respectively) of the SV40 virus, which contains the SV40-ori, was amplified by PCR (primers are available on request) and inserted into pBLUESCRIPT KSII+ using the *XbaI* site in the primers. This 219-nt fragment contains the SV40-OBR (viral nucleotide position 5,210/5,211)³⁰, which is located 44 nt from the viral *HindIII* site. The SV40-ori was cloned as a blunt *XbaI* fragment into the *AflIII*, *HindIII*, *EcoRI* or *BglI* (nucleotide 472, Fig. 1b) site of pBLUESCRIPT KSII+. For each clone, we confirmed the orientation of the SV40-ori by restriction analysis.

We used *dam*⁺ *E. coli* for large-scale plasmid preparations as described³⁶. Briefly, cells were collected and lysed with lysozyme (Gibco) and a detergent solution of 1% Brij 58 (Sigma) and 0.4% deoxycholate (Sigma). Plasmids were treated with RNase A and T1 (Sigma), extracted with phenol and purified twice by caesium chloride/ethidium bromide centrifugation.

Primate cell culture/transfections. COS-1 cells³⁹ were grown in DMEM supplemented with 10% (vol/vol) fetal bovine serum (FBS). A day before transfection, cells were plated to a confluency of 20% in 100-mm diameter dishes. On the following day, cells (at 40% confluency) were incubated for 4 h with 5 µg of plasmid DNA and 50 µl of Lipotaxi transfection reagent (Stratagene) in 4 ml of DMEM lacking FBS. We then added 5 ml of DMEM supplemented with 20% FBS and incubated the cells for 24 h, before replacing the medium with fresh DMEM (10% FBS). We extracted episomal DNAs 48 h after the addition of DNA as described⁴⁰ using 100 mM Tris-HCl (pH 7.5), 1 mM EDTA and 0.6% SDS. The DNA was purified with several phenol/chloroform extractions, ethanol precipitated and resuspended in 10 mM Tris-HCl (pH 7.6) and 1 mM EDTA.

Mutation analysis/STRIP assay. Mutation analysis was done with the STRIP assay as outlined in Fig. 2. Episomal DNAs isolated from the transfected cells were digested with 10 U of *DpnI*⁴¹ (New England Biolabs) for 3 h at 37 °C in 150 mM NaCl, 70 mM Tris-HCl (pH 7.6) and 1 mM dithiothreitol to eliminate unreplicated parental material. We monitored the effectiveness of *DpnI* digestion by transforming *DpnI*-digested bacterially grown plasmids into *E. coli*, which yielded no colonies. *DpnI*-digested Hirt's material⁴⁰ was transformed into *E. coli*⁴¹. We picked individual bacterial colonies, each representing individual products of primate replication, and cultured them for a limited growth period (6-h maximum, 4–6 generations) in 3 ml of Luria broth containing 30 µg ml⁻¹ ampicillin. All plasmid DNAs were isolated by Wizard SV Prep (Promega). We tested more than 15 bacterial strains, including SURE, Stbl2, HB101, XL1-Blue MR, XL1-Blue MUTS, NR8041 (*mutS101*), NR8039 (*mutH101*), ER1801, DH5α-mcr and DH5α-Mut10, for their ability to propagate stably the (CTG)_n-containing plasmids. To analyze the SV40 replication products, we selected DH5α-mcr

(Gibco) because of its ability to propagate stably (CTG)_n-containing plasmids (J.D.C. and C.E.P., unpublished results). We minimized the bacterial contribution for all clones by: first, having the (CTG)_n (CAG) repeat in the stable orientation relative to the unidirectional bacterial ColE1 origin of replication; second, a limited colony growth time (6 h); and third, bacterial strain selection. By using these stringent conditions, we can maintain stably as many as 255 CTG repeats in bacteria^{36,38}. In addition, all primate mutation frequencies were corrected to account for any bacterial background or length heterogeneity in a template preparation.

Expansion and deletion events were scored by analyzing the repeat tracts in miniprep DNA relative to the same starting material. Repeat-containing restriction fragments were resolved on high-resolution 4% polyacrylamide gels³⁸.

Corrected frequencies of expansion or deletion generated by primate replication (Table 1) were calculated by subtracting the background repeat length heterogeneities in the parental plasmid preparation. Because the degree and distribution of length heterogeneity can vary between different bacterial preparations of the same template, it is essential that the same template preparation that is used for primate cell transfection is used for bacterial background correction. We determined the background levels for each preparation of each template by direct bacterial transformation and analysis of single colonies (*n*>100). Only statistically significant differences (χ^2 , *P*<0.05) are reported.

The magnitudes of repeat length changes were determined by electrophoretic sizing of the repeat-containing fragments relative to the same starting length material and a known set of markers³⁸. To confirm that the changes observed were caused by changes in the repeat tract length, samples were electrotransferred (1.5 mA/cm², 45 min; Panther Semidry Electrobloater (HEP-1), Owl Separation Systems) to a Biotrans B membrane (0.45 µm, Pall Corporation). The membrane was probed with a ³²P-labeled (CTG)₁₅ oligonucleotide and exposed to film. We also analyzed selected products of expansion and deletion events by sequencing.

We compared the distribution of plasmid isomers in the primate cell-replicated DNAs with that of the parental DNAs by Southern-blot analysis. There was no difference between the two, which indicated that the replication products were monomeric (excluding recombination events taking place outside the repeat).

Electron microscopic analysis of replication products. We purified primate cell-replicated DNAs as described in the legend of Fig. 5 and mounted them for electron microscopy as described^{38,50}. Briefly, the indicated gel-purified DNAs were mixed in a buffer containing 2 mM spermidine, adsorbed to glow-charged carbon-coated grids, washed with a graded water/ethanol series and rotary shadow cast with tungsten. We examined samples on a Philips 420 electron microscope and took micrographs on sheet film. A Cohu CCD camera attached to a Macintosh computer programmed with IMAGE software (NIH) was used to form the images and to measure the contour length of DNA molecules.

PCR mutation analysis. Replication products were purified and treated as for STRIP analysis, except that the analysis of individual replication products was done by colony PCR amplification and not by bacterial culture. PCR reactions contained *Taq* DNA polymerase (Roche), the supplied buffer, 5% dimethyl sulfoxide and primers that flanked the repeat tracts (primers are available on request). The amplification conditions were 95 °C for 5 min, 25 cycles of 95 °C for 1 min, 65 °C for 1 min, 72 °C for 1 min, and 72 °C for 5 min. Products were resolved on a 1.5% agarose gel in 1 × TBE at 7V/cm for 3 h, in the presence of 50 ng ml⁻¹ ethidium bromide. The frequency of mutations were corrected for length heterogeneity in the starting template as described above.

Acknowledgments

We thank our colleagues at The Hospital for Sick Children for comments and support, and A. Paterson for statistical analysis. This work was supported by grants from the Muscular Dystrophy Association USA and the Canadian Institutes of Health Research (CIHR) to C.E.P. and from the NIH and the March of Dimes to Y.-H.W. J.D.C. is supported by Natural Sciences and Engineering Research Council of Canada. C.E.P. is a CIHR Scholar, a Canadian Genetic Disease Network Scholar and a Premier's Research Excellence Award scholar. Y.-H.W. is a Basil O'Connor March of Dimes Fellow.



Competing interests statement

The authors declare that they have no competing financial interests.

Received 20 December 2001; accepted 22 February 2002.

1. Cummings, C.J. & Zoghbi, H.Y. Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* **9**, 909–916 (2000).
2. La Spada, A.R. Trinucleotide repeat instability: genetic features and molecular mechanisms. *Brain Pathol.* **7**, 943–963 (1997).
3. Martorell, L., Monckton, D.G., Sanchez, A., Lopez De Munain, A. & Baiget, M. Frequency and stability of the myotonic dystrophy type 1 premutation. *Neurology* **56**, 328–335 (2001).
4. Richards, R.I. & Sutherland, G.R. Heritable unstable DNA sequences. *Nature Genet.* **1**, 7–9 (1992).
5. Richards, R.I. & Sutherland, G.R. Dynamic mutations: a new class of mutations causing human disease. *Cell* **70**, 709–712 (1992).
6. Wong, L.J., Ashizawa, T., Monckton, D.G., Caskey, C.T. & Richards, C.S. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am. J. Hum. Genet.* **56**, 114–122 (1995).
7. Martorell, L. et al. Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. *Hum. Mol. Genet.* **7**, 307–312 (1998).
8. Wohrle, D. et al. Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation *in vitro*: evidence for a casual involvement of methyl-directed DNA mismatch repair in triplet repeat stability. *Hum. Mol. Genet.* **4**, 1147–1153 (1995).
9. Ashizawa, T. et al. Instability of the expanded (CTG)_n repeats in the myotonin protein kinase gene in cultured lymphoblastoid cell lines from patients with myotonic dystrophy. *Genomics* **36**, 47–53 (1996).
10. Ashizawa, T. et al. Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. *Am. J. Hum. Genet.* **54**, 414–423 (1994).
11. Richards, R.I. et al. Dynamic mutation loci: allele distributions in different populations. *Ann. Hum. Genet.* **60**, 391–400 (1996).
12. Imbert, G., Kretz, C., Johnson, K. & Mandel, J.L. Origin of the expansion mutation in myotonic dystrophy. *Nature Genet.* **4**, 72–76 (1993).
13. Neville, C.E., Mahadevan, M.S., Barcelo, J.M. & Korneluk, R.G. High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum. Mol. Genet.* **3**, 45–51 (1994).
14. Biancalana, V. et al. Moderate instability of the trinucleotide repeat in spino bulbar muscular atrophy. *Hum. Mol. Genet.* **1**, 255–258 (1992).
15. Spiegel, R., La Spada, A.R., Kress, W., Fischbeck, K.H. & Schmid, W. Somatic stability of the expanded CAG trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Hum. Mutat.* **8**, 32–37 (1996).
16. Gourdon, G. et al. Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. *Nature Genet.* **15**, 190–192 (1997).
17. Monckton, D.G., Coolbaugh, M.I., Ashizawa, K.T., Siciliano, M.J. & Caskey, C.T. Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nature Genet.* **15**, 193–196 (1997).
18. Brock, G.J., Anderson, N.H. & Monckton, D.G. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum. Mol. Genet.* **8**, 1061–1067 (1999).
19. 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).
20. Samadashwily, G.M., Raca, G. & Mirkin, S.M. Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.* **17**, 298–304 (1997).
21. Freudenreich, C.H., Stavenhagen, J.B. & Zakian, V.A. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol.* **17**, 2090–2098 (1997).
22. Freudenreich, C.H., Kantrow, S.M. & Zakian, V.A. Expansion and length-dependent fragility of CTG repeats in yeast. *Science* **279**, 853–856 (1998).
23. Miret, J.J., Pessoa-Brandao, L. & Lahue, R.S. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **95**, 12438–12443 (1998).
24. Schweitzer, J.K. & Livingston, D.M. Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum. Mol. Genet.* **7**, 69–74 (1998).

25. Anderson, S. & DePamphilis, M.L. Metabolism of Okazaki fragments during simian virus 40 DNA replication. *J. Biol. Chem.* **254**, 11495–11504 (1979).
26. Burhans, W.C., Vassilev, L.T., Caddle, M.S., Heintz, N.H. & DePamphilis, M.L. Identification of an origin of bidirectional DNA replication in mammalian chromosomes. *Cell* **62**, 955–965 (1990).
27. Tseng, B.Y., Erickson, J.M. & Goulian, M. Initiator RNA of nascent DNA from animal cells. *J. Mol. Biol.* **129**, 531–545 (1979).
28. DePamphilis, M.L. & Wassarman, P.M. Replication of eukaryotic chromosomes: a close-up of the replication fork. *Annu. Rev. Biochem.* **49**, 627–666 (1980).
29. Waga, S. & Stillman, B. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67**, 721–751 (1998).
30. Hay, R.T. & DePamphilis, M.L. Initiation of SV40 DNA replication *in vivo*: location and structure of 5' ends of DNA synthesized in the *ori* region. *Cell* **28**, 767–779 (1982).
31. Nethanel, T., Reisfeld, S., Dinter-Gottlieb, G. & Kaufmann, G. An Okazaki piece of simian virus 40 may be synthesized by ligation of shorter precursor chains. *J. Virol.* **62**, 2867–2873 (1988).
32. Gordinin, D.A., Kunkel, T.A. & Resnick, M.A. Repeat expansion—all in a flap? *Nature Genet.* **16**, 116–118 (1997).
33. Spiro, C. et al. Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell* **4**, 1079–1085 (1999).
34. Henricksen, L.A., Tom, S., Liu, Y. & Bambara, R.A. Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. *J. Biol. Chem.* **275**, 16420–16427 (2000).
35. Roberts, J.D. & Kunkel, T.A. Fidelity of a human cell DNA replication complex. *Proc. Natl Acad. Sci. USA* **85**, 7064–7068 (1988).
36. Pearson, C.E. & Sinden, R.R. Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry* **35**, 5041–5053 (1996).
37. Pearson, C.E., Ewel, A., Acharya, S., Fishel, R.A. & Sinden, R.R. Human MSH2 binds to trinucleotide repeat DNA structures associated with neurodegenerative diseases. *Hum. Mol. Genet.* **6**, 1117–1123 (1997).
38. Pearson, C.E., Wang, Y.H., Griffith, J.D. & Sinden, R.R. Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)_n, (CAG)_n repeats from the myotonic dystrophy locus. *Nucleic Acids Res.* **26**, 816–823 (1998).
39. Gluzman, Y. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**, 175–182 (1981).
40. Hirt, B. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**, 365–369 (1967).
41. Vassilev, L. & Johnson, E.M. Evaluation of autonomous plasmid replication in transfected mammalian cells. *Nucleic Acids Res.* **16**, 7742 (1988).
42. 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).
43. Constantinou, A. et al. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep.* **1**, 80–84 (2000).
44. Seigneur, M., Bidnenko, V., Ehrlich, S.D. & Michel, B. RuvAB acts at arrested replication forks. *Cell* **95**, 419–430 (1998).
45. Postow, L., Crisona, N.J., Peter, B.J., Hardy, C.D. & Cozzarelli, N.R. Topological challenges to DNA replication: conformations at the fork. *Proc. Natl Acad. Sci. USA* **98**, 8219–8226 (2001).
46. Manley, K., Shirley, T.L., Flaherty, L. & Messer, A. Msh2 deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.* **23**, 471–473 (1999).
47. Moore, H., Greenwell, P.W., Liu, C.P., Arnheim, N. & Petes, T.D. Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl Acad. Sci. USA* **96**, 1504–1509 (1999).
48. Richard, G.F., Goellner, G.M., McMurray, C.T. & Haber, J.E. Recombination-induced CAG trinucleotide repeat expansions in yeast involve the MRE11-RAD50-XRS2 complex. *EMBO J.* **19**, 2381–2390 (2000).
49. Hay, R.T., Hendrickson, E.A. & DePamphilis, M.L. Sequence specificity for the initiation of RNA-primed simian virus 40 DNA synthesis *in vivo*. *J. Mol. Biol.* **175**, 131–157 (1984).
50. Griffith, J.D. & Christiansen, G. Electron microscope visualization of chromatin and other DNA-protein complexes. *Annu. Rev. Biophys. Bioeng.* **7**, 19–35 (1978).

