Expansions of CAG-CTG repeats in immortalized human astrocytes.

Lahue, Robert S.

2007


Oxford University Press

http://hdl.handle.net/10379/518

Downloaded 2018-12-21T15:51:59Z

Some rights reserved. For more information, please see the item record link above.
Expansions of CAG·CTG repeats in immortalized human astrocytes

David A. Claassen\textsuperscript{1} and Robert S. Lahue\textsuperscript{1,2,*}

\textsuperscript{1}Eppley Institute for Research in Cancer and Allied Diseases
University of Nebraska Medical Center
Box 986805
Omaha, NE 68198-6805 USA

\textsuperscript{2}Department of Biochemistry
National University of Ireland
Galway, Ireland

*Corresponding Author:
Telephone: +353 91 495756
Fax: +353 91 495504
E-mail: Bob.Lahue@nuigalway.ie
ABSTRACT

Expansions of trinucleotide repeats (TNRs) are the genetic cause for a number of neurodegenerative disorders. In some of these diseases, ongoing somatic expansions in the brain are thought to contribute to disease progression. Expansions can occur in both neurons and supporting glial cells, but little is known about molecular mechanisms of expansion in these cells, particularly glia. To help address this issue, a cultured human astrocyte cell line called SVG-A was tested for expansions of CAG•CTG repeats present on a shuttle vector. A quantitative genetic selection showed that +4 to +15 repeat expansions occur readily for starting alleles of 25 repeats, thereby spanning the important boundary between short stable repeats and longer more unstable CAG•CTG tracts. These expansions in glial cell culture, as in humans, were sequence and length dependent, and were inhibited by the presence of a sequence interruption within the triplet repeat tract. These findings suggest that the mutations seen in cell culture reflect at least some of the in vivo expansions seen in glia. Mechanistically, it was found that the direction of DNA replication through the TNR influenced the frequency of expansions, suggesting that either replication or a replication-associated process, such as DNA repair, contributes to CAG•CTG tract instability in SVG-A cells. This finding is consistent with the idea that replication-based mechanisms can be a source of TNR expansions in astrocytes, which, unlike neurons, retain proliferative capacity throughout life.

INTRODUCTION

Debilitating neurodegenerative disorders such as Huntington’s disease (HD), myotonic dystrophy type 1, and a number of the spinocerebellar ataxias are all caused by expansions of CAG•CTG repeat tracts (1, 2). Given the severity of these diseases, it is of considerable importance to understand the genetic mechanisms underlying trinucleotide repeat (TNR) instability (3). In virtually all CAG•CTG expansion diseases, germline mutations are frequent in affected individuals, and can occur in people carrying premutation alleles. Thus mutation frequency is strongly influenced by the starting allele length relative to a crucial threshold of about 30-40 repeats. In addition to inherited expansions, some diseases also show somatic instability of the TNR tract, and increasing evidence suggests the importance of such somatic instability in the pathology of triplet repeat disorders (4, 5). Within the brain, somatic expansions occur in non-dividing neurons of HD patients (4) and in HD transgenic mice (6). It is evident that neuronal instability must occur by a replication-independent mechanism, perhaps through a DNA repair- and/or transcription-dependent process (6-9). Somatic expansions also occur in glial cells (10, 4) and there is evidence that expanded alleles in glia affect neuron survival and thus TNR pathology (11, 12). Since glia retain the ability to proliferate throughout life, it is possible that replication-based mechanisms could explain TNR instability in these important neuron-supporting cells. There is little known about CAG•CTG mutagenesis in glial cells, however.

Work in cell lines from patients (typically lymphoblasts) has been conducted using small pool PCR and other physical techniques (13-17). These studies are helping to reveal the mechanism whereby long, unstable alleles become even longer in human
cells. A distinct issue, however, is how alleles near the threshold length initially expand into the disease-causing range. Studying this class of expansions is worthwhile because it may help understand the initiating germline mutational event, in addition to somatic instability. Also, it is possible that the mechanisms of mutation could be different near the threshold compared to long alleles. The experimental difficulty has been to detect the less frequent events that occur at TNR lengths near the germline instability threshold of ~30-40 repeats. To help better understand the mechanism of CAG·CTG instability in general, and in glial cell somatic mutations in particular, we reported a sensitive genetic assay for TNR contractions in an immortalized human astrocyte line called SVG-A (18). In the current study, we adapted this assay to monitor CAG·CTG expansions in SVG-A astrocytes. The results of this work demonstrate that expansions in cultured human glial cells share a number of important similarities with the genetics of triplet repeat diseases, and help provide mechanistic insights into replication-dependent CAG·CTG expansions in an important central nervous system (CNS) cell population.

RESULTS

To measure expansions of CAG·CTG repeats in SVG-A cells, the astrocytic subclone of the SVG cell line (19, 20), a new shuttle vector assay was designed. An overview of the expansion assay is shown in Figure 1A. Briefly, shuttle vector plasmids are transfected into SVG-A cells and allowed to replicate for two to three days. The plasmids then are recovered and treated with DpnI to remove unreplicated molecules. Upon transformation into yeast, genetic selection allows measurement of the expansion frequency that occurred during replication in cell culture. Thus yeast serves as a biosensor for plasmids with expansions. The major features of the shuttle vector plasmids (18, 21) are shown in Figure 1B and C. The presence of the SV40 origin of replication (ori) sequence drives replication in SVG-A cells, which express SV40 large T antigen. Additional features of the plasmids are described in Materials and Methods. The mechanism of the genetic selection utilizing the spacing sensitive yeast promoter is shown in Figure 1D. Expansion of the repeat tract leads to 5-fluoroorotic acid (5FOA) resistance; normalizing the frequency of 5FOA resistance to the total number of His \(^+\) transformants (Fig. 1A) provides a measurement of expansion frequency. Background expansions, defined as those occurring without exposure to SVG-A cells, are measured similarly (Fig. 1A). To confirm the presence of an expanded allele, and to measure its size, 5FOA resistant colonies are tested by PCR amplification of the repeat tract and separation of the products by denaturing PAGE. As detailed below, this assay allows sensitive and quantitative measurements of CAG·CTG tract expansions in cells derived from the human CNS.

CAG·CTG expansions occur in cultured human astrocytes

CAG·CTG tracts of 25 repeats were chosen for study, in part because they represent allele lengths just below the clinical and genetic threshold of ~30-40 repeats. We anticipated that 25 repeat tracts would expand at detectable levels, and that these expansions would represent initiating mutations where normal length alleles cross the threshold to generate more instability. In an individual human, such initiating events would increase the risk for additional, disease-causing mutations. Plasmids pBL252 and pBL253, containing the repeat tracts (CTG)\(_{25}\) and (CAG)\(_{25}\) respectively, were tested for
expansions in SVG-A cells. In our nomenclature, the sequence shown is situated within
the Okazaki fragment during DNA replication of the repeat tract, as deduced from the
relative locations of the SV40 origin and the TNR. Figure 2 is a representative gel
showing PCR confirmation of bona fide expanded alleles from (CTG)\textsubscript{25} (lanes 2-5) and
(CAG)\textsubscript{25} (lanes 6-8) starting tracts as well as an allele from a 5FOA resistant colony that
did not have a true expansion (lane1), likely resulting from mutagenic inactivation of the
URA\textsubscript{3} gene. Quantitative analysis showed that plasmids with the (CTG)\textsubscript{25} tract expanded
at a frequency of $3.2 \times 10^{-4}$ (Table 1 entry 9A). This frequency is significantly higher
than the background value of $2.8 \times 10^{-5}$ (Table 1 entry 12A, $P = 0.001$ by Student’s $t$-
test), suggesting that $>90\%$ of the observed expansions occurred in cell culture, not in E.
coli or yeast. (CAG)\textsubscript{25} tracts expanded with a frequency of $4.6 \times 10^{-5}$, which while a low
value, was significantly above the background expansion frequency for the tract (Table 1
entries 9B, 12B, and 13B). Clearly, the (CTG)\textsubscript{25} and (CAG)\textsubscript{25} tracts are unstable and
result in expansions.

It was formally possible that pre-existing expansions from production of the
plasmids in E. coli might be more highly replicated by the human cells, or might be
preferentially recovered from the astrocytes. To help rule out this possibility, a mixture
was created with approximately 50\% (CTG)\textsubscript{25} and 50\% (CTG)\textsubscript{33} plasmid, where the 33
repeat molecules were used as a representative for longer alleles. This plasmid mixture
was passaged through the human astrocytes, recovered and transformed into yeast. As a
control the same mixture was put directly into yeast. The transformed yeast were then
plated onto medium lacking histidine that selects for plasmid uptake, but that is non-
selective with regard to repeat tract length. Fifty colonies from the background plasmid
mixture and 50 colonies from plasmid recovered from SVG-A cells were lysed and had
their repeat tract measured by PCR and electrophoresis. The plasmid mixture that had
bypassed human cell culture yielded colonies with a 33 repeat:25 repeat ratio of 0.75.
Plasmids that had been passaged through the human cells gave a ratio of 0.49. Since the
ratio went down, not up, we conclude that plasmids containing longer repeats are not
favorably replicated in SVG-A cells, nor preferentially selected for in yeast.

**Expansion sizes in SVG-A cells**

PCR analysis was used to measure the expansion size ($\pm 2$ repeats) of 35
independently arising, bona fide expansions of the (CTG)\textsubscript{25} plasmid in SVG-A cells
(Figure 3). Expansions from (CTG)\textsubscript{25} ranged in size from $+4$ to $+15$ repeats, with a
median expansion size of $+10$ repeats. Similarly, the eight expansions from (CAG)\textsubscript{25}
ranged from $+4$ to $+15$ repeats with a median value of $+9$ to $+10$. In vivo somatic
expansions of 40 to 50 repeat CAG·CTG tracts measured in brain (22), peripheral blood
leukocytes and sperm (23, 24) in the spinocerebellar ataxia type 2, Huntington’s disease,
and spinocerebellar ataxia type I genes range in size from $+1$ to $+27$ repeat units. Thus
the expansion sizes seen in SVG-A cells overlap with in vivo expansions of multiple
CAG·CTG disease genes. We note that many expansions in humans are $+1$ to $+3$ repeats,
which our assay is not designed to detect. This fact may also help explain the lower
expansion frequency in SVG-A cells than in vivo. The spectrum of expansions seen in
plasmids from SVG-A cells was not statistically different from background expansions,
judged by two-tailed Student’s $t$-test (data not shown).
TNRs unable to form stable alternative secondary structures are stable in the shuttle vector assay

The finding that disease-related triplet repeat sequences have the ability to form stable non-B DNA structures in vitro (25, 26) led to models predicting that TNR expansions result from error-prone replication or repair of abnormal DNA structures that form at triplet repeat tracts (reviewed in 27-30). To test whether the ability of a TNR to form non-B structures alters its tendency to expand in SVG-A cells, we transfected cells with shuttle vectors containing either a (CTG)25 allele, predicted to form a hairpin under physiological conditions, or a (TAG)25 allele, which is not predicted to form alternative secondary structures. (TAG)25 plasmids were genetically stable in our system, as no expansion events were detected out of more than 430,000 plasmids, yielding an expansion frequency of less than $1.1 \times 10^{-6}$ (Table 1, entry C9). In contrast, (CTG)25-ori switched (described below, and in the same plasmid backbone as (TAG)25 reporter) was at least 100-fold more unstable at $1.1 \times 10^{-4}$ ($P=0.016$). This result supports the notion that secondary structure formation is important for expansion in our system.

CAG·CTG expansions in SVG-A cells show length dependence

One of the key features of TNR instability in human disease is the relative stability of shorter alleles compared to longer, more mutation-prone tracts. To test the hypothesis that shorter CAG·CTG tracts would expand less frequently than longer tracts in SVG-A cells, we created a (CTG)17+8 repeat insert with 17 repeats plus 24 random nucleotides appended, to normalize the spacing to 75 total nucleotides as in the other tracts. Thus expansions of four repeats or more will show up in the assay, regardless of the initial number of repeats. The (CTG)17+8 expanded at about one-half the rate of (CTG)25, supporting length-dependent mutability (Table 1, compare entry D9 and A9).

Although the (CTG)17+8 expansion frequency was not statistically significantly different from background it was lower than (CTG)25 with borderline statistical significance ($p=0.056$). The constraints of the shuttle vector assay prevent detection of expansions from starting tracts longer than 25, preventing a more thorough investigation of length dependence. Nonetheless, we conclude that shorter triplet repeat tracts are less expansion prone than their longer counterparts in our system.

Imperfect repeats are protected from expansion in cultured astrocytes

In humans, CAG·CTG tracts harboring short interrupting substitutions are not prone to expansions, unlike equivalent uninterrupted alleles (31-33). To determine whether interrupted TNRs in SVG-A cells are also less prone to expansions, we constructed a shuttle vector containing an interrupted repeat, (CTG)17(ATG)2(CTG)6, that mimics naturally occurring alleles of SCA1 (31) and compared its expandability to a perfect (CTG)25 control. The calculated frequency for interrupted repeats, $< 1.7 \times 10^{-5}$, was not detectable above the background frequency of $2 \times 10^{-5}$ (Table 1, row E), and was at least 6-fold lower (Table 1, entries E9 and F9) than its uninterrupted counterpart (CTG)25-ori switched (see next paragraph), suggesting that imperfect TNRs are protected from expansion in our system. Despite the fact that no expansions were observed from SVG-A cells, two authentic expansions were recovered from the background experiment (Table 1, entry E11). This proves that the interrupted plasmid can undergo expansion, so
the lack of mutability in the SVG-A cell experiment is most likely due to stabilization of the repeat tract per se.

Direction of replication influences the expansion frequency

Our finding (Table 1) that expansions of (CTG)$_{25}$ tracts occurred about seven times more frequently than (CAG)$_{25}$ is consistent with a model where CTG tracts in the Okazaki fragment are more likely to adopt secondary structure (Fig. 4A, left panel), and therefore more likely to expand, than CAG runs (Fig. 4A, right panel). The model also leads to the testable prediction that switching the direction of replication through otherwise identical repeats will alter the expansion frequency (Fig. 4B). The plasmid (CAG)$_{25}$-ori switched contains repeats in the same sequence context as the (CTG)$_{25}$ vector, but switching the position of the SV40 origin results in replication fork movement through the repeat in the opposite direction, such that CAG repeats comprise the Okazaki fragments (compare Fig. 4AB, left panels). The model predicts that this should reduce the expansion frequency and this was borne out experimentally (Fig. 4C and Table 1 entries 9A and 9G). The expansion frequency was sharply reduced for (CAG)$_{25}$-ori switched compared to (CTG)$_{25}$. By similar logic, (CAG)$_{25}$ can be switched to yield (CTG)$_{25}$-ori switched with the expectation of a higher expansion frequency (Fig. 4AB, right panels). (CTG)$_{25}$-ori switched had more than double the frequency of expansions compared to (CAG)$_{25}$, partially restoring the expandability (Figure 4C and Table 1 entries 9B and 9F; P=0.001). In the two ori-switched reporters, (CTG)$_{25}$-ori switched was substantially more unstable than (CAG)$_{25}$-ori switched (Fig. 4C and Table 1 entries 9F and 9G; P<0.001). Thus, for both cases tested, expansions are more likely when the Okazaki fragment is populated with CTG sequences rather than CAG repeats. These differences suggest that expansions in our system are dependent on DNA replication or on some other process, perhaps repair, that is stimulated by replication.

DISCUSSION

The expansion assay described in this study provides an experimentally tractable system for analyzing CAG-CTG tract instability. In particular, the TNR sequences can be easily manipulated and assayed for expansions. In a previous study, a complementary system was described to monitor contractions (18, 21). Thus both expansions and contractions can be quantified and examined molecularly. Importantly, the spectrum of expansions detected (+4 to +15, to final lengths of 29-40 repeats) cross the predicted boundaries between normal, moderate, and disease-causing alleles for most CAG-CTG repeat disease genes. The use of the SVG-A human astrocytic cell line provides an ex vivo model for studying instability in the glial cells of the brain. The selective system allowed detection of CAG-CTG repeat expansions in SVG-A cells at frequencies significantly above background, and near the critical threshold length, which can be difficult to study in non-selective assays. These are likely to have genuinely occurred in the astrocytes, as opposed to merely reflecting a favorable selection for pre-existing longer repeat tracts, based on the statistical significance of the 33 repeat/25 repeat plasmid mixing experiment. Taken together, the assay and the cell line provide a valuable tool for future studies and useful new information about mechanisms of somatic instability and initiating mutagenesis of CAG-CTG tracts in the human CNS.
The CAG-CTG expansion frequencies detected here lie in the middle relative to other systems with similar starting allele lengths. The expansions in SVG-A cells occurred from $3.2 \times 10^4$ for (CTG)$_{25}$ to $4.6 \times 10^5$ for (CAG)$_{25}$. The frequency of expansion in SVG-A cells was higher by approximately two orders of magnitude than that seen for (CAG)$_{31}$ alleles in Chinese hamster ovary cells (34). In contrast, expansions in germline cells tend to be more frequent. For example, 36 repeat tracts at the HD locus expand about 8% of the time, judged by single sperm analysis (35, 36). Besides the range of allele lengths (25, 31, and 36 repeats), these differences in expansion frequency may also reflect cell specific features and/or a difference between germline and somatic instability. Compared to expansions, contraction rates for (CTG)$_{25+8}$, in SVG-A were similar at a frequency of $1.9 \times 10^{-4}$ (18). This suggests that CTG tracts just below the threshold are about as likely to expand as contract. In contrast, the higher rate of (CAG)$_{25+8}$ contractions in SVG-A, $17 \times 10^{-4}$ (18), corresponds to a ~37-fold preponderance in favor of contractions. The basis for the difference between CTG and CAG repeats is not yet clear. The tendency toward contractions for alleles shorter than a threshold around 35 repeats is evident in the human germline as well (35, 36). Overall, the likelihood of expansion or contraction in SVG-A cells is lower than germline but higher than other cell culture systems.

Several additional features of human genetics are mimicked in SVG-A cells. The instability of a TNR sequence and its ability to cause human disease are correlated with the ability of the sequence to form hairpins in vitro (2, 25, 26). A TAG tract not predicted to form stable non-B DNA structures under physiological conditions did not expand at a detectable level. Another key feature of TNR instability in humans is length dependence, and we observed higher expansion frequencies for 25 compared to 17 repeat CTG runs. Interruptions in TNR tracts stabilize the repeats and decrease the likelihood of expansion in both spinocerebellar ataxia type 1 (31) and in yeast (37). A model has been proposed and tested in yeast in which interruptions prevent instability via mismatch repair-mediated processing of putative hairpins (38). That interrupted repeats were apparently protected from expansion in SVG-A cells recapitulates the human genetics observation and fits a model of interruption-mediated stability.

There is an active debate about TNR mutability in proliferating versus non-proliferating cells. In terminally-differentiated neurons, DNA replication cannot be a major source of expansions. In contrast, glial cells retain proliferative capacity throughout life, and thus replication could account for expansions. We observed 7- to 20-fold differences in expansion frequencies in SVG-A cells when tracts of CTG, versus CAG, were incorporated into the Okazaki fragment (Fig. 4). This orientation effect is reminiscent of a replication-dependent mutagenic mechanism. The reason for this orientation effect may be that CTG repeats are predicted to form more stable hairpins than CAG repeats, thus increasing the likelihood that an expansion would ensue. We note that reversing the direction of replication in (CTG)$_{25}$-ori switched did not fully restore the expansion frequency to as high a level as (CTG)$_{25}$ (Fig. 4C). Some expansions may be replication-independent, arising perhaps from error-prone DNA repair. Alternatively, replication-dependent instability could be influenced by the distance between the SV40
origin and the edge of the repeat tract, which varies in some of our constructs. Nonetheless, the finding of an orientation effect fits well with observations of differential expandability in *E. coli* (39) and yeast (40-42) depending on the orientation of replication through TNRs. Similarly, in human cell extracts (43) and in intact simian cells (44) an orientation effect was seen wherein CTGs in the Okazaki fragments were expansion prone while CAGs tended to contract. We conclude that replication or a replication-associated process, such as DNA repair, contributes to CAG•CTG tract instability in SVG-A cells.

**MATERIALS AND METHODS**

**Cells and Cell Culture**

SVG-A cells, an astrocytic subclone of the astroglial SVG cell line (19, 20), were obtained as a gift from Dr. Ashok Chauhaun, Johns Hopkins University. These cells are immortalized with SV40 and express replication competent SV40 Large T Antigen. Cells were grown adherently in DMEM (Mediatech, Inc.) supplemented with 10% fetal bovine serum, 50 U/ml Penicillin/Streptomycin and 2.5 mg/ml Amphotericin B (all from Invitrogen Corporation). Previous analysis (18) showed positive staining for glial fibrillary acidic protein in all ~1,000 cells examined, consistent with a homogeneous population of glial cells.

**Plasmids and Cloning**

All shuttle vectors (Figure 1B and C) were produced as described previously (18, 21). Briefly, these plasmids are derivatives of pRS313 (45). The presence of the SV40 *ori* enables plasmid replication in SVG-A cells. Yeast markers include a selectable promoter-TNR-*URA3* reporter construct for identification of TNR length changes (Figure 1D) (21), a *HIS3* gene allowing selection for plasmid molecules independent of TNR length, while the *ARS/CEN* sequence enables low-copy plasmid replication and segregation in yeast.

Two different shuttle vector backbones were used in this study. The first backbone is pBL245 (Figure 1B). This backbone contains the SV40 origin of replication cloned in the ApaI site of pRS313 to create pBL185 (9). SVG-A cells constitutively express SV40 large T antigen, allowing the exclusion of large T gene sequences found in the other backbone used in this study, pBL67 (Figure 1C). An advantage of excluding this material from pBL245 was that it eliminated an SphI site from the vector backbone that was present in the large T antigen gene, making unique the SphI site in the promoter-*URA3* cassette. The empty promoter-SphI-*URA3* cassette, containing no TNRs, was transferred into pBL185 after digestion with the EcoRI and BamHI, producing pBL245. SphI digestion (New England Biolabs, Inc.) of pBL245 enabled annealed TNR oligonucleotides containing SphI sticky ends to be cloned into the shuttle vector backbone directly. The second backbone used here, pBL67 (Figure 1C), contains the SV40 origin of replication and large T antigen gene cloned into the BamHI site of pRS313. The promoter-TNR-*URA3* cassettes were transferred into the vector backbone as previously described (18). First, complementary oligonucleotides containing TNRs and SphI sticky ends were annealed and cloned into the SphI site of pBL94 (42, 46). This produced promoter-TNR-*URA3* constructs which were then digested with EcoRI and XhoI (New England Biolabs, Inc.) and transferred en bloc to the EcoRI/XhoI sites of pBL67 (21).
Since cloning the SV40 origin of replication into the Apal site of pRS313 resulted in its placement on the opposite side of the repeat tract in pBL245 versus pBL67, a standard system of nomenclature was adopted to ensure consistency between experiments. In all experiments, the named repeat, such as (CTG)\textsubscript{25}, refers to a plasmid molecule where CTG repeats comprise the Okazaki fragments (Figure 4A). That is, CAG repeats comprise the template for lagging strand synthesis.

**Shuttle vector assay**

The shuttle vector assay (Figure 1A) was modified from previous studies (18, 21) to select for expansion events in yeast (46) and slightly altered to increase plasmid recovery. Briefly, the day before transfection, \(3 \times 10^5\) SVG-A cells were seeded in 60 mm tissue culture dishes. The next day, cells were transfected using Lipofectamine reagent (Invitrogen Corporation), according to the manufacturer’s protocol using 7 \(\mu\)g of shuttle vector DNA and 20 \(\mu\)l of Lipofectamine per dish. After transfection, cells were incubated for 72 h to enable plasmid replication. The cells were lysed and plasmid DNA was extracted using Hirt’s alkaline lysis technique (47). Lysates were concentrated by centrifugation through a Centricon 50 kDa MWCO filter (Millipore Corporation). Plasmid DNA was precipitated with ethanol and digested with DpnI (New England Biolabs, Inc). DpnI digestion eliminates any plasmid molecules that did not replicate in human cells. DpnI-resistant DNA was then transformed into yeast strain S1502B (\(MATa\ leu2^{-3},\ leu2^{-112}\ his3^{-}\Delta\ trp1^{-289}\ ura3^{-52}\)) using the lithium acetate method (48). 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 5FOA (synthetic complete, lacking histidine, with 5-fluoro-orotic acid) to select for expansions. Colonies on each plate were counted after 3 days of growth at 30°C. The observed frequency of expansion was determined as the number of colonies obtained on SC-His 5FOA divided by the total number of transformants on SC-His, with appropriate correction for dilution factors. Corrected frequencies took into account PCR confirmation, as described later. Statistical analyses were performed by applying a two-tailed Student’s t-test with equal variance to the relevant datasets. Comparisons between datasets observed in SVG-A cells were made with no correction for background. Full details of the assay results are given in Table 1.

A critical control for this assay involves determination of the background expansion frequency, defined as the percentage of expanded TNR tracts in the plasmid population that did not occur in human cells. The background was determined by transforming stock plasmid directly into yeast, bypassing replication in SVG-A cells (Figure 1A). The background provides quantitation of any TNR expansions that occurred in \(E.\ coli\) during preparation of stock plasmid, or in yeast before selection became effective. Background frequencies were calculated from at least three independent measurements and corrected background frequencies were determined using PCR data (see below).

**Identification of TNR expansions**

Yeast was used as a biosensor (21) for expansions arising in SVG-A cells (Figure 1D). Briefly, gene expression driven by the \(Schizosaccharomyces\ pombe adh1\) promoter is dependent on the proper spacing between the TATA box and the downstream
transcription initiation site (labeled ‘I’ in the figure). When a TNR containing 25 triplet repeats is inserted between the TATA box and ‘I’, yeast transcription initiates at the preferred initiation site, translation begins at the URA3 ATG start codon, making the cells Ura+. A Ura+ cell will incorporate 5FOA into its nucleotide pool, rendering the cell sensitive to media containing 5FOA. However, if an expansion to a TNR length of ≥30 repeats occurs in human cells, upon transfer into yeast transcription will initiate upstream of “I” causing an out of frame ATG to be encoded and a nonsense URA3 gene product to be produced. Such a cell will become phenotypically Ura- and resistant to 5FOA. Thus 5FOA resistant yeast colonies reflect expansion events that occurred in the SVG-A cells. Note that in some plasmids the initial TNR tract was <25 repeats. In these instances, genetically inert, randomized sequences were appended to make the total DNA length 75 nt. For example 17 CTG repeats +24 nt of randomized sequences are referred to as (CTG)\textsuperscript{17}+8.

**PCR analysis**

To authenticate expansions and to determine their size, individual 5FOA resistant yeast colonies were disrupted in 100μl of 50 mM DTT plus 0.5% Triton X-100, incubated at 95°C for ~6 min. and kept on ice thereafter. Subsequent PCR amplification by HotStart-IT Taq (USB Corporation) used Cy5-labeled primers flanking the repeat tract for 30 cycles (1 min at 95°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 standard containing repeats of known size. PCR confirmation of expansions is especially important since randomly occurring mutations within the URA3 gene itself could result in a 5FOA resistant phenotype. The final corrected expansion frequency for each template was determined by multiplying the fraction of plasmids with \textit{bona fide} alterations in TNR size times the raw expansion frequencies calculated above.

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge Lahue lab members for helpful discussion and criticism in manuscript preparation. We also thank Dr. Richard Pelletier and Dr. Brian Farrell for instruction in the procedures for the shuttle vector assay and for cloning several plasmids used here. This work was supported by National Institutes of Health grant GM61961 and by Science Foundation of Ireland grant 06/IN.1/B73 (to R.S.L.), by a graduate fellowship from the University of Nebraska Medical Center (to D.A.C) and by National Cancer Institute Cancer Center Support Grant P30 CA36727 (to the Eppley Institute).

**CONFLICT OF INTEREST**

None

**REFERENCES**


**TABLE LEGEND**

*Table 1.* Column 1 lists the trinucleotide repeat sequence comprising the Okazaki fragments (as defined in Materials and Methods). Column 2 gives each plasmid’s name for reference. Column 3 indicates the vector backbone in which each TNR sequence was cloned (see Fig. 1BC for reference). Column 4 indicates the total number of apparent expansions observed in transfections in SVG-A cells (the number of colonies counted on SC-His 5FOA plates). Column 5 lists the total number of plasmids in the population (the number of colonies counted on SC-His plates multiplied by the appropriate dilution factor). Column 6 lists the number of independent transfections performed for each sequence. Column 7 lists the expansion frequency observed for plasmids replicated in SVG-A cells (× 10^-4) ± 1 SEM (× 10^-4). Column 8 lists the fraction of 5FOA resistant colonies with *bona fide* expansions confirmed by PCR. Column 9 indicates the corrected expansion frequency obtained by multiplying the fraction of confirmed expansions times the observed expansion frequency in SVG-A cells. The corrected expansion frequency is used to produce the graphs in subsequent figures. Column 10 lists the observed background expansion frequency (× 10^-4) ± 1 SEM (× 10^-4) for plasmids that did not undergo SVG-A replication, as described in Materials and Methods. Column 11 indicates the fraction of background expansion events confirmed by PCR. Column 12 lists the corrected background expansion frequency calculated by multiplying the fraction of confirmed events times the observed background expansion frequency. Column 13 lists the p-value for each sequence vs. the background as determined by two-tailed Student’s t-test.

**FIGURE LEGENDS**

*Figure 1.* Assay outline and shuttle vector plasmids. Details are provided in Materials and Methods. (A) Each TNR-containing vector was transfected into SVG-A cells, the cells were cultured for 2-3 days, then the plasmid DNA was isolated, DpnI treated, and transformed into yeast for analysis. Expansion frequency was calculated by dividing the number of yeast colonies with a PCR-confirmed expansion by the total number of transformants. Background expansion frequencies were measured by transforming the plasmid stock directly into yeast and ascertaining the frequency of confirmed expansions. (B) The shuttle vector pBL245 contains several important genetic elements enabling analysis of TNR instability in cultured cells. Genetic elements tinted gray, such as the SV40 origin of replication, drive plasmid replication in SVG-A cells. A striped-pattern indicates yeast genetic elements for determination of expansion frequency by selection in *Saccharomyces cerevisiae*. Dashed bars indicate *E. coli* elements that enable replication and large-scale preparation of plasmid by plasmid maxi-kit (Qiagen Inc). (C) The vector
pBL67 follows the same pattern scheme, the only difference between the two vectors being the presence of the large T antigen gene and the relocation of the SV40 ori to the opposite side of the repeat tract. (D) The selection for expansions in yeast is based on spacing sensitivity of the S. pombe adh1 promoter to the distance between the TATA box and the preferred transcription initiation site, labeled “I”. Starting TNR lengths of 25 allow expression of the URA3 reporter gene and correspondingly yield a 5FOA<sup>S</sup> phenotype (46). TNR expansions that add at least four repeats alter promoter activity to begin transcription upstream of “I” and result in a silencing of URA3 expression and, hence, a 5FOA<sup>R</sup> phenotype.

Figure 2. PCR confirmation of CTG and CAG expansions in SVG-A cells. 5FOA resistant colonies from separate transfections were lysed and PCR amplified. Fragment sizes, given in repeat units, were deduced by comparison with molecular size standards (not shown). A control reaction from a plasmid with a starting tract length of 25 repeats is shown in lane 1. Lanes 2-5 are representative expansions from (CTG)<sub>25</sub>. Lanes 6-8 show expansions from (CAG)<sub>25</sub>. Allele lengths are estimated to within ±2 repeats.

Figure 3. Mutation spectrum for 35 genetically-independent expansions of (CTG)<sub>25</sub>. The number of observed events is plotted on the y-axis vs. the size of the expansion (as determined by PCR) on the x-axis.

Figure 4. Orientation of the TNR relative to the direction of replication influences CAG·CTG expansion frequencies in SVG-A cells. The direction of fork movement was deduced from the placement of the SV40 origin as shown in Fig. 1BC. (A) In the left panel, CTG repeats comprise the Okazaki fragments. In the right panel, the reverse orientation of the repeat tract places CAG repeats in the Okazaki fragments. Note that the replication fork proceeds from left to right. (B) Switching the position of the SV40 ori reverses the direction of the replication fork to right-to-left and thereby reconfigures the leading and lagging strand. The CAG<sub>ori switched</sub> reporter arises from reversing the replication fork through the CTG orientation, directly above. Similarly, CTG<sub>ori switched</sub> derives from the CAG orientation above it. (C) Expansion frequency (×10<sup>-4</sup>) for SVG-A replicated (filled bars) and background (open bars) plasmids is plotted as given in Table 1 (columns 9 and 12) for the indicated starting alleles. Error bars indicate ±1 SEM.