Destabilization of Simple Repetitive DNA Sequences by Transcription in Yeast

Monika Wierdl,* Christopher N. Greene,[†,‡] Abhijit Datta,[†,§] Sue Jinks-Robertson[†] and Thomas D. Petes*

**Department of Biology and Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280 and [†]Department of Biology, [‡]Graduate Program in Genetics and Molecular Biology, and [§]Graduate Program in Biochemistry and Molecular Biology, Emory University, Atlanta, Georgia 30322*

Manuscript received January 17, 1995
Accepted for publication March 11, 1996

ABSTRACT

Simple repetitive DNA sequences in the eukaryotic genome frequently alter in length. In wild-type strains, we find that transcription through a repetitive poly GT tract destabilizes the tract four- to ninefold. In mismatch repair-deficient yeast strains, simple repeats are very unstable. High levels of transcription in such strains destabilize repetitive tracts an additional two- to threefold.

SIMPLE repetitive sequences (microsatellites) are tandem repeats of a single base or a small number of bases. Such sequences are found in all eukaryotic genomes thus far examined (TAUTZ and RENZ 1984; TAUTZ *et al.* 1996). The most common simple repeats are 20–50-bp tracts of poly GT, poly A and poly GA (HAMADA *et al.* 1982; TAUTZ and RENZ 1984). Although the functions of these sequences (if any) are unclear, the high frequency with which repeats alter their length has a number of important consequences. First, the polymorphic nature of these repeats in human populations makes them valuable for genetic mapping (WEBER 1990). Second, length expansions within trinucleotide repeats cause several human genetic diseases (CASKEY *et al.* 1992; RICHARDS and SUTHERLAND 1994). Third, global microsatellite instability is associated with several types of cancer (AALTONEN *et al.* 1993; MERLO *et al.* 1994; MIRONOV *et al.* 1994; ORTH *et al.* 1994).

Two models have been proposed to explain the instability of simple repeats: unequal recombination and DNA polymerase slippage. Unequal recombination involves the pairing of repetitive tracts on sister chromatids (or homologous chromosomes) in a misaligned configuration. Crossovers between such misaligned repeats generate one larger repetitive tract and one tract with the reciprocal deletion (SMITH 1973). In the DNA polymerase slippage model (STREISINGER *et al.* 1966), a transient dissociation of the primer and template strands during DNA replication is followed by a misaligned reassociation of the strands leading to unpaired repeats (Figure 1). If replication resumes without repair of these intermediates, this mechanism will lead to an addition of repeats (if the unpaired repeat is in the primer strand) or deletion of repeats (if the unpaired repeat is in the template strand) in the newly synthesized strand.

Corresponding author: Thomas D. Petes, Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3280.
E-mail: tompetes@email.unc.edu

Most of the available evidence indicates that alterations of repetitive tracts by small numbers of repeats represent DNA polymerase slippage rather than recombination events. First, mutations that reduce recombination frequency do not affect the stability of simple repeats in *Escherichia coli* (LEVINSON and GUTMAN 1987) or in yeast (HENDERSON and PETES 1992). Second, the meiotic rate of repeat instability is about the same as the mitotic rate, although the frequency of meiotic recombination is ~1000-fold greater than the frequency of mitotic exchange (STRAND *et al.* 1993). Third, mutations in the mismatch repair genes elevate simple repeat instability in *E. coli* (LEVINSON and GUTMAN 1987), yeast (STRAND *et al.* 1993, 1995), and mammalian cells (reviewed by KARRAN and BIGNAMI 1994). This effect is consistent with the DNA polymerase slippage model, since the unpaired repeats would be expected to be a substrate for the DNA mismatch repair system. Failure to correct the unpaired repeats in mismatch repair-deficient mutants would result in elevated levels of tract instability.

Based on the polymerase slippage model, the rate of simple repeat instability is a function of the rate of DNA polymerase slippage and the rate of DNA mismatch repair. Two parameters that influence tract stability are the length and the "purity" of the tract (whether the tract contains variant repeats). In *E. coli* (LEVINSON and GUTMAN 1987; FREUND *et al.* 1989) and mammalian cells (WEBER 1990), long repetitive tracts are less stable than short tracts, presumably because the opportunity for a stable misaligned configuration is greater for the longer tracts. In addition, variant bases within repetitive tracts lead to increased stability in mammalian cells (CHONG *et al.* 1995).

In addition to the length and purity of repetitive tracts, there may be other factors that affect microsatellite stability. Transcription has been associated with both DNA replication and nucleotide excision repair

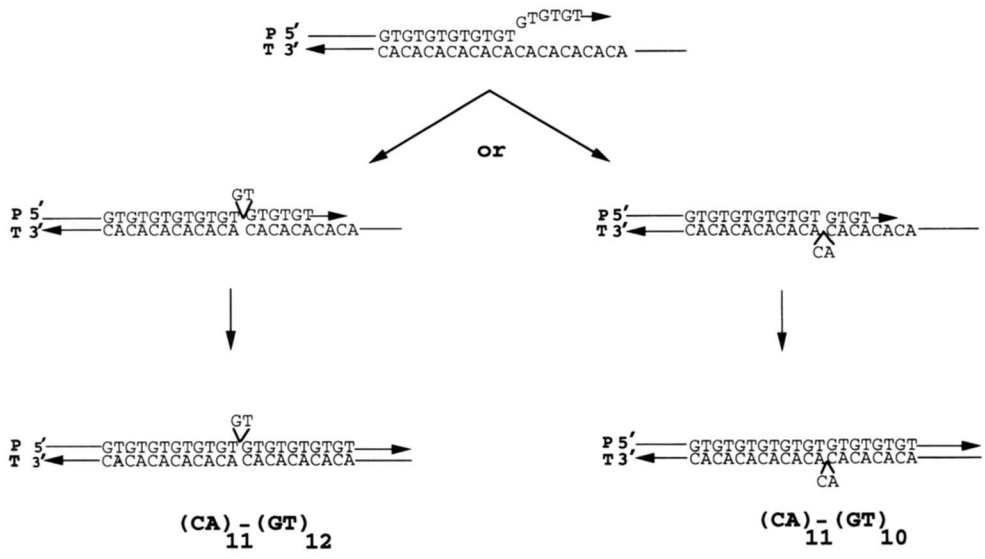

FIGURE 1.—Alterations in the length of a poly GT tract as a consequence of DNA polymerase slippage. During replication, there is a transient dissociation between the primer (P) and template (T) strands. The reassociation may result in DNA molecules in which a GT repeat is unpaired in either the primer (left side of figure) or template (right side of figure) strands. If this mismatch is not repaired, continued replication will lead to an addition of one repeat (displaced repeat in primer strand) or a deletion of one repeat (displaced repeat in template strand).

in a variety of experiments. The human transcription factor CTF is identical to NF-1, an initiation factor for adenovirus DNA replication (JONES *et al.* 1987). Several components of the nucleotide excision pathway are associated with transcription factor complexes in both prokaryotes and eukaryotes (reviewed by DRAPKIN *et al.* 1994). FOX *et al.* (1994) showed that transcription interfered with the function of the *E. coli MutY* mismatch repair gene. In addition, a number of researchers (WU *et al.* 1988; RAHMOUNI and WELLS 1989) found that transcription alters DNA superhelicity, which could influence either DNA replication or DNA repair.

DATTA and JINKS-ROBERTSON (1995) recently showed that transcription increases the level of reversion of a frameshift mutation in yeast ~30-fold. Since such reversion events often represent changes in short runs of the same base (C. GREENE and S. JINKS-ROBERTSON, unpublished data), this result suggests the possibility that transcription might also alter the stability of microsatellite sequences. Below, we show that poly GT tracts of 31 or 35 bp are destabilized four- to ninefold by high levels of transcription and that this destabilization results partly from an interference with DNA mismatch repair.

MATERIALS AND METHODS

Plasmid constructions: The plasmid pSR348 (Figure 2) contains a *pGAL10-LYS2-HIS4-URA3* translational fusion derived from pSR339 and pNKY48 (ALANI and KLECKNER 1987). pSR339 has a 1.2-kb *Xho*I-*Bam*HI *pGAL10-LYS2* fragment, generated by PCR amplification of pSR231, cloned into the *LEU2* integrating vector pRS305 (SIKORSKI and HIETER 1989). The *pGAL10-LYS2* fragment was amplified using the primers pgal10x (5′ CGGCTCGAGCGCTGATTAAATTACCCCAG) and lys2r (5′ CGCGGATCCTGGATGGATCGCTTAGCGCA). The primers pgal10x and lys2r have *Xho*I and *Bam*HI sites, respectively, at the 5′ ends. pNKY48 has a *HIS4-URA3* cassette that was removed as a 1148-bp *Bam*HI fragment and inserted into the *Bam*HI site of pSR339 to generate pSR348. Since yeast strains with a mutant *ura3* gene at the normal *URA3* locus and an integrated copy of pSR348 grow in medium lacking uracil, this construction results in a functional *URA3* fusion protein. pSR231 contains a *pGAL10-LYS2* fusion constructed by ligating a 685-bp *Bam*HI-*Eco*RI *GAL1-GAL10* promoter fragment from pBM150 (JOHNSTON and DAVIS 1984) with *Pst*I linkers into pSR229. pSR229 is a derivative of pDP6 (FLEIG *et al.* 1986) containing a promoter-less *LYS2* gene and was constructed by digestion with *Hind*III and *Eco*RV followed by creating flush ends with Klenow fragment and ligation of *Pst*I linkers.

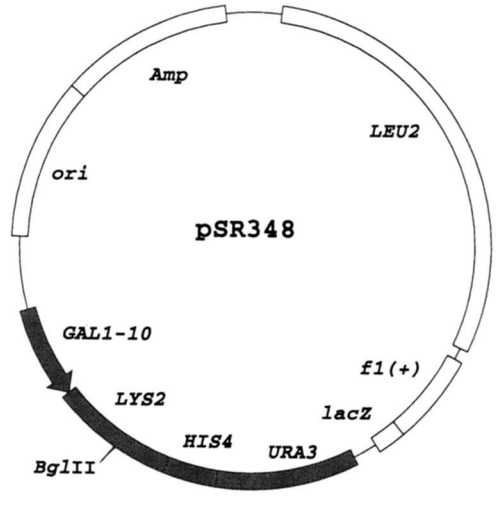

FIGURE 2.—Restriction map of plasmid pSR348. This plasmid contains a fusion gene (indicated by shading), in which synthesis of a fusion protein (with segments derived from the yeast *LYS2*, *HIS4* and *URA3* genes) is controlled by the *GAL1-10* promoter. This fusion protein results in a Ura$^+$ phenotype. Poly GT tracts were inserted into the unique *Bgl*II site within the fusion gene.

The plasmid pSR348 has a single *Bgl*II site within the *LYS2* sequence where annealed oligonucleotides containing the poly GT repeat were inserted. pMBW1 was derived from pSR348 by insertion of a 33-bp poly GT repeat at the unique *Bgl*II site. Two complementary oligonucleotides with cohesive *Bgl*II ends and the following sequences were synthesized: 5′ GATCGTCGACA$(TG)_{16}$TACTCGAG 3′ and 5′ GATCCTCGAGTA$(CA)_{16}$TGTCGAC 3′. These oligonucleotides were annealed by the method described by HENDERSON and PETES (1992), and ligated to *Bgl*II-digested pSR348. The ligated products were transformed into the *E. coli* strain DH5α. A positive clone was identified by restriction analysis and the DNA sequence confirmed by the method of KRAFT *et al.* (1988). In pMBW1, the poly GT tract was oriented such that the GT repeats were in the transcribed strand. The plasmid pMBW2 was constructed as described above, except the oligonucleotides had a 31-bp poly GT repeat in the opposite orientation within pSR348 compared with pMBW1.

The plasmid pSR376, used to monitor the level of induction of transcription from the *pGAL1-GAL10* promoter, is a *URA3*-marked integrating plasmid containing a *pGAL-lacZ* fusion. It was derived from pRY131 (YOCUM *et al.* 1984) by deleting a 1.9-kb *Spe*I fragment containing the 2-μm plasmid DNA replication origin. Linearizing pRS376 by *Stu*I digestion targets integration at *URA3*.

Media and growth conditions: Standard SD-complete and YEPD media were used (SHERMAN 1991). YEPGal plates were identical to YEPD, with galactose replacing glucose as a carbon source. Carbon sources for the various media were 2% glucose (YEPD and SD-complete); 2% galactose (YEPGal); 2% galactose, 1% raffinose (SG-complete); and 2% glycerol, 1% ethanol (YEP-GE). Cells were grown at 30°. For one protocol, cells were pregrown on plates containing SD-complete or SG-complete media for 3 days. In the other protocol, cells were pregrown for 5 days on plates containing YEP-GE medium.

Yeast strain constructions: All yeast strains used in this study were derived from AMY125 (*MAT*α *ade5 his7-2 leu2-3,112 trp1-289 ura3-52*; STRAND *et al.* 1993) by transformation with various plasmids. The strain MBW1-33 was created by transformation (SHERMAN *et al.* 1983) of AMY125 with *Eco*RI-digested pMBW1 DNA. The insertion of pMW1 at the *LEU2* locus was confirmed by Southern blot analysis (SAMBROOK *et al.* 1989). Since the poly GT insertion was in-frame with the fusion protein, MBW1-33 was Ura$^+$.

MBW1-31 and MBW1-35 strains were obtained as spontaneous 5-fluoroorotic acid-resistant (5FOA^r) derivatives of MBW1-33 and had poly GT tracts of 31 and 35 bp, respectively. Single colonies of MBW1-33 were spread on solid medium containing 5FOA (which selects for Ura$^-$ cells; BOEKE *et al.* 1984) and 5FOA^r papillae were identified. The lengths of the poly GT tracts in the Ura$^-$ isolates were determined by the hot PCR method described below. Strain MBW2-31 was isogenic to MBW1-31 except that the poly GT repeat had the opposite orientation. This strain was constructed by transformation of AMY125 with *Eco*RI-treated pMBW2.

The strain MS121 (provided by M. STRAND, Univ. of North Carolina) is isogenic to AMY125 except for a disruption of the *MSH2* gene by a *URA3* insertion. A 5FOA^r derivative of MS121 was transformed with pMBW1 to construct the strain MBW3-33. Southern analysis confirmed that the Leu$^+$ Ura$^+$ transformant contained a single integrated copy of the plasmid. The strain MBW3-35 was a 5FOA^r derivative of MBW3-33. By the PCR method described below, we showed that MBW3-35 contained a 35-bp poly GT tract (instead of the 33-bp tract of MBW3-33). Strain SJR482-35 is a *pms1* derivative of MBW1-35 constructed by two-step transplacement using the *Bst*XI-treated plasmid pSR211; this plasmid (provided by D. MALONEY, Univ. of California, Berkeley) contains a deletion mutant allele of *PMS1* with flanking sequences inserted in YIp5. The deletion in SJR482-35 was confirmed by PCR.

Strains SJR453-31 and SJR454-35 are *gal80* derivatives of MBW1-31 and MBW1-35, respectively. These strains were constructed by transformation of the parental strains with a *Bam*HI fragment containing a *gal80* gene disrupted by insertion of a *hisG-URA3-hisG* cassette, derived from plasmid pSR372 (provided by J. FRIDOVICH-KEIL, Emory University). The resulting Ura$^+$ transformants were then plated on medium containing 5FOA to select for loss of the *URA3* insertion within the *gal80* gene. The *gal80* mutation was confirmed by PCR and by testing the isolates for sensitivity to 2-deoxygalactose (PLATT 1984).

Measurements of simple sequence stability: The starting strains were phenotypically Ura$^-$ as a consequence of the out-of-frame insertion of the poly GT tract (either 35 or 31 bp). Consequently, we monitored alterations that restored the correct reading frame by measuring the frequency of Ura$^+$ colonies. Yeast strains were pregrown under conditions that lead to high or low transcription levels of the fusion gene using two different protocols (as described below). For ~20 independent colonies derived from each strain, we measured the frequency of Ura$^+$ cells in each colony by plating cells to medium (SG-uracil) that contained galactose (to induce expression of the fusion gene) and lacked uracil. Ura$^+$ colonies were counted after 2 days incubation at 30°. The rates of Ura$^+$ events per generation were calculated using the method of the median (LEA and COULSON 1949).

Two methods of elevating transcription levels were used. *GAL80* strains were pregrown on plates containing SD complete medium (glucose as carbon source, low rate of transcription of fusion gene) or on plates with SG complete medium (galactose as carbon source, high transcription rate of fusion gene). In other experiments, isogenic *GAL80* and *gal80* strains were pregrown on rich medium containing ethanol and glycerol (YEP-GE), resulting in low (*GAL80* strains) and high (*gal80* strains) levels of expression of the fusion gene. Individual colonies derived from the pregrown cells were examined for the frequency of Ura$^+$ cells as described above.

Analysis of the length of repetitive tracts by hot PCR: Lengths of poly GT tracts in Ura$^+$ colonies were measured by the method described by FARBER *et al.* (1994). PCR primers complementary to the *LYS2* coding sequences flanking the poly GT tract (5′ CCAACGTGGTCATTTAATGAGC and 5′ GCTTGAACTCGTCTAATTTG) were used to amplify an ~200-bp product. The 20 μl PCR reactions contained: 200–500 ng yeast DNA, 12.5 pM of the primers, 1 unit Taq polymerase (Perkin-Elmer), and 2.5 μCi ^{32}P-dATP in a solution of 10 mM Tris (pH 8.3), 50 mM KCl, 0.001% gelatin, 3 mM $MgCl_2$, and 0.25 mM dNTPs. PCR products were analyzed on 6% sequencing gels with control DNA samples containing poly GT tracts of 31, 33 and 35 bp.

Assay of transcription induction by the *GAL10* promoter: The plasmid pSR376 was transformed into MBW1-31 and SJR453-31 to create the strains SJR474-31 and SJR475-31. Single copy insertion at the *URA3* locus was confirmed by Southern analysis. These strains were grown in 5-ml cultures containing either 2% glycerol and 2% ethanol (YEP-GE) or cultures containing 2% galactose (YEPGal). Early log phase cultures were harvested by centrifugation. The cultures were diluted and permeabilized cells were assayed for β-galactosidase activity as described by AUSUBEL *et al.* (1987).

Statistical analyses: For each strain, the numbers of Ura$^+$ cells were determined for ~20 independent colonies per experiment, and each experiment was done at least twice. From the measured numbers, rates were calculated by the method of the median (LEA and COULSON 1949). The rates obtained for each experiment were then averaged for each strain or growth condition.

To determine whether the rates of instability were significantly different for one strain compared with another (or one growth condition compared to another), we converted the

number of Ura$^+$ cells to a rate measurement for each individual colony by using the number of Ura$^+$ cells for each colony as the median value and then employing the method of the median described by LEA and COULSON (1949). These rates were ranked for the two strains (or conditions) being compared in order from highest to lowest using the same number of colonies for each strain (or growth condition). We then determined by chi-square analysis whether one strain had significantly more colonies ranked in the top half of the rate values than the other strain. Chi-square values >3.85 ($P < 0.05$) were judged to be significant. For example, for strain MBW2-31, we measured 37 rates under conditions of low transcription and 37 under conditions of high transcription. When these 74 rates were ranked in a single list, 33 of the highest 37 rates were derived from the high-transcription experiments. If there is no significant difference in the rate of instability under the high- and low-transcription conditions, one would expect equal numbers (18.5) of the 37 top-ranked rates to be derived from cells under the two conditions. By chi-square analysis, we found that our observed result was a significant departure ($\chi^2 = 21$, $P < 0.001$) from that expectation.

Confidence limits (95%) were calculated for the rates shown in Tables 1 and 3. Using Table A-25a of DIXON and MASSEY (1969), we calculated 95% confidence limits on the median number of Ura$^+$ cells for each experiment. For example, for one of the three experiments in which we examined the rate of instability in MBW1-35 under low transcription conditions, we measured the number of Ura$^+$ cells in 15 independent colonies and these numbers were ranked. The median number of Ura$^+$ cells was 71. From Table A-25a, if the number of independent observations is 15, one can conclude that the 95% confidence intervals are between the 4th and 11th ranked values. For this experiment, these values were 58 and 88 Ura$^+$ cells. Using these numbers and the average number of cells per colony, we calculated rates using Table 3 of LEA and COULSON (1949), resulting in a rate estimate for this experiment of 1.1×10^{-5}/cell division with 95% confidence limits between 0.9 and 1.2×10^{-5} (Table 1).

RESULTS

Experimental system: To analyze the stability of simple repetitive sequences, we used a frameshift assay, similar to those described previously (HENDERSON and PETES 1992). We inserted an in-frame 33-bp tract of poly GT into the coding sequence of a *URA3* fusion gene in the plasmid pSR348 (Figure 2) to create the plasmid pMBW1. A yeast strain with a *ura3* mutation (AMY125) was transformed with this plasmid and Ura$^+$ transformants were isolated. Insertion of pMBW1 into the chromosome resulted in a strain (MBW1-33) in which the fusion gene is located between duplicated copies of *LEU2*. When we selected Ura$^-$ derivatives of MBW1-33 by using 5FOA, we found that some of these derivatives contained alterations of the poly GT tract leading to out-of-frame insertions; two such derivatives had 31-bp (MBW1-31) and 35-bp (MBW1-35) poly GT tracts. In addition, we found derivatives of MBW1-33 in which the fusion gene was lost as a consequence of recombination between the flanking *LEU2* genes. Consequently, rather than assessing tract instability by measuring the frequency of 5FOA^R cells derived from MBW1-33, we monitored the frequency of Ura$^+$ cells derived from the Ura$^-$ strains MBW1-31 and MBW1-35. As described below, DNA sequence analysis showed that all such derivatives had altered poly GT tracts that restored the correct reading frame for the fusion *URA3* gene.

Destabilization of simple repeats by transcription: The fusion *URA3* gene is transcribed from the galactose-inducible *GAL1-10* promoter, which is positively regulated by Gal4p and negatively regulated by Gal80p (JOHNSTON and CARLSON 1992). In *gal80* strains grown in medium containing glycerol and ethanol as carbon sources or in *GAL80* strains grown in medium containing galactose, transcription occurs at a high rate from this promoter. In *GAL80* strains grown in medium containing glucose or glycerol and ethanol, transcription occurs at a low rate. Consequently, to determine the effect of transcription on simple repeat instability, we used two different growth protocols. In the first protocol, we compared the rate of the appearance of Ura$^+$ colonies in a *GAL80* strain that was pregrown on medium containing glucose (low transcription) with the rate measured in the same strain pregrown on galactose (high transcription). In the second protocol, we compared the rate of Ura$^+$ colonies in a *GAL80* strain that was pregrown on medium containing ethanol and glycerol (low transcription), with the rate in an isogenic *gal80* strain pregrown on the same type of medium (high transcription).

Our Ura$^+$ rate measurements are summarized in Table 1. Similar results were obtained with both methods of inducing high transcription levels. For the *GAL80* strain with the 35-bp tract of poly GT (MBW1-35) pregrown on medium containing glucose (low transcription), we found a rate of Ura$^+$ derivatives of $\sim 1 \times 10^{-5}$/ mitotic division. Since we find (as described below) that all of these Ura$^+$ derivatives had alterations in length of the poly GT tract, this rate represents the rate of tract instability. When MBW1-35 was grown in galactose-containing medium (high transcription), the rate increased about sixfold, a statistically significant difference ($\chi^2 = 19.9$, $P < 0.001$). The rate of Ura$^+$ colonies was ninefold higher for the *gal80* strain SJR454-35 (high transcription) compared with the *GAL80* strain MBW1-35 (low transcription) when both strains were grown in medium containing glycerol and ethanol ($\chi^2 = 26.3$, $P < 0.001$).

In similar comparisons with strains containing poly GT insertions of 31 bp (strains MBW1-31 and SJR453-31), the high transcription levels elevated the rate of tract instability by about three- to fivefold (Table 1). Statistically significant differences were found comparing rates for MW1-31 pregrown on glucose- (low transcription) or galactose-containing (high transcription) medium ($\chi^2 = 8.7$, $P < 0.01$), and comparing rates for MW1-31 (low transcription) and SJR453-31 (high transcription) pregrown on YEPG medium ($\chi^2 = 24.5$, $P < 0.001$).

For both protocols, the induction of instability by high transcription levels was about twofold less for the

TABLE 1

Increased destabilization of simple repetitive sequences by high levels of transcription

Strain[a]	*GAL80* genotype/media[b]	Transcription	Rate of Ura$^+$ derivatives	Average rate of Ura$^+$ derivatives $\times 10^{-5}$	Fold increase by transcription
MBW1-35	*GAL80*/SD-complete	Low	1.1 (0.9–1.2), 0.9 (0.8–1.1)	1.0	1.0
MBW1-35	*GAL80*/SG-complete	High	5.3 (3.4–13.8), 7.1 (4.3–16.1)	6.2	6.2
MBW1-35	*GAL80*/YEP-GE	Low	0.5 (0.4–0.6), 0.2 (0.2–0.3), 0.4 (0.2–0.9), 0.8 (0.6–1.1)	0.5	1.0
SJR454-35	*gal80*/YEP-GE	High	4.3 (3.7–4.9), 4.5 (4.0–7.2)	4.4	8.8
MBW1-31	*GAL80*/SD-complete	Low	1.1 (1.0–1.3), 1.5 (1.1–2.5)	1.3	1.0
MBW1-31	*GAL80*/SG-complete	High	5.0 (3.1–7.3), 4.1 (3.1–5.5)	4.6	3.5
MBW1-31	*GAL80*/YEP-GE	Low	1.4 (1.0–2.0), 1.2 (0.9–1.7), 1.4 (1.1–2.8)	1.3	1.0
SJR453-31	*gal80*/YEP-GE	High	7.8 (5.6–9.3), 5.9 (5.1–8.5), 4.9 (3.4–7.0)	6.2	4.8
MBW2-31	*GAL80*/SD-complete	Low	1.9 (1.1–2.3), 2.5 (1.8–3.1)	2.2	1.0
MBW2-31	*GAL80*/SG-complete	High	8.7 (6.8–13.0), 4.5 (3.6–5.8)	6.6	3.0

[a] The numbers 31 and 35 after the strain name indicates the length of the GT tract in base pairs.
[b] SD-complete medium contains dextrose, SG-complete medium contains galactose, and YEP-GE medium contains glycerol and ethanol.
[c] Numbers in parentheses represent 95% confidence values for the rates measured in individual experiments.

31-bp tract than for the 35-bp tract. The difference appears to reflect a lower rate of instability for the 35-bp tract under conditions of low transcription, rather than a higher level of instability of the 35-bp tract under conditions of high transcription (Table 1). This issue will be discussed further below.

We also examined the effect of the orientation of the poly GT tract on transcription-stimulated tract instability. In MBW1-31, the transcript of the fusion *URA3* gene will contain a poly CA tract in the mRNA. The strain MBW2-31 is identical to MBW1-31, except for the orientation of the poly GT tract. High levels of transcription resulted in a 3.0-fold induction (statistically significant with $\chi^2 = 21$, $P < 0.001$) of tract instability in MBW2-31 (Table 1), similar to the 3.5-fold induction observed by the same protocol for MBW1-31.

Level of transcription induction: To confirm levels of high and low transcription from the *GAL10* promoter, we inserted a single copy of a *pGAL10-lacZ* reporter gene into strains MBW1-31 (*GAL80*) and SJR453-31 (*gal80*) to create the strains SJR474-31 and SJR475-31, respectively. Measurements of β-galactosidase activity of four independent cultures were done for each strain or growth condition. The *GAL80* strain SJR474-31 had 360-fold more β-galactosidase activity when grown in medium containing galactose (YEPGal) than when grown in medium with glycerol and ethanol (YEP-GE). The *gal80* strain SJR475-31, grown in YEP-GE, had 460-fold more β-galactosidase activity than the *GAL80* strain SJR474-31, grown under the same conditions.

Sequence analysis of altered tracts: In our previous studies of alterations in poly GT tracts, we found that most changes involve addition or deletion of one or two repeats with ~10% of the alterations representing larger deletions or insertions (HENDERSON and PETES 1992; STRAND *et al.* 1993, 1995). To confirm that the Ura$^+$ cells derived from the strains used in this study also resulted from tract alterations and to determine the types of alterations, we examined tract lengths by analyzing PCR products on DNA sequencing gels. These results are summarized in Table 2.

All Ura$^+$ derivatives examined had poly GT tract alterations that restored the correct reading frame of the fusion protein. In strains with 35-bp tracts, the most common alteration was a deletion of one repeat, although a few tracts with additions of two repeats were

TABLE 2

Tract length alterations in high and low transcription strains

Strain[a]	*GAL80* genotype/media[b]	Transcription	No. of tracts with additions or deletions of base pairs				
			−4	−2	+2	+4	Total
MBW1-35	*GAL80*/SD-complete	Low	0	19	0	4	23
MBW1-35	*GAL80*/SG-complete	High	0	21	0	2	23
SJR454-35	*gal80*/YEP-GE	High	0	15	0	1	16
MBW1-31	*GAL80*/SD-complete	Low	0	0	20	0	20
MBW1-31	*GAL80*/SG-complete	High	1	0	18	0	19
SJR453-31	*gal80*/YEP-GE	High	0	0	15	0	15
MBW2-31	*GAL80*/SD-complete	Low	0	0	15	0	15
MBW2-31	*GAL80*/SG-complete	High	0	0	15	0	15

[a] The numbers 31 and 35 after the strain name indicates the length of the GT tract in base pairs.
[b] SD-complete medium contains dextrose, SG-complete medium contains galactose, and YEP-GE medium contains glycerol and ethanol.

also observed. In strains with 31-bp tracts, almost all of the altered tracts had gained a single repeat; only one tract with a deletion of 4 bp was observed. Of 146 tract alterations examined, none contained an addition or deletion of more than two repeats. The types of changes observed with high transcription levels were not different from those observed with low transcription levels. In addition, the orientation of the poly GT tract did not affect the types of tract alterations.

Effects of transcription on simple repeat stability in mutants deficient for DNA mismatch repair: Mutations that reduce DNA mismatch repair dramatically increase simple repeat instability in prokaryotes (LEVINSON and GUTMAN 1987) and in eukaryotes (STRAND *et al.* 1993; HEMMINKI *et al.* 1994; WIND *et al.* 1995). In yeast, simple repeats are destabilized ~200-fold by mutations in the *PMS1*, *MSH2*, or *MLH1* genes (STRAND *et al.* 1993) and ~40-fold by mutations in *MSH3* (STRAND *et al.* 1995).

In strain MBW1-35, there is a six- to ninefold stimulation of repeat instability by transcription (Table 1). Strains MBW3-35 and SJR482-35 are isogenic with MBW1-35, except for mutations in the *msh2* and *pms1* genes, respectively. In both mismatch repair defective strains, high levels of transcription destabilized the poly GT repeats an additional two- to threefold (Table 3). The destabilizing effect of transcription, although smaller in the mismatch repair deficient strains than in wild-type strains, is still statistically significant: a chi-square value of 19 ($P < 0.001$) for the comparison of MBW3-35 under pregrowth conditions leading to high and low transcription and a chi-square value of 20 ($P < 0.001$) for the comparison of SJR482-35 under conditions leading to high and low transcription.

We also examined the types of tract alterations in the *msh2* strain MW3-35 and the *pms1* strain SJR482 (Table 4). Under conditions of high or low transcription, we found exclusively deletions of one repeat or additions of two repeats. The ratio between these two classes of events was not significantly affected by transcription.

DISCUSSION

We find that high levels of transcription stimulate instability of simple repetitive sequences, and that this effect is partly dependent on a functional DNA mismatch repair system. Below, we discuss the implications of these results.

As pointed out in RESULTS, the induction of tract instability by high levels of transcription is larger for strains with the 35-bp repeat (six- to ninefold) than for strains with the 31-bp repeat (three- to fivefold). This difference reflects a lower rate of instability in strains with the 35-bp repeat compared with strains with the 31-bp repeat under conditions of low transcription (Table 1). Previously, we observed that insertions of single GT repeats were more frequent than deletions by two- to threefold in wild-type strains under conditions of low transcription (HENDERSON and PETES 1992; STRAND *et al.* 1993, 1995). For the 35-bp repeat, deletion of 2 bp is required to restore the reading frame, whereas an addition of 2 bp is required to restore the reading frame for the 31-bp repeat. The difference in the rates of instability for strains with 35 and 31 bp repeats under conditions of low transcription may thus reflect a bias imposed by the selection system. Since strains with 35 and 31 bp repeats have similar rates of instability under conditions of high transcription, transcription-induced instability results in similar frequencies of additions and deletions.

In previous experiments, ~10% of the frameshift events detected in wild-type cells with a 33-bp poly GT tract were deletions or additions involving more than two repeats (STRAND *et al.* 1993, 1995). It is not clear whether these larger alterations also represent DNA polymerase slippage or a different mechanism (for example, recombination). In >100 tracts analyzed in our experiments, no large deletions or insertions were detected. Since the experiments reported here were done in the same genetic background as those of STRAND *et*

TABLE 3

Transcription destabilizes simple repeat tracts in the absence of mismatch repair

Strain[a]	Relevant genotype	Transcription	Rate of Ura$^+$ derivatives[b]	Average rate of Ura$^+$ derivatives	Relative rate of destabilization[c]
MBW1-35	Wild type	Low	1.1×10^{-5} (0.9–1.2) 0.9×10^{-5} (0.6–1.2)	1.0×10^{-5}	1.0
MBW1-35	Wild type	High	5.3×10^{-5} (3.4–13.8) 7.1×10^{-5} (4.3–16.1)	6.2×10^{-5}	6.2
MBW3-35	*msh2*	Low	1.1×10^{-3} (0.7–1.7) 1.2×10^{-3} (0.8–1.6) 1.6×10^{-3} (1.1–2.3)	1.3×10^{-3}	130 (1.0)
MBW3-35	*msh2*	High	2.6×10^{-3} (2.0–4.2) 2.8×10^{-3} (1.2–7.5) 3.7×10^{-3} (2.4–6.9)	3.0×10^{-3}	300 (2.3)
SJR482-35	*pms1*	Low	0.8×10^{-3} (0.5–0.9) 1.0×10^{-3} (0.5–1.2)	0.9×10^{-3}	90 (1.0)
SJR482-35	*pms1*	High	2.6×10^{-3} (1.9–2.9) 1.9×10^{-3} (1.5–2.8)	2.3×10^{-3}	230 (2.6)

[a] The numbers 31 and 35 after the strain name indicates the length of the GT tract in base pairs.
[b] Numbers in parentheses represent 95% confidence values for the rates measured in individual experiments.
[c] Rates are relative to MBW1-35 grown under low transcription conditions. Numbers in parentheses represent relative rates for the mismatch repair deficient strains grown under low and high transcription conditions.

al. (1993, 1995), the lack of large alterations does not reflect strain differences. It could be a consequence of the different fusion genes used to assay tract instability or a consequence of the location of the assayed gene (located within the chromosome in our experiments and on a plasmid in the experiments of STRAND *et al.* (1993, 1995).

The six- to ninefold induction of microsatellite instability by high levels of transcription is less than the ~30-fold induction of reversion observed for a *lys2* frameshift mutation in studies done with the same promoter (DATTA and JINKS-ROBERTSON 1995). It is likely that the induction of alterations by transcription in these two experiments has different causes. DATTA and JINKS-ROBERTSON (1995) suggested that high levels of transcription may lead to increased levels of DNA damage that is repaired (in part) by the error-prone *REV3*-encoded DNA polymerase. This effect increases the mutation rate from 10^{-9}/division to 3×10^{-8}/division. In our experiments, high levels of transcription elevate tract alterations from ~10^{-5}/division to 6×10^{-5}/division. Thus, the effect examined by DATTA and JINKS-ROBERTSON would probably not be detectable in our experiments.

The observed frequency of tract instability is likely to be a function of the error rate of DNA polymerase and the efficiency of DNA mismatch repair. The error rate of DNA polymerase reflects both the DNA polymerase slippage rate and the rate with which errors are corrected by the associated proofreading exonuclease activity. Thus, if the frequency with which DNA polymerase generates mismatches is *a*, and the frequency of mismatches escaping the mismatch repair system is *b*, then the observed frequency of simple repeat instability is *ab*, assuming that the frequencies of both processes are independent. Transcription could elevate simple repeat instability in a variety of ways: by decreasing the rate of mismatch repair, by increasing the error rate of DNA polymerase, by affecting both DNA polymerase and mismatch repair, or by activating a new pathway (for example, recombination) leading to simple repeat instability.

Our results limit the possible explanations for the

TABLE 4

Tract alterations observed in *msh2* (MBW3-35) and *pms1* (SJR482-35) mismatch repair mutants

Strain[a]	*GAL80* genotype/media[b]	Transcription	No. of tracts with additions or deletions of base pairs		
			−2	+4	Total
MBW3-35	*GAL80*/SD-complete	Low	28	6	34
MBW3-35	*GAL80*/SG-complete	High	28	9	37
SJR482-35	*GAL80*/SD-complete	Low	13	2	15
SJR482-35	*GAL80*/SG-complete	High	13	0	13

[a] The numbers 31 and 35 after the strain name indicates the length of the GT tract in base pairs.
[b] SD-complete medium contains dextrose, SG-complete medium contains galactose.

effects of transcription on tract instability. Since transcription destabilizes poly GT tracts even in mismatch repair-deficient strains, the effect of transcription in wild-type strains is not solely a consequence of reducing the efficiency of the *MSH2/PMS1* mismatch repair system. If the effect of transcription was solely on the DNA polymerase error rate, then one would expect that the effects of transcription and mismatch repair defects would be multiplicative. By this model, the expected level of instability for MBW3-35 (grown under conditions leading to high levels of transcription) should be ~800, instead of the observed value of 300 (Table 3). The third model that can be ruled out is the activation of a mechanism of tract instability that is independent of the DNA mismatch repair system. By this model, the effects of transcription would be additive with the effects of mutations of eliminating DNA mismatch repair. Thus, this model predicts a relative rate of tract instability for MBW3-35 (grown under conditions leading to high levels of transcription) of 136, a value considerably below the relative rate observed (300).

The conclusion that appears to fit the data best is that transcription has two effects, reducing the efficiency of DNA mismatch repair two- to threefold and increasing the error rate of DNA polymerase about two- to threefold. The net effect of these two processes is to increase tract instability in a wild-type strain about sixfold. One explanation for these effects is that the transcription complex may displace DNA polymerase and the mismatch repair proteins from DNA (FOX *et al.* 1994). The displacement of DNA polymerase could increase the rate of DNA polymerase slippage or reduce the efficiency of the proofreading exonuclease. Alternatively, the increased transcription could influence supercoiling in domains of the chromosome, reducing the efficiency of DNA-DNA polymerase interactions and DNA-mismatch repair protein interactions. Whatever the mechanism involved in transcription-induced destabilization of simple repeats, our results indicate that the rate of transcription is another parameter influencing microsatellite stability.

We thank M. STRAND, D. MALONEY, J. FRIDOVICH-KEIL and A. MORRISON for providing yeast strains or plasmids, E. SIA for comments on the manuscript, and D. GORDENIN and F. STUART for advice on the statistical analysis. This research was supported by grants from the National Institutes of Health (GM-50892 to T.D.P. and GM-38464 to S.J.-R.) and the National Science Foundation (MCB-9405953 to S.J.-R.).

LITERATURE CITED

AALTONEN, L. A., P. PELTOMAKI, F. S. LEACH, P. SISTONEN, L. PYLKKANEN *et al.*, 1993 Clues to the pathogenesis of familial colorectal cancer. Science **260:** 812–816.

ALANI, E., and N. KLECKNER, 1987 A new type of fusion analysis applicable to many organisms: protein fusions to the *URA3* gene of yeast. Genetics **117:** 5–12.

AUSUBEL, F. M., R. BRENT, R. E. KINGSTON, D. M. MOORE, J. G. SEIDMAN *et al.*, 1994 *Current Protocols in Molecular Biology*, vol. 2. John Wiley & Sons, Inc., New York.

BOEKE, J. D., F. LACROUTE and G. R. FINK, 1984 A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoroorotic acid resistance. Mol. Gen. Genet. **197:** 345–346.

CASKEY, C. T., A. PIZZUTI, Y.-H. FU, R. G. FENWICK, JR. and D. L. NELSON, 1992 Triplet repeat mutations in human disease. Science **256:** 784–789.

CHONG, S. S., A. E. MCCALL, J. COTA, S. H. SUBRAMONY, H. T. ORR *et al.*, 1995 Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. Nat. Genet. **10:** 344–350.

DATTA, A., and S. JINKS-ROBERTSON, 1995 Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science **268:** 1616–1619.

DIXON, W. J., and F. J. MASSEY, JR., 1969 *Introduction to Statistical Analysis.* McGraw-Hill Book Company, New York.

DRAPKIN, R., A. SANCAR and D. REINBERG, 1994 Where transcription meets repair. Cell **77:** 9–12.

FARBER, R. A., T. D. PETES, M. DOMINSKA, S. S. HUDGENS and R. M. LISKAY, 1994 Instability of simple sequence repeats in a mammalian cell line. Hum. Mol. Genet. **3:** 253–256.

FLEIG, U. N., R. D. PRIDMORE and P. PHILIPPSEN, 1986 Construction of *LYS2* cartridges for use in genetic manipulations of *Saccharomyces cerevisiae.* Gene **46:** 237–245.

FOX, M. S., J. P. RADICELLA and K. YAMAMOTO, 1994 Some features of base pair mismatch repair and its role in the formation of genetic recombinants. Experientia **50:** 253–260.

FREUND, A.-M., M. BICHARA and R. P. P. FUCHS, 1989 Z-DNA forming sequences are spontaneous deletion hot spots. Proc. Natl. Acad. Sci. USA **86:** 7465–7469.

HAMADA, H., M. G. PETRINO and T. KAKUNAGA, 1982 A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA **79:** 6465–6469.

HEMMINKI, A., P. PELTOMAKI, J.-K. MECKLIN, H. JARVINEN, R. SALOVAARA *et al.*, 1994 Loss of the wild type *MLH1* gene is a feature of hereditary nonpolyposis colorectal cancer. Nat. Genet. **8:** 405–410.

HENDERSON, S. T., and T. D. PETES, 1992 Instability of simple sequence DNA in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **12:** 2749–2757.

JOHNSTON, M., and M. CARLSON, 1992 Regulation of carbon and phosphate utilization, pp. 200–227 in *The Molecular and Cellular Biology of the Yeast Saccharomyces,* Ed. 2. edited by E. W. JONES, J. R. PRINGLE and J. R. BROACH. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

JOHNSTON, M., and R. W. DAVIS, 1984 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **4:** 1440–1448.

JONES, K. A., J. T. KADONAGA, P. J. ROSENFELD, T. J. KELLY and R. TJAIN, 1987 A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell **48:** 79–89.

KARRAN, P., and M. BINGAMI, 1994 DNA damage tolerance, mismatch repair and genome instability. BioEssays **16:** 833–839.

KRAFT, R. J., K. S. TARDIFF, K. S. KRAUTER and L. A. LEINWARD, 1988 Using miniprep plasmid DNA for sequencing double stranded templates with Sequenase. Biotechniques **6:** 544–545.

LEA, D. E., and C. A. COULSON, 1949 The distribution of the number of mutants in bacterial populations. J. Genet. **49:** 264–285.

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.

MERLO, A., M. MABRY, E. GABRIELSON, R. VOLLMER, S. B. BAYLIN *et al.*, 1994 Frequent microsatellite instability in primary small cell lung cancer. Cancer Res. **54:** 2098–2100.

MIRONOV, N. M., M. A.-M. AUGELON, G. I. POTAPOVA, Y. OMORI, O. V. GORBUNOV *et al.*, 1994 Alterations of (CA)n DNA repeats and tumor supressor genes in human gastric cancer. Cancer Res. **54:** 41–43.

ORTH, K., J. HUNG, A. GAZDAR, A. BOWCOCK, J. M. MATHIS *et al.*, 1994 Genetic instability in human ovarian cancer cell lines. Proc. Natl. Acad. Sci. USA **91:** 9495–9499.

PLATT, T., 1984 Toxicity of 2-deoxygalactose to *Saccharomyces cerevisiae* cells constitutively synthesizing galactose-metabolizing enzymes. Mol. Cell. Biol. **4:** 994–996.

RAHMOUNI, A. R., and R. D. WELLS, 1989 Stabilization of Z DNA *in vivo* by localized supercoiling. Science **246:** 358–363.

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

SAMBROOK, J., E. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SHERMAN, F., 1991 Getting started with yeast, pp. 12–18 in *Guide to Yeast Genetics and Molecular Biology,* edited by C. GUTHRIE and G. R. FINK. Academic Press, San Diego.

SHERMAN, F., G. R. FINK and J. B. HICKS, 1983 Methods in yeast genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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.

SMITH, G. P., 1973 Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. **38:** 507–513.

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.

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.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA *et al.*, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

TAUTZ, D., and M. RENZ, 1984 Simple sequences are ubiquitous components of eukaryotic genomes. Nucleic Acids Res. **12:** 4127–4138.

TAUTZ, D., M. TRICK and G. A. DOVER, 1996 Cryptic simplicity in DNA is a major source of genetic variation. Nature **322:** 652–656.

WEBER, J. L., 1990 Informativeness of human (dC-dA)n-(dG-dT)n polymorphism. Genomics **7:** 524–530.

WIND, N. D., M. DEKKER, A. BERNS, M. RADMAN and H. T. RIELE, 1995 Inactivation of the mouse *Msh2* gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell **82:** 321–330.

WU, H.-Y., S. SHYY, J. C. WANG and L. F. LIU, 1988 Transcription generates positively and negatively supercoiled domains in the template. Cell **53:** 433–440.

YOCUM, R. R., S. HANLEY, R. WEST, JR. and M. PTASHNE, 1984 Use of *lacZ* fusions to delimit regulatory elements of the inducible divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae.* Mol. Cell. Biol. **4:** 1985–1998.

Communicating editor: P. L. FOSTER