Microsatellite Instability in Yeast: Dependence on the Length of the Microsatellite

Monika Wierdl, Margaret Dominska and Thomas D. Petes

Department of Biology, Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280

Manuscript received December 18, 1996
Accepted for publication March 4, 1997

ABSTRACT

One of the most common microsatellites in eukaryotes consists of tandem arrays [usually 15–50 base pairs (bp) in length] of the dinucleotide GT. We examined the rates of instability for poly GT tracts of 15, 33, 51, 99 and 105 bp in wild-type and mismatch repair-deficient strains of *Saccharomyces cerevisiae*. Rates of instability increased more than two orders of magnitude as tracts increased in size from 15 to 99 bp in both wild-type and *msh2* strains. The types of alterations observed in long and short tracts in wild-type strains were different in two ways. First, tracts ≥51 bp had significantly more large deletions than tracts ≤33 bp. Second, for the 99- and 105-bp tracts, almost all events involving single repeats were additions; for the smaller tracts, both additions and deletions of single repeats were common.

REGIONS of DNA in which a single base or a small number of bases is repeated multiple times are common in most eukaryotic genomes. Such repetitive tracts (microsatellites) are unstable (LEVINSON and GUTMAN 1987; HENDERSON and PETES 1992), frequently undergoing changes in tract length. These changes result in polymorphisms that are useful in genetic mapping studies (WEBER 1990). In addition, microsatellite instability is related to human diseases in two different ways. First, locus-specific expansion of trinucleotide tracts is associated with a number of human genetic disorders (reviewed by ASHLEY and WARREN 1995). Second, genome-wide microsatellite instability is diagnostic of certain types of human tumors (reviewed by DE LA CHAPELLE and PELTOMAKI 1995).

Two mechanisms have been proposed to explain the instability of simple repeats: DNA polymerase slippage and unequal recombination. As shown in Figure 1, the first model involves dissociation of replicating DNA strands, followed by their misaligned reassociation (STREISINGER *et al.* 1966). This process results in unpaired repeat units (loops) on either the primer or on the template strand. If DNA replication resumes without the repair of these loops, the number of repeats on the newly synthesized DNA strand will be either greater (if the loop was on the primer strand) or smaller (if the loop was on the template strand) than the original number of repeats. In the second model (SMITH 1973), crossing over or gene conversion between misaligned repeated sequences on sister chromatids or on homologous chromosomes results in tracts with altered numbers of repeats. A reciprocal crossover results in one longer and one shorter microsatellite, whereas gene conversion could change tract length on only one of the participating DNA molecules.

A number of arguments indicate that changes in tract length involving small numbers of repeats are likely to reflect DNA polymerase slippage, although the recombination model cannot be completely excluded. (1) Mutations that reduce or eliminate most types of recombination in *Escherichia coli* and yeast do not affect microsatellite instability (LEVINSON and GUTMAN 1987; HENDERSON and PETES 1992). (2) The rate of tract instability is much greater than expected for standard mitotic recombination events involving small regions of sequence homology (AHN *et al.* 1988; HENDERSON and PETES 1992). (3) Mutations in genes affecting DNA mismatch repair greatly destabilize simple repetitive DNA sequences in *E. coli* (LEVINSON and GUTMAN 1987), yeast (STRAND *et al.* 1993, 1995) and humans (reviewed by KOLODNER 1996). These results are consistent with the DNA polymerase slippage model, since the intermediates of slippage events (unpaired loops on the primer or template strands) would be expected to be substrates for the mismatch repair system.

If microsatellite alterations occur by DNA polymerase slippage, a number of different types of mutations would be expected to affect microsatellite stability: mutations of DNA polymerase affecting processivity or proofreading exonuclease activity, mutations affecting the activity of helicases involved in strand dissociations and alterations in genes required for DNA mismatch repair. As described above, mutations in genes required for DNA mismatch repair dramatically elevate microsat-

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

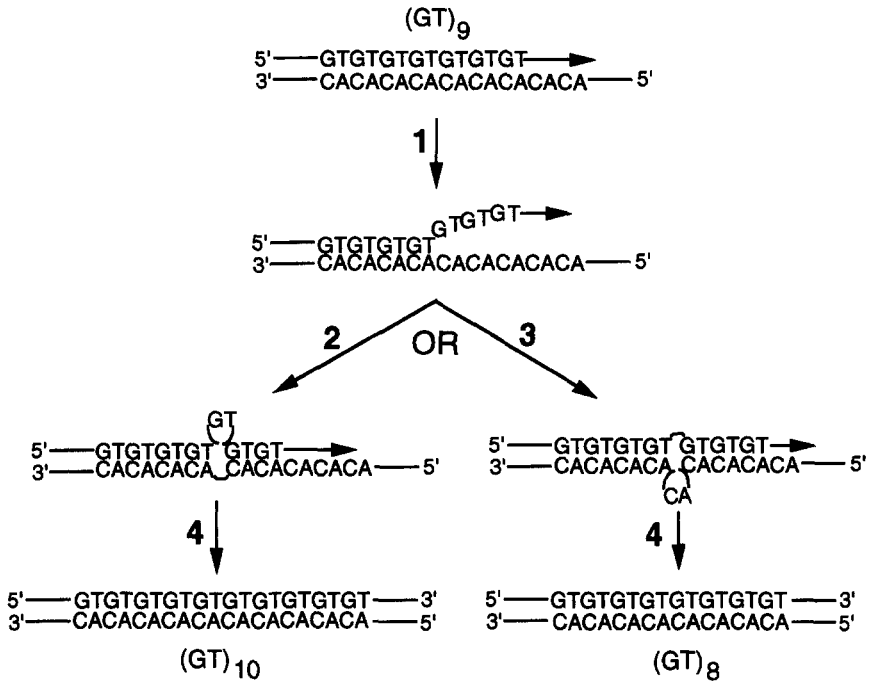

FIGURE 1.—Alterations in the length of poly GT tracts by DNA polymerase slippage. During DNA replication of an 18-bp poly GT tract, the elongating DNA strand and the template strand temporarily dissociate (1). Reassociation of the strands occurs, leaving either a mismatched repeat in the elongating (2) or in the template (3) strand. Continued DNA synthesis (4) results in a tract of 20 bp (left) or 16 bp (right).

ellite instability, whereas mutations affecting the proofreading exonuclease of DNA polymerase have a more modest effect (STRAND *et al.* 1993). In addition, the rate of instability could be affected by the microsatellite substrate: the size of the repeat units, the base composition of the repeat, the length of the repetitive tract, the transcriptional state of the microsatellite and the "purity" of the repetitive tract.

Several of these parameters have been examined. For example, microsatellites are stabilized by variant repeats within the tract in *E. coli* (BICHARA *et al.* 1995), human cells (CHONG *et al.* 1995) and yeast (T. PETES, P. GREENWELL and M. DOMINSKA, unpublished data). Elevated levels of transcription result in microsatellite destabilization by decreasing the efficiency of DNA mismatch repair and increasing DNA polymerase slippage (WIERDL *et al.* 1996). The stability of some triplet repeat microsatellites is affected by the orientation of the microsatellite with respect to the replication origin (KANG *et al.* 1995; MAURER *et al.* 1996), although some dinucleotide microsatellites do not show this effect (HENDERSON and PETES 1992). Finally, as described below, the length of a repetitive tract affects microsatellite stability.

The effect of tract length on the frequency of microsatellite instability was first studied in *E. coli* (LEVINSON and GUTMAN 1987; FREUND *et al.* 1989). Examining a poly GT tract located within the single-stranded M13 phage, LEVINSON and GUTMAN (1987) found frequencies of instability of 0.3×10^{-2} for a 22-base pair (bp) tract and 1.2×10^{-2} for a 40-bp tract. Almost all of the tract alterations were additions or deletions of single repeats, with deletions exceeding additions by about fivefold. FREUND *et al.* (1989) analyzed the stability of microsatellite sequences in double-stranded plasmid vectors. These workers found that increasing the size of a poly GT tract from 24 to 48 bp elevated the frequency of tract alterations from 4.4×10^{-4} to 6.1×10^{-3}. For the 24-bp tract, about half of the alterations involved addition or deletion of one or two repeats and half involved larger additions or deletions. For the 48-bp tract, ~90% had large (>10 bp) deletions. FREUND *et al.* (1989) suggest that the tendency for the long poly GT tracts to acquire deletions might be related to a propensity of long poly GT tracts to adopt the Z-form DNA conformation.

The effect of tract length on tract stability in eukaryotes has been examined in only a small number of studies. In yeast it has been shown that long (150–250 bp) poly CAG tracts are unstable, frequently undergoing large deletions (MAURER *et al.* 1996; J. MIRET, L. PESSOA-BRANDAO and R. S. LAHUE, personal communication). In addition, the rate of instability of poly A tracts between 4 and 12 bp varied by more than two orders of magnitude (Tran *et al.* 1997). In human populations long poly GT tracts tend to be more polymorphic than short tracts, suggesting that long tracts are more unstable (WEBER 1990). By PCR analysis of single sperm LEEFLANG *et al.* (1995) found the frequency of size alterations for a triplet CAG tract located near the Huntington's disease gene increased with tract length as follows: 0.6% (15–18 repeats), 11% (30 repeats), 53% (36 repeats) and 92–99% (38–51 repeats).

In summary, although a number of experiments suggest that microsatellite instability increases as repetitive tracts become longer, systematic studies of the effects of altering tract length on rates of tract stability have

not been done in these systems. Below we analyze the rates of instability and the types of alterations in poly GT tracts varying in size between 15 and 105 bp in wild-type yeast strains and strains with an *msh2* mutation.

MATERIALS AND METHODS

Plasmid constructions: All plasmids used in this study are derivatives of pSH44 (HENDERSON and PETES 1992). This plasmid contains an in-frame insertion of a 33-bp poly GT tract within the coding sequence of a gene encoding a fusion protein with *URA3* activity. Yeast cells with this plasmid are Ura$^+$ and sensitive to the drug 5-fluoro-orotate (5FOA) (BOEKE *et al.* 1984). Thus, alterations in tract length that cause a frameshift can be selected on 5FOA. Plasmids p15GT, p51GT, p99GT and p105GT are identical to pSH44 except for the size of the poly GT tract. Plasmids p15GT and p99GT were constructed by insertion of annealed oligonucleotides (5′TCGACA(TG)$_n$TAC3′ and 5′TCGAGTA(CA)$_n$TG3′ with n = 7 or 49) into *Xho*I/*Sal*I-treated pSH44. The protocol for annealing the complementary oligonucleotides was described previously (HENDERSON and PETES 1992). The sequence and orientation of the insertions were confirmed by DNA sequencing. In pSH44 and all derivatives used in this study, the poly GT repeats were on the nontranscribed strand. Plasmids p51GT and p105GT were derived in two steps (initial selection for 5FOA-resistance, followed by selection for Ura$^+$ phenotype) as products of spontaneous *in vivo* frameshift events in yeast of plasmids pSH44 and p99GT, respectively.

Yeast strain constructions: Yeast strains PD3/p15, SH44 (PD3/pSH44), PD3/p51, PD3/p99 and PD3/p105 were derived from PD3 (*MATα trp1 arg4-17 tyr7-1 ade6 ura3 his4-Sal*) by transformation with the plasmids described above; PD3 is a *his4* derivative of the haploid AS4 (NAG and PETES 1990). PD3Δ*msh2* was constructed from PD3 by the disruption of the *MSH2* gene using the plasmid pII-2/Tn10-LUK (REENAN and KOLODNER 1992); in this plasmid the *MSH2* gene is disrupted by a *URA3* insertion. The plasmid was treated with *Spe*I and transformed into PD3. Ura$^+$ transformants were examined by Southern analysis to confirm disruption of *MSH2*. We then selected a Ura$^-$ derivative of PD3Δ*msh2* using 5FOA. This strain (MBW10) was transformed with the plasmids described above to create strains MBW10/p15, MBW10/p33, MBW10/p51 and MBW10/p99. The strain SH52 is a *rad52* derivative of PD3 previously described (HENDERSON and PETES 1992). In the strain SH52*/p51, the p51GT plasmid replaces pSH44.

The diploid strain MBW20 (*MAT***a**/*MATα trp1Δ1/trp1 ura3-52/ura3 lys2-801/lys2 ARG4/arg4-17 TYR7/tyr7-1 ADE6/ade6 HIS4/his4-Sal*) was constructed by mating a *lys2* derivative of PD3 to TP45 (*MAT***a** *trp1Δ1 ura3-52 lys2-801*). The diploid strain was transformed with p99GT to generate MBW20/p99.

Media and growth conditions: All yeast strains were grown at 30° using standard media (SHERMAN 1991). Since all assay plasmids contained the *TRP1* marker, strains were grown in medium lacking tryptophan to force retention of the plasmid. To estimate rates of tract instability, we plated cells on medium containing 0.1% 5FOA (BOEKE *et al.* 1984) and lacking tryptophan, leucine and threonine (HENDERSON and PETES 1992); strains with high backgrounds on plates with 0.1% 5FOA were reexamined on plates with 0.2% 5FOA. Media and conditions for sporulation were standard (SHERMAN 1991).

Measurements of rates of microsatellite instability: As described above, strains used in this study were Ura$^+$ due to the presence of an assay plasmid with an in-frame poly GT insertion within the coding sequence of the *URA3* reporter gene. Previous studies (HENDERSON and PETES 1992; STRAND *et al.* 1993, 1995) indicated that the rate of 5FOA^R derivatives reflected the rate of tract alterations. To determine the rate at which 5FOA^R derivatives appeared, we first grew the yeast strains on solid medium lacking tryptophan for 3 days. On the third day, ~20 colonies were suspended in water, and dilutions were plated on medium containing 5FOA but lacking leucine, threonine and tryptophan (to measure the frequency of 5FOA-resistant cells) and on the same medium without 5FOA (to monitor viable cells). The plates were incubated at 30° for 3 days and the numbers of colonies on each plate were counted. At least two independent experiments involving 12–20 cultures per experiment were done for each strain.

Frequency data were converted to rate estimates using the method of the median (LEA and COULSON 1949). For this calculation, we determined the median number of 5FOA^R derivatives in the cultures of a single experiment (r_0) and the average number of viable cells per culture (x). From the r_0 value and Table 3 of Lea and Coulson (1949) the mean number of mutational events per culture (m) can be calculated. By dividing m by x, we calculated a rate of 5FOA^R mutations/cell division. From 25 or more independent 5FOA^R derivatives of each strain, we analyzed the size of the microsatellite by DNA sequencing or PCR (as described below). For all strains except those containing p15GT >80% of the tracts analyzed had altered lengths. To calculate a rate of tract instability, we multiplied the rate of appearance of 5FOA^R derivatives by the ratio of number of tracts with alterations to the total number of tracts analyzed for the strain.

For comparisons of the rates of mitotic and meiotic instability in MBW20/p99, the strain was grown for 3 days on solid medium lacking tryptophan. Colonies were then resuspended in 5 ml of liquid presporulation medium (SHERMAN 1991) and grown for 24 hr. Cells were harvested by centrifugation. Half of the cultures were analyzed for microsatellite instability as described above, whereas the remaining half was suspended in 5 ml of sporulation medium (SHERMAN 1991). After 30–34 hr in sporulation medium, the frequency of 5FOA^R cells was determined as described above. Since strain MBW20/p99 was heteroallelic for mutations at the *LYS2* locus, we also measured the frequency of Lys$^+$ cells in the same cultures as a control for the frequency of meiotic recombination.

The frequency of Lys$^+$ and 5FOA^R cells in sporulated cultures represent both the frequency of mitotic events that occurred in the culture before sporulation and the frequency of meiotic events. If the frequency of the meioitic events is not high relative to the frequency of mitotic events, there will be little difference in the pre- and postsporulation cultures. We found little difference in the frequencies of 5FOA^R in pre- and postsporulation cultures, and a large difference in the frequencies of Lys$^+$ cells in the two types of cultures. For example, in one experiment, the median frequencies of 5FOA^R cells were 1.1×10^{-3} for the presporulation culture and 1.6×10^{-3} for the postsporulation culture, an insignificant difference. Since these results suggest that most of the 5FOA^R cells were generated during mitotic growth, these data were converted to rates by the method of the median (LEA and COULSON 1949). The median frequencies of Lys$^+$ cells in this experiment were 1.3×10^{-5} for the presporulation culture and 2.3×10^{-3} in the sporulated culture. Since this result indicates that most of Lys$^+$ cells in the sporulated culture were generated in meiosis, the rate of Lys$^+$ in the sporulated culture is the same as the frequency, 2.3×10^{-3}/meiotic division. The mitotic rate of Lys$^+$ cells was calculated by the method of the median as described above.

Analysis of the lengths of poly GT tracts: Two methods were used to measure the length of poly GT tracts in 5FOA^R cells. One procedure involved analyzing labeled PCR fragments that spanned the microsatellite (FARBER *et al.* 1994). Genomic DNA was isolated and PCR was performed with primers with sequences 5′GTTTTCCCAGTCACGAC3′ and 5′CCAATAGGTGGTTAGCAATCG3′, generating a DNA fragment of ~200 bp. The PCR reaction conditions were described previously by WIERDL *et al.* (1996). The PCR products were run on 6% sequencing gels with appropriate control samples. To confirm the length observed in PCR gels, we also rescued plasmids from yeast into *E. coli* (strain DH5α) and sequenced them using standard methods.

Statistical analysis: To determine whether the types of events were different in different strains, we compared numbers of events in each relevant class (*e.g.*, deletions and additions) using the Fisher exact test (computer program InStat 1.12).

RESULTS

Experimental design: To analyze the instability rates of poly GT tracts of different lengths in *Saccharomyces cerevisiae*, we used the plasmid frameshift assay described previously (HENDERSON and PETES 1992). The plasmids used in this assay contain a centromere, *TRP1*, and a fusion gene in which the *LEU2* promoter controls synthesis of a hybrid protein with wild-type *URA3* activity. Poly GT tracts with in-frame insertions of 15, 33, 51, 99 or 105 bp within the coding sequence were constructed. Strains containing these plasmids are Ura$^+$. Ura$^-$ derivatives selected with 5FOA usually contain plasmids in which the poly GT tract has altered in length, generating an out-of-frame insertion. We measured the rates of appearance of 5FOA resistance in strains with poly GT tracts of different lengths. We also analyzed the lengths of poly GT tracts in 5FOA^R derivatives of each strain. These experiments allowed us to determine how the rate of alterations and the types of alterations varied as a function of microsatellite length.

Tract length instability and types of tract alterations in wild-type yeast strains: Using methods described in MATERIALS AND METHODS, we measured the rate of tract instability in wild-type strains with poly GT tracts of 15, 33, 51, 99 and 105 bp (Table 1). In Figure 2A we show the rate of instability as a function of tract length. We found that the 15-bp GT tract was extremely stable. Most of the plasmids sequenced in 5FOA^R strains derived from PD3/p15 did not have altered tracts (Table 1). With increasing tract length, the rate of instability increased rapidly (Figure 2A, thick lines). Although the 105-bp tract is only seven times larger than the 15-bp tract, microsatellite instability is 500-fold greater for the larger tract. Although we were unable to fit the observed curve of Figure 2A to either a simple linear or exponential function, the increase in rate is clearly greater than linear. The rate of instability per repeat for the 15-bp tract is 4.2×10^{-8} ($3.2 \times 10^{-7}/7.5$), but 3.2×10^{-6} ($1.7 \times 10^{-4}/52.5$) for the 105-bp tract.

Using PCR or DNA sequencing, we analyzed the length of tract alterations in ≥25 5FOA^R derivatives of each strain; these data are summarized in Table 1. Most of the alterations were either small changes in tract length (additions or deletions of one or two repeat units) or large deletions (deletions of more than two repeats). Only four large additions were observed. As expected, almost all (250/253) of the alterations caused loss of the reading frame required to make a functional *URA3* gene product. The exceptions were the 6-bp deletion in PD3/p15 and two 24-bp deletions in PD3/p51. In addition to the observed frameshift, these plasmids presumably contained a mutation elsewhere in the *URA3* gene coding sequence. Using the data of Table 1, we also calculated rates of instability of events involving additions or deletions of one repeat as a function of tract length. As shown in Figure 2B (thick lines), although the rate of additions of single repeats increases dramatically with tract length, the rate of deletions of single repeats for the 99-bp tract is less than for the 51-bp tract. The implications of this result will be discussed below.

The increase in repeat length had two effects on the spectra of tract alterations. First, there is a transition between tract sizes of 33 and 51 bp in the tendency for tracts to undergo large deletion events (Figure 3A). The 51-, 99- and 105-bp GT tracts had significantly (Fisher exact test P values between 0.0008 and 0.03) more large deletions compared to small (one or two repeat deletions or additions) alterations than the 33-bp tract. In addition, the single "large" deletion observed in the 33-bp GT repeat involved only four repeat units, compared to the deletions of eight or more repeats commonly observed with the 51- and 99-bp poly GT tracts. A second length-dependent property of tract alterations is the ratio of one repeat insertions to one repeat deletions. As shown in Table 1 and Figure 3A, for tracts of ≤51 bp, both additions and deletions of one repeat are common, with additions exceeding deletions. No deletions of single repeats were observed in the 99- and 105-bp in wild-type strains. The altered ratio of single repeat additions and deletions was significantly different (P values by Fisher exact test varying from <0.0001 to 0.003) for all comparisons of the 99- and 105-bp tracts with the 15-, 33- and 51-bp tracts.

The lack of 2-bp deletions is not likely to represent a detection problem for several reasons. First, 2-bp deletions from 99- and 105-bp repeats would be in the same reading frame as 2-bp deletions from the 15-, 33- and 51-bp repeats, which were readily detectable. Second, in experiments that will be discussed in more detail below, in 5FOA^R derivatives of MBW10/p99 (an *msh2* strain with a 99-bp poly GT tract), we found one strain with a 97-bp tract and several with 103-bp tracts.

Tract length instability and types of tract alterations in *msh2* yeast strains: Mutations in DNA mismatch repair genes significantly increase simple sequence insta-

TABLE 1

Rates of instability and types of alterations for poly GT tracts of different lengths in wild-type yeast strains

GT tract (bp)	Rate of tract instability[b]	No. of tracts with additions or deletions of base pairs[a]				Total
		−2	0	+2	Others	
15	3.2×10^{-7}	7	47	9	1 (−6)	64
33	5.9×10^{-6}	16	5	35	3 (−4, −8, +38)	59
51	2.1×10^{-5}	39	1	54	28 (2×−4, 10×−16, 4×−20, 5×−22, 2×−24, 4×−32, +20)	122
99	1.5×10^{-4}			29	7 (−16, −40, −50, −52, −62, 2×+14)	36
105	1.7×10^{-4}			20	5 (−4, −16, −22, −28, −52)	25

[a] From independent 5FOA^R colonies, we analyzed the poly GT tract length using either a PCR procedure (described in MATERIALS AND METHODS) or by DNA sequence analysis of rescued plasmids. All tracts with deletions or additions >4 bp were examined by DNA sequencing. The × designation in the Others column signifies multiple independent tracts of the same size.

[b] From measuring the frequencies of 5FOA^R derivatives in multiple independent cultures and using the method of the median (LEA and COULSON 1949), we calculated the rate of the appearance of 5FOA^R colonies. We multiplied this rate by the fraction of the 5FOA^R derivatives that had altered tracts in order to obtain the rate of tract instability.

bility in prokaryotes and eukaryotes (LEVINSON and GUTMAN 1987; STRAND *et al.* 1993, 1995; reviewed by KOLODNER 1996). A mutation in the *MutS* homologue *MSH2* results in destabilization of a 33-bp poly GT tract by ~100- to 200-fold (STRAND *et al.* 1993, 1995; JOHNSON *et al.* 1996). In addition, in *msh2* strains the ratio of deletions to additions for single repeats was higher than observed in wild-type strains.

In wild-type strains, both the rate of instability and the types of alterations observed are affected by the length of the microsatellite. As discussed in the Introduction, in wild-type strains the rate of instability can be affected by the rate of DNA polymerase slippage (which results in DNA mismatches) and by the efficiency with which these DNA mismatches are corrected. To determine whether the effects of tract length on stability and the types of alterations in wild-type strains reflected length-dependent changes in the properties of DNA polymerase slippage or length-dependent changes in the properties of DNA mismatch repair (or both), we examined tract length effects in strains with an *msh2* mutation. These data are shown in Table 2 and

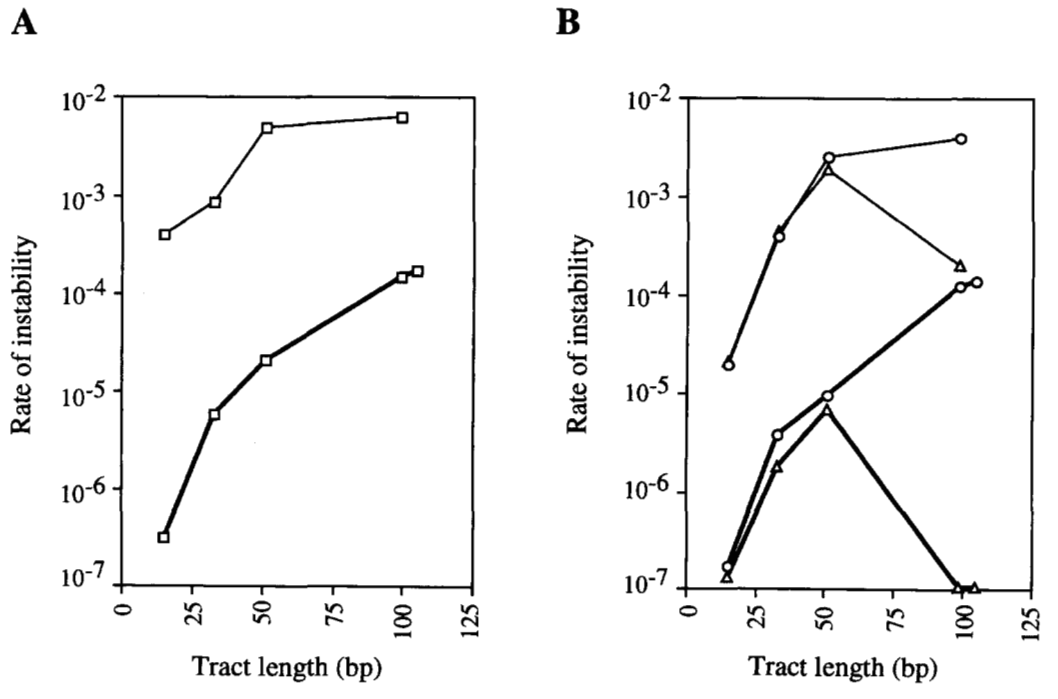

FIGURE 2.—Rates of length alterations in poly GT tracts of different sizes in wild-type and *msh2* strains. These diagrams are based on data in Tables 1 and 2. (A) Rates of length alterations including all sizes of additions and deletions in wild-type (thick lines) and *msh2* (thin lines) strains. (B) Rates of length alterations involving one repeat additions (○) and one repeat deletions (△) in wild-type (thick lines) and *msh2* (thin lines) strains.

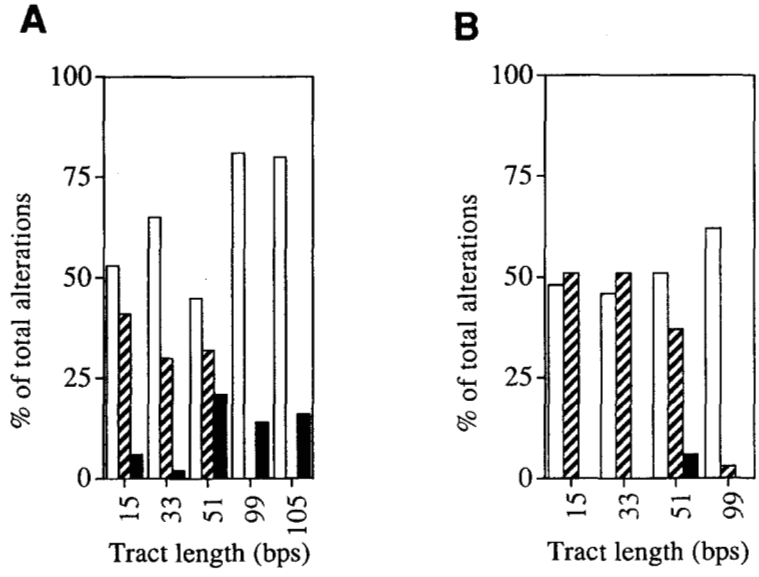

FIGURE 3.—Classes of length alterations (excluding the small number of large additions) found with poly GT tracts of different lengths in wild-type and *msh2* strains. (A) Tract alterations observed in wild-type strains. The white, cross-hatched and black rectangles represent one repeat additions, one repeat deletions and large (greater than two repeat) deletions, respectively. (B) Tract alterations observed in *msh2* strains. The same key is used as in A.

Figure 2. As observed in wild-type strains, tract instability increases dramatically with tract length in *msh2* strains (Figure 2A, thin lines). This result suggests that the increase in tract instability with increasing tract length primarily reflects length-dependent effects on DNA polymerase slippage rather than length-dependent effects on DNA mismatch repair.

The sizes of tract alterations in the mismatch repair-deficient *msh2* strains were also examined (Table 2). The spectra of alterations observed in *msh2* strains were different from those observed in wild-type strains for all poly GT tracts (P values between <0.001 and 0.05). For the 15-bp tract, 47 of 64 sequenced tracts had no alterations in the wild-type strain, whereas only 7 of 40 tracts had no alterations in the *msh2* strain, a significant ($P = 0.0001$) difference. For the 33-bp tract, there were significantly ($P = 0.05$) more deletions of single repeats compared to additions of single repeats in the *msh2* strain than in the wild-type strain; a similar effect was noted previously for the 33-bp tract (STRAND *et al.* 1995). For the 51- and 99-bp tracts, the number of large deletions relative to other alterations was significantly (P values of 0.006 and 0.04, respectively) reduced in the *msh2* strains compared to the wild-type strains. As observed with the 99-bp tract in the wild-type strain, the 99-bp tract in the *msh2* strain underwent many more insertions of single repeats than deletions of single repeats. This difference was significant for comparisons of the 99-bp tract with the 15-, 33- and 51-bp tracts (P values between <0.0001 to 0.0015). In Figure 2B (thin lines), we show the rates of one-repeat additions and deletions in *msh2* strains as a function of tract length (calculated from the data of Table 2). As observed in wild-type strains the rate of additions increases with tract length, whereas the rate of deletions is not directly related to tract length.

The data in Tables 1 and 2 can be used to estimate the efficiency of the DNA mismatch repair system in correction of mismatches caused by DNA polymerase slippage. If the rate of instability in wild-type strains is denoted R_{WT} and the rate in the *msh2* strain is denoted R_{M2}, then the *in vivo* efficiency of error correction (EC) of the Msh2p-dependent mismatch repair system (expressed as a percentage) can be approximated by the following equation: $EC = (R_{M2} - R_{WT})/R_{M2} \times 100$ (SIA *et al.* 1997). When this calculation is performed for strains with poly GT tracts of different sizes (Table 3) EC is >99% for all tracts except for the 99-bp poly

TABLE 2

Rates of instability and types of alterations for poly GT tracts of different lengths in *msh2* yeast strains[a]

		No. of tracts with additions or deletions of base pairs				
GT tract (bp)	Rate of tract instability	−2	0	+2	Others	Total
15	4.0×10^{-5}	17	7	16		40
33	8.5×10^{-4}	19		17	1 (+8)	37
51	4.9×10^{-3}	13		18	4 (−4, +4, −10, −26)	35
99	6.3×10^{-3}	1	5	23	13 (6×−4, 6×+4, +10)	42

[a] The methods used in obtaining data for this table are the same as specified in Table 1.

TABLE 3

The percentage efficiency of correction of DNA mismatches dependent on Msh2p[a]

Tract length (bp)	EC (all changes) (%)	EC (2-bp deletions) (%)	EC (2-bp additions) (%)
15	99.2	99.4	99.1
33	99.3	99.6	99.0
51	99.6	99.6	99.6
99	97.6	100	96.9

EC, efficiency of correction.

[a] As discussed in the text, from measurements of the rates of instability of poly GT tracts in wild-type and *msh2* yeast strains, one can obtain an estimate of the minimal efficiency of correction of DNA mismatches by the Msh2p-dependent system. If the rates of instability are R_{WT} and R_{M2} in wild-type and *msh2* strains, respectively, we calculate EC = $(R_{M2} - R_{WT})/R_{M2} \times 100$.

GT tract. In addition, the EC values can be calculated separately for additions and deletions of single repeats (Table 3). These calculations indicate that EC is lower for events resulting in an addition in the 99-bp poly GT tract.

The relationship between recombination and large alterations in microsatellite length: As discussed in the Introduction, deletions or additions of single repeats are likely to be a consequence of DNA polymerase slippage. It is less clear that large deletions or additions reflect the same process. An alternative possibility is that large alterations represent recombination events. As shown in Figure 4A unequal crossovers between misaligned tandem arrays could result in large deletions or additions of repeat units. An alternative model (Figure 4B) is loss of repeats by single-strand annealing. This mechanism, observed in a variety of eukaryotic systems (reviewed by HABER 1992; KLEIN 1995), involves a double-strand DNA break within the repetitive tract, excision of single strands of the broken ends, followed by reannealing and repair. In addition to the loss of repeats diagrammed in Figure 4B, repair of the broken ends from a different poly GT tract could lead to an increase in tract length.

We investigated the role of recombination in the generation of large deletions in two ways. First, we examined whether the rate of large microsatellite alterations was elevated in meiosis. In *S. cerevisiae,* most recombination events are elevated about three orders of magnitude in meiosis compared to mitosis (PETES *et al.* 1991). Previously, STRAND *et al.* (1993) showed that the meiotic and mitotic stabilities of a 29-bp poly GT microsatellite were approximately the same. In these experiments, since the poly GT tract was small, few of the alterations involved large deletions. Consequently, we reexamined the effects of meiosis using a diploid strain (MBW20/p99) containing a plasmid with a 99-bp poly GT tract. Since this strain also had two different mutant *lys2* alleles, we monitored meiotic recombination by measuring the frequency of Lys^+ derivatives in sporulating cells. We found that the rates of mitotic and meiotic instability of the 99-bp tract were very similar, 1.4×10^{-4} and 1.7×10^{-4}, respectively. In contrast, the mitotic and meiotic rates of Lys^+ derivatives were different by three orders of magnitude: 2.7×10^{-6} and 2×10^{-3}, respectively. We also analyzed the types of tract alterations in plasmids derived from mitotic and meiotic cells. Of 19 plasmids derived from $5FOA^R$ mitotic cells, 14 had additions of single repeats and five had large deletions (8, 38, 40, 46 and 56 bp); these types of alterations are similar to those observed in wild-type haploid strains for 99-bp tracts (Table 1). Of 18 plasmids derived from $5FOA^R$ meiotic cells, we found 16 with additions of single repeats and two with large deletions (38 and 52 bp). Although the sample sizes were not large, these results indicate that the frequency of large deletions is not substantially elevated in meiosis.

In *S. cerevisiae,* most (although not all) types of recombination require the *RAD52* gene product (PETES *et al.* 1991). Previously, we showed that *rad52* strains had approximately the same rate of instability for a 33-bp poly GT tract as wild-type strains (HENDERSON and PETES 1992). To determine whether *rad52* affected the rate of large deletions, we examined the stability of a 51-bp poly GT tract in isogenic wild-type (PD3/p51) and *rad52* mutant (SH52*/p51) strains. The rates of instability were approximately the same in the two strains, 2.1×10^{-5} for PD3/p51 and 2.7×10^{-5} for SH52*/p51. As shown in Table 1, 26 of 122 plasmids rescued from $5FOA^R$ derivatives of PD3/p51 had large deletions (25 tracts) or additions (one tract). In 57 plasmids rescued from $5FOA^R$ derivatives of SH52*/p51, we found the following classes: single repeat additions (36), single repeat deletions (12), no change (2), two repeat deletions (2) and large deletions (5; 3 × 16, 20, 26 bp). The numbers of large alterations compared to small alterations for the *RAD52* and *rad52* strains were not significantly different ($P = 0.06$). From these data, it is clear that large deletions in microsatellites can occur in the absence of the *RAD52* gene product. In summary, our results indicate that either large alterations in microsatellite length do not occur as a consequence of

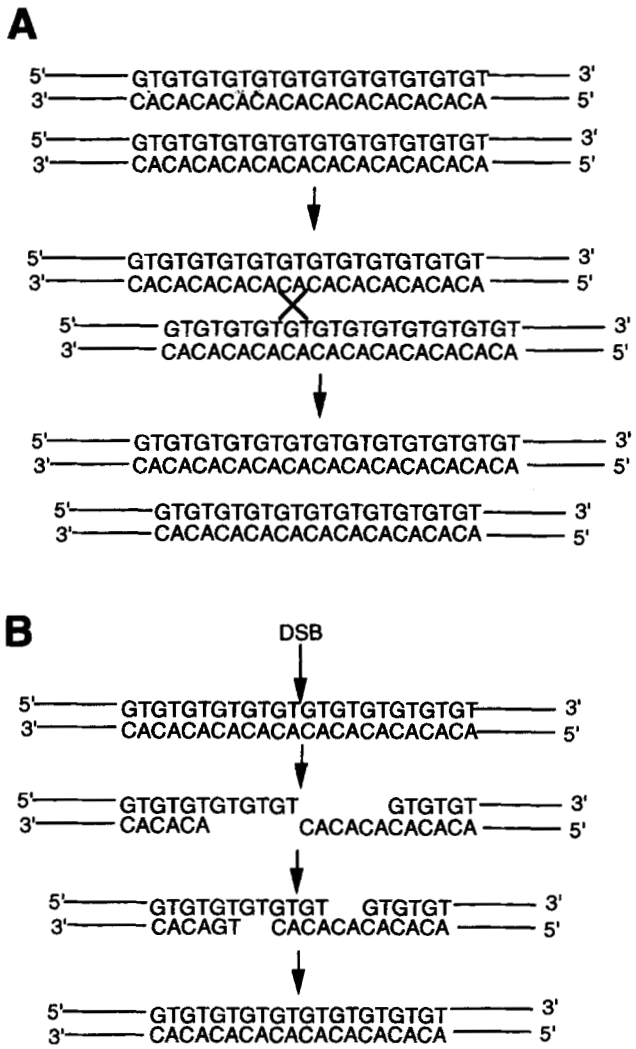

FIGURE 4.—Alteration of poly GT tract length by recombination. (A) Alteration in tract length by unequal crossing-over (SMITH 1973). Repetitive tracts on different DNA molecules pair in a misaligned configuration and crossover, generating one shorter and one longer tract. (B) Alteration in tract length by single-strand annealing model (HABER 1992; KLEIN 1995). In this model, a double-strand DNA break (DSB) results in DNA ends that are processed to yield 3′ protruding strands. Reannealing of these strands, followed by DNA repair synthesis, results in a tract with a deletion.

recombination or these alterations occur by a recombination process that is not elevated in meiosis and is not affected by the *RAD52* gene product.

DISCUSSION

The results of these studies can be summarized by the following statements: (1) the rate of microsatellite instability for the poly GT tract increases with increasing tract length in both wild-type and *msh2* strains, (2) the rate of single repeat additions, but not the rate of single repeat deletions, increases continuously as tracts increase in size in both wild-type and *msh2* strains, (3) there are significantly more large (greater than two repeat) deletions in tracts ≥51 bp in length than in tracts ≤33 bp, (4) the rate of large deletions is not elevated in meiotic cells and is not significantly reduced in cells with a *rad52* mutation, (5) the fraction of alterations representing large deletions is reduced in *msh2* strains and (6) in wild-type strains, deletions involving eight repeats or more are more common than those involving two to seven repeats. Each of these findings will be discussed below.

In vitro studies indicate that DNA polymerase slippage increases as repetitive tracts in the substrate become longer (KUNKEL 1985; BEBENEK and KUNKEL *et al.* 1990; KROUTIL *et al.* 1996). In our *in vivo* studies in *S. cerevisiae*, we find that the instability of poly GT tracts increases with increasing tract length, consistent with previous *in vivo* observations of tract instability in *E. coli* (STREISINGER and OWEN 1985; LEVINSON and GUTMAN 1987; FREUND *et al.* 1989; BICHARA *et al.* 1995; KANG *et al.* 1995), in yeast (TRAN *et al.* 1997) and in humans (WEBER 1990; LEEFLANG *et al.* 1995).

If microsatellite instability reflects DNA polymerase slippage events, then the rate of tract changes will be affected by the following parameters: the rate of DNA strand dissociation during DNA replication, the probability of reassociation in a misaligned configuration, the frequency of correction of displaced loops by the DNA proofreading exonuclease and the frequency of correction of displaced loops by the DNA mismatch repair system. Since we observe an increase in tract instability with increasing tract length in both wild-type and *msh2* strains, our results cannot be explained solely as a consequence of the diminished efficiency of DNA mismatch repair with increasing tract length. Of the other parameters, the two most likely factors leading to elevated tract instability in longer tracts are an increase in the probability of reassociation of DNA strands in a misaligned configuration and a decrease in the efficiency of the proofreading exonuclease. The number of potential misaligned configurations in which one repeat is unpaired would be expected to increase as a linear function of tract length. We found previously that a 33-bp poly GT tract was destabilized ~10-fold by a mutation in the proofreading exonuclease domain of DNA polymerase δ, whereas mutations in the DNA mismatch repair genes had a much larger effect (STRAND *et al.* 1993). We suggested that mismatched repeats resulting from misaligned reannealing events might be relatively immune to surveillance by the proofreading exonuclease, since these repeats would be expected to be displaced from the end of the elongating DNA strand. Subsequent *in vitro* (KROUTIL *et al.* 1996) and *in vivo* (TRAN *et al.* 1997) studies showed that the destabilizing effect of proofreading exonuclease, relative to the DNA mismatch repair mutations, was much greater for short tracts than long tracts. Thus, two ef-

fects may be responsible for the greater-than-linear increase in tract instability as a function of tract length in wild-type strains: (1) a linear increase in misalignments as a function of tract length and (2) relatively inefficient removal of mismatched repeats from the long tracts by the DNA polymerase proofreading exonuclease.

This simple model, however, fails to explain the observation that changes in tract length involving single repeats in long tracts are strongly biased toward additions (Figure 3). This result indicates that DNA polymerase slippage events in long poly GT tracts have a length-dependent asymmetry in the behavior of primer and template strands. The mismatched repeats in long poly GT tracts, but not short tracts, almost always involve the elongating DNA strand rather than the template strand (Figure 1). The reason for this preference is not clear but it may involve a length-dependent property of the single-stranded DNA sequence of the microsatellite (poly GT and poly CA). In *E. coli,* the types of tract alterations in trinucleotide microsatellites are influenced by the sequence of the repeat and the direction of replication of the tract (KANG *et al.* 1995); deletion events involving intermediates with extensive single-stranded regions occur preferentially on the lagging strand in both *E. coli* and yeast (TRINH and SINDEN 1991; TRAN *et al.* 1995).

An alternative possibility is that the events that lead to addition of single repeats in the long poly GT tracts occur by a mechanism different from DNA polymerase slippage during replicative DNA synthesis. One such mechanism is nick-directed slippage. Some acridine-induced frameshift mutations in T4 appear to involve processing of nicks generated by topoisomerase (reviewed by RIPLEY 1990). After addition or deletion of bases from the 3′ end of the nicked strand, the strand is religated, leading to an insertion of a base or deletion of a base adjacent to the nick. The increased frequency of single repeat additions in long tracts is consistent with a model in which long poly GT tracts are nicked more frequently than short tracts and in which these nicks are usually processed by addition of single repeats adjacent to the nick. A final possibility is an *MSH2*-independent DNA repair system that either corrects mismatches on the elongating strand and loses repair efficiency as tracts get longer or corrects mismatches on the template strand with increased efficiency as tracts get longer.

We found that large deletions increase as a function of tract length in wild-type strains (Figure 3A). Similar effects have been observed in *E. coli* (FREUND *et al.* 1989; KANG *et al.* 1995). Large deletions could result from DNA polymerase slippage events involving the formation of large loops on the template strand or recombination events such as those shown in Figure 4. Although the evidence is not definitive, we favor the first explanation for several reasons. First, a recombination model involving reciprocal unequal crossovers (Figure 4A) predicts equal recovery of large additions and large deletions, and we find a strong preference for deletions. Second, *rad52* reduces or eliminates most types of recombination, including both mechanisms shown in Figure 4. In most assays involving direct repeats in yeast, recombination is reduced ~10-fold by *rad52* (reviewed by KLEIN 1995). We observe no significant effect of *rad52* on the rate of formation of large deletions within the poly GT tract. In addition, the frequency of large deletions within the poly GT tract is not increased in meiotic cells.

The increase in large deletions in long microsatellites observed in *E. coli* (FREUND *et al.* 1989; KANG *et al.* 1995) was explained as reflecting alterations in the conformation of DNA as the length of the tracts increased. From studies of the dependence of large deletions on the direction of replication of a trinucleotide microsatellite, KANG *et al.* (1995) suggested that single-stranded poly CTG DNA on the lagging template strand adopted a duplex structure resulting in large deletions by DNA polymerase slippage; they hypothesized that the complementary poly CAG sequence formed a less stable secondary structure. In addition, in yeast, long poly CAG tracts have shown orientation-dependent effects on the rate of large deletions (MAURER *et al.* 1996). Previously, we showed that the orientation of a 33-bp poly GT tract had no effect on the rate of alterations or types of changes (HENDERSON and PETES 1992). Since 33-bp tracts have low rates of large deletions, however, these results do not rule the possiblity that large poly GT tracts might be subject to orientation-specific effects on the rate of large contractions.

The fraction of large deletions relative to small additions and deletions is reduced in *msh2* strains (Figure 3B). This result suggests that mismatches involving small loops (2–4 bp) are efficiently recognized by the mismatch repair system, but large loops are not. In a study of the effects of *msh2* on the repair of microsatellites with various repeat lengths, we found that loops of 1–8 bp were efficiently repaired, whereas loops of ≥16 bp were not (SIA *et al.* 1997). These conclusions fit well with the size distribution of large deletions observed in our studies (Figure 5). In wild-type strains, we find very few deletions between 4 and 14 bp. The rarity of this class of deletions probably reflects two factors: (1) the relative rarity of large slippage events compared to small slippage events and (2) efficient repair for loops < 16 bp.

In humans, the frequency of large expansions of trinucleotide repeats increases with increasing tract size (ASHLEY and WARREN 1995). From PCR analysis of single sperm with a $(CAG)_{51}$ tract near the Huntington's disease gene, LEEFLANG *et al.* (1995) found that 97% of the tracts had expansions with a mean expansion size

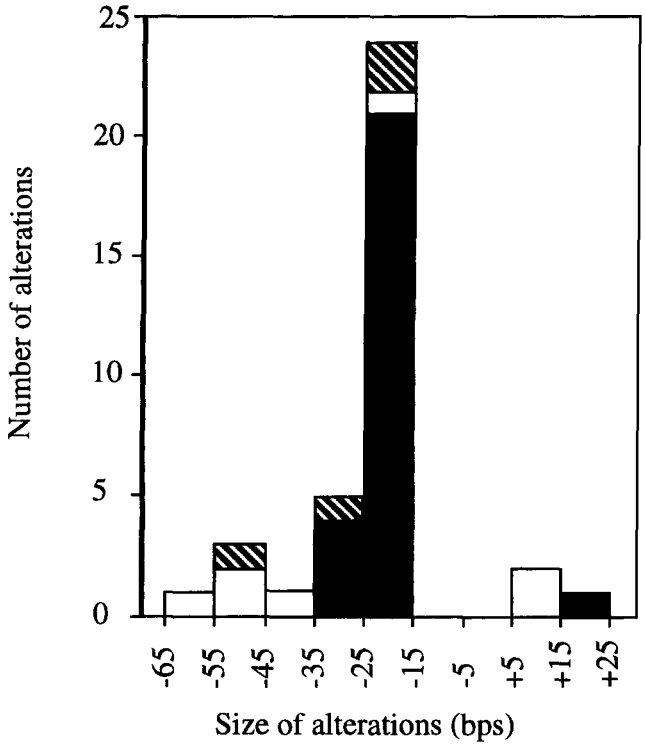

FIGURE 5.—Size distributions of large (>4 bp) alterations in wild-type strains with 51-, 99- or 105-bp poly GT tracts. This figure is based on data of Table 1. The black, white and cross-hatched rectangles represent deletions from 51-, 99- and 105-bp tracts, respectively.

of 21 repeats. It is not clear whether the low frequency of large additions in our studies reflects the type of microsatellite studied, the total length of the microsatellite, species-specific differences or other factors.

In summary, we find that the rate of instability for the poly GT microsatellite increases as the length of the repetitive tract increases. In addition, the types of changes observed in repetitive tracts alter with increasing tract length. It is not clear whether these effects reflect length-dependent changes in the types of DNA polymerase slippage events or new mechanisms for altering tract length. The mechanisms responsible for these events may be clarified by the isolation and characterization of additional mutants that affect the rate of microsatellite instability.

We thank D. KIRKPATRICK, R. KOKOSKA and E. SIA for comments on the manuscript and L. RIPLEY for suggesting alternative mechanisms of frameshift mutagenesis. The research was supported by National Institutes of Health grant GM-52319.

LITERATURE CITED

AHN, B.-Y., K. J. DORNFELD, T. J. FAGRELIUS and D. M. LIVINGSTON, 1988 Effect of limited homology in a *Saccharomyces cerevisiae* plasmid recombination system. Mol. Cell. Biol. **8:** 2442–2448.

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

BEBENEK, K., and T. A. KUNKEL, 1990 Frameshift errors initiated by nucleotide misincorporation. Proc. Natl. Acad. Sci. USA **87:** 4946–4950.

BICHARA, M., S. SCHUMACHER and R. P. P. FUCHS, 1995 Genetic instability within monotonous runs of CpG sequences in *E. coli*. Genetics **140:** 897–907.

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.

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.

DE LA CHAPELLE, A., and P. PELTOMAKI, 1995 Genetics of hereditary colon cancer. Annu. Rev. Genet. **29:** 329–348.

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.

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.

HABER, J. E., 1992 Exploring the pathways of homologous recombination. Curr. Opin. Cell Biol. **4:** 401–412.

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

JOHNSON, R. E., G. K. KOVVALI, L. PRAKASH and S. PRAKASH, 1996 Requirement of the yeast *MSH3* and *MSH6* genes for *MSH2*-dependent genomic stability. J. Biol. Chem. **271:** 7285–7288.

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*. Nature Genet. **10:** 213–218.

KLEIN, H. L., 1995 Genetic control of intrachromosomal recombination. BioEssays **17:** 147–159.

KOLODNER, R., 1996 Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev. **10:** 1433–1442.

KROUTIL, L. C., K. REGISTER, K. BEBENEK and T. A. KUNKEL, 1996 Exonucleotic proofreading during replication of repetitive DNA. Biochemistry **35:** 1046–1053.

KUNKEL, T. A., 1985 The mutational specificity of DNA polymerase-beta during *in vitro* DNA synthesis. Production of frameshift, base substitution, and deletion mutations. J. Biol. Chem. **260:** 5787–5796.

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

LEEFLANG, E. P., L. ZHANG, S. TAVARE, R. HUBERT, J. SRINIDHI *et al.*, 1995 Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. Hum. Mol. Genet. **4:** 1519–1526.

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.

MAURER, D. J., B. L. O'CALLAGHAN and D. M. LIVINGSTON, 1996 Orientation dependence of trinucleotide CAG repeat instability in yeast. Mol. Cell. Biol. **16:** 6617–6622.

NAG, D. N., and T. D. PETES, 1990 Genetic evidence for preferential strand transfer during meiotic recombination in yeast. Genetics **125:** 753–761.

PETES, T. D., R. E. MALONE and L. S. SYMINGTON, 1991 Recombination in yeast, pp. 407–521 in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, Vol. 1, edited by J. BROACH, E. W. JONES and J. R. PRINGLE. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

REENAN, A. G., and R. D. KOLODNER, 1992 Characterization of insertion mutations in the *Saccharomyces cerevisiae MSH1* and *MSH2* genes: evidence for separate mitochondrial and nuclear functions. Genetics **132:** 975–985.

RIPLEY, L. S., 1990 Frameshift mutation: determinants of specificity. Annu. Rev. Genet. **24:** 189–213.

SHERMAN, F., 1991 Getting started with yeast. Methods Enzymol. **194:** 3–21.

SIA, E., R. KOKOSKA, M. DOMINSKA, P. GREENWELL and T. D. PETES,

1997 Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell. Biol. **17:** 2851–2858.

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., and J. E. OWEN, 1985 Mechanisms of spontaneous and induced frameshift mutations in bacteriophage T4. Genetics **109:** 633–659.

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.

TRAN, H. T., N. P. DEGTYAREVA, N. N. KOLOTEVA, A. SUGINO, H. MASUMOTO *et al.*, 1995 Replication slippage between distant short repeats in *Saccharomyces cerevisiae* depends on the direction of replication and the *RAD50* and *RAD52* genes. Mol. Cell. Biol. **15:** 5607–5617.

TRAN, H. T., J. D. KEEN, M. KRICKER, M. A. RESNICK and D. A. GORDENIN, 1997 Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. Mol. Cell. Biol. **17:** 2859–2865.

TRINH, T. Q., and R. R. SINDEN, 1991 Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli.* Nature **352:** 544–548.

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

WIERDL, M., C. N. GREENE, A. DATTA, S. JINKS-ROBERTSON and T. D. PETES, 1996 Destabilization of simple repetitive DNA sequences by transcription in yeast. Genetics **143:** 713–721.

Communicating editor: M. JOHNSTON