Cis-Elements Governing Trinucleotide Repeat Instability in *Saccharomyces cerevisiae*

Michael L. Rolfsmeier,*,[1,2] Michael J. Dixon,*,†,[1] Luis Pessoa-Brandão,*,[3] Richard Pelletier,* Juan José Miret*,[4] and Robert S. Lahue*,†

**Eppley Institute for Research in Cancer and Allied Diseases and* †*Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska 68198-6805*

Manuscript received October 20, 2000
Accepted for publication January 18, 2001

ABSTRACT

Trinucleotide repeat (TNR) instability in humans is governed by unique *cis*-elements. One element is a threshold, or minimal repeat length, conferring frequent mutations. Since thresholds have not been directly demonstrated in model systems, their molecular nature remains uncertain. Another element is sequence specificity. Unstable TNR sequences are almost always CNG, whose hairpin-forming ability is thought to promote instability by inhibiting DNA repair. To understand these *cis*-elements further, TNR expansions and contractions were monitored by yeast genetic assays. A threshold of ~15–17 repeats was observed for CTG expansions and contractions, indicating that thresholds function in organisms besides humans. Mutants lacking the flap endonuclease Rad27p showed little change in the expansion threshold, suggesting that this element is not altered by the presence or absence of flap processing. CNG or GNC sequences yielded frequent mutations, whereas A-T rich sequences were substantially more stable. This sequence analysis further supports a hairpin-mediated mechanism of TNR instability. Expansions and contractions occurred at comparable rates for CTG tract lengths between 15 and 25 repeats, indicating that expansions can comprise a significant fraction of mutations in yeast. These results indicate that several unique *cis*-elements of human TNR instability are functional in yeast.

THE genetic behavior of trinucleotide repeats (TNRs) in humans is governed by several unusual features, as judged by the inheritance pattern of families afflicted with Fragile X syndrome, Huntington's disease, or other diseases caused by mutations in endogenous TNRs (reviewed in Richards and Sutherland 1994; Ashley and Warren 1995; Gusella and MacDonald 1996; Paulson and Fischbeck 1996; Cummings and Zoghbi 2000). Although each disease has its own genetic characteristics, one common theme is that tract length is a very important indicator of the tendency toward mutation. Tracts at or above a crucial threshold length of about 35 repeats (Paulson and Fischbeck 1996) are much more prone to expansion than shorter tracts. In addition to the sharp increase in mutability at the threshold, longer tracts exhibit a more gradual tendency toward additional expansions. A second characteristic of TNR diseases is that unstable triplet repeats have been reported only for the sequences CNG (where N is any nucleotide) or, in the sole case of Friedreich's ataxia, the sequence GAA (Campuzano *et al.* 1996). Finally, for TNRs at or above the threshold level, transmissions in humans exhibit a strong proclivity toward expansions rather than contractions, with a ratio of 3- to 175-fold, depending on the disease gene and other factors (McMurray 1995).

Physical analysis of TNR-containing DNA strands showed that they readily adopt unusual secondary structures, such as hairpins, triplexes, quadruplexes, and slipped-strand conformations (summarized in Sinden 1999). Hairpins have been directly observed by NMR and melting studies or inferred by molecular modeling for the sequences CNG and GNC (Gacy *et al.* 1995; Mitas 1997). The restriction to CNG and GNC is due to energy requirements for adequate hydrogen bonding and stacking interactions. The genetic features of TNR diseases and the ability of these sequences to form alternative secondary structures led many groups to suggest that hairpin intermediates formed during DNA replication are the source of TNR instability (Richards and Sutherland 1994; Kang *et al.* 1995; Mitas *et al.* 1995; Freudenreich *et al.* 1997; Gordenin *et al.* 1997; Schweitzer and Livingston 1997; Miret *et al.* 1998). These models envision that expansions arise from hairpin formation on the Okazaki fragment, followed by escape of the hairpin from DNA repair and subsequent

Corresponding author: Robert Lahue, Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Box 986805, Omaha, NE 68198-6805. E-mail: rlahue@unmc.edu

[1] These authors contributed equally to this work.

[2] *Present address:* University of California, Division of Biological Sciences, Section of Microbiology, 1 Shields Ave., Davis, CA 95616.

[3] *Present address:* Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, 4200 E. Ninth Ave., Box B121, Denver, CO 80262.

[4] *Present address:* OSI Pharmaceuticals Inc., 106 Charles Lindbergh Blvd., Uniondale, NY 11553-3649.

fixation as a mutation in the next round of synthesis. Contractions are proposed to occur when the lagging strand template folds into a hairpin, resulting in bypass synthesis and loss of repeats. Variations on this idea suggest that DNA synthesis associated with double-strand break repair (Richard *et al.* 2000) or gene conversion (Jakupciak and Wells 1999) are mutagenic within triplet repeat regions. Again, a hairpin structure is thought to be formed during synthesis and then subsequently to escape DNA repair. A different type of model for expansions (but not for contractions), called reiterative synthesis (Sinden 1999), suggests TNR-dependent, slipped-strand structures that block DNA polymerase progression. The polymerase idles at the block, generating very long nascent strands of TNR sequence. The relative contribution of these models to TNR instability remains to be determined.

Although unusual TNR secondary structures have not been directly demonstrated *in vivo*, a number of experiments support the idea that hairpins or other aberrant structures are key intermediates in mutation. For example, some sequence specificity for instability has been reported both in *Escherichia coli* (Ohshima *et al.* 1996) and yeast (Miret *et al.* 1998). There are numerous reports of orientation effects, in which greater instability is observed for one TNR than its complement (*e.g.*, CTG *vs.* CAG; Kang *et al.* 1995; Freudenreich *et al.* 1997; Schweitzer and Livingston 1997; Miret *et al.* 1998; Balakumaran *et al.* 2000), an effect generally attributed to differences in hairpin strengths. TNRs that can fold into strong hairpins are resistant to DNA repair in yeast (Moore *et al.* 1999; Spiro *et al.* 1999), in *E. coli* (Sarkar *et al.* 1998), and *in vitro* (Spiro *et al.* 1999; Henricksen *et al.* 2000). Replicational pausing occurs at CGG/CCG and CTG/CAG sequences in *E. coli* (Samadashwily *et al.* 1997). These findings support the hypothesis that TNR instability results from replication errors that are triggered by unusual secondary structures (McMurray 1999).

To address more directly the idea that thresholds and sequence are key mediators of TNR instability, this study utilizes a series of TNR constructs in yeast. We took advantage of yeast genetic assays for TNR expansions and contractions that are selective, sensitive, and quantitative to determine the modulation of TNR instability by important *cis*-elements. We find that thresholds exist in yeast and that TNR instability can usually be predicted on the basis of length of the repeat and its sequence. This study extends the current knowledge of how *cis*-acting elements influence TNR instability in eukaryotes.

MATERIALS AND METHODS

Strains: The *E. coli* strain DH5α [*endA1 hsdR17* (rk^- mk^+) *supE44 thi-1 recA1 gyrA* (*nal*^{r}) *relA1* Δ (*lacI ZYA-argF*) *U169 deoR*] was used for plasmid construction of the TNR and large-scale plasmid preparations. The *Saccharomyces cerevisiae* strains used were MW 3317-21A (*MAT*α Δ*trp1 ura3-52 ade2*Δ *ade8 hom3-10 his3-KpnI met4 met13*; Kramer *et al.* 1989) and CH1585 (*MAT***a** *leu2*Δ*1 trp1*Δ*63 ura3-52 his3-200*; obtained from S. Mirkin, University of Illinois at Chicago). The *rad27*Δ mutant strain in the MW 3317-21A background was constructed as described previously (Spiro *et al.* 1999). TNR-containing plasmids were directed to integrate at *LYS2* or *URA3* by *Bsu*36I or *Stu*I digestion, respectively, followed by transformation via the lithium acetate protocol (Schiestl and Gietz 1989). Single integration of the TNR sequence was confirmed by Southern hybridization.

Plasmids: All plasmids were constructed using the pBL94 vector, described previously as pURA with a 1.1-kb *LYS2* fragment in the "forward" direction (Miret *et al.* 1998). pBL94 contains the *URA3* gene driven by the *Schizosaccharomyces pombe adh1* promoter, with a unique *Sph*I site separating the two elements. TNR plasmids containing 25 (Miret *et al.* 1998), 50 (Miret *et al.* 1997), or 33 repeats (this work) were constructed by insertion of oligonucleotide duplexes into the *Sph*I site of pBL94. The 25-repeat derivatives contained, on the sense strand from 5′ to 3′, the following: $(CAG)_{25}$, $(CTG)_{25}$, $CAG(TAG)_{24}$, $(CTA)_{24}CTG$, $(CCG)_{25}$, $(CGG)_{25}$, $C(TTC)_{25}$, $C(GTC)_{25}G$, $C(GAC)_{25}G$, and $C(GAA)_{25}G$. The 50-repeat derivatives contained the following: $(CTG)_{50}$, $(CAG)_{50}$, $(CCG)_{50}$, $(CGG)_{50}$, $CAG(TAG)_{49}$, $CAC(GAC)_{48}GAG$, and $CTC(GTC)_{48}GTG$. Randomized sequence controls containing the equivalent of 25 repeats (Miret *et al.* 1998) or 50 repeats (Miret *et al.* 1997) have been described previously. The sequence of the randomized control equivalent to 33 repeats was CCGGTGTGGTGCCC GCTTGCGGTTCCGTCGGCGGTTTTCCGTTTTCGCT GGCCGCTTGCGTTGCGTCGTTGCTCTTTGGCGGTCCTTG CGCGGCCCGCG. All oligonucleotides contained a CATG 3′ extension to provide a compatible end for the *Sph*I site.

The threshold repeat derivatives contained a scrambled (randomized) sequence in addition to the perfect repeat tract. The scrambled sequence on the sense strand, read 5′ to 3′, contains (CGGCGACGCAACGAACCGAAACGCGGGCCGCGCAAG GACCGCCAA) or subsets thereof. For example the 15-repeat threshold plasmid contained the repeats plus the 3′-most 30 base pairs of the scrambled sequence (starting CCGAAACGC, underlined above). Likewise the 17-repeat threshold plasmid contained the repeats in addition to the 3′-most 24 bp from the scrambled sequence (starting CGCGGGCCG). The final 24 randomized base pairs (beginning CGCGGGCCG) were used to generate a "25+8" repeat plasmid for certain contraction assays. Plasmids were transformed into DH5α using the Hanahan procedure (Hanahan 1983) or by electroporation at 2.5 mV using a Bio-Rad (Hercules, CA) *E. coli* pulser. Plasmids were recovered using a QIAspin miniprep kit (QIAGEN, Valencia, CA) following the manufacturer's protocol. Plasmids were sequenced to confirm the accuracy of the cloned sequence, prior to integration into yeast.

Fluctuation analysis: Fluctuation analysis was performed as previously described (Miret *et al.* 1998). The rates of TNR instability were determined by the method of the median (Lea and Coulson 1948). Briefly, single yeast colonies harboring the TNR sequences were resuspended in water and appropriate dilutions were plated onto nonselective media (YPD). After 24–40 hr of growth at 30°, 7–10 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 but containing 1 mg/ml 5-fluoroorotic acid (5-FOA) for determining expansion rates or selective complete media lacking histidine and uracil to measure the contraction rates. To ensure reproducibility at least three independently isolated clones were tested.

PCR of independent expansion or contraction events: As a safeguard to minimize microheterogeneity of the starting tract

size, a portion of each colony was examined by PCR to ensure that the starting tract contained 25 or 50 repeats prior to fluctuation analysis (Miret *et al.* 1998). To analyze colonies that arose from fluctuation analysis, template DNA was released from single colonies by heating in a solution containing dithiothreitol and Triton X-100. PCR, usually in the presence of 0.25 μCi of [α-^{32}P]dCTP, was performed with primers that flank the triplet repeat tract. The products of the PCR reactions were analyzed on 6% denaturing polyacrylamide gels and product sizes (±1–2 repeats) were determined by comparison of the reaction products with a M13 DNA sequence ladder. For strains bearing CGG/CCG repeats, tract sizes were measured by Southern blotting. Colonies were grown in liquid (YPD) cultures for 2 days, the genomic DNA was subsequently extracted and digested with *Sph*I to release the TNR plus a four-nucleotide overhang on each strand. The DNA fragments were resolved on 6% denaturing PAGE gels, electroblotted to a nylon membrane, and hybridized to a radiolabeled probe derived from a $(CGG/CCG)_{25}$ control fragment. CGG and CCG repeat tract sizes were deduced by comparison with appropriate molecular weight markers.

The fraction of *bona fide* expansions and contractions, judged by PCR analysis or Southern blotting, was used as a correction factor for rate measurements. At least 25 genetically independent colonies were tested for each strain. For example, if 24 of 30 5-FOA resistant colonies showed expansion by PCR analysis, then the rate of 5-FOA resistance was multiplied by 80% to generate the rate of expansion. For certain constructs, the estimated mutation rate was below the detectable limit of $\sim 3 \times 10^{-8}$ per cell generation. These low rates arose from one of two circumstances. Either there were no colonies on the selective plates (5-FOA-containing media for expansions or media lacking uracil for contractions); or colonies arose on selective media but PCR analysis showed that none of the colonies had altered TNR tracts. To estimate the upper limit of the mutation rate under these circumstances, it was assumed that the next cell plated would have been a *bona fide* expansion or contraction.

RESULTS

Rationale: If mechanisms of TNR instability show conservation among species, then some or all of the unusual genetic features of TNR behavior in humans may be manifested in other organisms. This investigation examined if important features of human trinucleotide repeat diseases are recapitulated in *S. cerevisiae*. Specifically, the hypothesis under scrutiny was that TNR instability in yeast is strongly dependent on two factors. One factor is the length of the repeat tract *vs.* its threshold. By definition, TNRs longer than the threshold are genetically unstable whereas tracts shorter than the threshold are stable. The TNR sequence is the second strong determinant of instability. Based on observations from human genetics and from structural considerations, it was predicted that the G-C rich sequences C*N*G (where *N* is any nucleotide), GAC, and GTC would mutate most frequently, provided that the tract length met or exceeded a threshold level. These characteristics were investigated using selective and sensitive genetic assays capable of identifying expansions and contractions.

Genetic assays: Rates of TNR expansions or contractions were assessed using quantitative genetic assays in *S. cerevisiae* (Miret *et al.* 1997, 1998). These assays allow the identification of expanded or contracted TNR alleles based on phenotypic changes. For expansions (Figure 1A), a starting tract of 25 repeats allows expression of the *URA3* reporter gene with concomitant sensitivity of the cells to the cytotoxic drug 5-fluoroorotic acid. Expansions of the tract to lengths of ≥30 repeats inactivate the *URA3* reporter and the cells are accordingly resistant to 5-FOA. This assay was designed specifically to reveal expansions of ≥5 repeats because this size class is among the most frequent in the polyglutamine class of human TNR diseases (Ashley and Warren 1995; Gusella and MacDonald 1996; Paulson and Fischbeck 1996). For contractions (Figure 1B), a starting tract of 50 repeats was typically used. This tract length prevents expression of the downstream reporter gene *URA3* and the cells are accordingly Ura$^-$ (require uracil for growth). Contractions to final lengths of ≤28 repeats allow functional expression of the *URA3* reporter, leading to a phenotypic change from Ura$^-$ to Ura$^+$ (able to grow without exogenous uracil). Control experiments with a 28-repeat construct confirmed this prediction. The rate of contraction is therefore proportional to the number of Ura$^+$ colonies. Later, we describe an alternative assay that allows identification of smaller contraction events.

These yeast assays also allow the use of single-colony PCR (Miret *et al.* 1997, 1998) to characterize individual expansion or contraction events. *Bona fide* TNR expansions and contractions are manifested as changes in the size of the PCR product, relative to the starting tract control. For expansions, increased PCR fragment size distinguishes true expansions from other mutations that confer a 5FOA^R phenotype, such as inactivating mutations within *URA3*. Thus expansions can be differentiated from other mutations. A similar logic applies to TNR contractions. All rates reported here refer to confirmed expansions and contractions.

CTG expansions exhibit a threshold: Although thresholds have been identified or inferred for a number of human TNR expansion diseases, this feature has not been demonstrated directly in model organisms. Our genetic assay is suitably sensitive to detect a threshold effect. From previous work (Miret *et al.* 1998), it was known that $(CTG)_{25}$ tracts are unstable in our system, expanding at rates of 1×10^{-5} per cell generation, compared to a value of $<5 \times 10^{-8}$ for a scrambled control sequence of equal length. This range of several hundredfold formed the basis for an experiment in which the length of the repeating tract was varied from 10 to 25 repeats. Scrambled sequences, containing no repeating nature, were added to equalize the total tract length to 75 bp in all cases. For example, the $(CTG)_{10}$ construct also contained 45 bp of scrambled sequences. Each construct was integrated into the yeast genome at the *LYS2* locus and the expansion rate was determined.

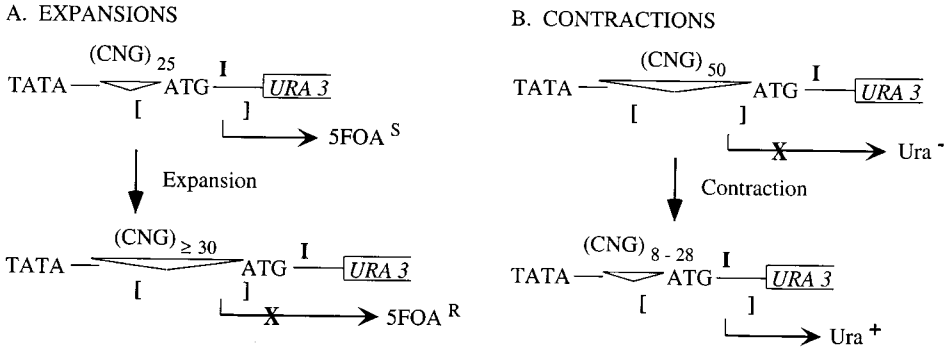

FIGURE 1.—A genetic assay to monitor TNR alterations in yeast. The regulatory region controlling expression of the reporter gene *URA3* is shown. The important features include the TATA box, the trinucleotide region (marked with an inverted triangle), an out-of-frame ATG initiator codon, the preferred transcription initiation site **I** (CCACA sequence), and the start of the *URA3* structural gene. (A) The starting construct with anticipated transcription (right-angle arrow) initiating within 55–125 bp (brackets) from TATA is shown. Initiation at **I** results in functional expression of *URA3* and sensitivity to 5-FOA. If the TNR expands to ≥30 repeats (A, bottom), 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 the TNR contains 50 repeats. Therefore in B the top diagram 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 22–42 repeats (to a final tract length of 8–28 repeats), bottom diagram, initiation will begin at **I** and *URA3* will be expressed. The cells with the contracted alleles will change to Ura^+ phenotype, that is, be capable of growth on media lacking uracil. For both, the top strand (*i.e.*, the sense strand of the *URA3* gene) is the lagging strand template.

Expansions of CTG repeats in wild-type yeast exhibit a threshold effect (Figure 2A, solid circles). The 0-repeat, randomized control construct defines the baseline. The 10-repeat construct gave the same baseline rate. Since the assay is very sensitive (the detection limit is 3×10^{-8}), even rare events would show up. Therefore expansions of 0- or 10-CTG tract are extremely infrequent. In contrast, the expansion rate rose sharply (about 100-fold) as the initial tract size was increased to 20 repeats. Thus a doubling of the CTG tract from 10 to 20 repeats resulted in two orders of magnitude increase in instability. There is a further, less dramatic rate increase when the tract is lengthened from 20 to 25 repeats. The midpoint of the transition, judged as the halfway point between the lowest and highest expansion rates, fell at ~15 repeats. We conclude that CTG expansions in wild-type yeast are governed by a threshold of about 15 repeats. If the expansion risk were a simple stochastic function of tract length, there would not be the sharp upward inflection seen in Figure 2A.

PCR analysis indicated that the expansions were limited by the size of the initial repeat tract. In 153 of the 154 expansion events arising from the 13- to 20-repeat constructs, PCR sizing showed that the expanded allele size was no more than two times the size of the initial repeat. For example, all expansions examined from the 13-repeat construct ranged from +5 to +11 repeats. These expanded allele sizes are consistent with the idea that hairpin-mediated, aberrant replication events were the source of these expansions, since models for single hairpin intermediates (GORDENIN *et al.* 1997; SPIRO *et al.* 1999) predict that final repeat length should be no more than double the original tract size. In contrast, if reiterative synthesis were responsible for the expansions, a higher fraction of the expanded alleles would be larger than double the starting tract.

To address the possibility that position effects might influence thresholds, a subset of the constructs used in Figure 2A were integrated at another locus, *URA3.* TNR expansions at *URA3* were assessed using a spot test (Figure 2B). Instability is judged by the number of $5FOA^R$ colonies that appear as the cell suspension is serially diluted by fivefold. Figure 2B (top) shows growth on media containing 5-FOA. In the spot test, a $(CTG)_{25}$ tract exhibited the highest instability, followed by a repeat containing 17 CTGs. The 17-repeat construct was included as an example near the threshold length; note in Figure 2A that the expansion rate of 15- and 17-repeat constructs is very similar. The difference in expansion rate between 25 and 17 repeats in the spot test is similar to the eightfold quantitative rate difference seen earlier (Figure 2A). In Figure 2B, the $(CTG)_{10}$ tract was as stable as the randomized control. The lower plate contains no 5-FOA, and all strains grew similarly under these control conditions. These results at *URA3* closely parallel the quantitative analysis of expansions at *LYS2*, suggesting that the threshold for CTG expansions is similar at both integration sites.

If thresholds define a minimal number of repeats necessary for hairpins to persist *in vivo*, then the flap endonuclease encoded by *RAD27* might provide a *trans*-acting function that alters the threshold. Although long TNR sequences can fold into hairpins that are strong enough to resist Rad27p processing, the enzyme more readily cleaves shorter TNRs that have limited hairpin-forming capacity (SPIRO *et al.* 1999; HENRICKSEN *et al.* 2000) and therefore prevents expansions. If Rad27p affects thresholds, shorter TNRs in *rad27*Δ mutants should exhibit a stronger tendency to expand than in wild type. To address this possibility, the experiment was repeated in a *rad27*Δ strain (open circles in Figure 2A). For repeat lengths of 0 or 10 repeats, there were

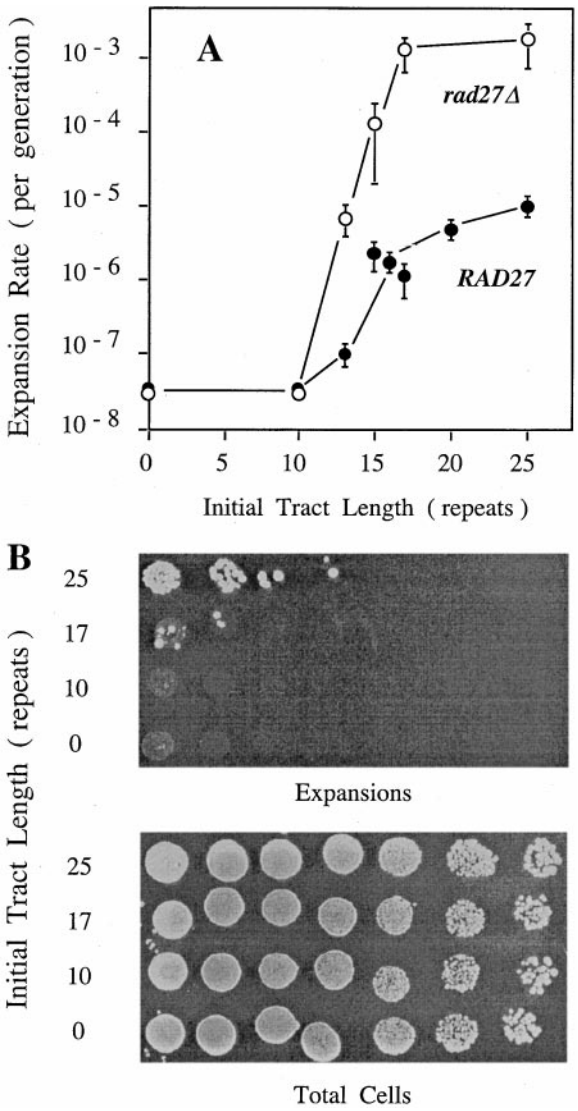

FIGURE 2.—Threshold length for CTG repeats in yeast. (A) Shown is the CTG expansion rate in wild type (solid circles) and *rad27*Δ (open circles) cells as a function of the initial repeat length. Note that the rates are expressed on a logarithmic scale. The alleles tested had a total length of 25 repeats with the nonperfect portion of the tract filled with randomized sequence to generate a total tract length of 75 bp. Error bars indicate ±1 standard deviation. The rates shown for 0 and 10 repeats are upper limits, both for *RAD27* and *rad27* strains. The quantitative values are $<3 \times 10^{-8}$. For more information regarding the threshold result, please consult http://www.unmc.edu/Eppley/publications/chart_lah.html. B shows a qualitative spot test for expansions. Wild-type cells, containing the indicated number of starting CTG repeats, were resuspended in water and serially diluted by fivefold. A small aliquot of each suspension was spotted on the plates. Expansions are scored by the top plate, which contained synthetic complete (SC) media lacking histidine and containing 5-FOA. The bottom plate contained SC media lacking histidine but without 5-FOA and serves as a control for total cell number.

no detectable expansions in *rad27*Δ cells. Intermediate repeat lengths of 13, 17, and 20 were 50- to 100-fold more unstable in the mutant strain compared to wild type, consistent with previous demonstrations of *rad27*Δ destabilization of TNR tracts (FREUDENREICH *et al.* 1998; SCHWEITZER and LIVINGSTON 1998; SPIRO *et al.* 1999). Like the wild-type cells, the *rad27*Δ mutant showed an apparent threshold, based on the sharp upward inflection of the response curve. The threshold in the *rad27*Δ background, judged as the midpoint of the transition, was about 13 repeats. Thus there seems to be little change in the apparent threshold in the *rad27*Δ mutant compared to wild type. PCR analysis showed that about 85% of the expansions in the mutant were limited to gains equal to or less than the original tract length, consistent with a replicational mechanism of instability (SPIRO *et al.* 1999). We conclude that changes in the threshold are unlikely to explain the increased CTG instability in the *rad27*Δ background.

Sequence dependence of expansion and contraction rates: There has been no extensive sequence analysis of TNR instability in eukaryotic systems, so it was informative to test different repeats side-by-side in yeast. If predictions from human genetics and from *in vitro* studies hold true, we expected that CNG and GNC sequences would be the most unstable, based on their ability to form hairpins. To address this hypothesis, expansion rates for 25-repeat tracts were measured for six variations of the G-C rich sequences and compared to four A-T rich TNRs with limited ability to fold into hairpins. Similarly, contractions from 50-repeat runs were measured for the same six G-C rich sequences plus two TNRs with limited hairpin-forming capacity. Scrambled controls provided estimates of the baseline for both expansions and contractions. It is important to note the nomenclature for these experiments. For expansions, the cited sequence resides on the lagging daughter strand, since most models for TNR expansions envision hairpin formation in the Okazaki fragment (RICHARDS and SUTHERLAND 1994; KANG *et al.* 1995; MITAS *et al.* 1995; FREUDENREICH *et al.* 1997; GORDENIN *et al.* 1997; SCHWEITZER and LIVINGSTON 1997; MIRET *et al.* 1998). The lagging daughter strand corresponds to the antisense strand of *URA3*. In contrast, contractions are thought to occur by folding of the TNR in the lagging template strand; therefore we use the convention that the template strand contains the indicated sequence. These assignments are possible because the direction of DNA replication through the integration locus is known (FREUDENREICH *et al.* 1997).

Examination of the expansion and contraction rates of the different triplet repeats revealed two different groups. These results are depicted graphically in Figure 3, to help clarify the groupings. Precise values and experimental errors are presented in Table 1. The first group, which exhibited high rates for both expansions and contractions, contained the five G-C rich triplets CTG, CGG, CCG, GTC, and GAC. Compared to the scrambled controls, expansion rates for this group ranged from 50- to 3000-fold higher and contraction rates were 400

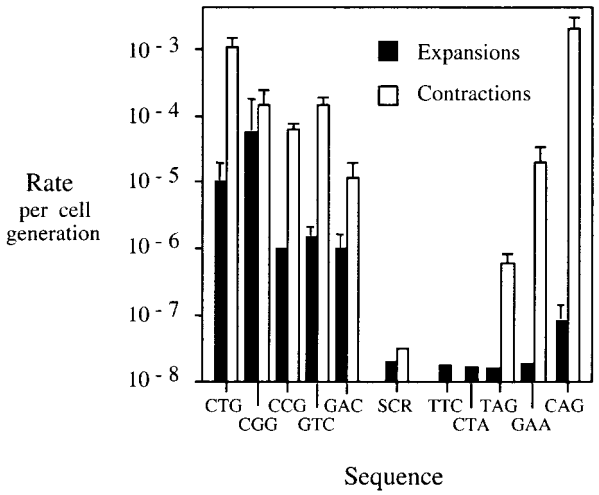

FIGURE 3.—Expansion and contraction rates of TNR repeats. For this histogram the *x*-axis indicates the different trinucleotides tested. SCR refers to the scrambled control sequences. The *y*-axis represents the expansion or contraction rates of the individual trinucleotide expressed on a logarithmic scale. Precise values of these rates are provided in Table 1. Most triplets are represented by two bars, expansion rates (solid bars) and contraction rates (open bars). Error bars represent ±1 standard deviation. Expansion rates for CTG, GAC, CAG, CTA, and scrambled sequence were taken from MIRET *et al.* (1998) and SPIRO *et al.* (1999) and are included here for comparison. The expansion rate for $(CCG)_{25}$ from Table 1 is $>1 \times 10^{-6}$. In this figure, the value is represented as 1×10^{-6} to reflect the minimum rate. It is possible that the actual rate may be somewhat higher.

to 40,000 times greater. For all sequences within this group, contractions occurred about 3–100 times more frequently than expansions, although the difference in starting repeat length (50 compared to 25) may explain some of the tendency toward contractions. Further evidence in support of a replication model for TNR expansions in group one is that there is virtually no alteration in the expansion rate in a *rad52* strain for GAC expansions (4.0×10^{-7} *vs.* 1.0×10^{-6} in wild type), consistent with previous findings for CTG (MIRET *et al.* 1998). Since *RAD52* is required for most mitotic recombination in yeast (PETES *et al.* 1991), the lack of a phenotype on GAC and CTG expansions indicates that recombination is unlikely to play a major role in the expansion process. PCR analysis of individual expansion events from wild-type cells demonstrated that the range of expansions were overlapping for GAC repeats (range 12–20, median 16), GTC repeats (range 9–25, median 13), and CTG repeats (range 10–22, median 15; MIRET *et al.* 1998). The group one repeats have final expansion sizes restricted to 2-fold or less of the initial repeat tract (63 of 63 events), consistent with the expectations of expansion from replicational errors (GORDENIN *et al.* 1997; SPIRO *et al.* 1999).

Sequences in group two (TTC, CTA, TAG, GAA, and CAG) exhibited expansion rates that were at or close to the scrambled control baseline (Figure 3). These sequences therefore show little tendency to expand in our system when the tract length is 25 repeats. When contractions were assayed, $(TAG)_{50}$ showed a modest but detectable rate, suggesting some hairpin-forming

TABLE 1

Expansion and contraction rates for trinucleotide repeat sequences

		Mutations/cell division (mean ± SD)	
Group	Sequence[a]	Expansion[b]	Contraction[b]
1	CTG[c]	$1.0 \pm 0.3 \times 10^{-5}$	$1.1 \pm 0.4 \times 10^{-3}$
	CGG	$7.8 \pm 2.1 \times 10^{-5}$	$1.7 \pm 0.7 \times 10^{-4}$
	CCG[d]	$\geq 1 \times 10^{-6}$	$6.4 \pm 0.4 \times 10^{-5}$
	GTC	$1.5 \pm 0.4 \times 10^{-6}$	$1.4 \pm 0.2 \times 10^{-4}$
	GAC	$1.0 \pm 0.3 \times 10^{-6}$	$1.2 \pm 0.7 \times 10^{-5}$
	Control scrambled	$\leq 2 \times 10^{-8}$	$\leq 3 \times 10^{-8}$
2	TTC	$\leq 1.7 \times 10^{-8}$	ND[e]
	CTA	$\leq 1.6 \times 10^{-8}$	ND
	TAG	$\leq 1.3 \times 10^{-8}$	$6.0 \pm 2.5 \times 10^{-7}$
	GAA	$\leq 1.8 \times 10^{-8}$	$2.1 \pm 1.1 \times 10^{-5}$
	CAG	$8.0 \pm 6.0 \times 10^{-8}$	$2.0 \pm 1.0 \times 10^{-3}$

[a] Sequence refers to the lagging daughter strand for expansions and to the lagging template strand for contractions.

[b] Expansions were determined for initial tract sizes of 25 repeats, whereas contractions were measured for initial tract lengths of 50 repeats.

[c] Expansion rates for CTG, GAC, CAG, CTA, and TAG were taken from MIRET *et al.* (1998) and SPIRO *et al.* (1999) and are included here for comparison.

[d] The expansion rate for $(CCG)_{25}$ could not be determined more accurately due to technical limitations.

[e] ND, not determined.

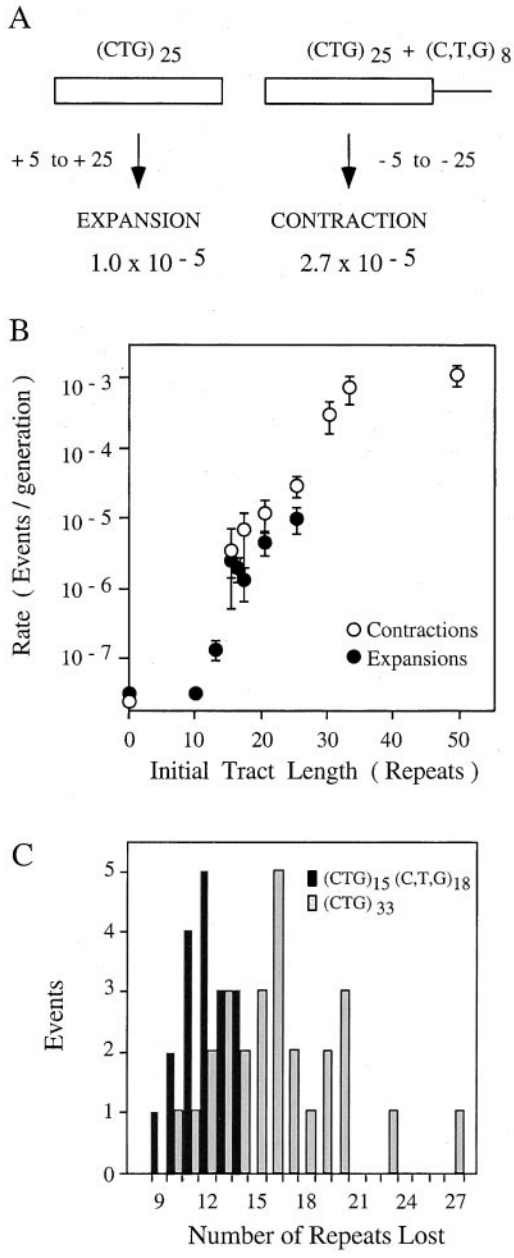

FIGURE 4.—Direct comparison of expansion to contraction rates. (A) A schematic diagram of the alleles used for determination of expansion and contraction rates for $(CTG)_{25}$. In these schematics the open boxes represent the perfect repeat tract and the line indicates randomized sequence. The contraction construct, denoted 25+8, differs from the expansion allele by the addition of 24 bp of scrambled DNA to the perfect repeat tract. For expansions, alterations that increase the tract size by +5 to +25 repeats yield $5FOA^R$ colonies. Similarly, contractions of the 25+8 construct that lose 5 to 25 repeats are scorable as Ura^+ colonies. Rates for both expansions and contractions are expressed as events per cell generation. (B) Rate comparison of expansions (solid circles) and contractions (open circles). Error bars designate ±1 standard deviation. The data for expansions are recapitulated from Figure 2A for purposes of comparison. Contraction rates for initial CTG tracts of 0 to 33 repeats were measured as described in A and in the text. The contraction rate for 50 repeats was taken from Figure 3. (C) Distribution of contraction sizes for the 15+18 CTG construct (solid bars) and for the 33-CTG repeat reporter. On the *x*-axis is the number of repeats lost, with respect to the starting tract length. A change of 10 means a final allele size of 23 repeats for contractions of the 33-repeat starting tract or a final allele size of 5 repeats for the 15-CTG starting tract. On the *y*-axis is the observed number of tract alteration events of that size.

ability and consistent with observations from meiotic recombination experiments (MOORE *et al.* 1999). However, the TAG contraction rate was still 20- to 2000-fold lower than for the sequences in group one. Contractions of TTC and CTA were not tested, as these sequences show little propensity to form secondary structures (GACY *et al.* 1995; MITAS 1997). GAA repeats are genetically unstable in Friedreich's ataxia (CAMPUZANO *et al.* 1996) and are thought to change length via triplex formation (GACY *et al.* 1998; SAKAMOTO *et al.* 1999). In our contraction assay, $(GAA)_{50}$ yielded a rate similar to the lowest values in group one, suggesting some ability of GAA to form alternate secondary structures at tract lengths of 50. CAG repeats gave unexpected results, as noted earlier (MIRET *et al.* 1997, 1998) and confirmed here. Contraction rates were as high as for group one, but expansion rates were low, similar to group two. We conclude that the hairpin-forming sequences in the first group are indeed genetically unstable. All sequences in group two are genetically stable at tract lengths of 25, consistent with most predictions of their hairpin-forming capabilities. However, several of these sequences show moderate to high capacity to contract at tract lengths of 50.

Direct comparison of expansions and contractions for CTG tracts of 15 to 25 repeats: The results from Figure 3 suggest that contraction rates are substantially higher than expansion rates in our system. However, the comparison is not direct, due in part to the difference in tract length (25 compared to 50 repeats). Also, the changes detectable for expansions (gains of ≥5 repeats) were different from that for contractions (losses of 22–42 repeats). A more direct comparison was performed by monitoring expansions and contractions of the same size tract, $(CTG)_{25}$, as diagrammed in Figure 4A. Expansions of ≥5 repeats were detected by the scheme described earlier. To measure contractions, the assay was adapted slightly by using a $(CTG)_{25}$ tract to which 24 bp of scrambled sequences, equivalent to 8 repeats, were added ("25+8" configuration). The total tract length is therefore equivalent to 33 repeats, which prevents expression of *URA3* and thus confers a Ura^- phenotype. Contractions that remove 5–25 repeats are sufficient to restore functional expression of *URA3* and thereby generate the Ura^+ phenotype, as confirmed by control experiments with 8-repeat and 28-repeat constructs. Contractions of 5–25 repeats can be detected from the 25+8 construct, while expansions of the same range are scorable for $(CTG)_{25}$. Therefore, the results

from expansion and contraction assays can be compared directly.

The contraction rate for the 25+8 reporter was 2.7×10^{-5} per cell generation (Figure 4A), about 3-fold higher than the expansion rate for $(CTG)_{25}$. This result suggests that for $(CTG)_{25}$, contractions outnumber expansions, but not by nearly as much as for longer CTG repeats in yeast. Other groups have shown that runs of 64–130 triplet repeats contract at a minimum of 20-fold more often than they expand (Maurer *et al.* 1996; Freudenreich *et al.* 1997; Schweitzer and Livingston 1997, 1999). It appears that shorter repeat tracts in yeast more closely resemble, but do not duplicate, the situation in human TNR transmissions, where expansions are the more common event.

The similarity in expansion and contraction rates for $(CTG)_{25}$ encouraged us to examine tract lengths with other repeat lengths. There were two goals to this experiment: first, to see if the contraction rates could help verify the threshold of about 15 repeats that was observed for expansions. If expansions and contractions occur through similar, hairpin-based intermediates, then perhaps the thresholds would be similar. The second goal was to see if expansions and contractions occur with similar rates for other tract lengths near the threshold, as seen for 25 repeats. The results are shown in Figure 4B. The solid circles are the expansion data, reiterated here for comparison. The open circles between 0 and 33 repeats represent contraction data using an approach analogous to that used in Figure 4A. In other words, the 33-repeat contraction reporter was $(CTG)_{33}$, or a "33+0" configuration, and the 0-repeat reporter had all scrambled sequence, or "0+33." As an example of intermediate length tracts, the 15-CTG contraction construct had a "15+18" configuration. Since the overall length of each reporter was equivalent to 33 total repeats, all rates measured contractions of 5 repeats or larger. Therefore these contraction data over this size range can be compared directly to expansion rates, since all experiments detect similar size events.

The contraction rate for 0 CTG repeats was below detection ($<3 \times 10^{-8}$; Figure 4B). In contrast, contraction rates for 15- to 33-repeat tracts increased over 100-fold in a nearly linear fashion on this semilogarithmic plot. An approximate doubling of the initial tract length (from 15 to 33 repeats) led to over two orders of magnitude change in contraction rates. Extending the contraction analysis to tracts shorter than 15 was judged to be noninformative, since short repeat tracts would artificially limit the observable contractions to very small lengths. This technical limitation of the experiment prevented us from extending the contraction analysis to lengths below 15 repeats, where additional information might have helped define a potential contraction threshold. Two other features of the results in Figure 4B are noteworthy. First, there is considerable similarity with the expansion rates between 15 and 25 repeats. The rate ratios (contractions/expansions) ranged from one to five, with an average ratio of three. Thus expansions and contractions occur with comparable rates over this range of tract lengths. Second, the expansion and contraction rates between 13 and 33 repeats form a nearly continuous line. Using the midpoint analysis described earlier, a threshold value of about 17 repeats was estimated from the combined data points. Although the contraction experiment does not permit a precise threshold measurement, the very similar threshold estimates for CTG expansions (15 repeats) and for the combined CTG expansion and contraction rates (17 repeats) are consistent with the hypothesis that expansions and contractions are governed by similar threshold lengths.

Single-colony PCR of contractions was used to determine the spectrum of actual changes in this experiment. Figure 4C shows the analysis for initial repeat tracts of 15 and 33 repeats. From the starting tract of 15 CTGs, the contracted alleles clustered from −9 to −14 repeats. Contractions of the 33-repeat reporter were generally larger (−10 to −27), but they overlapped a portion of the first spectrum. Although control experiments showed that the assay system is capable of detecting contractions ranging from −5 to −25 repeats, no contractions in the −5 to −8 range were observed for any of the tracts examined. The reason for this is not known. The data in Figure 4C help rule out an alternative interpretation of the nearly 200-fold difference in rate between the 15- and 33-repeat reporters. The alternative explanation is that the two reporters measure different size contractions and therefore are not directly comparable. By this scenario, the 33-repeat construct yields only large contractions, possibly greater than the entire 15 CTG tract in the shorter reporter. However, the results in Figure 4C show substantial overlap in the 10–14 contraction range for the two reporters. In fact, 9 of 27 contractions observed for the 33-repeat tract were losses of 10–14 repeats. Therefore only a 3-fold reduction in contraction rate can be attributed to possible deletion size differences, whereas the rate analysis shows an almost 200-fold change.

DISCUSSION

The results of this investigation reveal that several major *cis*-elements that govern human TNR instability are operative in the model system *S. cerevisiae*. First, CTG expansions in yeast show a threshold of ~15 repeats, indicating the first direct demonstration of a functional threshold in an organism besides humans. CTG contractions gave results consistent with, although not definitive for, a threshold of about 17 repeats. Our threshold experiments help to further differentiate the behavior of TNRs compared with other simple elements. For example, mutation rates in yeast increase with the length of the tract for runs of mononucleotides (Tran

et al. 1997), dinucleotides (Wierdl *et al.* 1997), and TNRs (Maurer *et al.* 1996; Freudenreich *et al.* 1998; Richard *et al.* 2000; this work). Length dependence *per se* is therefore not unique to TNRs, but thresholds are. Second, TNR instability in yeast is dependent on sequence. Most G-C rich sequences, which have the capacity to form hairpins *in vitro*, expand and contract frequently in yeast. In contrast, A-T rich sequences were substantially more stable, especially at shorter tract lengths. Our results represent the most comprehensive examination of sequence specificity in eukaryotes and thereby provide important additional support for a hairpin-mediated mechanism of TNR instability in yeast. A third similarity is that interruptions stabilize TNR tracts (Rolfsmeier and Lahue 2000). Clearly there are numerous parallels of TNR instability in yeast and in humans. In particular, the rates and sizes of TNR expansions in yeast (this study and Miret *et al.* 1998) are most similar to the expansions in human polyglutamine disease genes, such as Huntington's and spinocerebellar ataxia 1. There are also several interesting differences regarding TNR instability in humans and yeast, including the very high mutation frequency in some human disorders. Identifying the additional factors that influence the formation and removal of hairpins during DNA synthesis will clearly be important for understanding the mechanistic features of TNR instability.

This study provides a direct demonstration of a TNR threshold of about 15 repeats for CTG expansions. The most compelling evidence is that the rate of expansions increased over 100-fold when the CTG tract length was increased from 10 to 20 repeats (Figure 2). It is inherently clear that a simple dependence on tract length cannot explain the sharp upward increase in expansion rate between 10 and 20 repeats. There was no evidence for a genetic position effect to explain this observation, since similar thresholds were observed at two different integration sites. Therefore we conclude that the threshold value of 15 is *bona fide*. A parallel approach defined an apparent threshold for CTG contractions of 17 repeats. The similarity in the two threshold values is consistent with the hypothesis that similar mechanisms govern TNR expansions and contractions in our system. In human diseases, the generally accepted threshold value is about 35 repeats (Paulson and Fischbeck 1996). If thresholds were due solely to hairpin folding energies, there should be approximately equal thresholds observed in yeast and humans. Our results with CTG expansions suggest otherwise. We considered two possible explanations for this difference. The first possibility is that there is another threshold in humans at 15 repeats, which might be difficult to detect in familial transmissions. More sensitive detection methods, using single-sperm analysis of the human HD (Leeflang *et al.* 1995), AR (Zhang *et al.* 1994), and MJD-1 (Takiyama *et al.* 1997) loci, provide no evidence of a threshold below 35 repeats. Second, cellular proteins may influence thresholds. Although the ability to form secondary structure is clearly a central element of TNR instability (McMurray 1999), perhaps thresholds are influenced by *trans*-acting factors, which might participate in hairpin formation or removal *in vivo*. We conclude that differential interactions with cellular proteins provide the most reasonable explanation for the difference in CTG expansion threshold between yeast and humans.

Although 35 repeats has been estimated from *in vitro* experiments to provide maximum hairpin strength (Gacy *et al.* 1995), at least two findings suggest that hairpins form and survive *in vivo* when TNRs are shorter than 35. Moore *et al.* (1999) showed that $(CTG)_{10}$ and certain other sequences capable of secondary structure formation frequently defeat DNA repair during meiotic recombination in yeast. We found that +10 expansions are among the most frequent events starting from $(CTG)_{25}$ tracts (Miret *et al.* 1998). Both results are consistent with *in vitro* analysis of hairpins indicating that $(CAG)_{10}$ and $(CTG)_{10}$ melt at temperatures around 50° (Gacy and McMurray 1998), well above cell growth temperature. Therefore, relatively short repeats can achieve adequate hairpin strength and be manifested as expansions. Together with the observation of a CTG threshold at about 15 repeats in yeast, these findings make it appear unlikely that thresholds correspond to maximum hairpin strength.

We examined thresholds in a *rad27*Δ background because several reports (Freudenreich *et al.* 1998; Schweitzer and Livingston 1998; Spiro *et al.* 1999) showed that TNRs are destabilized in this mutant. We tested the specific hypothesis that destabilization was manifested in a threshold-dependent manner. However, the threshold in the mutant background, about 13 repeats, represented little change from the estimated 15-repeat threshold in wild type. We conclude that the CTG threshold is not substantially altered by the absence of the flap-processing enzyme and therefore the destabilizing influence of the *rad27*Δ mutation appears not to manifest itself by reducing CTG thresholds. The relatively short threshold of 15 repeats in wild-type backgrounds may have limited our ability to observe reduction of the threshold in the *rad27*Δ strain. It is possible that threshold changes in *rad27*Δ cells may be more apparent for sequences that exhibit a longer threshold in wild-type cells. Testing this possibility awaits the identification of such sequences. The large majority of CTG expansions from starting tract sizes of 13–25 repeats in the *rad27*Δ strain are likely due to aberrant replication, based on the sizes of the expanded alleles. The remaining ~15% of events in the *rad27*Δ strain may arise from recombination (Spiro *et al.* 1999) or from reiterative synthesis (Sinden 1999).

In humans, nearly all TNR expansion diseases arise from CNG sequences (Cummings and Zoghbi 2000) and these triplets are capable of hairpin formation *in vitro* (Gacy *et al.* 1995; Mitas 1997). Structural analysis

also suggested that GAC and GTC repeats form strong hairpins (Gacy *et al.* 1995; Mitas 1997) and therefore should expand readily. However, the human genome is underrepresented for long runs of GAC/GTC repeats. A BLASTN search of the human genome (http://www.ncbi.nlm.nih.gov on 6-13-2000) showed that the longest run of perfect GAC/GTC repeats was only 20 base pairs, or 6.67 repeats. If a single interrupting nucleotide is permitted, the longest repeat was only 8.67 repeats. This underrepresentation of long GAC/GTC repeats suggests a reason why expansions of these sequences have not been reported in human disease.

Genetic analysis of 10 different TNRs in our yeast assay separated the triplets into two groups with a rough correlation between genetic instability and hairpin-forming capacity. Group one triplets (CTG, CGG, CCG, GAC, and GTC), which form hairpins in solution, showed high expansion and contraction rates. The contraction rates for CTG and CAG integrated at *URA3* (this article) were virtually identical to values seen when these sequences were integrated at *ADE8* or *HIS3* loci (Miret *et al.* 1997). These similarities further support the idea that results from our assays are typical of the entire genome and cannot be ascribed to position effects. Besides the sequence specificity, additional observations for the group one repeats support aberrant replication as the source of expansions. These observations include the lack of an effect on GAC and CTG expansion rates in a *rad52* background and the finding that expanded alleles were within twofold of the starting tract sizes, consistent with predictions from replicational models of expansion (Gordenin *et al.* 1997; Spiro *et al.* 1999). Group two triplets (TTC, CTA, TAG, GAA, and CAG) showed little tendency to expand when the tract length was short, but moderate to high contraction rates from longer tract lengths. Experiments in *E. coli* are generally in agreement with our data regarding sequence specificity and the replicational source of expansions (Kang *et al.* 1996; Ohshima *et al.* 1996; Sarkar *et al.* 1998).

The TNR repeat length influences the ratio of contractions to expansions. Unlike transmission in humans, microbial systems have not mimicked the proclivity toward expansions (Kang *et al.* 1995; Freudenreich *et al.* 1997; Hirst and White 1998; Wells *et al.* 1998; Schweitzer and Livingston 1999; White *et al.* 1999; Balakumaran *et al.* 2000). When instability results from recombination mechanisms rather than replication, there are some indications of increased expansion frequencies (Jakupciak and Wells 1999; Jankowski *et al.* 2000; Richard *et al.* 2000). Most microbial experiments use long tract lengths in order to generate enough events to be seen by nonselective methods. Using a selective assay, we showed (Figure 4) that expansion and contraction rates were similar (within about threefold) for CTG tracts of 15–25 repeats. Shorter TNR tracts in yeast show a reduced ratio of contractions to expansions, compared to longer tracts. Analysis of human male germline events indicates that TNR with lengths near or below the threshold tend to yield as many or more contractions than expansions, whereas long TNRs show a strong tendency toward expansions (Zhang *et al.* 1994; Leeflang *et al.* 1995, 1999; Mornet *et al.* 1996; Kunst *et al.* 1997; Takiyama *et al.* 1997; Monckton *et al.* 1999). This interesting propensity in humans toward excess expansions at long repeat lengths is not yet understood at the molecular level. In an interesting exception, sperm analysis of 500–800 CTG repeat tracts at the SCA8 locus displayed a strong tendency toward contractions, not expansions (Moseley *et al.* 2000). Clearly there is still much to be learned about expansions and contractions.

This work was supported by National Institutes of Health award GM-61961, by funds from the Eppley Institute, and by a research grant from the Muscular Dystrophy Association (all to R.S.L.), by National Cancer Institute (NCI) Training Grant T32 CA09476 (M.L.R. and M.J.D.), by a Hereditary Disease Foundation postdoctoral fellowship (J.J.M.), and by NCI Cancer Center Support Grant P30 CA36727 (Eppley Institute).

LITERATURE CITED

Ashley, C. T., Jr., and S. T. Warren, 1995 Trinucleotide repeat expansion and human disease. Annu. Rev. Genet. **29:** 703–728.

Balakumaran, B. S., C. H. Freudenreich and V. A. Zakian, 2000 CGG/CCG repeats exhibit orientation-dependent instability and orientation-independent fragility in *Saccharomyces cerevisiae*. Hum. Mol. Genet. **9:** 93–100.

Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee *et al.*, 1996 Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science **271:** 1423–1427.

Cummings, C. J., and H. Y. Zoghbi, 2000 Fourteen and counting: unraveling trinucleotide repeat diseases. Hum. Mol. Genet. **9:** 909–916.

Freudenreich, C. H., J. B. Stavenhagen and V. A. Zakian, 1997 Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Mol. Cell. Biol. **17:** 2090–2098.

Freudenreich, C. H., S. M. Kantrow and V. A. Zakian, 1998 Expansion and length-dependent fragility of CTG repeats in yeast. Science **279:** 853–856.

Gacy, A. M., and C. T. McMurray, 1998 Influence of hairpins on template reannealing at trinucleotide repeat duplexes: a model for slipped DNA. Biochemistry **37:** 9426–9434.

Gacy, A. M., G. Goellner, N. Juranic, S. Macura and C. T. McMurray, 1995 Trinucleotide repeats that expand in human disease form hairpin structure in vitro. Cell **81:** 533–540.

Gacy, A. M., G. M. Goellner, C. Spiro, X. Chen, G. Gupta *et al.*, 1998 GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Mol. Cell **1:** 583–593.

Gordenin, D. A., T. A. Kunkel and M. A. Resnick, 1997 Repeat expansion—all in a flap? Nat. Genet. **16:** 116–118.

Gusella, J. F., and M. E. MacDonald, 1996 Trinucleotide instability: a repeating theme in human inherited disorders. Annu. Rev. Med. **47:** 201–209.

Hanahan, D., 1983 Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:** 557–580.

Henricksen, L. A., S. Tom, Y. Liu and R. A. Bambara, 2000 Inhibition of flap endonuclease 1 by flap secondary structure and relevance to repeat sequence expansion. J. Biol. Chem. **275:** 16420–16427.

Hirst, M. C., and P. J. White, 1998 Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instabil-

ity suggestive of *in vivo* lagging strand secondary structure. Nucleic Acids Res. **26:** 2353–2358.

Jakupciak, J. P., and R. D. Wells, 1999 Genetic instabilities in (CTG-CAG) repeats occur by recombination. J. Biol. Chem. **274:** 23468–23479.

Jankowski, C., F. Nasar and D. K. Nag, 2000 Meiotic instability of CAG repeat tracts occurs by double-strand break repair in yeast. Proc. Natl. Acad. Sci. USA **97:** 2134–2139.

Kang, S., A. Jaworski, K. Ohshima and R. D. Wells, 1995 Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. Nat. Genet. **10:** 213–218.

Kang, S., K. Ohshima, A. Jaworski and R. D. Wells, 1996 CTG triplet repeats from the myotonic dystrophy gene are expanded in *Escherichia coli* distal to the replication origin as a single large event. J. Mol. Biol. **258:** 543–547.

Kramer, B., W. Kramer, M. S. Williamson and S. Fogel, 1989 Heteroduplex DNA correction in *Saccharomyces cerevisiae* is mismatch specific and requires functional *PMS* genes. Mol. Cell. Biol. **9:** 4432–4440.

Kunst, C. B., E. P. Leeflang, J. C. Iber, N. Arnheim and S. T. Warren, 1997 The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. J. Med. Genet. **34:** 627–631.

Lea, D. E., and C. A. Coulson, 1948 The distribution of the number of mutants in bacterial populations. J. Genet. **49:** 264–284.

Leeflang, E. P., L. Zhang, S. Tavare, R. Hubert, J. Srinidhi *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.

Leeflang, E. P., S. Tavare, P. Marjoram, C. O. S. Neal, J. Srinidhi *et al.*, 1999 Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. Hum. Mol. Genet. **8:** 173–183.

Maurer, D. J., B. L. O'Callaghan and D. M. Livingston, 1996 Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **16:** 6617–6622.

McMurray, C. T., 1995 Mechanisms of DNA expansion. Chromosoma **104:** 2–13.

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.

Miret, J. J., L. Pessoa-Brandão and R. S. Lahue, 1997 Instability of CAG and CTG trinucleotide repeats in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **17:** 3382–3387.

Miret, J. J., L. Pessoa-Brandão and R. S. Lahue, 1998 Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **95:** 12438–12443.

Mitas, M., 1997 Trinucleotide repeats associated with human disease. Nucleic Acids Res. **25:** 2245–2253.

Mitas, M., A. Yu, J. Dill, T. J. Kamp, E. J. Chambers *et al.*, 1995 Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: (CTG)15. Nucleic Acids Res. **23:** 1050–1059.

Monckton, D. G., M. L. Cayuela, F. I. Gould, G. J. R. Brock, R. de Silva *et al.*, 1999 Very large (CAG)n DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. Hum. Mol. Genet. **8:** 2473–2478.

Moore, H., P. W. Greenwell, C.-P. Liu, N. Arnheim and T. D. Petes, 1999 Triplet repeats form secondary structures that escape DNA repair in yeast. Proc. Natl. Acad. Sci. USA **96:** 1504–1509.

Mornet, E., C. Chateau, M. C. Hirst, F. Thepot, A. Taillandier *et al.*, 1996 Analysis of germline variation at the *FMR1* CGG repeat shows variation in the normal-premutated borderline range. Hum. Mol. Genet. **5:** 821–825.

Moseley, M. L., L. J. Schut, T. D. Bird, M. D. Koob, J. W. Day *et al.*, 2000 SCA8 CTG repeat: en masse contractions in sperm and intergenerational sequence changes may play a role in reduced penetrance. Hum. Mol. Genet. **9:** 2125–2130.

Ohshima, K., S. Kang and R. D. Wells, 1996 CTG triplet repeats from human hereditary diseases are dominant gene expansion products in *Escherichia coli*. J. Biol. Chem. **271:** 1853–1856.

Paulson, H. L., and K. H. Fischbeck, 1996 Trinucleotide repeats in neurogenetic disorders. Annu. Rev. Neurosci. **19:** 79–107.

Petes, T. D., R. E. Malone and L. S. Symington, 1991 Recombination in yeast, pp. 407–521 in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, edited by J. R. Broach, J. R. Pringle and E. W. Jones. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Richard, G.-F., G. M. Goellner, C. T. McMurray and J. E. Haber, 2000 Recombination-induced CAG trinucleotide repeat expansions in yeast involve the *MRE11-RAD50-XRS2* complex. EMBO J. **19:** 2381–2390.

Richards, R. I., and G. R. Sutherland, 1994 Simple repeat DNA is not replicated simply. Nat. Genet. **6:** 114–116.

Rolfsmeier, M. L., and R. S. Lahue, 2000 Stabilizing effects of interruptions on trinucleotide repeat expansions in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **20:** 173–180.

Sakamoto, N., P. D. Chastain, P. Parniewski, K. Ohshima, M. Pandolfo *et al.*, 1999 Sticky DNA: self-association properties of long GAA-TTC repeats in R-R-Y triplex structures from Friedreich's ataxia. Mol. Cell **3:** 465–475.

Samadashwily, G. M., G. Raca and S. M. Mirkin, 1997 Trinucleotide repeats affect DNA replication *in vivo*. Nat. Genet. **17:** 298–304.

Sarkar, P. S., H.-C. Chang, F. B. Boudi and S. Reddy, 1998 CTG repeats show bimodal amplification in *E. coli*. Cell **95:** 531–540.

Schiestl, R. H., and D. Gietz, 1989 High efficency transformation of intact yeast cells by single stranded nucleic acids as carrier. Curr. Genet. **16:** 339–346.

Schweitzer, J. K., and D. M. Livingston, 1997 Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. Hum. Mol. Genet. **6:** 349–355.

Schweitzer, J. K., and D. M. Livingston, 1998 Expansions of CAG repeat tracts are frequent in a yeast mutant defective in Okazaki fragment maturation. Hum. Mol. Genet. **7:** 69–74.

Schweitzer, J. K., and D. M. Livingston, 1999 The effect of DNA replication mutations on CAG tract stability in yeast. Genetics **152:** 953–963.

Sinden, R. R., 1999 Biological implications of the DNA structures associated with disease-causing triplet repeats. Am. J. Hum. Genet. **64:** 346–353.

Spiro, C., R. Pelletier, M. L. Rolfsmeier, M. J. Dixon, R. S. Lahue *et al.*, 1999 Inhibition of FEN-1 processing by DNA secondary structure at trinucleotide repeats. Mol. Cell **4:** 1079–1085.

Takiyama, Y., K. Sakoe, M. Soutome, M. Namekawa, T. Ogawa *et al.*, 1997 Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (*MJD1*): evidence for non-Mendelian transmission of the *MJD1* gene and for the effect of the intragenic CGG/GGG polymorphism on the intergenerational instability. Hum. Mol. Genet. **6:** 1063–1068.

Tran, H. T., J. D. Keen, M. Kricker, M. A. Resnick and D. A. Gordenin, 1997 Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol. Cell. Biol. **17:** 2859–2865.

Wells, R. D., P. Parniewski, A. Pluciennik, A. Bacolla, R. Gellibolian *et al.*, 1998 Small slipped register genetic instabilities in *Escherichia coli* in triplet repeat sequences associated with hereditary neurological diseases. J. Biol. Chem. **273:** 19532–19541.

White, P. J., R. H. Borts and M. C. Hirst, 1999 Stability of the human Fragile X (CGG)n triplet repeat array in *Saccharomyces cerevisiae* deficient in aspects of DNA metabolism. Mol. Cell. Biol. **19:** 5675–5684.

Wierdl, M., M. Dominska and T. D. Petes, 1997 Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics **146:** 769–779.

Zhang, L., E. P. Leeflang, J. Yu and N. Arnheim, 1994 Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. Nat. Genet. **7:** 531–535.

Communicating editor: N. Arnheim