Published online February 24, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 4 ***1289–1297***
DOI: 10.1093/nar/gkh292

DNA elements important for CAG·CTG repeat thresholds in *Saccharomyces cerevisiae*

Michael J. Dixon and Robert S. Lahue*

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

Received October 24, 2003; Revised January 6, 2004; Accepted January 28, 2004

ABSTRACT

Trinucleotide repeat (TNR) instability is of interest because of its central role in human diseases such as Huntington's and its unique genetic features. One distinctive characteristic of TNR instability is a threshold, defined as a minimal repeat length that confers frequent mutations. While thresholds are well established, important risk determinants for disease-causing mutations, their mechanistic analysis has been delayed by the lack of suitably tractable experimental systems. In this study, we directly compared for the first time three DNA elements—TNR sequence, purity and flanking sequence—all of which are suggested in the literature to contribute to thresholds. In a yeast model system, we find that CAG repeats require a substantially longer threshold to contract than CTG tracts, indicating that the lagging template repeat sequence helps determine the threshold. In contrast, ATG interruptions within a CTG run do not inhibit contractions via a threshold mechanism, but by altering the likelihood of forming a hairpin intermediate. The presence of a GC-rich flanking sequence, similar to a haplotype found in some Huntington's patients, does not detectably alter expansions of Okazaki fragment CTG tracts, suggesting no role for this flanking sequence on thresholds. Together these results help better define TNR thresholds by delineating sequence elements that modulate instability.

INTRODUCTION

The inheritance patterns of trinucleotide repeats (TNRs) have attracted considerable attention because the unstable nature of these microsatellites profoundly influences a number of neurological disorders, such as Huntington's disease (HD). TNRs expand at high frequency in affected families (reviewed in 1–5), with TNR length as an especially important risk factor. Triplet repeats become unstable in humans once they achieve a minimum allele size, called a threshold, which is usually about 30–40 repeats long (summarized in 1). Repeat lengths below the threshold are relatively stable, whereas repeats longer than the threshold are unstable. For example, single sperm analysis of the HD locus (6) showed short tracts of 15–18 CAG repeats mutated at a frequency of 0.6%. The mutation frequencies increased to 11% and 53%, respectively, for intermediate repeat tracts of 30 and 36 repeats, and rose further to 92–99% for diseased alleles of 38–51 repeats. Most mutations above the threshold are expansions (1–5), but contractions also occur, and they make up a significant fraction of events for short tracts (6). A threshold, therefore, is a unique feature of TNR genetics that distinguishes short, genetically stable alleles from long, unstable alleles. For the polyglutamine class of TNR diseases in particular, the particularly acute dependence of expansion risk on repeat lengths near the threshold means that allelic changes of just a few repeats can strongly influence the likelihood of disease for the individual, as well as the chance of expansion in subsequent generations (1,3). This reality helps to emphasize that thresholds are very important and more than just a genetic phenomenon.

The molecular nature of thresholds remains unclear. One possibility is that the threshold is largely determined by the formation of aberrant DNA structures. These structures are widely believed to be key mutagenic intermediates that can arise during replication, repair and/or gene conversion (4,5,7–9). Regardless of the mechanistic source of instability, the formation of unusual secondary structure is common to all models of TNR mutability. Structural studies of TNR sequences associated with disease showed that these sequences form secondary structures such as hairpins (when single-stranded), triplexes and slipped-strand duplexes (10–18). Biophysical analysis from Gacy *et al.* (13) led to the hypothesis that the threshold corresponds to the minimum number of repeats required for stable secondary structure formation. Taking this model one step further, Mitas (15) reported that CTG repeats can adopt a hairpin configuration with fewer repeats than its complementary sequence, CAG. The difference between the two repeat tracts is the T nucleotide from CTG stacks better within the helix than does the A of CAG. Experiments by Pearson *et al.* (17) showed that loops of slipped-strand duplexes are more structured for CTG sequences than for CAG repeats. These observations support the idea that TNR sequence might influence thresholds because of structural/energetic considerations. Support also comes from biological evidence in bacteria, yeast and human cells. Expansions and contractions are more frequent on the DNA strand containing the CTG

*To whom correspondence should be addressed. Tel: +1 402 559 4619; Fax: +1 402 559 8270; Email: rlahue@unmc.edu repeat compared with the CAG run (19–25), where 'strandedness' is determined by the nearest replication origin.

The purity of the TNR tract is another element that might impact thresholds. Normal TNR alleles at the *SCA1*, *SCA2* and *FRAXA* loci are punctuated with one to three base-pair interruptions, but expanded TNR alleles at these loci contain fewer or no interruptions (26–28). For example, the *SCA1* locus in 98% of normal individuals contains CAT interruptions within the CAG repeat tract, but families afflicted with SCA1 show expanded CAG tracts that are contiguous (26). These observations are usually interpreted to mean that interruptions stabilize (prevent expansions) at these loci. Three molecular explanations are possible for this stabilization. First, *in vitro* analyses of interrupted TNRs showed that interruptions weaken secondary structures (13,29,30). The second interpretation is that interruptions break up the repeat tract into shorter, subthreshold domains (28,31–33). Rather than being incorporated into a hairpin or other structure, perhaps the interruptions somehow physically demarcate where structures can form, consistent with the finding that interruptions limit the number of slipped-strand isomers that form *in vitro* (30). The third interpretation is that interruptions lead to mismatched bases in or near a TNR hairpin, thus providing a target for DNA mismatch repair. In yeast, it was shown that CTG expansions of interrupted repeats were inhibited primarily *in trans* by mismatch repair (34). This mechanism has not yet been demonstrated in other systems.

Flanking DNA sequence is another candidate factor for influencing TNR instability, possibly by altering thresholds. Evidence from HD genetics suggests that the flanking polymorphic CCG repeats exert a haplotype effect on CAG expansion risk (35–37). The CCG tract is itself stable and does not change when CAG expansions occur, however, indicating that any modulation of expansion risk must be indirect. Structural studies examining the hairpin-forming ability of the HD CAG repeat, as well as several other expandable TNRs, indicated that flanking sequences could participate in and strengthen hairpin structures (13). Further modeling studies (35) indicated that a specific polymorphism found in the CCG repeat in one family would further stabilize hairpin formation by the HD sequence. It has not been possible to assess these ideas directly using human genetics.

Thresholds are of considerable interest for the reasons summarized above. Although thresholds have been known in the literature for some time, the lack of appropriate experimental systems has delayed the ability to directly test the DNA elements that influence them. Here we present a study in yeast that examines threshold-dependent TNR instability. Our results provide clear-cut support for the idea that TNR sequence directly influences thresholds, at least for CAG·CTG repeats, but that interruptions and flanking elements do not.

MATERIALS AND METHODS

Strains

The *Escherichia coli* strain DH5α [*endA1 hsdR17* (*rk*$^-$ *mk*$^+$) *supE44 thi-1 recA1 gyrA* (*nal*^{r}) *relA* Δ (*lacI ZYA-argF*) *U169 deoR*] was used for plasmid constructions and large-scale plasmid preparations. The *Saccharomyces cerevisiae* strain used was MW 3317–21A [*MAT*α Δ*trp1 ura3-52 ade2*Δ *ade8 hom3-10 his3-KpnI met4 met13* (38)]. An isogenic derivative containing a disruption in the *PMS1* gene was described previously (34). TNR-containing plasmids were directed to integrate at either *LYS2* or *ura3* by *Bsu*36I or *Stu*I digestion, respectively, followed by transformation via the lithium acetate protocol (39). Single integrations of the TNR sequence were confirmed by Southern hybridization.

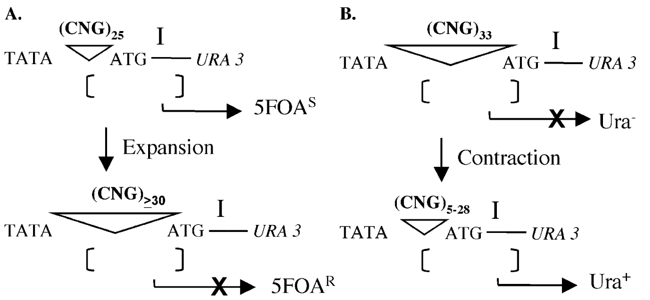


Figure 1. A genetic assay to monitor TNR alterations in yeast. The regulatory region controlling expression of the reporter gene *URA3* is shown. The important features include the following: the TATA box; the trinucleotide region (marked with an inverted triangle); an out-of-frame ATG initiator codon; the preferred transcription initiation site I (CCACA sequence); and the start of the *URA3* structural gene. The sense strand of *URA3* corresponds to the lagging strand template (22). (**A**) shows the starting construct with anticipated transcription (right-angle arrow) initiating within 55–125 base pairs (square parentheses) from TATA. Initiation at I results in functional expression of *URA3* and sensitivity to 5-FOA. If the TNR expands to ⩾30 repeats [lower half of (A)], the window of allowed transcription no longer includes I. Transcription initiation upstream of I will include the out-of-frame ATG, resulting in translational incompetence (indicated by X) and resistance to 5-FOA. (**B**) Includes the same important features as (A), except that the TNR contains 33 repeats. The top diagram therefore illustrates a situation where transcription will initiate upstream of I and include the out-of-frame ATG which will result in translational incompetence. The failure to express the functional *URA3* gene leads to the inability to grow without uracil (Ura$^-$). If the TNR loses five repeats or more (to a final tract length of ⩽28 repeats) (lower diagram), initiation will begin at I and *URA3* will be expressed. The cells with the contracted alleles will change to a Ura$^+$ phenotype, i.e. be capable of growth on media lacking uracil. In both panels, the 'top' strand (i.e. the sense strand of the *URA3* gene) is the lagging strand template.

Plasmids

All plasmids were constructed using the pBL94 vector described previously (23). Briefly, this vector contains the *URA3* gene driven by the *Schizosaccharomyces pombe adh1* promoter, with a unique *Sph*I site separating the TATA box and the preferred transcription initiation site (22; see also Fig. 1). Oligonucleotide duplexes containing the triplet repeat region of interest were created with a CATG 3′ extension to allow for insertion into the *Sph*I site. Many of the TNR derivatives used in this study were reported earlier (22,23,34). Constructs unique to this study include the $(CAG)_{15+18}$ and $(CAG)_{20+13}$ repeat contraction derivatives, which contain randomized nucleotide sequences to normalize all tracts to the equivalent of 33 repeats. For example, the 54 randomized bases in $(CAG)_{15+18}$ are 5′ CGCGGGCCGCGCAAGG-ACCGCCAAAGAGCAACGACGCAACGCAAGCGGCCA-GAG 3′. In $(CAG)_{20+13}$ and $(CAG)_{25+8}$, the randomized region corresponds to the 5′ 39 and 24 bases shown above. The plasmids used to examine the influence of flanking DNA on

CTG expansions were created by insertion of the oligonucleotide $(CGG)_8CCGCGG(CTG)_{15}CATG$ annealed to $(CAG)_{15}CCGCGG(CCG)_8CATG$ into the *Sph*I site of pBL94. All plasmids were electroporated into DH5α using a Bio-Rad (Hercules, CA) *E.coli* pulser. Plasmids were recovered using a QIAspin miniprep kit (Qiagen, Valencia, CA) and sequenced to confirm the accuracy of the cloned repeat. Before integration in yeast, large-scale plasmid stocks were obtained using a Qiagen Plasmid Maxi kit and re-sequenced to ensure sequence integrity.

Fluctuation analysis

Fluctuation analysis was performed as described previously (22,23,40). The rates of TNR instability were determined by the method of the median (41). Briefly, single yeast colonies harboring the TNR sequence of interest were resuspended in water and appropriate dilutions were plated onto non-selective media [yeast extract/bactopeptone/dextrose (YPD)]. After 24–48 h of growth at 30°C, seven to 10 colonies were resuspended in water and an appropriate dilution was plated on YPD for total cell counts. The remaining suspension was plated on selective complete media lacking histidine and uracil to measure the contraction rates, or media lacking histidine but containing 1 mg/ml 5-fluoroorotic acid (5-FOA) for determining expansion rates. To ensure reproducibility, at least three independently isolated clones were tested. The level of detection in these assays is ~2×10^{-8} per cell generation.

Determination of mutational spectra for contractions and expansions

To analyze colonies that arose from fluctuation analysis, PCR, usually in the presence of [α-^{32}P]dCTP, was performed with primers that flank the triplet repeat tract, as described previously (22). The products of the PCR reactions were analyzed on 6% denaturing polyacrylamide gels, and repeat tract sizes (±1–2 repeats) were determined by comparison with a M13 DNA sequence ladder as described previously (23). In some cases, restriction analysis of the PCR product was used. Briefly, radioactive PCR amplification of yeast cells containing a contracted or expanded TNR was purified via QIAquick PCR purification kit according to the manufacturer's protocol. After purification, 10 000–30 000 c.p.m. of the PCR products were subjected to restriction analysis with *Sfa*NI, to test for retention of ATG interruptions, or *Sac*II, to determine the location of the expansion (New England Biolabs). The restriction digest was visualized on a 6% denaturing polyacrylamide gel and fragment sizes were inferred by comparison with an M13 DNA sequence ladder.

RESULTS

Experimental rationale

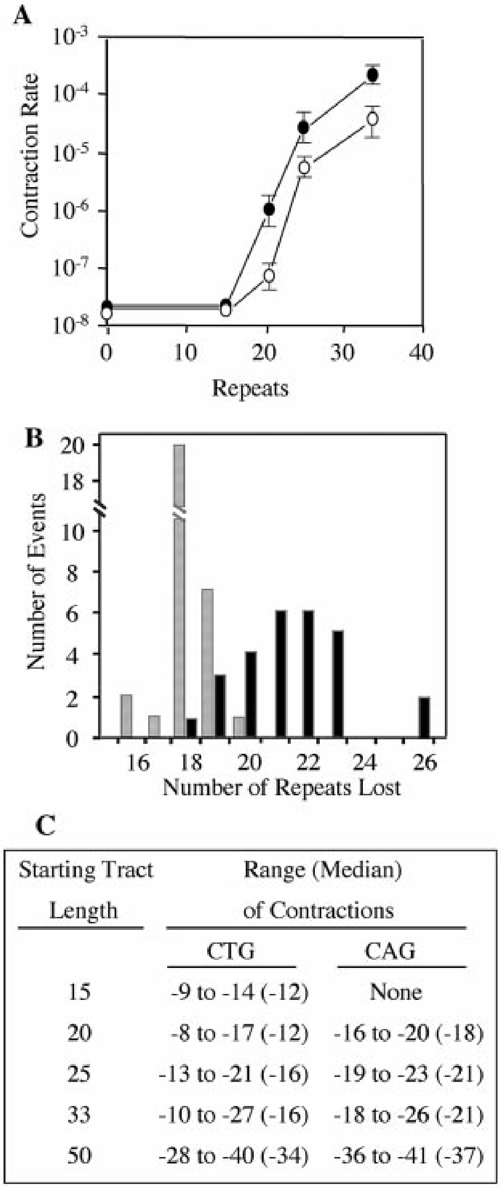
To help define how thresholds are affected by TNR sequence, purity and flanking sequences, we utilized a yeast genetic system for TNR instability that has been described previously (22,23). This system offers three significant advantages for answering important questions about thresholds. First, genetic assays for TNR expansions and contractions (Fig. 1) are suitably sensitive to reveal changes in instability. Secondly, direct assessments of thresholds can be achieved through straightforward alterations of the TNR motif (see Materials and methods). Thirdly, evidence for thresholds in yeast, as summarized below, has already been established, therefore potential changes in the thresholds can be readily detected.

CAG repeats exhibit a different contraction threshold than CTG tracts

We examined the possibility that the threshold for CAG repeat contractions might be higher than for CTG runs. Contractions were chosen for this experiment because CAG sequences are known to undergo frequent contractions in our system, provided that the repeat tract is suitably long (22,23). Therefore the sensitivity in the assay, up to 10 000-fold over background, is appropriate to look for a CAG threshold. The basis for examining CAG·CTG repeats stems from studies using bacteria (19), yeast (20–23) and human cells (24,25), which showed that CTG tracts are usually more unstable than their complementary CAG partners. (Strandedness for contractions is defined by convention as the sequence occupying the template for lagging strand DNA replication.) Furthermore, we previously showed that CTG contractions are governed by a threshold of ~17 repeats (23), so the data already exist for CTG tracts.

In the current study, we used a chimeric approach where CAG alleles from zero to 33 repeats were appended with genetically inert, scrambled sequences to normalize the overall sequence length. Contraction rates were then measured as a function of each CAG repeat length. Rate data (Fig. 2A) are consistent with a threshold for CAG contractions. Similar results are observed when the reporter is integrated at either of two loci, *ura3* on chromosome V or *LYS2* on chromosome II. Reporter tracts containing zero or 15 repeats show no detectable contractions and are therefore assigned the baseline value. A sharp upward break in both curves was observed for $(CAG)_{20}$ and $(CAG)_{25}$ repeats. The contraction rate for $(CAG)_{33}$ was even higher, although the curve appears to flatten out somewhat. The shape of the curve suggests a sharp demarcation between stable and unstable repeat lengths predicted for a threshold, and contrasts with the smooth upward trend expected if contraction rates had a stochastic dependence on tract length. The CAG contraction threshold, estimated as the midpoint between the highest and lowest rates, corresponded to 22 (at *ura3*) or 23 repeats (at *LYS2*). Therefore analysis of CAG repeat contractions at two loci yielded similarly shaped response curves and nearly identical estimates of the threshold. We noted an apparent position effect, in that contraction rates at the *ura3* locus were always 5–15-fold higher than for the *LYS2* integration site. The molecular basis for this position effect is unknown at this time.

There is an alternative explanation for the sharp increase (200–600-fold) in contraction rates between $(CAG)_{20}$ and $(CAG)_{33}$. Perhaps the two reporters measure different size contractions and therefore the rates are not directly comparable. In this scenario, the 33 repeat tract would yield frequent, large contractions, whereas the 20 repeat allele would be constricted to shorter, less frequent events. If so, there should be no overlap in the contraction spectra in the two cases. In contrast to this prediction, the results in Figure 2B show that both spectra include deletions of similar sizes (–18 to –20 repeats). Eight of the 27 contractions (30%) observed for $(CAG)_{33}$ overlapped with those seen for $(CAG)_{20}$. Therefore,



Starting Tract Length	Range (Median) of Contractions	
	CTG	CAG
15	-9 to -14 (-12)	None
20	-8 to -17 (-12)	-16 to -20 (-18)
25	-13 to -21 (-16)	-19 to -23 (-21)
33	-10 to -27 (-16)	-18 to -26 (-21)
50	-28 to -40 (-34)	-36 to -41 (-37)

Figure 2. Evidence for a CAG contraction threshold in yeast. (**A**) Contraction rates per cell generation are shown for CAG repeat alleles integrated at *ura3* (filled circles) or at *LYS2* (unfilled circles). Note that the rates are expressed on a logarithmic scale. The alleles tested had a total length equivalent to 33 repeats, with the non-repeating portion of the tract filled with randomized sequence to generate a total tract length of 99 base pairs. Error bars indicate ±1 standard deviation. The rates shown for 0 and 15 repeats are upper limits ($<2 \times 10^{-8}$) for both integration sites. (**B**) Distribution of contraction sizes arising from the $(CAG)_{20}$ and $(CAG)_{33}$ starting tracts (hatched and filled bars, respectively). On the *x*-axis is the number of repeats lost with respect to the starting tract length. A change of 18 repeats means a final allele size of 15 repeats for contractions of the 33-repeat starting tract, or a final allele size of two repeats for the 20-CTG starting tract. The *y*-axis shows the number of tract alteration events of that size. (**C**) Contraction sizes from various lengths of CTG and CAG starting tracts, in repeat units. No contraction events were observed for $(CAG)_{15}$.

even if the other 70% of the larger contractions events were discounted, there would only be an ~3-fold reduction in the contraction rate as opposed to the 200–600-fold differences seen.

Another difference between CTG and CAG repeat deletions became clear when the contraction size range was compared from various starting tract lengths (Fig. 2C). The lower end of the range and the median were always smaller for CTG than for CAG, whereas deletions at the upper end of the range tended to be similar. Since contraction size is thought to stem directly from the size of the hairpin intermediate, this finding indicates that CTG tracts can form shorter hairpins more readily than CAG repeats, consistent with the idea that CTG hairpins in our system are more energetically stable than CAG hairpins. For longer hairpins (and thus longer contractions), there is no noticeable difference between CTG and CAG.

Analysis of the mechanism of stabilization for interrupted TNR contractions

Do interruptions stabilize TNR repeats by altering the threshold? Both in humans (26,29,32) and in yeast (34,42), TNR tracts containing one to three interruptions (base pair substitutions) are considerably more stable than uninterrupted, perfect repeats. Clearly the presence of the interruption stabilizes the tract, but different mechanisms are possible. One idea is that interruptions might reduce the likelihood of hairpin formation by effectively breaking up a long TNR into shorter, more stable alleles (Fig. 3A, scheme 1). By this model, interruptions alter instability by creating TNR subdomains that might be at or below the threshold. In effect, this model describes a change in threshold due to the interruption. *In vitro*, interruptions have been shown to decrease the formation of slipped-strand DNA (S-DNA) structures and also to limit the number of different S-DNA isomers (30). These findings are consistent with the idea that interruptions somehow physically demarcate the positions where perfect repeats start and stop. Alternatively, as shown in scheme 2, interruptions might be incorporated into TNR hairpins, thereby causing the hairpin stem to be weakened by the base–base mismatches arising from the interruptions (13,29,30). We sought to distinguish these ideas using data from two previous observations: the threshold for CTG contractions is ~17 repeats (23) and contraction rates are reduced ~10-fold when a 25 repeat tract is interrupted to create subdomains of 17 and six repeats (34). Although the two existing studies showed similar contraction rates for $(CTG)_{17}$ and the interrupted tract (7.0×10^{-6} and 3×10^{-6}, respectively), this information was consistent with either model. Additional tests were needed to distinguish them.

We tested the two possibilities in Figure 3A by examining the sizes of the contracted alleles, rather than their rate of formation. Scheme 1 (subthreshold theory) predicts that contractions from the interrupted tracts should be restricted to ⩽17 repeats, whereas scheme 2 (hairpin destabilization) allows for both short and long contractions. As standards, Figure 3B provides the size spectra for uninterrupted $(CTG)_{17+16}$ and $(CTG)_{25+8}$ repeats. As expected, there is overlap but the shorter starting tract tends toward shorter contractions (median value –12) and vice versa (median value –16 for the longer starting repeat). A key result (Fig. 3C) was that contractions from interrupted alleles tended to be large (median –18), with 21 out of 37 (57%) losing ⩾18 repeats; thus, the majority of contractions must correspond to scheme 2 of Figure 3A.

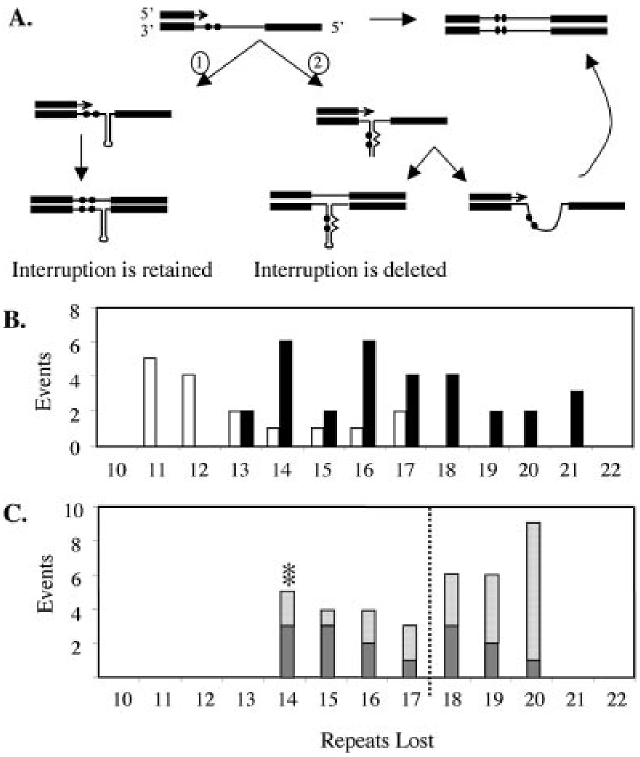


Figure 3. Model of interruption-mediated stabilization of contractions. In (**A**), the black rectangles represent the non-repeating flanking DNA, the black line denotes the triplet repeat, the tandem circles correspond to the interruptions, the arrowheads mark Okazaki fragment replication moving 5′ to 3′, and the carats symbolize mismatches within the hairpin. In scheme 1, it is assumed that the interruptions act to break the repeat tract into two smaller subdomains. Therefore secondary structures would only form in the perfect repeat region, contractions would be limited to –17 or less, and the interruptions will be retained. Scheme 2 allows for secondary structure formation to include the interruptions, thereby producing extra mismatches. If contractions occur, they can be any size up to –33 repeats and the interruptions will be lost. Alternatively, the mismatches may favor hairpin melting and allow another chance at accurate replication of the tract. (**B**) Distribution of contraction sizes for $(CTG)_{25+8}$ (filled bars) and $(CTG)_{17+16}$ (open bars). On the *x*-axis is the number of repeats lost and the *y*-axis shows the observed number of contraction events of that size. (**C**) Distribution of interrupted TNR contraction sizes in wild-type cells. The results are the compilation of contractions from two interrupted alleles, $(CTG)_{17}ATGATG(CTG)_6$ (hatched bars) and $(CTG)_6ATGATG(CTG)_{17}$ (darker bars). The asterisks indicate the two cases (out of 37 in total) where the interruption was retained.

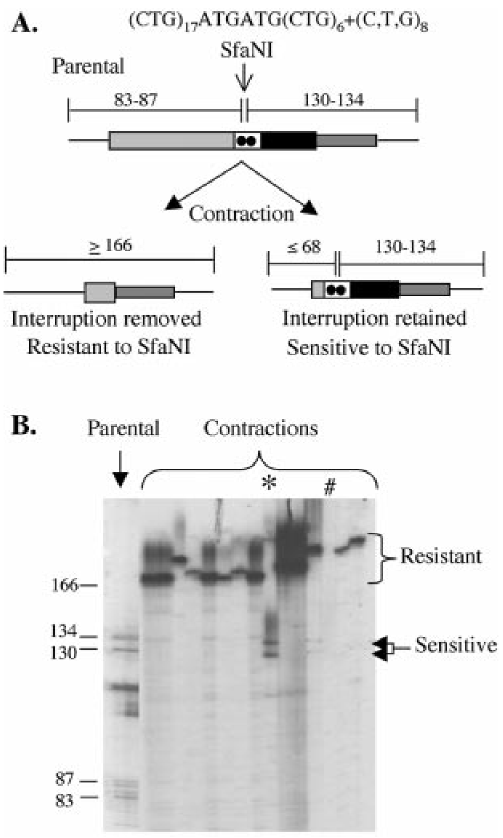


Figure 4. Classification of contractions with or without interruptions. (**A**) The schematic diagram depicts the parental 3′ interrupted repeat tract and the two possible digestion patterns. The thin lines represent non-repetitive flanking sequence, the light gray and black boxes symbolize the perfect CTG repeat sequence, the circles are the ATG interruptions, and the dark gray boxes are the scrambled (C,T,G) repeats used to normalize the overall tract to 33 repeats. The expected fragments after SfaNI digestion of these PCR products are shown (in nucleotides) above each allele. Cleavage requires the presence of at least one interrupting ATG. The cut site for SfaNI is displaced by five to nine nucleotides from the recognition sequence. The resulting four-nucleotide overhang results in two bands differing in size by four bases. If the interruption is lost, there will be no digestion and the PCR product will be ⩾166 nucleotides. Only repeat tracts that lost 17 repeats (51 nucleotides) or fewer were tested using this method, so the maximum possible reduction is 217 nucleotides (starting tract size) minus 51 nucleotides, i.e. 166 nucleotides. If the interruption is retained and the contraction took place in the $(CTG)_{17}$ repeat region, the larger band will contain 130–134 nucleotides and the smaller product will have 32–68 nucleotides, depending on the size of the contraction. (**B**) An example autoradiograph showing contractions from the interrupted allele that have been amplified and digested with SfaNI. The first two lanes are controls from the parental 3′ interrupted construct. The restriction digest products at 130–134 and 83–87 appear as predicted. Shadow bands, like the ones below the 130 base-pair arrow, are due to polymerase 'chatter' and are commonly seen during PCR of TNRs. The next 16 lanes examine contractions of the interrupted construct. The asterisk indicates the only sample to exhibit sensitivity to SfaNI. Since the 130–134 base-pair band appeared, the contraction must have taken place in the 17-perfect repeat region. In this case, the large contraction only left a few repeats and therefore the smaller digestion products have run off the bottom of the gel. In the lane marked with the hash sign, there was no PCR signal. In some samples (e.g. lanes 3–4), the upper smeared band is presumed to be a PCR artifact. The lower, more defined band was used to measure contraction sizes.

For the remaining contractions that deleted ⩽17 repeats, the two models can be distinguished by asking whether the interruptions are retained (scheme 1) or lost (scheme 2). The ATG interruptions within the CTG tract provide a recognition site for the restriction enzyme SfaNI (Fig. 4A). Retention of at least one interruption will yield an SfaNI-sensitive PCR product. Alternatively, if a contracted allele removes both interruptions then the PCR product will be resistant to the cleavage. An example of this analysis is shown in Figure 4B. We interpreted the gel as follows: if the interruption is retained, SfaNI digestion will create products of <166 nucleotides. For example, the PCR products from two parental yeast colonies show SfaNI digestion products of the predicted sizes, indicating that the technique works properly. Of the 14 contracted alleles tested in Figure 4B, only one (marked with an asterisk) showed sensitivity to SfaNI. For the other 13 samples shown, the presence of a band at ⩾166 nucleotides indicates SfaNI resistance and therefore loss of the interruption. Of 16 contraction events analyzed overall, only two showed sensitivity to this enzyme and these were both in the shortest contraction observed (Fig. 3C, asterisks). Combining results from the long (–18 or more) and short (–17 or fewer) contractions from interrupted alleles, 95% (35 out of 37) deleted the interruption, consistent with scheme 2. These results provide the first direct evidence that interruption-mediated stabilization of contractions is due to the thermodynamic weakening of secondary structure formation, and not a physical demarcation of the TNR allele. Therefore a threshold-based model does not explain the observed stabilization effects.

One possible complication of this conclusion is mismatch repair (MR). A previous report (34) showed that MR accounts for ~4-fold of the 10-fold stabilization for contractions of interrupted alleles. MR was proposed to recognize mismatches in the hairpin caused by the interruptions and to excise the hairpin, thus preventing instability. Thus, MR may eliminate some contractions and thereby skew the size analysis described above. We found, however, that among contractions from a *pms1* strain, deficient in MR, almost all (34 out of 35) lacked the interruption. Therefore MR does not influence the distribution of contractions from interrupted alleles.

Influence of CG-rich flanking DNA on the threshold for CTG expansions

Does the DNA sequence immediately flanking the TNR affect thresholds? In HD, the expanding CAG repeat is flanked by a polymorphic $(CGG)_{7–10}$ sequence (35–37,43). There is some evidence for CCG haplotype effects on CAG expansion risk (35–37,44). However, the CCG tract length is not known to change when CAG expansions occur; therefore any effect of the CCG repeats on CAG instability must be transient. To account for these findings, we hypothesized that flanking sequences might alter thresholds. In this scenario, a short CAG·CTG tract might show unexpectedly high instability in the presence of a flanking CCG·CGG repeat. One mechanism to account for this idea comes from the molecular modeling studies of Gacy *et al.* (13), which predicted the flanking CCG repeats may improve hairpin energetics and therefore lead to more expansions. Another study (35) showed that one naturally existing HD haplotype creates a sequence $(CAG)_{35}(CCG)_7$ with even better folding energies than the more common haplotypes.

To test the threshold hypothesis, we created a yeast reporter to mimic the situation at the human HD locus. The expansion threshold for CTG repeats in yeast is ~15 (23), so this repeat length was chosen as a starting point because alterations in expansion rate, either up or down, can be detected. A flanking sequence of 10 CCG/CGG repeats was appended to form the test sequence, $(CGG)_8\underline{CGGCCG}$+$(CTG)_{15}$. (This sequence is based on the complementary strand of the HD haplotype described above. The slight variation from perfect CGG sequences was intentional, as it creates a SacII site for later analysis.) Expansions of five repeats or more can be scored in our assay. We found the expansion rate of this construct was 3.4 (±1.5) × 10^{-6} events per cell generation. A control reporter, $(CTG)_{15}$+$(C,T,G)_{10}$, with a scrambled repeat flanking sequence yielded nearly as high an expansion rate: 2.6 (±1.0) × 10^{-6} events per cell generation. The similarity of these rates indicates that the flanking repeats tested do not alter the CTG threshold for expansions in yeast. Does the flanking CCG·CGG repeat influence expansion sizes? This question addresses the idea that the flanking DNA might be actively involved in hairpin formation (13). If so, then some expansions might be larger than the original size of the starting CTG repeat tract. Expansions of >15 repeats were not observed, however. For $(CGG)_8$CGGCCG+$(CTG)_{15}$ the range of expansions was +5 to +11 repeats (median +8 repeats), and for $(CTG)_{15}$+$(C,T,G)_{10}$ the range was +7 to +15 repeats (median +9 repeats).

Do the flanking CCG·CGG repeats expand in this experiment? Modeling studies (13) suggest the presence of the CCG tract within the hairpin, therefore in principle they might expand in our system. Alternatively, the expansions in our experiments might be confined to the CTG·CAG repeats, as in HD. This issue was resolved by SacII analysis of expanded alleles (Fig. 5A). SacII digestion patterns indicate whether the expansion occurred only in the CTG repeat tract, only in the CGG flanking sequence, or both (legend to Fig. 5A). The results of three independent experiments shown in Figure 5B indicate that expansions took place only in the CTG repeat region. The number of CGG flanking repeats was unchanged. We conclude that, in the situation we tested, a CGG tract flanking a CTG repeat in yeast does not influence the expansion rate or the spectrum of expanded alleles.

DISCUSSION

In a number of human TNR disease genes, the threshold can profoundly influence the risk of incurring the disease, and also of instability in subsequent generations. We sought to elucidate better the molecular nature of thresholds by examining DNA elements suggested in the literature to be potential modulators. By taking advantage of sensitive genetic assays, it was possible to examine directly the influence of each element in a yeast model system, and therefore provide new information about CAG·CTG repeat thresholds. This is the first time that repeat sequence, repeat purity and flanking sequences have been compared directly with regard to thresholds. Our findings indicate that nucleotide composition strongly influences the rate of contractions by changing the threshold. In contrast, interruptions do not stabilize contractions by creating TNRs with subthreshold lengths; instead the interruptions reduce the likelihood of hairpin formation. When a GC-rich flanking sequence similar to an HD haplotype was tested, no effects were detected for expansions of an allele near the threshold length, suggesting no role for this flanking sequence in yeast. Together these results help better define thresholds by delineating sequence elements that do or do not modulate instability of CAG·CTG repeats. The implications of these findings are discussed below.

In a previous study, we demonstrated CTG expansion and contraction thresholds in yeast (23). In the current work, this approach was extended to provide strong evidence that CAG repeats are also governed by a contraction threshold in yeast (Fig. 2A). We observed a dramatic rate increase (up to 7000-fold over background) for CAG contractions that cannot be explained by a simple length dependence; instead, the shape of the curve and the magnitude of the rate changes are consistent with a threshold. Interestingly, the apparent CAG contraction threshold (22–23 repeats) is greater than the value for its

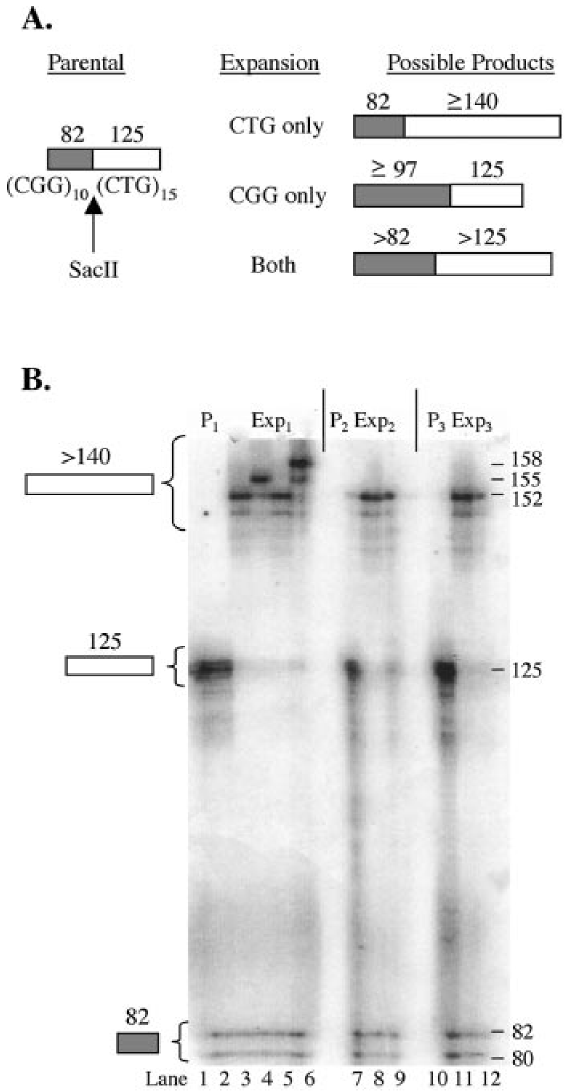


Figure 5. CGG-flanking DNA does not participate directly in CTG expansions. (**A**) The experimental design. PCR products of expanded $(CGG)_8$CGGCCG+$(CTG)_{15}$ tracts were generated and digested with the restriction enzyme SacII, which recognizes the underlined sequence. Digestion of the parental allele with SacII creates two products. The 80–82 nucleotides contain the CGG repeat flanking DNA (SacII generates a two-nucleotide overhang, therefore the cleaved samples display two bands differing in size by two bases). The second product, of 125 nucleotides, contains the $(CTG)_{15}$ repeats. Three categories of expansion events are possible. If the expansion occurs solely within the CTG repeat region, the 125-nucleotide fragment will increase to ⩾140 nucleotides. (Since the expansion must contain five or more extra repeats to appear in our assay, the minimum expansion event is 15 nucleotides.) If the expansion is constrained to the CGG tracts, the 82-nucleotide band will increase to ⩾97 nucleotides. If an expansion includes both repeating sequences, both bands will increase in length. (**B**) A representative gel with eight expansions from three genetically independent isolates. Lanes 1 and 2 are from the first parental isolate, while lanes 3–6 are from expanded alleles. Similarly, lanes 7–9 and 10–12 show parental and expanded alleles from the other two isolates. Rectangles to the left of the gel indicate the locations of key band sizes, and numbers to the right show the measured product sizes in nucleotides.

complementary sequence, CTG [~17 repeats (23)]. This conclusion is also supported by contraction rate data at specific key allele lengths. At 15 repeats CAG contractions were undetectable, whereas $(CTG)_{15}$ contracted at a rate at least 120-fold higher (23). Since 15 repeats is significantly shorter than the CAG threshold but near that for CTG, this finding confirms the sharp sequence-dependence on instability conferred by the threshold. Another prediction is that CTG and CAG contraction rates should become similar at repeat lengths above the threshold. At 33 repeats, contraction rates for the two sequences are within ~10-fold of each other, and at 50 repeats the rates are indistinguishable (23,45). Therefore, once the threshold has been achieved, the difference in nucleotide composition does not affect the rate of contraction. We note that other studies have shown higher contraction rates for CTG tracts than for CAG tracts, even at relatively long repeat lengths (19–21), suggesting that a much higher CAG contraction threshold might be operative under different conditions. In summary, our yeast assay provides direct evidence that thresholds govern CTG expansions and contractions (23), and CAG contractions (this work). The remaining possibility—CAG expansion threshold—is not known about because rates are too low for the repeat lengths (up to 25) testable in our system.

The results of Figure 2 indicate that one of the most important components of the threshold is the repeat sequence. The finding of different CAG and CTG thresholds adds new information to that gained from previous studies investigating hairpin strengths of CTG and CAG repeats. The literature is not unanimous about sequence effects on hairpin strengths. Some *in vitro* work (13,46) concluded that CTG and CAG tracts form hairpins approximately equally well, and *in vivo* repair assays under certain conditions support this finding (47). If both repeats form hairpins equally well in our system, it implies that thresholds may be dictated, at least in part, by something other than folding energetics. For example, cellular proteins may play a role in thresholds. In contrast, the bulk of the evidence from physical (15,17,48–50) and biological (19–25) studies indicates that CTG hairpins are more energetically favorable than CAG hairpins. Our results correspond more closely to this second interpretation. For example, we observed a consistent trend where CTG repeats form smaller hairpins than CAG tracts (Fig. 2C). Our results also provide a direct correlation between threshold length and hairpin energetics (i.e. the weaker hairpin corresponds to the longer threshold). This correlation implies that thresholds are due in part to hairpin folding energy, as proposed earlier (13). Other factors in addition to hairpin strength must play important roles in determining thresholds, since the CTG expansion threshold in yeast is ~15 repeats (23), compared with 30–40 repeats in humans (summarized in 1). If folding energy were solely responsible for determining the threshold, one would expect thresholds to be much closer in yeast and humans than the observed values.

The stabilizing influence of interruptions on expansions and contractions has been attributed to three possible sources: the creation of smaller subthreshold tracts (28,31–33), reducing hairpin formation (13,29,30) and/or DNA mismatch repair (34). Previous results from a CTG expansion study indicate that mismatch repair is the primary stabilizing force (34), presumably due to the ability of mismatch repair-provoked excision to remove the hairpin intermediate on the lagging daughter strand. Thus, for expansions, mismatch repair is the major stabilizing force, whereas hairpin weakening is thought to play a minor role (34). Our current results on contractions

(Figs 2 and 3) as well as previous data (34) suggest a different mechanism: contractions are inhibited primarily by reducing the likelihood of hairpin formation, with mismatch repair playing only a modest role. We showed (Figs 3 and 4) that contractions of interrupted alleles almost always (⩾95%) removed both interruptions, indicating that the interruptions are incorporated into a hairpin during the mutagenic event. The reduced contraction rates are therefore directly correlated with hairpin weakening (Fig. 3A, scheme 2), not with the creation of subthreshold domains (scheme 1). In another yeast study, Maurer *et al.* (51) used repeat tracts containing $(CTG)_{96}$ alleles interrupted in the middle by a single ATG repeat. They observed a large number of contraction events in both wild-type and mismatch repair-deficient (*msh2*) backgrounds. In the majority of these events, the interruption was lost, indicating that large hairpins can overcome the minor destabilization of the interruption. In agreement with our results, they showed (51) that the mismatch repair ability of the cell does not greatly influence the number of contraction events or their sizes.

The DNA sequence immediately flanking the HD CAG repeat, as well as other unstable TNRs, may influence expansion risk (35–37). One possibility is that thresholds are involved, particularly if secondary structure formation is a key feature for determining the threshold and if the flanking sequence facilitates hairpin formation (13,35). If true, one would expect that expansion rates would increase for a tract near the threshold when the flanking sequence is present. We examined this hypothesis in yeast for the first time using $(CTG)_{15}$ tracts with or without a flanking, HD-like CGG haplotype (Fig. 5). As in HD, there was no change in the flanking repeat, but there was also no discernible change in the expansion rate or the size of the expanded alleles compared with a control sequence. This is the first time, to our knowledge, that a flanking sequence similar to one found in an unstable human TNR gene has been evaluated in detail in a model system.

ACKNOWLEDGEMENTS

This work was supported by a National Institutes of Health award (GM61961 to R.S.L.), by a NCI Training Grant (T32 CA09476 to M.J.D.) and by National Cancer Institute Cancer Center Support Grant (P30 CA36727 to the Eppley Institute).

REFERENCES

1. Paulson,H.L. and Fischbeck,K.H. (1996) Trinucleotide repeats in neurogenetic disorders. *Annu. Rev. Neurosci.*, **19**, 79–107.
2. McMurray,C.T. (1999) DNA secondary structure: a common and causative factor for expansion in human disease. *Proc. Natl Acad. Sci. USA*, **96**, 1823–1825.
3. Cummings,C.J. and Zoghbi,H.Y. (2000) Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.*, **9**, 909–916.
4. Usdin,K. and Grabczyk,E. (2000) DNA repeat expansions and human disease. *Cell. Mol. Life Sci.*, **57**, 914–931.
5. Cleary,J.D. and Pearson,C.E. (2003) The contribution of CIS-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet. Genome Res.*, **100**, 25–55.
6. 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.
7. Gordenin,D.A., Kunkel,T.A. and Resnick,M.A. (1997) Repeat expansion—all in a flap? *Nature Genet.*, **16**, 116–118.
8. Kovtun,I., Goellner,G. and McMurray,C.T. (2001) Structural features of trinucleotide repeats associated with DNA expansion. *Biochem. Cell Biol.*, **79**, 325–336.
9. Lahue,R.S. and Slater,D.L. (2003) DNA repair and trinucleotide repeat instability. *Front. Biosci.*, **8**, S553–S565.
10. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA*, **91**, 4950–4954.
11. Chen,X., Mariappan,S.V.S., Catasti,P., Ratliff,R., Moyzis,R.K., Laayoun,A., Smith,S.S., Bradbury,E.M. and Gupta,G. (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl Acad. Sci. USA*, **92**, 5199–5023.
12. Usdin,K. and Woodford,K.J. (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.*, **23**, 4202–4209.
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. Pearson,C.E. and Sinden,R.R. (1996) Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. *Biochemistry*, **35**, 5041–5053.
15. Mitas,M. (1997) Trinucleotide repeats associated with human disease. *Nucleic Acids Res.*, **25**, 2245–2253.
16. Usdin,K. (1998) NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases. *Nucleic Acids Res.*, **26**, 4078–4085.
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. 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.
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. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. Chung,M.-Y., Ranum,L.P.W., Duvick,L.A., Servadio,A., Zoghbi,H.Y. and Orr,H.T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genet.*, **5**, 254–258.
27. Choudhry,S., Mukerji,M., Srivastava,A.K., Jain,S. and Brahmachari,S.K. (2001) CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum. Mol. Genet.*, **10**, 2437–2446.
28. Snow,K., Tester,D.J., Kruckeberg,K.E., Schaid,D.J. and Thibodeau,S.N. (1994) Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum. Mol. Genet.*, **3**, 1543–1551.
29. Pulst,S.-M., Nechiporuk,A., Nechiporuk,T., Gispert,S., Chen,X.-N., Lopes-Cendes,I., Pearlman,S., Starkman,S., Orozco-Diaz,G., Lunkes,A.

et al. (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nature Genet.*, **14**, 269–276.

30. Pearson,C.E., Eichler,E.E., Lorenzetti,D., Kramer,S.F., Zoghbi,H.Y., Nelson,D.L. and Sinden,R.R. (1998) Interruptions in the triplet repeats of SCAI and FRAXA reduce the propensity and complexity of slipeed strand DNA (S-DNA) formation. *Biochemistry*, **37**, 2701–2708.
31. Richards,R.I. and Sutherland,G.R. (1992) Dynamic mutations: a new class of mutations causing human disease. *Cell*, **70**, 709–712.
32. Eichler,E.E., Holden,J.J.A., Popovich,B.A., Reiss,A.L., Snow,K., Thibodeau,S.N., Richards,C.S., Ward,P.A. and Nelson,D.L. (1994) Length of uninterrupted CGG repeats determines instability of the *FMR1* gene. *Nature Genet.*, **8**, 88–94.
33. Hirst,M.C., Grewal,P.K. and Davies,K.E. (1994) Precursor arrays for triplet repeat expansion at the fragile X locus. *Hum. Mol. Genet.*, **3**, 1553–1560.
34. Rolfsmeier,M.L., Dixon,M.J. and Lahue,R.S. (2000) Mismatch repair blocks expansions of interrupted trinucleotide repeats in yeast. *Mol. Cell*, **6**, 1501–1507.
35. Goldberg,Y.P., McMurray,C.T., Zeisler,J., Almqvist,E., Sillence,D., Richards,F., Gacy,A.M., Buchanan,J., Telenius,H. and Hayden,M.R. (1995) Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. *Hum. Mol. Genet.*, **4**, 1911–1918.
36. Yapijakis,C., Vassilopoulos,D., Tzagournisakis,M., Maris,T., Fesdjian,C., Papageorgiou,C. and Plaitakis,A. (1995) Linkage disequilibrium between the expanded (CAG)n repeat and an allele of the adjacent (CCG)n repeat in Huntington's disease patients of Greek origin. *Eur. J. Hum. Genet.*, **3**, 228–234.
37. Hecimovic,S., Klepac,N., Vlasic,J., Vojta,A., Janko,D., Skarpa-Prpic,I., Canki-Klain,N., Markovic, d., Bozikov,J., Relja,M. *et al.* (2002) Genetic background of Huntington disease in Croatia: molecular analysis of CAG, CCG and Delta2642 (E2642del) polymorphisms. *Hum. Mutat.*, **20**, 233.
38. Kramer,B., Kramer,W., Williamson,M.S. and Fogel,S. (1989) Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional *PMS* genes. *Mol. Cell. Biol.*, **9**, 4432–4440.
39. Schiestl,R.H. and Gietz,D. (1989) High efficency transformation of intact yeast cells by single stranded nucleic acids as carrier. *Curr. Genet.*, **16**, 339–346.
40. Dixon,M.J. and Lahue,R.S. (2002) Examining the potential role of DNA polymerases η and ζ in triplet repeat instability in yeast. *DNA Rep.*, **1**, 763–770.
41. Lea,D.E. and Coulson,C.A. (1948) The distribution of the number of mutants in bacterial populations. *J. Genet.*, **49**, 264–284.
42. Rolfsmeier,M.L. and Lahue,R.S. (2000) Stabilizing effects of interruptions on trinucleotide repeat expansions in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **20**, 173–180.
43. The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, **72**, 971–983.
44. Chong,S.S., Almqvist,E., Telenius,H., LaTray,L., Nichol,K., Bourdelat-Parks,B., Goldberg,Y.P., Haddad,B.R., Richards,F., Sillence,D. *et al.* (1997) Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analysis. *Hum. Mol. Genet.*, **6**, 301–309.
45. 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.
46. 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.
47. Moore,H., Greenwell,P.W., Liu,C.-P., Arnheim,N. and Petes,T.D. (1999) Triplet repeats form secondary structures that escape DNA repair in yeast. *Proc. Natl Acad. Sci. USA*, **96**, 1504–1509.
48. Mitas,M., Yu,A., Dill,J., Kamp,T.J., Chambers,E.J. and Haworth,I.S. (1995) Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. *Nucleic Acids Res.*, **23**, 1050–1059.
49. Smith,G.K., Jie,J., Fox,G.E. and Gao,X. (1995) DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucleic Acids Res.*, **23**, 4303–4311.
50. Mariappan,S.V.S., Garcoa,A.E. and Gupta,G. (1996) Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy. *Nucleic Acids Res.*, **24**, 4775–4783.
51. Maurer,D.J., O'Callaghan,B.L. and Livingston,D.M. (1998) Mapping the polarity of changes that occur in interrupted CAG repeat tracts in yeast. *Mol. Cell. Biol.*, **18**, 4597–4604.