Nucleic Acids Research, 2005, Vol. 33, No. 17 ***5667–5676***
doi:10.1093/nar/gki880

Mechanistic features of CAG•CTG repeat contractions in cultured cells revealed by a novel genetic assay

Richard Pelletier, Brian T. Farrell[1], Juan José Miret and Robert S. Lahue*

Eppley Institute for Research in Cancer and Allied Diseases and [1]Department of Pathology and Microbiology, University of Nebraska Medical Center, Box 986805, Omaha, NE 68198-6805, USA

Received August 16, 2005; Revised and Accepted September 16, 2005

ABSTRACT

Trinucleotide repeats (TNRs) undergo high frequency mutagenesis to cause at least 15 neurodegenerative diseases. To understand better the molecular mechanisms of TNR instability in cultured cells, a new genetic assay was created using a shuttle vector. The shuttle vector contains a promoter-TNR-reporter gene construct whose expression is dependent on TNR length. The vector harbors the SV40 *ori* and large T antigen gene, allowing portability between primate cell lines. The shuttle vector is propagated in cultured cells, then recovered and analyzed in yeast using selection for reporter gene expression. We show that $(CAG•CTG)_{25-33}$ contracts at frequencies as high as 1% in 293T and 293 human cells and in COS-1 monkey cells, provided that the plasmid undergoes replication. Hairpin-forming capacity of the repeat sequence stimulated contractions. Evidence for a threshold was observed between 25 and 33 repeats in COS-1 cells, where contraction frequencies increased sharply (up 720%) over a narrow range of repeat lengths. Expression of the mismatch repair protein Mlh1 does not correlate with repeat instability, suggesting contractions are independent of mismatch repair in our system. Together, these findings recapitulate certain features of human genetics and therefore establish a novel cell culture system to help provide new mechanistic insights into CAG•CTG repeat instability.

INTRODUCTION

Trinucleotide repeats (TNRs) are microsatellite sequences whose genetic instability causes at least 15 human neurological disorders, including Huntington's disease, myotonic dystrophy and fragile X syndrome (1–4). TNR instability shows an unusual dependence on the length of the tract. Alleles up to ~30–35 repeats are stably inherited in humans (5), however beyond that size the repeats display much more frequent instability with a strong tendency towards expansions (6). This pattern can be attributed to two characteristics of TNRs that are not seen with other microsatellites. First, the 30–35 repeat length constitutes a genetic threshold, defined as a repeat span that demarcates short, relatively stable tracts from long, unstable alleles. The allele length relative to the threshold is therefore a key indicator of the likelihood of mutation. Second, once the threshold is reached, mutation spectra are shifted towards expansions rather than contractions (6).

A key factor in avoiding disease therefore depends on preventing TNRs from attaining the threshold. In humans, CAG•CTG alleles in the normal (subthreshold) range tend to contract more often, or at least as often, as they expand, based on single sperm typing at the *HD* locus (7), at the androgen receptor gene and the *DM1* locus (8), and at the *MJD1* gene (9). This observation also holds for subthreshold CGG•CCG repeats at the fragile X locus (10,11). A reasonable conclusion from these data is that contractions of alleles below the threshold are important for maintaining short, genetically stable alleles that do not confer disease and that are unlikely to expand in later generations.

The molecular foundation for TNR instability is not well understood, but the key role of DNA structure is widely accepted (2,3). Structural studies (12) revealed that disease-causing TNRs adopt unusual configurations, such as single-strand hairpins, slipped-strand structures and triplex DNA, whose thermodynamic stability correlates with the length of the repeat tracts (13–17). These observations, coupled with the knowledge that new repeats need to be synthesized (for an expansion) or deleted (for a contraction), generated the

*To whom correspondence should be addressed. Tel: +1 402 559 4619; Fax: +1 402 559 8270; Email: rlahue@unmc.edu
Present addresses:
Richard Pelletier, Unite de Recherche en Genetique Humaine, CHUL, 2705 Boul. Laurier RC-9300, Ste-Foy, Quebec G1V 4G2, Canada
Juan José Miret, Discovery Technology Center, Global Research and Development, Cambridge USA Laboratories, 620 Memorial Drive, Cambridge, MA 02139, USA

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

proposal that TNR-induced structures lead to aberrant DNA synthesis (2–4,12). One prevalent line of thought is that hairpin formation on the newly synthesized strand promotes expansions. The corollary is that hairpins on the template strand are the precursors for contractions. Thus both types of TNR mutations can be ascribed to models where structural perturbations, such as hairpins, promote error-prone DNA synthetic events (2–4,12). It should be noted however that hairpins have only been observed in the test tube; there has been no direct detection of hairpin formation in cells. This difficulty can be overcome by comparing genetic instability of different TNR sequences with different hairpin-forming capacities. Nearly always, the better a TNR sequence is at forming a hairpin, the higher its instability [e.g. (18)].

Evidence from bacterial and yeast model systems supports the hypothesis that instability results from three synthesis-dependent sources, namely DNA replication (19–31), repair [summarized in (32)], and recombination (33–36). Similarly, instability in mammalian systems is probably derived from more than one cellular mechanism. For example, replication slippage as a source of CAG•CTG instability was suggested in several studies (7,37–40). Also, several human cell lines were shown to have active replication initiation regions within 3.6 kb or less of *SCA-7*, *SBMA*, and both normal and expanded alleles of *HD* loci (41). A third line of evidence is that replication inhibitors also modulate instability at the endogenous *DM1* gene in human fibroblasts (42). In contrast, at least two reports (43,44) found no simple association between the rate of cell division and instability in transgenic mice or in tissue culture cell lines established from transgenic mice.

Aberrant DNA repair in mammals has also been proposed as an important contributor to instability. Available evidence, however, does not suggest a unifying model by which repair accounts for TNR mutations because several pathways have been implicated. Mismatch repair was inferred in five papers showing decreased expansion frequencies for CAG•CTG repeats in transgenic mice that were also deficient in the mismatch repair genes *Msh2* (45–48) or *Msh3* (49). For contractions, four of these papers found little or no apparent change in contraction frequencies (45–47,49), whereas the fifth study showed a dramatic shift towards contractions in Msh2-deficient animals (48). In *Pms2* knockouts, expansion frequencies were decreased by about one-half relative to wild-type littermates, but contractions were largely unaffected (50). In contrast to the mouse findings, studies in human cell lines deficient for Msh2 or Mlh1 showed no apparent instability for endogenous (normal-length) alleles of DM or the CGG•CCG repeat at FRAXA (51). Recombinational repair, perhaps by single-strand annealing, was deduced from a study of maternally derived *SCA1* expansions (52). Subsequently, CTG repeats were shown to markedly influence the outcome of genetic recombination in CHO cells (53). Repair of double-strand breaks and/or stalling of replication forks can also induce instability of some CTG•CAG tracts in CHO cells (40) or in HeLa cells (54). It was recently shown that changes in DNA methylation status also modulate instability (55–57), indicating an epigenetic influence on the mutagenic process.

The studies summarized above suggest multiple mechanisms of CAG•CTG repeat instability in mammalian systems. To understand the potential mechanisms more fully, improvements in detection methods will be required. Physical analyses of the repeat tracts (e.g. PCR monitoring) would be nicely complemented by selective genetic assays for TNR instability. In this study, we established an assay that combines the relevance and convenience of cultured primate cells with quantitative genetic selection in yeast to better understand CAG•CTG repeat mutability. We found that instability in cultured cells recapitulates a number of the characteristic features from humans, suggesting this new system will be useful in helping understand the mechanisms causing TNR mutations.

MATERIALS AND METHODS

Cell lines

Human 293 and 293T (human embryonic kidney) cells, and COS-1 (African green monkey kidney) cells were cultured in DMEM supplemented with 10% fetal bovine serum, 50 U/ml penicillin-streptomycin (Gibco BRL) and anti-PPLO agent (1× final concentration; Gibco BRL) or Fungizone amphotericin B (2.5 μg/ml final concentration) (Invitrogen).

Plasmids

The shuttle vector backbone (pBL67) was constructed by subcloning the HpaII–BamH1 fragment (with ends filled-in) of the SV40 viral DNA (Gibco BRL), encompassing the origin of replication (ori) and the large T antigen gene, into the BamH1 site (blunt ended) of the pRS313 vector (58). The ∼9.8 kb shuttle vector includes the SV40 viral origin of replication (ori) and the large T antigen gene. Yeast markers include a reporter system to detect TNR contractions, which occur in human cells, via subsequent expression of the URA3 gene (18). Also present on the plasmid are the HIS3 gene, to identify yeast transformants, and the genetic elements autonomously replicating sequence and CEN, a centromere element that controls plasmid number in yeast to 1–2 copies per cell. The vector contains an *Escherichia coli* origin of replication and an ampicillin resistance gene. To create individual shuttle vectors $(CAG)_{25+8}$, $(CTG)_{25+8}$, $(C,T,G)_{33}$, and $(CAG)_{33}$, cassettes (for explanation of the plasmid nomenclature see Results section) containing the P_{adh1} promoter, the TNR tract and the URA3 reporter gene were excised from plasmids carrying the relevant TNR tracts (as described previously in (18)] using XhoI and EcoRI enzymes, and subcloned into the corresponding restriction sites of pBL67. The reporters $(CAG)_{25+8}$ and $(CTG)_{25+8}$ differ only in the cloning orientation of the repeat tracts. A vector backbone was also constructed in which the SV40 ori is on the opposite, or 5′, side of the TNR tract as compared with the vectors described above (3′ side in pBL67). In this vector, pBL185, a 443 bp SfaNI fragment from SV40 was subcloned into the ApaI (filled-in) of pRS313. To create plasmid $(CAG)_{25+8}$ new ori, a cassette encompassing the $(CAG)_{25+8}$ tract from pBL145 was subcloned into the XhoI–EcoRI sites of pBL185. The $(TAG)_{25+8}$ reporter was created by replacing the $(CAG)_{25+8}$ triplet repeat tract of $(CAG)_{25+8}$ new ori with an oligonucleotide duplex containing the TAG repeat sequence. Finally, removal of all SV40 sequences (ori and large T antigen) from pBL67 yielded plasmid pBL207. Subcloning the XhoI–EcoRI cassette from pBL145 into the corresponding restriction sites of the ori^- shuttle vector yielded $(CTG)_{25+8}$ ori^-. Each plasmid DNA

was prepared by the Qiagen method (plasmid maxi-kit, Qiagen Inc.) from bacterial cultures grown to no more than mid-log phase, to minimize background TNR contractions (59).

Mammalian cell transfections

Supercoiled DNA (15 μg) was used to transfect exponentially growing mammalian cells (~1–2 × 10^6 cells per 75 cm^2 flask, one flask per transfection) by the calcium phosphate co-precipitation method (60). After 2–3 days of growth at 37°C, the cells were lysed with 3 ml of Hirt's lysis buffer (61), and the lysate was pooled into a 15 ml Corex tube. NaCl (5 M) (750 μl) was added to each lysate and stored overnight at 4°C. The lysate was then centrifuged in a Sorval SS34 rotor at 10 000 r.p.m. for 15 min at 4°C. The supernatant, containing the plasmid DNA, was collected into a fresh tube. To remove proteins, the lysate was spun on a Centriplus YM-100 column (Millipore) for 1–1.5 h at 4°C in a Sorval SS34 rotor at 5000 r.p.m., following the manufacturer's instructions. The retentate was recovered by inverting the column into a collection tube and spinning an additional 3–4 min at 4000 r.p.m. The DNA was then precipitated with two volumes of ethanol, and suspended in 100 μl of 10 mM Tris–HCl and 1 mM EDTA, pH 7.5.

Yeast transformations

DpnI digestion was used to eliminate unreplicated plasmid molecules that retained the bacterial methylation pattern. The plasmid DNA (50 μl), prepared as described above, were digested for 1 h with ~60 U of DpnI (New England Biolabs) plus 10 μg RNase A in the buffer provided by the manufacturer at a final volume of 60 μl. In the case of the SV40 *ori*$^-$ control plasmid, which does not replicate in human cells, the DpnI digestion was omitted. A fraction of the digest (between 50 and 100%) was used to transform yeast S150-2B cells (MATa *leu2-3,112 ura3-52 his3Δ trp1-289*) via a modified version (62) of the lithium acetate method. A fraction of each transformation mixture (typically 0.5%) was plated onto SC-His plates (synthetic complete, lacking histidine), and the remainder onto SC-His-Ura (synthetic complete, lacking histidine and uracil) to score for contractions. Colonies on each plate were counted after 3 days of growth at 30°C. The frequency of contraction was determined as the number of colonies obtained on SC-His-Ura divided by the total number of transformants on SC-His, with appropriate correction for dilution factors.

PCR analysis was performed to determine the fraction of colonies that had taken up a plasmid with a bona fide change in TNR size. Isolated colonies were disrupted in 100 μl of 50 mM DTT plus 0.5% Triton X-100, incubated at 37°C for 30 min, heated at 95°C for 6 min and kept on ice thereafter. A portion of this material was used as a template for PCR amplification as described previously (23), with the following modifications. Amplification in the presence of ~1.25 μCi of [α-^{32}P]dCTP was performed for 30 cycles (1 min at 94°C, 1 min at 60°C and 1 min at 72°C), plus a final extension at 72°C for 5 min. The products were separated on a 6% denaturing polyacrylamide gel. PCR product sizes (±2 repeat units) were determined by comparison with a DNA sequencing ladder. The final, reported mutation frequency for each template was determined by multiplying the fraction of plasmids with bona fide alterations in TNR size (between 80 and 100%) times the raw mutation frequencies calculated above. Additional details regarding the assay results can be viewed in Supplementary Data.

The experimental protocol for detecting TNR mutations in cultured cells is shown in Figure 1A. By measuring frequencies, where the denominator reflects the total number of plasmid molecules available for testing, each experiment is internally controlled for DNA recovery, transformation efficiency and so on. A critical concern with this assay is to ensure that the TNR mutation events occurred in the tissue culture cells, and not during plasmid propagation in *E.coli* (59) or immediately following transformation in yeast, before selection becomes effective. To determine the background frequency at which TNR tracts mutate in yeast and/or *E.coli*, 0.5–5 μg of each DNA plasmid preparation was used to transform yeast cells directly, without passaging through mammalian cells (Figure 1A). The cells were plated on selective media as described above and the background mutation frequency was calculated by dividing the number of colonies on selective media by the total number of transformants. *E.coli* strain HB101 and yeast strain S150-2B provided the lowest achievable background frequencies among several strains tested. As noted by others (19), we found instability in *E.coli* varied with the repeat orientation. This explains why our background values are different between the CAG and CTG orientations of the reporter.

Western blot analysis

Mlh1 protein expression was analyzed in 293 and 293T cells and compared with HeLa controls by western blotting. Cell lysates (40 μg protein per sample) were separated by 8% SDS–PAGE, transferred to a nitrocellulose membrane (Amersham), and blocked overnight in Odyssey Blocking Buffer (LI-COR). Mlh1 protein was detected with mouse anti-human Mlh1 monoclonal antibody (PharMingen), followed by secondary antibody AlexaFluor 680 goat anti-mouse IgG (Molecular Probes), then the membrane was scanned using the Odyssey System according to the manufacturer's specifications.

RESULTS

The shuttle vector assay

A new genetic assay was established that combines the availability of many useful tissue culture lines with yeast genetics. A DNA shuttle vector is transfected into cultured mammalian cells, the cells are propagated for a few days to allow contractions of the TNR tract to occur, and then the plasmid DNA is recovered and transformed into yeast (Figure 1A). Since the vector contains a selective reporter system, described in the figure legend (Figure 1B), only those yeast cells that have incorporated a plasmid missing at least five repeats from the TNR tract survive on selective medium. Thus contractions that occur in the mammalian cells are revealed selectively in yeast. While the current study focuses on contractions, future work will include experiments to examine TNR expansions. This system also complements existing, frameshift-based assays for microsatellite instability in tissue culture cells (63,64) by

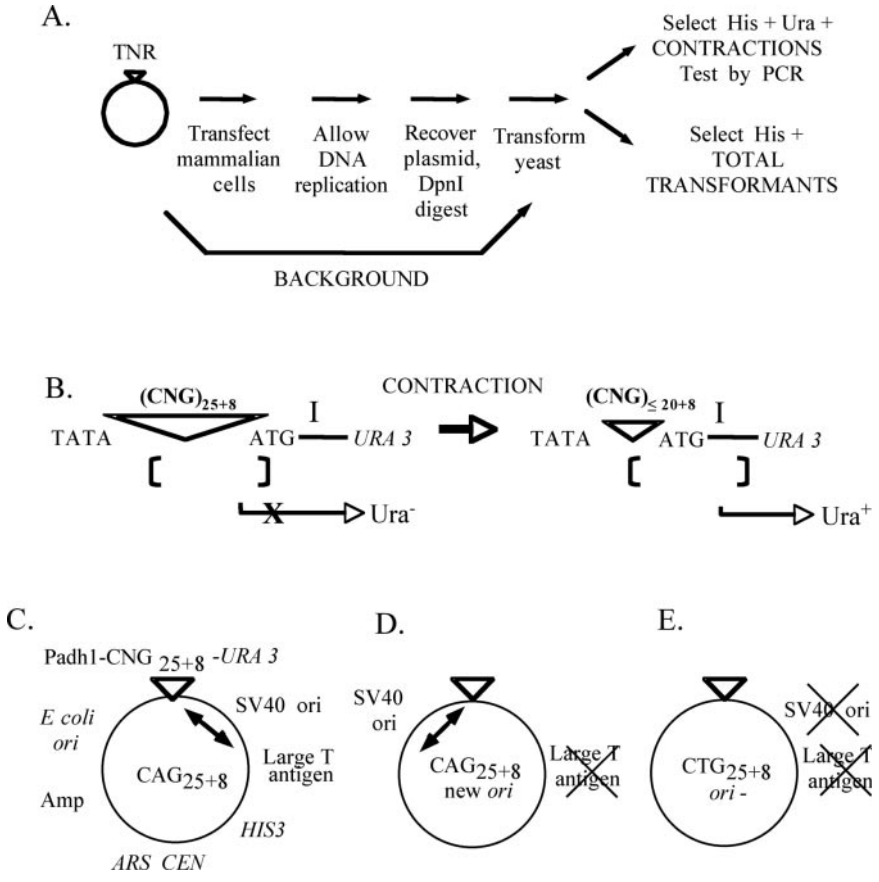

Figure 1. Experimental procedure to detect TNR tract changes in human tissue cultured cells and the DNA shuttle vector. Details are provided in Materials and Methods. (**A**) Each TNR-containing vector was transfected into mammalian cells, the cells were cultured for 2–3 days, then the plasmid DNA was transformed into yeast for analysis. Contraction frequency was calculated by dividing the number of yeast colonies with a contraction by the total number of transformants. Background contraction frequencies were measured as described in Materials and Methods. (**B**) The selection for contractions in yeast is based on spacing sensitivity of the yeast promoter from the preferred transcription initiation site, labeled 'I'. Starting TNR lengths of 25 + 8 (e.g. 25 repeats of CTG plus non-repeating DNA equivalent in length to 8 repeats) inhibit expression of the *URA3* reporter gene and yield a Ura$^-$ phenotype (18). TNR contractions that remove at least five repeats restore promoter activity and result in a Ura$^+$ phenotype. See Supplementary Figure S1 for detailed sequence information of the TNR-containing region. (**C**) The ~9.8 kb shuttle vector allows replication in primate cells, yeast and *E.coli*. The TNR-*URA3* reporter provides a readout of TNR tract length via *URA3* expression, as described above. **(D)** Plasmid $(CAG)_{25+8}$ new *ori* is a variant with the SV40 origin moved to the opposite side of the TNR tract, plus deletion of the large T antigen gene (denoted by the cross). **(E)** $(CTG)_{25+8}$ *ori*$^-$ has deletions of both the SV40 origin and the large T antigen gene. Details of plasmid constructions are provided in Materials and Methods.

providing a way to select for mutations that affect 3 bp repeating units.

The essential features of the shuttle vector are shown in Figure 1C. The plasmid can replicate by virtue of an SV40 viral origin of replication (*ori*), which drives high-copy production of the plasmid in just a few days. Host factors from many cell types can mediate replication from this origin; the only required exogenous protein is SV40 large T antigen, which is encoded by the shuttle vector. The plasmid is therefore portable between different primate cell lines. The vector also contains a TNR-*URA3* reporter system that allows detection of contractions once the plasmid is subsequently transferred into yeast (Figure 1B). In most cases, the starting TNR tract had 25 repeats appended to 24 nt of randomized sequence ('25 + 8', total length equivalent to 33 repeats). The perfect 25 repeat length represents a subthreshold allele, and the addition of the randomized sequence allows the yeast selection to function as designed (Figure 1B). Three classes of sequences were assayed: CAG and CTG repeats that form strong hairpin structures (13,15–17) and which are relevant to human disease; a TAG repeat which forms hairpins less well (13) to test the influence of structure on instability; and scrambled sequences lacking any repetitive nature, as a negative control. Because the bidirectional SV40 *ori* (65,66) is located on one side of the repeats at close distance (Figure 1C), it is possible to assign a particular repeat sequence to either the leading or lagging strand of replication. All triplet constructs as written represent the sequence on the lagging template strand. The plasmid also has elements for selection in yeast, plus *E.coli* sequences for cloning and propagation purposes. Therefore it is a trifunctional shuttle vector, capable of replicating in primate cells, in yeast and in *E.coli*.

CAG•CTG repeat instability in human 293T cells

The TNR vectors were first tested in human 293T embryonic kidney cells, which express SV40 large T antigen constitutively. These cells were chosen because they are easily transfected by the calcium phosphate co-precipitation method, and provide a high replicative yield of SV40 based

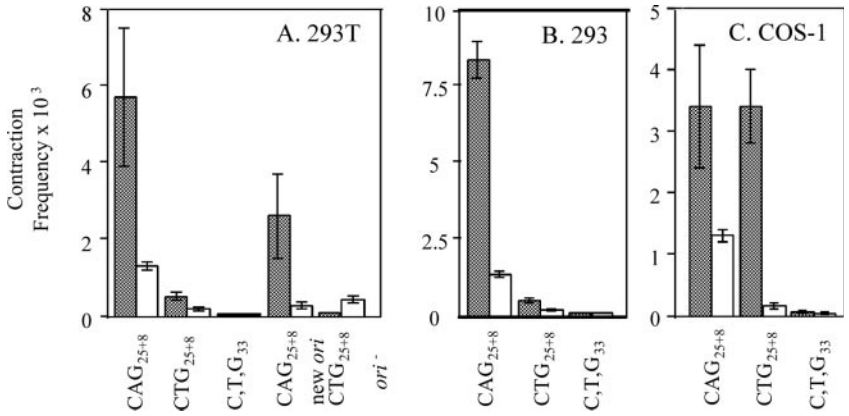

Figure 2. CAG•CTG contraction frequencies in mammalian cells. For each graph, the stippled bars represent the average frequency of contractions (±SEM) after passage through the human cells, while the open bars represent the background mutation frequency. Each transfection was performed 3–9 times. The *X*-axis denotes the TNR sequence on the lagging template strand relative to the SV40 origin of replication on the vector. (**A**) Results from 293T cells. (**B**) Frequencies using 293 cells. (**C**) Contraction frequencies in COS-1 cells. Note that the scale on the *Y*-axis is different in the three graphs. The frequency data are available in the Supplementary Data.

plasmids (67). Furthermore, the kidney is a site of substantial somatic instability in myotonic dystrophy transgenic cell lines (44). Figure 2A shows that contractions of $(CAG)_{25+8}$ occurred at a frequency of 5.7×10^{-3}. This value is significantly above the background level of 1.3×10^{-3} ($P = 0.047$ by Student's *t*-test), indicating that the majority of the contraction events occurred in the human cells, not in *E.coli* or yeast. The frequency indicates that nearly 0.5% of all plasmid molecules contracted upon passage through 293T cells. Surprisingly the $(CTG)_{25+8}$ reporter gave only about one-tenth as many contractions (0.5×10^{-3}) as the CAG reporter. While a low value, the $(CTG)_{25+8}$ contraction frequency was significantly above background value of 0.15×10^{-3} ($P = 0.021$). This orientation effect was statistically significant [$P = 0.029$, comparing the data from $(CAG)_{25+8}$ with $(CTG)_{25+8}$]. The orientation of the tract is therefore important in 293T cells. The $(C,T,G)_{33}$ scrambled control, containing a 99 nt scrambled sequence with no repetitive nature, was inert in this assay ($<0.01 \times 10^{-3}$), as expected.

The observed orientation effect, where CAG repeats are more unstable than CTG runs, leads to the specific, testable prediction that reversing the direction of replication should reverse the instability pattern. Figure 3A shows the deduced replication fork as it moves past the TNR repeat sequence in $(CAG)_{25+8}$. The position of the CAG repeats is on the lagging template strand. Similarly, Figure 3B shows that the $(CTG)_{25+8}$ reporter gives rise to CTG sequences as the template for lagging strand synthesis. If the direction of DNA replication for $(CTG)_{25+8}$ is reversed, such that the fork progresses in the opposite direction, then the lagging template strand now contains CAG repeats (Figure 3C). This situation was accomplished by moving the SV40 *ori* in the vector backbone from 3′ of the CTG tract to a 5′ location to create $(CAG)_{25+8}$ new *ori* (Figure 1D). If CAG repeats are more probable to act as the template for contraction events, then $(CAG)_{25+8}$ new *ori* should be more unstable than $(CTG)_{25+8}$. Figure 2A shows an increase of ~5-fold in contraction frequency for $(CAG)_{25+8}$ new *ori* (2.6×10^{-3}), compared with $(CTG)_{25+8}$ ($P = 0.015$). The increase implies that the direction of DNA replication through the repeat tract influences instability. If true, then we also predict that removal of the

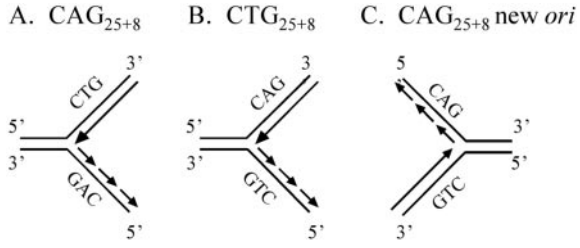

Figure 3. Schematic representation of replication forks proceeding through TNR regions. Each panel shows a replication fork as it passes through the TNR site. The direction of replication was deduced from the position of the SV40 ori sequences relative to the TNR (Figure 1C and D). The 5′ and 3′ polarities are indicated for the template strands. The leading strand product is shown as a long continuous arrow, while the newly replicated lagging strand is shown as discontinuous Okazaki fragments prior to processing and ligation. (**A**) For the substrate $(CAG)_{25+8}$, replication moves from right to left. This places the CAG repeat sequence on the lagging template strand. (**B**) When the TNR is cloned in the opposite orientation, denoted as $(CTG)_{25+8}$, the CTG repeat now occupies the lagging template position. (**C**) The SV40 origin in $(CTG)_{25+8}$ was moved to the opposite side of the TNR to create $(CAG)_{25+8}$ new *ori*, as diagrammed in Figure 1D. Accordingly, replication now proceeds from left to right. The change in fork direction makes the 'top' strand, with CAG repeats, the lagging template, hence the designation as a CAG plasmid.

SV40 origin (Figure 1E) should preclude replication and therefore contractions. This loss of instability was seen for $(CTG)_{25+8}$ *ori*$^-$ (Figure 2A), which yielded the low value of $<0.08 \times 10^{-3}$. The $(CAG)_{25+8}$ new *ori* and *ori*$^-$ data indicate that instability requires replication through the triplet repeat tract, however it does not prove definitively that replication *per se* is the source of CAG•CTG contractions. In summary, CAG•CTG contractions in human 293T cells occur at relatively high frequencies, contractions are dependent on the triplet repeat sequence, they require replication of the repeat DNA, and there are ~10-fold more contractions for runs of CAG rather than CTG on the lagging strand template.

Contraction sizes in 293T cells

PCR analysis of 48 contracted alleles from the $(CAG)_{25+8}$ plasmid was performed to determine the extent of individual contraction events in 293T cells (Figure 4). For the $(CAG)_{25+8}$

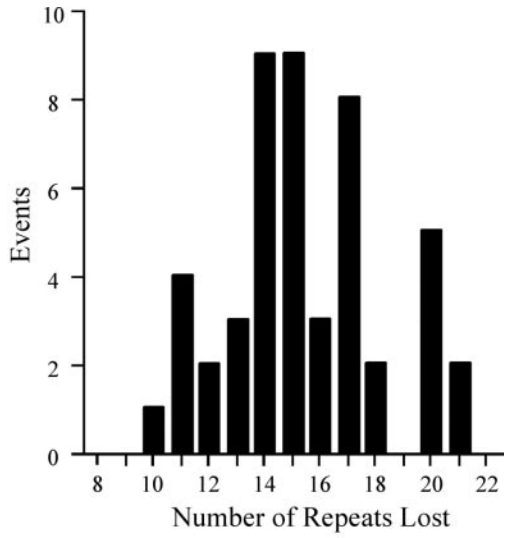

Figure 4. Summary of contraction events for $(CAG)_{25+8}$ in 293T cells. Contraction sizes were determined by single-colony PCR amplification of the repeat tract, followed by size measurements on DNA sequencing gels. Sizes of the contractions were determined (±1–2 repeats) by comparison with an M13 DNA sequencing ladder.

reporter, deletions from −5 to −25 repeats are identifiable with our assay (18). Contracted alleles were observed from −10 to −21 repeats, with a median of −15 (Figure 4). The majority of these events must be genetically independent because the results represent several independent transfections from three different plasmid preparations, and because most contractions from a given experiment were of distinct sizes. In contrast, if the majority of contractions resulted from a few early events that were subsequently replicated, there would be predominance of a few allele sizes, rather than the dispersed pattern observed. There is a possibility that a particular plasmid molecule contracted more than once, thus giving multiple size products, but this seems unlikely based on our frequency estimates. Taken together, these factors indicate that genetically independent contractions make up the majority of observed events.

In two other tissue culture systems for TNR instability (40,54), deletion of the flanking sequences sometimes accompanies contractions. To evaluate the possibility of flanking sequence deletions in our system, PCR amplification was performed on 24 yeast colonies (from four separate 293T cell transfection experiments). The PCR primer sites end 41 nt upstream and 53 nt downstream from the ends of the CTG tract. In 23 of 24 samples, a shortened PCR product was clearly identified, indicating that the primer sites were intact. The one sample that failed to amplify at the TNR site was also negative for amplification at the *URA3* and *HIS3* genes on the plasmid. We conclude that the DNA template preparation from that one sample was defective, and therefore it was excluded from further consideration. Among the 23 amplified samples, DNA sequence analysis was performed on the PCR products from eight randomly chosen representatives (corresponding to all four independent 293T transfections). In all eight cases, the deletion was entirely within the CTG tract; there were no examples of deletions extending into flanking sequences, including the randomized $(C,T,G)_8$ sequence immediately

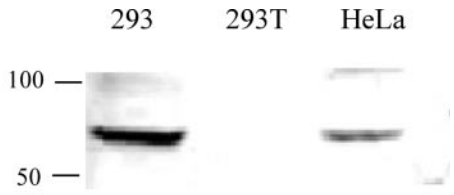

Figure 5. Western blot analysis of Mlh1 in several human cell lines. Extracts were prepared from the indicated cell lines and 40 μg of total protein were subjected to SDS–PAGE separation and immunoblotting, as described in Materials and Methods. The primary antibody was anti-human Mlh1. The notations at the left indicate the position of the 50 and 100 kDa molecular weight markers run on the same gel.

adjacent to the repeat. All deletions precisely removed 3 bp CTG units, ranging from −10 to −18 repeats. This size range is consistent with the spectrum seen in Figure 4. We conclude that contractions within the CTG repeats comprise the large majority of deletions in our system.

Mlh1 expression does not correlate with TNR instability

It was reported that expression of the human mismatch repair gene *Mlh1* is impaired in 293T cells owing to promoter hypermethylation (68). Extracts from 293T cells are accordingly deficient for *in vitro* mismatch repair activity, but this deficiency is reversed when the cells are transfected with a wild-type *Mlh1* gene (68). These results indicate that the mismatch repair deficiency in 293T cells is solely owing to lack of Mlh1 protein expression. Previous reports from transgenic mouse studies implicated the mismatch repair genes *Msh2*, *Msh3* and *Pms2* in TNR instability (45–50). It is conceivable that Mlh1 expression levels, and therefore mismatch repair status, might influence contraction frequencies, orientation effects or both. Also, based on a recent study with *Msh2*-deficient mice (48), contraction frequencies may also be heavily dependent on the mismatch repair status of the cell. To determine if differences in Mlh1 expression affected contraction frequencies or orientation effects, we examined Mlh1 expression by western blot analysis. The results in Figure 5 reveal that Mlh1 was undetectable in the 293T cell extract, while it is clearly expressed in 293 and in control HeLa extracts. The apparent molecular weight corresponds well to the predicted mass of 84.5 kDa. Because the frequencies of CAG•CTG instability in 293T cells were similar to that of 293 cells (see next section), these data suggest that the presence or absence of Mlh1 does not detectably alter CAG or CTG contraction frequencies, nor do differences in Mlh1 steady-state levels correlate with the orientation effects described above.

Instability in other cell lines

Our shuttle vector replicates efficiently in 293T cells, which constitutively express the large T antigen necessary for initiation of replication. Because the gene for the T antigen is present on the shuttle vector, its expression should be sufficient to drive replication of the plasmid to levels high enough to detect TNR contractions. The assay should therefore be portable in cell lines that do not express T antigen. To test this idea, we chose human 293 embryonic kidney cells, which are immortalized with adenovirus 5 DNA instead of T antigen (60). We transfected these cells with a subset of shuttle vectors

and measured their contraction frequencies. The genetic instability in 293 cells (Figure 2B) is reproduced at levels similar to the 293T cells. Although less plasmid DNA (in absolute terms) was recovered from the 293 cells, the yield is sufficient to reveal genetic instability at 8.3×10^{-3} for the vector bearing the $(CAG)_{25+8}$ tract (Figure 2B). This value was significantly above the background value of 1.3×10^{-3} ($P = 0.001$). As seen previously in 293T cells, the $(CTG)_{25+8}$ tract showed a low but significant level of contractions (0.42×10^{-3}) compared with background frequency of 0.15×10^{-3} ($P = 0.023$). The $(CTG)_{25+8}$ reporter in 293 cells was therefore more stable than its $(CAG)_{25+8}$ complement ($P < 0.001$) by ~20-fold. This finding suggests that the orientation dependence is not a unique feature of the 293T cells tested earlier, and the orientation effect can be reproduced in 293 cells as well. As was the case previously, a plasmid bearing the scrambled control sequence $(C,T,G)_{33}$ was genetically stable in 293 cells.

To determine if TNR instability occurs in a non-human cell type, African green monkey kidney cells (COS-1) were chosen. These cells are easily transfectable by the calcium phosphate method, and they also express T antigen constitutively. The vector bearing $(CAG)_{25+8}$ displayed contraction levels (Figure 2C) similar to 293 and 293T cells (3.4×10^{-3}, $P = 0.030$ compared with background of 1.3×10^{-3}). However, the TNR tract in the opposite orientation, $(CTG)_{25+8}$, was equally unstable as the CAG test tract (also 3.4×10^{-3}; $P = 0.034$ compared with background of 0.15×10^{-3}). Clearly there is no orientation effect in COS-1 cells [$P = 0.996$ comparing $(CAG)_{25+8}$ with $(CTG)_{25+8}$]. The scrambled repeat vector, $(C,T,G)_{33}$, was stable in the monkey cells as expected. The control repeat $(TAG)_{25+8}$, predicted to have reduced hairpin-forming capacity, showed a lower frequency of contractions [1.2×10^{-3}; $P = 0.02$ versus $(CAG)_{25+8}$]. This result indicates that the ability of CAG and CTG repeats to form hairpins increases their instability compared with TAG repeats.

CAG contractions exhibit an apparent threshold

Although thresholds have been identified or inferred for a number of human TNR expansion diseases, this feature has not been directly demonstrated in mammalian cells. Our genetic assay seemed suitably sensitive to detect a threshold effect. To address this issue, we created a series of reporter constructs with CAG repeat lengths from 0 to 33 repeats. Scrambled sequences, containing no repeating nature, were added to equalize the total tract length to 99 bp in all cases. For example, the $(CAG)_{15}$ construct also contained 54 bp of scrambled sequences (equivalent in length to 18 repeats). Each construct was tested for instability in COS-1 cells, background frequencies were also determined, and the difference (contraction frequency in COS-1 minus background frequency) is shown in Figure 6. Contractions were observed at low levels for repeat lengths of 15 and 17, with a somewhat higher value observed at 25 repeats. However, the response increased sharply for starting tracts of 30 and 33 repeats. Compared with the frequency at 25 repeats (2.1×10^{-3}), contractions occurred 3.2-fold more often for 30 repeats, and 7.2-times more often for 33 repeats. Thus, lengthening the starting tract from 25 to 33 repeats (a 32% increase in DNA sequence) resulted in a 720% increase in instability. This result is inconsistent with a simple stochastic function of tract length. Instead, the findings in Figure 6 lead us to conclude that CAG contractions in COS-1 cells are governed by a threshold between 25 and 33 repeats. This value is in concordance with the accepted range of 30–35 repeats in human disease genes (5). Additional experiments with our system will be necessary to define the threshold more definitively.

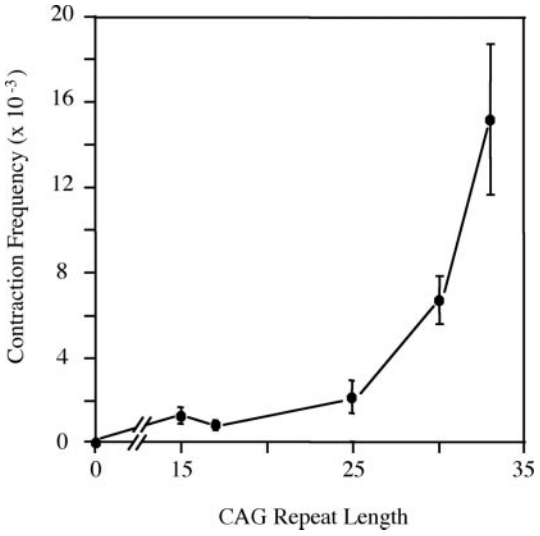

Figure 6. Response of CAG repeat contraction frequency on initial tract length. Contraction frequencies were measured in triplicate in COS-1 cells for the following triplet repeat reporters: 0 repeats, $(C,A,G)_{33}$; 15 repeats, $(CAG)_{15+18}$; 17 repeats, $(CAG)_{17+16}$; 25 repeats, $(CAG)_{25+8}$; 30 repeats, $(CAG)_{30+3}$; and 33 repeats, $(CAG)_{33}$. Background frequencies were also measured, and the difference (COS-1 frequency minus background frequency) are plotted. Error bars indicate ±1 SEM.

DISCUSSION

The shuttle vector assay described here provides a new way to investigate TNR contractions in cultured primate cells. The assay combines convenient features (genetic selection, easy manipulation of the input TNR sequence and molecular analysis of the mutated tract) and thereby adds to the repertoire of tools available to study mechanisms of TNR instability. The shuttle vector assay is also demonstrably portable between primate cell lines. We confirmed replication-dependent instability described in other studies, plus our experiments discovered an apparent threshold for contractions of CAG•CTG tracts that coincides with the expansion threshold in humans. Furthermore, we found that contractions apparently occur independently of the mismatch repair protein Mlh1. Each of these findings is discussed in more detail below.

The threshold for TNR instability is a genetic hallmark that distinguishes triplet repeats from other microsatellites. While the threshold is well established in human genetics, previous studies in cell culture systems have not reported a clear threshold effect. Our finding of an apparent threshold was made possible by the sensitive nature of the assay plus the ability to manipulate the input TNR tracts. It is of interest that the estimated threshold value in COS-1 cells, between 25 and 33 repeats, coincides with the range of 30–35 repeats generally

observed in human TNR disease genes (5). This concordance implies that the mechanism of TNR instability may be accurately reflected in tissue culture. This also may suggest that whatever molecular features dictate thresholds may apply to both expansions and contractions for CAG•CTG repeats.

The experiments presented here were designed to model the situation of TNR alleles just below the threshold length in mammalian cells. The findings suggest that contractions play an important, if perhaps under appreciated, role in disease prevention by maintaining TNR lengths below the critical threshold where expansions become frequent and predominant. A similar conclusion was drawn from transgenic mouse studies (69), where deletions in early development help safeguard the genome against expansions. In our studies, $(CAG)_{25}$ subthreshold alleles contract at relatively high frequencies of 0.3–0.8% in three separate cell lines. These frequency measurements are in agreement with studies of similar length alleles measured by single sperm analysis at the *HD* locus (7), at the androgen receptor gene and *DM1* locus(8), at the *MJD1* gene (9) and at the fragile X locus (10,11). When results from the sperm analyses are considered together, contraction frequencies ranged from 0.2 to 3% for CAG•CTG alleles of 15–31 repeats (7–9), and from 0.2 to 2% for fragile X CGG•CCG tracts of 19–29 repeats (10,11). The sizes of the contractions we observed (−10 to −21 repeats) tend to be larger than what is seen in human sperm, in part owing to the fact that our assay requires loss of at least five repeats, whereas PCR assays in human tissue samples can record smaller changes as well. We cannot rule out that our plasmid-based system might have some differences from TNRs in a chromosomal location.

TNR instability in our assay is dependent on many of the same elements as in human families, including TNR sequence, length and the repeating nature of the tract. This dependence suggests that the contractions seen in our system reasonably mimic TNR instability in humans. Hairpins are implicated as important intermediates by the fact that CAG and CTG repeats, which form hairpins readily, also contract at appreciable frequencies. In contrast, TAG repeats contract less frequently and scrambled sequences are genetically stable. The data also indicate that DNA replication is the most probably mechanistic source for contractions in our system, since removal of the SV40 *ori* sequence from the reporter plasmid eliminated contractions. It is also possible that contractions result from an event following replication, such as recombination between newly replicated regions of the plasmid. Clarification on this point awaits further investigation, such as the use of cell lines deficient in specific recombination or repair pathways. Possible effects on p53, which is inactivated by large T antigen, are largely ruled out by previous studies that found no connection between p53 status and TNR instability (70,71).

The orientation dependence seen in 293T and 293 cells also supports a replication model of TNR contractions, since the instability of the tract depends on the direction of the replication fork. The most persuasive evidence was that relocation of the SV40 origin from one side of the TNR tract to the other resulted in a 5-fold increase in contraction frequency. Thus a relatively stable tract was converted to a more unstable configuration when replication occurred from the opposite direction. Orientation effects for TNR instability have been reported from numerous assays performed in *E.coli*, yeast and mammalian cells (19,20,22,23,30,38), although some reports indicated no obvious orientation effect (72). In these other studies, however, placement of the CTG repeat on the lagging template produces more instability than when CAG repeats occupy the lagging template. The surprise in our 293T and 293 cell experiments was that CAG template sequences trigger more contraction events than for CTG repeat templates. In contrast, about equal levels of instability were observed in COS-1 cells. The orientation effect did not correlate with level of vector replication, as both high-replicating 293T cells and low-replicating 293 cells displayed the orientation effect, whereas high-replicating COS-1 cells did not. Cleary *et al.* (38) found an orientation effect in COS-1 cells, while we observed this phenomenon only in 293/293T cells. The different repeat lengths tested in the two laboratories might contribute to this disparity. Another difference was that we did not systematically vary the distance of the replication *ori* relative to the TNR in our vectors. We offer two possible explanations for our orientation data. First, the threshold for $(CTG)_{25+8}$ tract instability may be lower in COS-1 cells than in 293 or 293T cells. In other words, the minimum number of CTG repeats necessary to bring about instability may be lower in COS-1 cells than in 293/293T cells. This way, while the $(CTG)_{25+8}$ tract is genetically unstable in the former cell line, it is too short to trigger contractions in the latter two lines. Second, specific *trans*-acting factors may exist in COS-1 or 293/293T cells that can modulate TNR instability, in addition to the other factors common to both cell types. (Because the repeats are located on an identical shuttle vector used for all cell lines, *cis*-acting modifiers are unlikely.) These cellular factors may help define the genetic threshold, for example. Further investigation is needed to identify such *trans*-acting modifiers in COS-1 and 293T cells using our assay.

One factor that did not correlate with instability was the expression of the important mismatch repair protein Mlh1. Both instability and replication orientation effects were very similar in 293 cells that show substantial steady-state levels of Mlh1 protein, and in 293T cells, which have undetectable levels of Mlh1 and are accordingly deficient in mismatch repair activity (68). To our knowledge, our data is the first to assess a potential role of Mlh1 in TNR instability. In mismatch repair, heterodimers containing Mlh1 are important modulators of the mismatch binding properties of the Msh2/Msh3 complex [summarized in (73–75)]. Thus transgenic mice lacking Mlh1 will provide an important test in helping understand the role of mismatch repair on TNR instability. If it turns out that TNR mutagenesis in mice is unaffected by loss of Mlh1, it would imply that whatever role Msh2/Msh3 heterodimer plays in TNR instability (45–49) is separable from Mlh1 and therefore unlikely to be mismatch repair *per se*.

How does our selection system compare with others? Two recent publications described genetic selections in mammalian cells for CAG•CTG contractions (40) and expansions (76). When embedded into an intron of either the APRT or HPRT gene, CAG tracts of ~34 or more repeats interfere with mRNA splicing and therefore yield an $APRT^-$ or $HPRT^-$ phenotype (40). Selection for $APRT^+$ or $HPRT^+$ cells showed that the majority of such isolates had contracted the CAG allele to 33 repeats or less. In some cases, deletion of flanking sequences was also observed. The sensitivity of this assay approaches

10^{-6} per cell generation. One limitation of this assay is that it only works for CAG repeats; when CTG repeats were tested, the phenotype of the cells was always $APRT^{+}$. Expansions that increase a $(CAG)_{31}$ tract to 39 or more repeats can be identified by selection to $APRT^{-}$ (76). Thus our genetic assay and that of Wilson and co-workers (40,76) provide alternative, complementary methods, each with different strengths and limitations.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grant GM61961 (to R.S.L.), by a graduate fellowship from the University of Nebraska Medical Center and National Institutes of Health fellowship F30 NS046135 (to B.T.F.), and by National Cancer Institute Cancer Center Support Grant P30 CA36727 (to the Eppley Institute). Funding to pay the Open Access publication charges for this article was provided by NIH GM61961.

Conflict of interest statement. None declared.

REFERENCES

1. Cummings,C.J. and Zoghbi,H.Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet*., **9**, 909–916.
2. Usdin,K. and Grabczyk,E. (2000) DNA repeat expansions and human disease. *Cell. Mol. Life Sci*., **57**, 914–931.
3. Kovtun,I., Goellner,G. and McMurray,C.T. (2001) Structural features of trinucleotide repeats associated with DNA expansion. *Biochem. Cell Biol*., **79**, 325–336.
4. Wells,R.D., Dere,R., Hebert,M.L., Napierala,M. and Son,L.S. (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res*., **33**, 3785–3798.
5. Paulson,H.L. and Fischbeck,K.H. (1996) Trinucleotide repeats in neurogenetic disorders. *Annu. Rev. Neurosci*., **19**, 79–107.
6. McMurray,C.T. (1995) Mechanisms of DNA expansion. *Chromosoma*, **104**, 2–13.
7. Leeflang,E.P., Zhang,L., Tavare,S., Hubert,R., Srinidhi,J., MacDonald,M.E., Myers,R.H., de Young,M., Wexler,N.S., Gusella,J.F. *et al*. (1995) Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum. Mol. Genet*., **4**, 1519–1526.
8. Zhang,L., Leeflang,E.P., Yu,J. and Arnheim,N. (1994) Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. *Nature Genet*., **7**, 531–535.
9. Takiyama,Y., Sakoe,K., Soutome,M., Namekawa,M., Ogawa,T., Nakano,I., Igarashi,S., Oyake,M., Tanaka,H., Tsuji,S. *et al*. (1997) Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (*MJD1*): evidence for non-Mendelian transmission of the *MJD1* gene and for the effect of the intragenic CGG/GGG polymorphism on the intergenerational instability. *Hum. Mol. Genet*., **6**, 1063–1068.
10. Mornet,E., Chateau,C., Hirst,M.C., Thepot,F., Taillandier,A., Cibois,O. and Serre,J.-L. (1996) Analysis of germline variation at the *FMR1* CGG repeat shows variation in the normal-premutated borderline range. *Hum. Mol. Genet*., **5**, 821–825.
11. Kunst,C.B., Leeflang,E.P., Iber,J.C., Arnheim,N. and Warren,S.T. (1997) The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. *J. Med. Genet*., **34**, 627–631.
12. Sinden,R.R., Potaman,V.N., Oussatcheva,E.A., Pearson,C.E., Lyubchenko,Y.L. and Shlyakhtenko,L.S. (2002) Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J. Biosci*., **27**, 53–65.
13. Gacy,A.M., Goellner,G., Juranic,N., Macura,S. and McMurray,C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structure *in vitro*. *Cell*, **81**, 533–540.
14. Gacy,A.M. and McMurray,C.T. (1998) Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. *Biochemistry*, **37**, 9426–9434.
15. Pearson,C.E., Wang,Y.-H., Griffith,J.D. and Sinden,R.R. (1998) 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.
16. Gacy,A.M., Goellner,G.M., Spiro,C., Chen,X., Gupta,G., Bradbury,E.M., Dyer,R.B., Mikesell,M.J., Yao,J.Z., Johnson,A.J. *et al*. (1998) GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. *Mol. Cell*, **1**, 583–593.
17. Pearson,C.E., Tam,M., Wang,Y.-H., Montgomery,S.E., Dar,A.C., Cleary,J.D. and Nichol,K. (2002) Slipped-strand DNAs formed by long (CAG)•(CTG) repeats: slipped-out repeats and slip-out junctions. *Nucleic Acids Res*., **30**, 4534–4547.
18. Rolfsmeier,M.L., Dixon,M.J., Pessoa-Brandao,L., Pelletier,R., Miret,J.J. and Lahue,R.S. (2001) *Cis*-elements governing trinucleotide repeat instability in *Saccharomyces cerevisiae*. *Genetics*, **157**, 1569–1579.
19. Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) 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.
20. Maurer,D.J., O'Callaghan,B.L. and Livingston,D.M. (1996) Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. *Mol. Cell. Biol*., **16**, 6617–6622.
21. Samadashwily,G.M., Raca,G. and Mirkin,S.M. (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet*., **17**, 298–304.
22. Freudenreich,C.H., Stavenhagen,J.B. and Zakian,V.A. (1997) Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol. Cell. Biol*., **17**, 2090–2098.
23. Miret,J.J., Pessoa-Brandao,L. and Lahue,R.S. (1998) Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA*, **95**, 12438–12443.
24. Freudenreich,C.H., Kantrow,S.M. and Zakian,V.A. (1998) Expansion and length-dependent fragility of CTG repeats in yeast. *Science*, **279**, 853–856.
25. Schweitzer,J.K. and Livingston,D.M. (1998) Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. *Hum. Mol. Genet*., **7**, 69–74.
26. Schweitzer,J.K. and Livingston,D.M. (1999) The effect of DNA replication mutations on CAG tract stability in yeast. *Genetics*, **152**, 953–963.
27. Spiro,C., Pelletier,R., Rolfsmeier,M.L., Dixon,M.J., Lahue,R.S., Gupta,G., Park,M.S., Chen,X., Mariappan,S.V.S. and McMurray,C.T. (1999) Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. *Mol. Cell*, **4**, 1079–1085.
28. White,P.J., Borts,R.H. and Hirst,M.C. (1999) Stability of the human Fragile X $(CGG)_n$ triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism. *Mol. Cell. Biol*., **19**, 5675–5684.
29. Ireland,M.J., Reinke,S.S. and Livingston,D.M. (2000) The impact of lagging strand replication mutations on the stability of CAG repeat tracts in yeast. *Genetics*, **155**, 1657–1665.
30. Iyer,R.R., Pluciennik,A., Rosche,W.A., Sinden,R.R. and Wells,R.D. (2000) DNA polymerase III proofreading mutants enhance the expansion and deletion of triplet repeat sequences in *Escherichia coli*. *J. Biol. Chem*., **275**, 2174–2184.
31. Pelletier,R., Krasilnikova,A.S., Samadishwily,G.M., Lahue,R.S. and Mirkin,S.M. (2003) Replication and expansion of trinucleotide repeats in yeast. *Mol. Cell. Biol*., **23**, 1349–1357.
32. Lahue,R.S. and Slater,D.L. (2003) DNA repair and trinucleotide repeat instability. *Front. Biosci*., **8**, s553–565.
33. Jakupciak,J.P. and Wells,R.D. (2000) Gene conversion (recombination) mediates expansions of CTG•CAG repeats. *J. Biol. Chem*., **275**, 40003–40013.
34. Jankowski,C., Nasar,F. and Nag,D.K. (2000) Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. *Proc. Natl Acad. Sci. USA*, **97**, 2134–2139.
35. Richard,G.-F., Goellner,G.M., McMurray,C.T. and Haber,J.E. (2000) Recombination-induced CAG trinucleotide repeat expansions in yeast involve the *MRE11–RAD50–XRS2* complex. *EMBO J*., **19**, 2381–2390.

36. Richard,G.-F., Cyncynatus,C. and Dujon,B. (2003) Contractions and expansions of CAG/CTG trinucleotide repeats occur during ectopic gene conversion in yeast, by a *MUS81*-independent mechanism. *J. Mol. Biol.*, **326**, 769–782.
37. Leeflang,E.P., Tavare,S., Marjoram,P., Neal,C.O.S., Srinidhi,J., MacDonald,M.E., de Young,M., Wexler,N.S., Gusella,J.F. and Arnheim,N. (1999) Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum. Mol. Genet.*, **8**, 173–183.
38. Cleary,J.D., Nichol,K., Wang,Y.-H. and Pearson,C.E. (2002) Evidence of *cis*-acting factors in replication-mediated trinucleotide repeat instability in primate cells. *Nature Genet.*, **31**, 37–46.
39. Panigrahi,G.B., Cleary,J.D. and Pearson,C.E. (2002) *In vitro* (CTG)/(CTG) expansions and deletions by human cell extracts. *J. Biol. Chem.*, **277**, 13926–13934.
40. Gorbunova,V., Seluanov,A., Dion,V., Sandor,Z., Meservy,J.L. and Wilson,J.H. (2003) Selectable system for monitoring the instability of CTG/CAG triplet repeats in mammalian cells. *Mol. Cell. Biol.*, **23**, 4485–4493.
41. Nenguke,T., Aladjem,M.I., Gusella,J.F., Wexler,N.S., Project,T.V.H. and Arnheim,N. (2003) Candidate DNA replication initiation origins at human trinucleotide repeat disease loci. *Hum. Mol. Genet.*, **12**, 1021–1028.
42. Yang,Z., Lau,R., Marcadier,J.L., Chitayat,D. and Pearson,C.E. (2003) Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type I disease locus in human cells. *Am. J. Hum. Genet.*, **73**, 1092–1105.
43. Fortune,M.T., Vassilopoulos,C., Coolbaugh,M.I., Siciliano,M.J. and Monckton,D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum. Mol. Genet.*, **9**, 439–445.
44. Gomes-Pereira,M., Fortune,M.T. and Monckton,D.G. (2001) Mouse tissue culture models of unstable triplet repeats: *in vitro* selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. *Hum. Mol. Genet.*, **10**, 845–854.
45. Manley,K., Shirley,T.L., Flaherty,L. and Messer,A. (1999) *Msh2* deficiency prevents *in vivo* somatic instability of the CAG repeat in Huntington disease transgenic mice. *Nature Genet.*, **23**, 471–473.
46. Kovtun,I.V. and McMurray,C.T. (2001) Trinucleotide expansion in haploid germ cells by gap repair. *Nature Genet.*, **27**, 407–411.
47. Wheeler,V.C., Lebel,L.-A., Vrbanac,V., Teed,A., te Riele,H. and MacDonald,M.E. (2003) Mismatch repair gene *Msh2* modifies the timing of early disease in HdhQ111 striatum. *Hum. Mol. Genet.*, **12**, 273–281.
48. Savouret,C., Brisson,E., Essers,J., Kanaar,R., Pastink,A., te Riele,H., Junien,C. and Gourdon,G. (2003) CTG repeat instability and size variation timing in DNA repair-deficient mice. *EMBO J.*, **22**, 2264–2273.
49. van den Broek,W.J.A.A., Nelen,M.R., Wansink,D.G., Coerwinkel,M.M., te Riele,H., Groenen,P.J.T.A. and Wieringa,B. (2002) Somatic expansion behaviour of the $(CTG)_n$ repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. *Hum. Mol. Genet.*, **11**, 191–198.
50. Gomes-Pereira,M., Fortune,M.T., Ingram,L., McAbney,J.P. and Monckton,D.G. (2004) Pms2 is a genetic enhancer of trinucleotide CAG•CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. *Hum. Mol. Genet.*, **13**, 1815–1825.
51. Kramer,P.R., Pearson,C.E. and Sinden,R.R. (1996) Stability of triplet repeats of myotonic dystrophy and fragile X loci in human mismatch repair cell lines. *Hum. Genet.*, **98**, 151–157.
52. Kaytor,M.D., Burright,E.N., Duvick,L.A., Zoghbi,H.Y. and Orr,H.T. (1997) Increased trinucleotide repeat instability with advanced maternal age. *Hum. Mol. Genet.*, **6**, 2135–2139.
53. Meservy,J.L., Sargent,R.G., Iyer,R.R., Chan,F., McKenzie,G.J., Wells,R.D. and Wilson,J.H. (2003) Long CTG tracts from the myotonic dystrophy gene induce deletions and rearrangements during recombination at the *APRT* locus in CHO cells. *Mol. Cell. Biol.*, **23**, 3152–3162.
54. Marcadier,J.L. and Pearson,C.E. (2003) Fidelity of primate cell repair of a double-strand break within a (CTG)•(CAG) tract. Effect of slipped DNA structures. *J. Biol. Chem.*, **278**, 33848–33856.
55. Gorbunova,V., Seluanov,A., Mittelman,D. and Wilson,J.H. (2004) Genome-wide demethylation destabilizes CTG•CAG trinucleotide repeats in mammalian cells. *Hum. Mol. Genet.*, **13**, 1–11.
56. Gomes-Pereira,M. and Monckton,D.G. (2004) Chemically induced increases and decreases in the rate of expansion of a CAG*CTG triplet repeat. *Nucleic Acids Res.*, **32**, 2865–2872.
57. Edamura,K.N., Leonard,M.R. and Pearson,C.E. (2005) Role of replication and CpG methylation in Fragile X syndrome CGG deletions in primate cells. *Am. J. Hum. Genet.*, **76**, 302–311.
58. Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19–27.
59. Bowater,R.P., Rosche,W.A., Jaworski,A., Sinden,R.R. and Wells,R.D. (1996) Relationship between *Escherichia coli* growth and deletions of CTG-CAG triplet repeats in plasmids. *J. Mol. Biol.*, **264**, 82–96.
60. Graham,F.L. and Van Der Eb,A.J. (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, **52**, 456–467.
61. Hirt,B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.*, **26**, 365–369.
62. Gietz,D., St Jean,A., Woods,R.A. and Schiestl,R.H. (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.*, **20**, 1425.
63. Farber,R.A., Petes,T.D., Dominska,M., Hudgens,S.S. and Liskay,R.M. (1994) Instability of simple sequence repeats in a mammalian cell line. *Hum. Mol. Genet.*, **3**, 253–256.
64. Lee,J.S., Hanford,M.G., Genova,J.L. and Farber,R.A. (1999) Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. *Hum. Mol. Genet.*, **8**, 2567–2572.
65. Hay,R.T. and DePamphilis,M.L. (1982) Initiation of SV40 DNA replication *in vivo*: location and structure of 5′ ends of DNA synthesized in the ori region. *Cell*, **28**, 767–779.
66. Li,J.J. and Kelly,T.J. (1985) Simian virus 40 DNA replication *in vitro*: specificity of initiation and evidence for bidirectional replication. *Mol. Cell. Biol.*, **5**, 1238–1246.
67. DuBridge,R.B. and Calos,M.P. (1988) Recombinant shuttle vectors for the study of mutagenesis. *Mutagenesis*, **3**, 1–9.
68. Trojan,J., Zeuzem,S., Randolph,A., Hemmerle,C., Brieger,A., Raedle,J., Plotz,G., Jiricny,J. and Marra,G. (2002) Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology*, **122**, 211–219.
69. Kovtun,I.V., Thornhill,A.R. and McMurray,C.T. (2005) Somatic deletion events occur during early embryonic development and modify the extent of CAG expansion in subsequent generations. *Hum. Mol. Genet.*, **13**, 3057–3068.
70. Gennarelli,M., Lucarelli,M., Amicucci,P., Soddu,S., Novelli,G. and Dallapiccola,B. (1998) Genomic instability associated with myotonic dystrophy does not involve p53 expression and activity. *Cell Biochem. Funct.*, **16**, 117–122.
71. Fleming,K., Riser,D.K., Kumari,D. and Usdin,K. (2003) Instability of the fragile X syndrome repeat in mice: the effect of age, diet and mutations in genes that affect DNA replication, recombination and repair proficiency. *Cytogenet. Genome Res.*, **100**, 140–146.
72. Miret,J.J., Pessoa-Brandao,L. and Lahue,R.S. (1997) Instability of CAG and CTG trinucleotide repeats in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **17**, 3382–3387.
73. Buermeyer,A.B., Deschenes,S.M., Baker,S.M. and Liskay,R.M. (1999) Mammalian DNA mismatch repair. *Ann. Rev. Genet.*, **33**, 533–564.
74. Nakagawa,T., Datta,A. and Kolodner,R.D. (1999) Multiple functions of MutS- and MutL-related heterocomplexes. *Proc. Natl Acad. Sci. USA*, **96**, 14186–14188.
75. Harfe,B.D. and Jinks-Robertson,S. (2000) DNA mismatch repair and genetic instability. *Ann. Rev. Genet.*, **34**, 359–399.
76. Lin,Y., Dion,V. and Wilson,J.H. (2005) A novel selectable system for detecting expansion of CAG•CTG repeats in mammalian cells. *Mutat. Res.*, **572**, 123–131.