MOLECULAR AND CELLULAR BIOLOGY, June 1997, p. 3382–3387 Vol. 17, No. 6
0270-7306/97/$04.00+0

Instability of CAG and CTG Trinucleotide Repeats in *Saccharomyces cerevisiae*

JUAN JOSÉ MIRET,† LUIS PESSOA-BRANDÃO,† AND ROBERT S. LAHUE*

Department of Biochemistry and Molecular Biology, University of Massachusetts Medical Center, Worcester, Massachusetts 01655

Received 17 January 1997/Returned for modification 11 March 1997/Accepted 25 March 1997

A quantitative genetic assay was developed to monitor alterations in tract lengths of trinucleotide repeat sequences in *Saccharomyces cerevisiae*. Insertion of $(CAG)_{50}$ or $(CTG)_{50}$ repeats into a promoter that drives expression of the reporter gene *ADE8* results in loss of expression and white colony color. Contractions within the trinucleotide sequences to repeat lengths of 8 to 38 restore functional expression of the reporter, leading to red colony color. Reporter constructs including $(CAG)_{50}$ or $(CTG)_{50}$ repeat sequences were integrated into the yeast genome, and the rate of red colony formation was measured. Both orientations yielded high rates of instability (4×10^{-4} to 18×10^{-4} per cell generation). Instability depended on repeat sequences, as a control harboring a randomized $(C,A,G)_{50}$ sequence was at least 100-fold more stable. PCR analysis of the trinucleotide repeat region indicated an excellent correlation between change in color phenotype and reduction in length of the repeat tracts. No preferential product sizes were observed. Strains containing disruptions of the mismatch repair gene *MSH2*, *MSH3*, or *PMS1* or the recombination gene *RAD52* showed little or no difference in rates of instability or distributions of products, suggesting that neither mismatch repair nor recombination plays an important role in large contractions of trinucleotide repeats in yeast.

Tandemly repeated simple sequences are widely distributed throughout the genomes of many eukaryotic organisms. Simple repeats are typically arranged in tracts with 1 to 5 nucleotides (nt) per repeating unit (33). The mutability of these sequences in several human genetic diseases has engendered special interest in the cellular factors governing repeat stability. Mismatch repair provides an important role in stabilization of mononucleotide and dinucleotide repeats. Bacteria, yeasts, and mammalian cells deficient in mismatch repair exhibit increases of up to 700-fold in the rate of dinucleotide repeat instability (1, 10, 18, 31, 32, 34). The mutational spectrum of altered dinucleotide tracts in mismatch repair mutants is dominated by expansions and contractions of one or two repeat units. In wild-type backgrounds, about 10% of the alterations are contractions of five to eight repeats. The observation of dinucleotide repeat instability in *Saccharomyces cerevisiae* mismatch repair mutants (31, 32) and in cells derived from certain human colon tumors (1, 10, 34) was one of the key clues leading to the discovery that hereditary nonpolyposis colon cancer is due to defects in mismatch repair (4, 22).

The genetic stability of trinucleotide repeats differs from that of dinucleotide repeats, as inferred from the mutational basis of several human diseases. Expansions that increase the lengths of naturally occurring trinucleotide tracts are the genetic basis for Huntington's disease, fragile X syndrome, myotonic dystrophy, and several other diseases (reviewed in references 2, 7,and 23). Among the affected kindreds, expansions are frequent events that lead to the large numbers of trinucleotide repeats associated with the disease state. Although disease-causing mutations were originally believed to arise solely during meiosis, it is now thought that relevant expansions occur during mitotic cell growth at early postzygotic stages (reference 20 and references therein). Cellular factors affecting trinucleotide repeat stability are unknown. However, it has been proposed (summarized in reference 20) that expansions occur via replicational slippage in which single-stranded repeat sequences adopt a hairpin structure (6). Thus DNA metabolic pathways that act in forming or metabolizing hairpin intermediates might play an important role in trinucleotide repeat mutability.

The study of trinucleotide repeat contractions also has important implications for genetic diseases because the ultimate length of a trinucleotide repeat tract is determined by the sum of contractions and expansions. The large repeat sizes observed for human diseases can be explained as a greater tendency to expand and/or a reduced tendency to contract. The length of the trinucleotide repeat is one factor that influences the outcome of the mutational process. Examination of single sperm from affected and unaffected individuals with different numbers of CAG repeats in the Huntington's disease gene (17) indicates that, at low numbers of repeats, both expansions and contractions are observed. At high numbers of repeats, the overall mutation frequency is increased and expansions predominate. In contrast, sperm typing for unaffected individuals with normal to high levels of CAG repeats in the androgen receptor gene showed a greater tendency towards contractions at longer repeat lengths (37). The tract lengths examined were not as long as those for the Huntington's disease study (17), which might account for the different results. As noted elsewhere (37), the observation for the androgen receptor gene is in contrast to expansions typically observed for diseased families, prompting the authors to suggest that expansions and contractions may have different mutational origins. It is important to note, however, that no consensus has been reached on this important point, and so examination of contraction events also provides valuable insights into the mutational mechanism(s). Also of interest are those rare cases of CTG contractions in the myotonic dystrophy gene that occur upon paternal

* Corresponding author. Present address: Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, 600 South 42nd St., Omaha, NE 68198-6805. Phone: (402) 559-4619. Fax: (402) 559-4651. E-mail: rlahue@mail.unmc.edu.

† Present address: Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, NE 68198-6805.

transmission from affected individuals (summarized in reference 3). The offspring in these cases were asymptomatic, indicating that contractions in the trinucleotide repeat length can eliminate the disease.

To date, the only experimentally tractable assays for trinucleotide repeat instability involve physical monitoring of tract length alterations. In *Escherichia coli* (11–13) and in *S. cerevisiae* (19), long trinucleotide repeats are unstable. Unlike affected human kindreds, large contractions are the overwhelming majority of products in these assay systems; expansions are rarely observed. Instability is strongly dependent on CTG repeats in a particular orientation relative to an origin of DNA replication. Little evidence is available for these systems regarding the roles of specific cellular factors. Mutations in the *E. coli mutSLH* mismatch repair pathway significantly stabilize the repeats (11), suggesting that mismatch repair plays an active role in the mutational process for trinucleotide repeats. Such a role would be in contrast to the stabilizing influence of mismatch repair on dinucleotide repeats. In mammalian cell lines deficient in mismatch repair, the trinucleotide repeat lengths at the fragile X and myotonic dystrophy loci were within normal ranges, suggesting that mutations in *hMLH1* or *hMSH2* do not contribute to instability of the CGG or CTG tracts in these genes (15). Cell lines deficient in another human mismatch repair gene, *hPMS2*, do exhibit instability of a CTT microsatellite sequence, but only small numbers of repeats appear to be altered in a given event (25). Thus there is no consensus on the effects of mismatch repair on trinucleotide stability.

The lack of a genetic assay for trinucleotide repeat stability in eukaryotes has hindered experimentation to identify cellular factors that influence these repeats. We describe here a quantitative in vivo assay to measure trinucleotide repeat instability in yeast. This assay reveals that trinucleotide repeat sequences undergo large contractions at very high rates. However, mutations in genes important for mismatch repair or recombination do not affect the rate of appearance of large deletions within trinucleotide tracts.

MATERIALS AND METHODS

Strains. *E. coli* GM4257 (F^- *lacI*^{q} *lacZ proAB*$^+$/*ara thi* Δ(*gpt-lac*) *mutS215*::Tn*10* [36]) was used for cloning and large-scale plasmid purifications. The *S. cerevisiae* strain MW3317-21A (*MAT*α Δ*trp1 ura3-52 ade2* Δ*ade8 hom3-10 his3-KpnI met4 met13* [14]) or its isogenic derivatives were used for fluctuation assays. Gene disruptions to yield Δ*pms1*, Δ*leu2*, *msh2*::Tn*10LUK*, or *msh3*::*TRP1* derivatives were performed by single-step or two-step protocols (26) as described elsewhere (29). Disruption of *RAD52* was accomplished with plasmid pSM20 (L. Symington, Columbia University). Integration of plasmids containing trinucleotide repeats was accomplished by linearization of the plasmid with *Blp*I (for targeting to the *ADE8* locus) or *Nde*I (for targeting to *HIS3*) followed by lithium acetate-mediated transformation (28). All constructs were confirmed for single integration events at the desired site by Southern blotting.

Plasmid constructions. Plasmid pBL24, the vector used for all trinucleotide repeat constructs, was prepared in four steps. First, the promoter region of the *Schizosaccharomyces pombe adh1* gene was PCR amplified from plasmid pADHpo1 (27). The primers were oBL77 (CCGCTCGAGCCTACAACAACT AAGAAAATGGC, corresponding to coordinates −806 to −783 of the *adh1* promoter [4a]) and oBL78 (GCGGCATGCCCACTATTTATACCATGGGA GG, complementary to coordinates −101 to −123 [27]). The underlined sequences provide sites for the enzymes *Xho*I and *Sph*I, respectively. Following amplification by Vent polymerase (25 cycles of 1 min at 94°C, 2 min at 55°C, and 2 min at 72°C), the PCR product was phenol extracted, precipitated with ethanol, and treated with *Xho*I and *Sph*I. The resulting 711-bp fragment was cloned into pSL1180 (Pharmacia), which had been cleaved with the same enzymes, to create pBL21. The second step was to insert the duplex oligonucleotide created by annealing together oBL79 (CTCTAACCACATAATGGCCAT) and oBL80 (CGATGGCCATTATGTGGTTAGAGCATG; underlined sequences designate a site for *Msc*I) into pBL21, which had been cleaved with *Sph*I and *Cla*I, to yield pBL22. In the third step, pBL22 was treated with *Xho*I and *Cla*I to release a ~730-bp fragment which was cloned into the corresponding sites of pRS303 (30) to create pBL23. Finally, a ~1-kb *Msc*I fragment of pHD5 (20a) containing the *ADE8* gene (35) from position 5 of the coding sequence to approximately 300 bp beyond the 3′ end of the structural gene was inserted into the polylinker *Msc*I site of pBL23 to yield pBL24. The DNA sequence of pBL24 was correct from coordinates −111 to +101 (relative to the *ADE8* structural gene), with the exception of the loss of a single T residue at position −23 (see next paragraph).

Variants of pBL24 that harbored trinucleotide repeats or repeat variants were prepared by inserting oligonucleotide duplexes into the *Sph*I site of the vector. These derivatives contained on the sense strand either $C(CAG)_{50}CATG$ or $C(CTG)_{50}CATG$, where the nonrepeat sequences provide compatible ends for the *Sph*I site. A random sequence control, containing 50 residues each of C, A, and G on the sense strand (CACGGCGGCCAAACCAACGGGCGAACGCCAA GACCAAGACGGAACGAGCGCAAAGGGCAAACGGACCGGGCCGACC GGCGACGCAACGAAACGAAACGCGGGCCGAGCAAGGACCGCCAAAG AGCAACGACGCAACGCAAGCGGCCAGCGCATG) was cloned analogously. Derivatives containing 25 repeats were produced by insertion of $C(CAG)_{25}CATG$ or $C(CTG)_{25}CATG$. The primary structures of all variants were confirmed by DNA sequence analysis. Insertion of any of these sequences reiterates the *Sph*I site (GCATGC) and hence provides an ATG codon that is out of frame with respect to the initiator ATG of *ADE8*. The DNA sequence of the promoter region, starting at the TATA element and ending at the initiator ATG of *ADE8* (see Fig. 1), is TATAAATGTGGGCAT (repeats) GCATGCTCTAACCACAT AATG.

Fluctuation analysis. Rates of trinucleotide repeat instability were determined by the method of the median (16). Briefly, single colonies of yeast cells harboring the various test constructs were resuspended in water and appropriate dilutions were plated onto nonselective (yeast extract-peptone-dextrose [YPD]) plates. After 24 to 36 h of growth at 30°C to yield small colonies, 10 colonies were resuspended independently in water and replated onto YPD (four to five plates per colony). Following 5 days of growth, colonies were scored as either white, red, or white/red sectored. Only red colonies were scored as trinucleotide repeat alterations. White/red sectored colonies result from contraction events during growth on the final YPD plate and hence do not appear in the numerator of the calculation. In a typical experiment, 100 to 1,000 red colonies were observed per strain tested. To ensure reproducibility, a minimum of three repetitions of the fluctuation assay were performed per strain and at least three independently isolated clones were tested.

Single-colony PCR analysis. Isolated colonies of homogeneous color arising from the fluctuation tests were picked, resuspended in 100 μl of water, heated to 100°C for 5 min, and chilled on ice for 5 min. A portion (10 μl) of this material was used as a template for PCR analysis with primers oBL91 (AAACTCGGT TTGACGCCTCCCATG; coordinates −54 to −31 of pBL24) and oBL99 (AG AGACAATATGAGCATCCTCGCC; complementary to coordinates +104 to +81 of *ADE8* [35]). The primer oBL99 contained a mixture (33:1) of unlabelled and radioactive molecules that had been 5′ end labelled with [γ-^{32}P]ATP by polynucleotide kinase. The amplified products (35 cycles of 1.5 min at 93°C, 2 min at 55°C, and 1.5 min at 72°C) were analyzed on 6% denaturing polyacrylamide gels. PCR product sizes were determined by comparison with a DNA sequencing ladder generated with M13 DNA and the "−20" primer (U. S. Biochemicals). The error in measurement of PCR product size by this method is estimated to be ±2 repeat units. In some cases, the change in PCR product size was shown to reside in the trinucleotide repeats by performing the amplification with unlabelled primers but in the presence of [α-^{32}P]dCTP, followed by purification of the fragment on a 6% nondenaturing gel, cleavage by *Sph*I, and display of the products on a sequencing gel (see Fig. 2).

RESULTS

Development of an in vivo assay to monitor trinucleotide repeat instability. Understanding the genetic features governing trinucleotide repeat instability in eukaryotic cells has been hampered by the lack of an experimentally tractable assay. Here, a sensitive and simple genetic assay, based on a colony color phenotype, was developed to monitor trinucleotide repeat instability in yeast. The basis for this assay is that the *S. pombe adh1* promoter exhibits specific spacing requirements for function in *S. cerevisiae* (5). For this promoter, the permissive distance between the TATA element and the transcription initiation site is 55 to 125 bp; shorter or longer distances result in transcription initiation at alternate sites (5). We fused the *adh1* promoter to the reporter gene *ADE8* to take advantage of a colony color assay (Fig. 1). Yeast strains harboring *ade2* mutations are red due to the accumulation of a colored intermediate in adenine biosynthesis. Double mutants, *ade2 ade8*, are white because the *ade8* mutation acts upstream of *ade2* and blocks formation of the colored pigment. Expression of wild-type *ADE8* in an *ade2 ade8* background produces red colonies. Thus, function of the *adh1* promoter in this system is easily

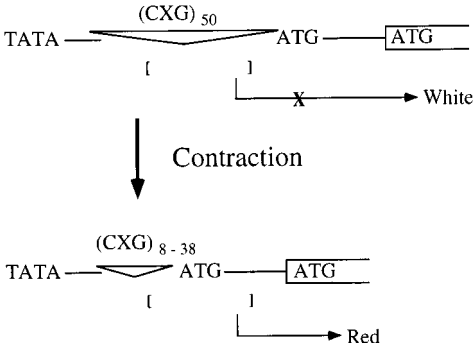

FIG. 1. A colony color assay to monitor trinucleotide repeat instability in yeast. The regulatory region that controls expression of the reporter gene *ADE8* is shown. The important features include the following: the TATA box, to provide promoter recognition; the trinucleotide repeat region (marked with an inverted triangle), where X = A or T; an out-of-frame ATG initiator codon immediately after the trinucleotide repeat element; and the beginning of the *ADE8* structural gene. The promoter also contains a CCACA transcription initiation site (not shown) between the ATG codons. The top diagram illustrates the starting construct, with anticipated transcription (right-angled arrow) initiating within 55 to 125 bp (square brackets) downstream of TATA. Translation of this transcript should initiate out of frame, resulting in white colony color in this genetic background. If a contraction event occurs to reduce the trinucleotide repeat number to 8 to 38 (bottom diagram), the window of allowed transcription now extends beyond the out-of-frame ATG and includes the transcriptional start sequence CCACA. Translation can proceed from the correct initiator codon to provide functional *ADE8* product and hence red colony color.

assayed as either off (white colonies) or on (red colonies). One extra safeguard was built into the construction to prevent spurious but functional *ADE8* expression due to transcription events initiating at upstream sites. A second ATG was placed upstream of, and out of frame with, the genuine ATG. Thus any upstream initiation events will be translationally incompetent and no Ade8 protein can be synthesized. Only transcription events that initiate between the ATGs will yield functional product.

Insertion of CAG or CTG trinucleotide repeats between the TATA box and the transcription initiation site makes expression of *ADE8* (and hence colony color) dependent on the length of the trinucleotide repeat tract (Fig. 1). The presence of too few or too many repeats puts the distance between the TATA box and the transcription initiation site outside the 55- to 125-bp window (5), resulting in failure to express *ADE8* and white colony color. In contrast, intermediate numbers of trinucleotide repeats allow functional *ADE8* expression and a red colony phenotype. Based on the amount of DNA sequence in our constructs, we estimated that tract lengths of 9 to 33 repeats should result in red colonies. Experimental findings are in satisfactory agreement with predicted values. Tests with reporter constructs containing 25 repeats yielded red colonies. Additional evaluations (see below) showed that the actual range for red colonies is 8 to 38 repeats. In contrast, test constructs harboring tracts of 0 or 50 repeats gave rise to white colonies and derivatives with 40 to 59 repeats (see below) are also white. Thus expression of *ADE8* in this system is dependent on the repeat length in a predictable manner.

We first measured contractions of the trinucleotide repeat number in starting constructs with $(CAG)_{50}$ or $(CTG)_{50}$ repeats. The tract length of 50 was chosen because it falls in or near the critical threshold range in several human diseases (reviewed in references 2, 7, and 23). In this assay, contractions to a final length of 8 to 38 repeats are scorable, whereas other contractions or expansions are phenotypically undetectable. CAG and CTG sequences were chosen as representative repeats because the complementary nature of these sequences facilitates construction of strains with repeats in both orientations. Furthermore, direct comparisons can be made with instability of CAG and CTG repeats in *E. coli* (11, 12).

TABLE 1. Contractions of trinucleotide repeats in wild-type cells

Integration locus	Tract alterations (10^{-4})/cell division (mean ± SD)			Repeat tract shortening (no. of events/total no. of events)	
	$(CAG)_{50}$	$(CTG)_{50}$	$(C,A,G)_{50}$[a]	$(CAG)_{50}$	$(CTG)_{50}$
ADE8	8 ± 3	10 ± 6	<0.05	18/18	10/10
HIS3	4 ± 1	18 ± 5	ND[b]	9/9	10/10

[a] $(C,A,G)_{50}$ refers to a random sequence of 50 C, A, and G residues with no repeat units, as described in Materials and Methods.
[b] ND, not determined.

Trinucleotide repeats are unstable in yeast. To ensure that the $(CAG)_{50}$ or $(CTG)_{50}$ repeats would be subject to DNA metabolic events typical of the entire genome, repeat plasmids were chromosomally integrated by homology-based targeting. Rates of red colony formation for these strains are shown in Table 1. Whether integrated at the *ADE8* locus or the *HIS3* locus, both $(CAG)_{50}$ and $(CTG)_{50}$ repeats exhibited rates of altered colony color of 4×10^{-4} to 18×10^{-4} per cell generation. Because the rates were largely unaffected by chromosomal location or by orientation of the repeats, this instability is likely an intrinsic property of the repeats. Furthermore, the magnitude of trinucleotide repeat alterations indicates that these events are among the most frequent that have been observed in yeast. In contrast, typical spontaneous mutation rates in yeast are approximately 10^{-8} per cell generation (21).

Several additional experiments support the validity of the assay. Experiments were performed with at least three independent isolates of each strain, and similar rates were observed in each case (as reflected by the rate averages and standard deviations [Tables 1 and 2]). Furthermore, the instability is typical of the majority of cells in the population, as judged by secondary rate analysis. In other words, white colonies from the fluctuation test were used in a second round of instability measurements. In four of four cases, the second round of rate measurements was essentially identical to the first (10×10^{-4} to 22×10^{-4} per cell generation). This control discounts the possibility that instability is due to heterogeneity in the population, with a small percentage of cells exhibiting extremely high rates. As a third control, a construct with a randomized, nonrepeating $(C,A,G)_{50}$ sequence was tested. Strains harboring this randomized sequence exhibited a rate of red colony formation of $<0.05 \times 10^{-4}$ per cell generation (Table 1), or more than 100-fold more stable than that with $(CAG)_{50}$. No

TABLE 2. Contractions of trinucleotide repeats in mismatch repair-deficient or recombination-deficient cells

Relevant genotype	Tract alterations (10^{-4})/cell division (mean ± SD)		Repeat tract shortening (no. of events/total no. of events)	
	$(CAG)_{50}$	$(CTG)_{50}$	$(CAG)_{50}$	$(CTG)_{50}$
Wild type	8 ± 3	10 ± 6	18/18	10/10
msh2	19 ± 6	10 ± 4	16/16	18/18
msh3	10 ± 3	17 ± 7	18/18	25/25
pms1	13 ± 5	10 ± 7	26/26	29/29
rad52	12 ± 1	25 ± 2	20/20	20/20

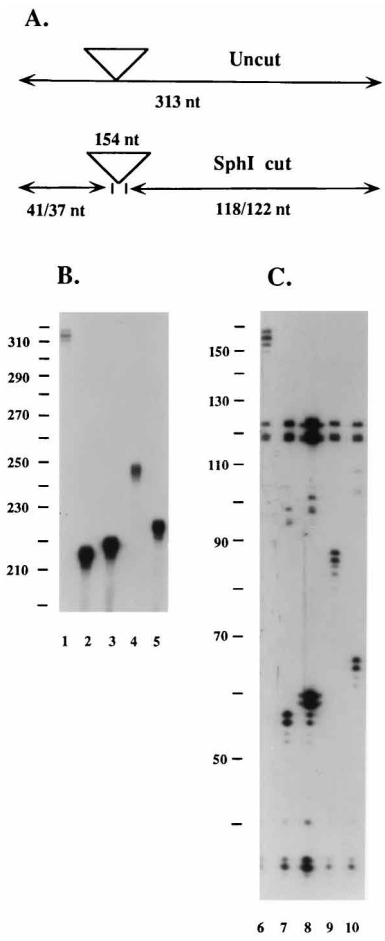

FIG. 2. PCR analysis of contraction events. PCR products were generated from individual yeast colonies in which [α-^{32}P]dCTP was included in the amplification reaction and the products were subsequently purified on a 6% nondenaturing gel, digested appropriately, and then displayed on a 6% sequencing gel. (A) Predicted sizes of PCR products from the region containing 50 repeats of CAG or CTG (inverted triangle). Undigested product is predicted to yield a 313-nt product. Digestion with *Sph*I, which has two sites (short vertical lines), generates several fragments: 41 and 37 nt from the 5′ flanking region, 154 nt for a repeat of 50 trinucleotides, and 118 and 122 nt from the 3′ flanking region. These size estimates allow for the 4-nt overhanging ends generated by *Sph*I. (B and C) Uncut and *Sph*I-cut PCR products, respectively. Size markers were derived from an M13 sequencing ladder (not shown). Lanes 1 and 6 are uncut and *Sph*I-cut products from a starting (white) colony, respectively. Lanes 2 and 7, 3 and 8, 4 and 9, and 5 and 10 show products from individual red colonies.

red colonies have yet been observed with the randomized sequences, so that the value shown represents the upper rate limit. These three controls provide further evidence that instability depends on repeat sequences.

Molecular analysis confirms repeat tract shortening. The in vivo assay is based on the assumption that changes in colony color result from molecular alterations in the number of trinucleotide repeats. To test this premise, the length of the repeat region was measured by single-colony PCR on a randomly chosen subset of white and red colonies. To assure that alterations in colony color represented genetically independent events, only one red colony from a given culture was examined. The PCR primers flank the region of the trinucleotide repeat. An example of the radiolabelled products is shown in Fig. 2. The predicted length of the PCR product containing 50 trinucleotide repeats is 313 nt (Fig. 2A). Lane 1 of Fig. 2B shows the product from a representative white colony. The size of this band corresponds to the predicted length. (Heterogeneity of PCR products from repeated sequences is commonly observed and is thought to result from polymerase stuttering through the repeated element.) Examples from red colonies are shown in lanes 2 to 5. Products from the red colonies are smaller, consistent with contractions to yield 18, 19, 28, and 21 repeats, respectively. All told, we have observed PCR product sizes in red colonies ranging from 8 to 38 repeats. In every case examined (47 of 47 [Table 1]), red colonies have yielded shortened PCR products, supporting the idea that change in colony color is due to contraction of trinucleotide repeats. Attempts to sequence the PCR products were unsuccessful, due in part to the heterogeneous nature of the DNA.

Restriction digests of the PCR products confirmed that the molecular change is in the repetitive element. The repeat sequences are flanked by *Sph*I sites (Fig. 2A). Cleavage with this enzyme followed by denaturing gel electrophoresis is predicted to yield fragments of 41 and 37 nt from the 5′ flanking sequences, 154 nt for a trinucleotide repeat of 50, and 118 and 122 nt from the 3′ flanking sequences. (All size predictions include the 4-nt overhang generated by *Sph*I.) Lane 6 shows the *Sph*I products of the PCR fragment from lane 1 (white colony). Prominent products in lane 6 are observed at 154, 122, 118, and 37 nt, as predicted. The 41-nt product is sometimes difficult to see because there is only one position for incorporation of a radiolabelled C residue. Lanes 7 to 10 show the cleavage products of the PCR fragments corresponding to lanes 2 to 5 (red colonies). In each case, the 122-, 118-, 41-, and 37-nt bands are present, indicating that the flanking sequences are unchanged. In contrast, the fragment corresponding to the trinucleotide repeat region is shortened in each case to 58, 61, 88, and 67 nt (lanes 7 to 10, respectively). The sizes of these fragments (including the 4-nt *Sph*I overhang) correspond well to the predicted repeat sizes from Fig. 2B. Bands at about 100 nt (lanes 7 and 8) are due to incomplete digestion. Thus we conclude that contractions within the trinucleotide repeat region are the mutational events that result in red colony formation.

Large contractions of trinucleotide repeat are not affected by mismatch repair or *RAD52*. The mismatch repair system plays an important role in the stability of mononucleotide and dinucleotide repeat sequences. We tested the effect of *msh2*, *msh3*, and *pms1* mismatch repair mutants on the stability of trinucleotide repeats in yeast. Deficiency in mismatch repair had little or no effect on large trinucleotide repeat contractions (Table 2). The rates observed in the mutants (10×10^{-4} to 25×10^{-4} per cell generation) were essentially unchanged from those of wild type. In all 132 cases examined, PCR analysis indicated that occurrence of red colonies correlated perfectly with shortened repeat tracts. Thus there is no evidence that mismatch repair in yeast affects the large contraction events measured in this assay. It is important to note that contractions involving small numbers of repeats could occur without detection in our assay. Thus we cannot rule out a role for mismatch repair in small contraction events. Clearly, however, large contractions occur by a mechanism that is independent of *MSH2*, *MSH3*, and *PMS1*.

Another possible mechanism of repeat instability involves unequal recombination. If this is occurring in our system, mutations that reduce or eliminate recombination should decrease the rate of instability. Trinucleotide repeat instability was tested in a strain harboring a disruption of *RAD52*, whose function is required for most mitotic recombination activity in yeast (reviewed in ref. 24). As shown in Table 2, the rates of alteration in a *rad52* background are unchanged or possibly slightly enhanced (12×10^{-4} to 25×10^{-4} per cell generation).

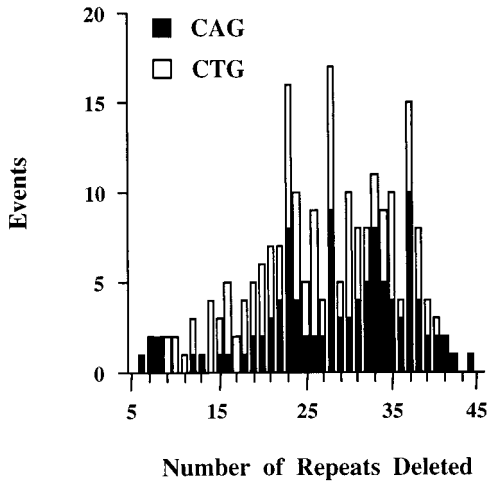

FIG. 3. Summary of PCR analysis. A summary histogram of the distribution of contraction products is shown. The change in PCR product size was determined for 64 independent genetic events from wild-type, *pms1*, *msh2*, *msh3*, and *rad52* strains. Deletion sizes from the starting (white) colony to the contracted (red) colony were estimated on denaturing gels (see Materials and Methods). Filled bars represent 33 contraction events from (CAG) repeats, and open bars correspond to 31 contractions of (CTG) repeats.

Changes in colony color were due to shortening of the PCR product in 40 of 40 examples tested (Table 2). Thus loss of *RAD52* function does not significantly alter trinucleotide repeat instability either quantitatively or qualitatively.

PCR analysis reveals additional data on contraction sizes. Examination of the PCR products revealed three additional interesting points. First, we observed some microheterogeneity among the white colonies. In other words, 23 of 219 white colonies harbored repeat lengths ranging from 40 to 59 repeats. This microheterogeneity could be traced to 5 of 23 independent isolates examined. Since all strains were derived from a common plasmid construct that had 50 repeats, this result suggests some instability of trinucleotide tracts during workup of the strains. This supports the idea that random contractions and expansions frequently occur within the trinucleotide repeat tract. Therefore, repeat instability is even greater than can be measured by our assay. Second, among contraction events scored by the switch in color phenotype, changes in size of the PCR products were observed across the possible range (Fig. 3). No evidence for a strong hotspot was readily apparent when results from all strains were combined (Fig. 3) or for individual strains (data not shown). Thus there is no apparent preference for specific contraction sizes, even though this assay would have revealed changes of approximately 20 repeats observed in *E. coli* mismatch repair mutants (11). Third, an experiment with a white variant containing 40 repeats yielded rates of red colony formation in two independent measurements of 4×10^{-4} and 6×10^{-4} per cell generation. Furthermore, large reductions in tract lengths (10 to 30 repeats) were observed in 10 of 10 independent events. Thus a tract of 40 repeats is similar to the 50-repeat constructs in rate of contraction and type of events.

DISCUSSION

This study reveals three major points about trinucleotide repeat mutability in yeast. First, a genetically tractable assay has been developed to measure trinucleotide repeat stability in a eukaryotic system. Genetic and molecular evidence established the validity of this approach for contractions. Second, tracts of 50 trinucleotide repeats frequently undergo contractions in yeast. Instability is intrinsic to the repeats, as control sequences lacking repeats were at least 100-fold more stable. The mutation rates are about 4 orders of magnitude higher than the spontaneous mutation rate (21) and about 1 order of magnitude higher than reported rates for dinucleotide repeats (9, 31, 32). It is not yet clear whether the difference in rates between the trinucleotide and dinucleotide repeat experiments is quantitatively meaningful or whether the differences are due to the assay systems. For example, the trinucleotide tracts examined here are 150 bp in length whereas published observations are available for dinucleotide tracts only up to about 50 bp (8). It is possible that longer dinucleotide tracts will exhibit higher rates of alteration. Third, large contractions within trinucleotide repeat tracts are not subject to control by the *MSH2/MSH3/PMS1* mismatch repair system, suggesting that large changes within trinucleotide repeats are regulated by different cellular factors than mononucleotide or dinucleotide runs. Contractions of trinucleotide repeats also occur readily in a *rad52* background, suggesting that the mutational mechanism is unlikely to occur via recombination. These results indicate that yeast differs from mammalian cells in that trinucleotide repeats undergo large length changes at greater frequency in yeast.

One prediction of our results is that CAG and CTG repeats in the yeast genome should be relatively short due to the tendency to contract. The yeast sequence database was inspected for perfect CAG and CTG trinucleotide tracts by a BLASTN search. The longest tract contained only 11 repeats and is located on chromosome XI. Thus the tendency towards contraction of CAG and CTG trinucleotide repeats in yeast is supported by this analysis. In contrast, the human genome contains significantly longer CAG and CTG trinucleotide tracts. For example, normal individuals exhibit polymorphisms of up to 34 CAG repeats in the Huntington's disease gene and up to 37 CTG repeats in the myotonic dystrophy gene (2, 7, 23). In diseased individuals, repeat tracts increase to as many as 121 and >2,000, respectively.

The quantitative genetic results presented here generally agree with qualitative physical assessments of trinucleotide instability in *E. coli* (12, 13) and yeast (19). Trinucleotide repeats in these assay systems frequently undergo large contractions, and instability increases with tract length. In both organisms, loss of a major recombination factor, *recA* (11) or *RAD52* (Table 2), did not detectably affect the results, suggesting that recombination is not involved in the mutational process or that alternative recombination pathways are responsible.

Two noteworthy differences exist between our work and that of others. Our assay demonstrated no orientation dependence for instability whereas other groups have reported strong orientation effects (12, 19). When orientation effects have been observed, they depend on the close proximity of an origin of DNA replication and on placement of CTG repeats on a particular strand (leading versus lagging). In the case of our system, it is possible that origins of replication may be distant from the trinucleotide repeat tract. If passage of a replication fork were similarly probable from either the right or the left, one would not expect to see orientation dependence. To the best of our knowledge, no information is available regarding active origins of replication in the vicinities of *ADE8* (chromosome IV) or *HIS3* (chromosome XV), the two loci used for integration of trinucleotide repeats in our experiments.

The role of mismatch repair in trinucleotide repeat instability provides another important point of interest. Two possibil-

ities have been put forth to account for observed mismatch repair effects on trinucleotide repeat stability. Wells and colleagues suggested that the *mutSLH* system of *E. coli* enhances instability of (CTG) repeats (11), based on the observation that (CTG) tract contractions are much less frequent in *mutS*, *mutL*, or *mutH* backgrounds. They proposed (11) that the combination of large hairpins formed by single-stranded repeat tracts (6) plus the formation of large gaps resulting from *mutSLH* action could explain enhancement of repeat tract instability. Alternatively, hairpins formed by single-stranded repeats (6) might be refractory to repair. In support of this possibility, mammalian cell lines deficient in the mismatch repair gene *hMLH1* or *hMSH2* harbor normal lengths of CGG tracts at the fragile X locus and CTG repeats at the myotonic dystrophy loci, suggesting that mutations in mismatch repair do not contribute to trinucleotide repeat instability in these genes (15). In an *hPMS2* mutant cell line, destabilization of a CTT microsatellite was observed, but alterations appeared limited to only small numbers of repeats (25). Therefore, the role of mismatch repair in trinucleotide repeat stability was unclear. Our genetic assay in yeast provides quantitative evidence that defects in mismatch repair are not involved in large contractions of trinucleotide repeats and are in agreement with the observations for mammalian cells.

ACKNOWLEDGMENTS

This work was supported by a research grant from the Muscular Dystrophy Foundation, by the Translational Grant Program of the University of Massachusetts Medical Center, and by a postdoctoral fellowship from the Hereditary Disease Foundation (to J.J.M.).

We thank Rick Baker and Martin Marinus for helpful discussions, Ben Hall and Lorraine Symington for plasmids, Elizabeth Furter-Graves for sequence information, and Tom Petes for communicating results prior to publication.

REFERENCES

1. **Aaltonen, L. A., P. Peltomaki, F. S. Leach, P. Sistonen, L. Pylkkanen, J.-P. Mecklin, H. Jarvinen, S. M. Powell, J. Jen, S. R. Hamilton, G. M. Petersen, K. W. Kinzler, B. Vogelstein, and A. de la Chapelle.** 1993. Clues to the pathogenesis of familial colorectal cancer. Science **260:**812–816.
2. **Ashley, C. T., Jr., and S. T. Warren.** 1995. Trinucleotide repeat expansion and human disease. Annu. Rev. Genet. **29:**703–728.
3. **Brook, J. D.** 1993. Retreat of the triplet repeat? Nat. Genet. **3:**279–281.
4. **Fishel, R., M. K. Lescoe, M. R. Rao, N. G. Copeland, N. A. Jenkins, J. Garber, M. Kane, and R. Kolodner.** 1993. The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer. Cell **75:**1027–1038.

4a. **Furter-Graves, E.** Personal communication.

5. **Furter-Graves, E. M., and B. D. Hall.** 1990. DNA sequence elements required for transcription initiation of the *Schizosaccharomyces pombe* ADH gene in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **223:**407–416.
6. **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.
7. **Gusella, J. F., and M. E. MacDonald.** 1996. Trinucleotide instability: a repeating theme in human inherited disorders. Annu. Rev. Med. **47:**201–209.
8. **Heale, S. M., and T. D. Petes.** 1995. The stabilization of repetitive tracts of DNA by variant repeats requires a functional DNA mismatch repair system. Cell **83:**539–545.
9. **Henderson, S. T., and T. D. Petes.** 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**2749–2757.
10. **Ionov, Y., M. A. Peinado, S. Malkhosyan, D. Shibata, and M. Perucho.** 1993. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature **363:**558–561.
11. **Jaworski, A., W. A. Rosche, R. Gellibolian, S. Kang, M. Shimizu, R. P. Bowater, R. R. Sinden, and R. D. Wells.** 1995. Mismatch repair in *Escherichia coli* enhances instability of $(CTG)_n$ triplet repeats from human hereditary diseases. Proc. Natl. Acad. Sci. USA **92:**11019–11023.
12. **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.
13. **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.
14. **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.
15. **Kramer, P. R., C. E. Pearson, and R. R. Sinden.** 1996. Stability of triplet repeats of myotonic dystrophy and fragile X loci in human mismatch repair cell lines. Hum. Genet. **98:**151–157.
16. **Lea, D. E., and C. A. Coulson.** 1948. The distribution of the number of mutants in bacterial populations. J. Genet. **49:**264–284.
17. **Leeflang, E. P., L. Zhang, S. Tavare, R. Hubert, J. Srinidhi, M. E. MacDonald, R. H. Myers, M. de Young, N. S. Wexler, J. F. Gusella, and N. Arnheim.** 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.
18. **Levinson, G., and G. A. Gutman.** 1987. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucleic Acids Res. **15:**5323–5338.
19. **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.
20. **McMurray, C. T.** 1995. Mechanisms of DNA expansion. Chromosoma **104:** 2–13.

20a. **Miret, J., and R. Lahue.** Unpublished data.

21. **Morrison, A., A. L. Johnson, L. H. Johnston, and A. Sugino.** 1993. Pathway correcting DNA replication errors in *Saccharomyces cerevisiae*. EMBO J. **12:**1467–1473.
22. **Parsons, R., G. M. Li, M. J. Longley, W. H. Fang, N. Papadopoulos, J. Jen, A. de la Chapelle, K. W. Kinzler, B. Vogelstein, and P. Modrich.** 1993. Hypermutability and mismatch repair deficiency in RER^+ tumor cells. Cell **75:**1227–1236.
23. **Paulson, H. L., and K. H. Fischbeck.** 1996. Trinucleotide repeats in neurogenetic disorders. Annu. Rev. Neurosci. **19:**79–107.
24. **Petes, T. D., R. E. Malone, and L. S. Symington.** 1991. Recombination in yeast, p. 407–521. *In* J. R. Broach, J. R. Pringle, and E. W. Jones (ed.), The molecular and cellular biology of the yeast *Saccharomyces*, vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
25. **Risinger, J. I., A. Umar, J. C. Barrett, and T. A. Kunkel.** 1995. A *hPMS2* mutant cell line is defective in strand-specific mismatch repair. J. Biol. Chem. **270:**18183–18186.
26. **Rose, M. D., F. Winston, and P. Hieter.** 1988. Methods in yeast genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. **Russell, P. R., and B. D. Hall.** 1983. The primary structure of the alcohol dehydrogenase gene from the fission yeast *Schizosaccharomyces pombe*. J. Biol. Chem. **258:**143–149.
28. **Schiestl, R. H., and D. Gietz.** 1989. High efficiency transformation of intact yeast cells by single stranded nucleic acids as carrier. Curr. Genet. **16:**339–346.
29. **Selva, E. M., L. New, G. F. Crouse, and R. S. Lahue.** 1995. Mismatch correction acts as a barrier to homeologous recombination in *Saccharomyces cerevisiae*. Genetics **139:**1175–1188.
30. **Sikorski, R. S., and P. Hieter.** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics **122:**19–27.
31. **Strand, M., M. C. Earley, G. F. Crouse, and T. D. Petes.** 1995. Mutations in the *MSH3* gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **92:** 10418–10421.
32. **Strand, M., T. A. Prolla, R. M. Liskay, and T. D. Petes.** 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365:**274–276.
33. **Sutherland, G. R., and R. I. Richards.** 1995. Simple tandem DNA repeats and human genetic diseases. Proc. Natl. Acad. Sci. USA **92:**3636–3641.
34. **Thibodeau, S. N., G. Bren, and D. Schaid.** 1993. Microsatellite instability in cancer of the proximal colon. Science **260:**816–819.
35. **White, J. H., J. F. DiMartino, R. W. Anderson, K. Lusnak, D. Hilbert, and S. Fogel.** 1988. A DNA sequence conferring high postmeiotic segregation frequency to heterozygous deletions in *Saccharomyces cerevisiae* is related to sequences associated with eucaryotic recombination hotspots. Mol. Cell. Biol. **8:**1253–1258.
36. **Wu, T.-H., and M. G. Marinus.** 1994. Dominant negative mutator mutations in the *mutS* gene of *Escherichia coli*. J. Bacteriol. **176:**5393–5400.
37. **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.