Copyright © 2000 by the Genetics Society of America

Sequence Composition and Context Effects on the Generation and Repair of Frameshift Intermediates in Mononucleotide Runs in *Saccharomyces cerevisiae*

Brian D. Harfe and Sue Jinks-Robertson

Department of Biology, Emory University, Atlanta, Georgia 30322

Manuscript received January 19, 2000
Accepted for publication June 19, 2000

ABSTRACT

DNA polymerase slippage occurs frequently in tracts of a tandemly repeated nucleotide, and such slippage events can be genetically detected as frameshift mutations. In long mononucleotide runs, most frameshift intermediates are repaired by the postreplicative mismatch repair (MMR) machinery, rather than by the exonucleolytic proofreading activity of DNA polymerase. Although mononucleotide runs are hotspots for polymerase slippage events, it is not known whether the composition of a run and the surrounding context affect the frequency of slippage or the efficiency of MMR. To address these issues, 10-nucleotide (10N) runs were inserted into the yeast *LYS2* gene to create +1 frameshift alleles. Slippage events within these runs were detected as Lys^+ revertants. 10G or 10C runs were found to be more unstable than 10A or 10T runs, but neither the frequency of polymerase slippage nor the overall efficiency of MMR was greatly influenced by sequence context. Although complete elimination of MMR activity (*msh2* mutants) affected all runs similarly, analyses of reversion rates in *msh3* and *msh6* mutants revealed distinct specificities of the yeast Msh2p-Msh3p and Msh2p-Msh6p mismatch binding complexes in the repair of frameshift intermediates in different sequence contexts.

TRACTS of simple repetitive DNA are hotspots for frameshift mutations, and this behavior has been attributed to the propensity of DNA polymerase to "slip" in repetitive sequences during DNA replication (Streisinger *et al.* 1966; Streisinger and Owen 1985). Slippage is believed to occur by the transient dissociation of the nascent and template strands during DNA synthesis, followed by out-of-register annealing of the two strands and subsequent extension of the misaligned nascent strand. The unpaired nucleotide(s) in slippage intermediates can be repaired either by the exonucleolytic proofreading activity of DNA polymerase or by the postreplicative mismatch repair (MMR) machinery. If a slippage event is not repaired before the next round of DNA replication, it will give rise to a deletion if the unpaired nucleotide is on the template strand or to an insertion if the unpaired nucleotide is on the nascent strand.

In mononucleotide runs, the frequency of DNA polymerase slippage is positively correlated with the length of the run (Kroutil *et al.* 1996; Tran *et al.* 1997), but there is little information about how the sequence of a run (run composition) affects polymerase slippage. In the *Escherichia coli lacZ* gene, loss of a C from an 8C run occurred at a 10-fold higher rate than loss of an A from an 8A run (Sagher *et al.* 1999). Although this suggests that G or C runs may be inherently more unstable than A or T runs, it should be noted that the runs were not at identical positions within the *lacZ* gene. Previous studies in yeast have examined slippage within an 18G run (Sia *et al.* 1997) or within poly(A) tracts varying in length from 4 to 14 nucleotides (Tran *et al.* 1997). Given the different strain backgrounds and the different assay systems used in these studies, however, it has not been possible to deduce the effect that run composition may have on polymerase slippage in yeast.

Although the frequency of DNA polymerase slippage is positively correlated with mononucleotide run length, the efficiency of removing the corresponding frameshift intermediate via DNA polymerase proofreading activity is inversely related to the length of the run (Kroutil *et al.* 1996; Tran *et al.* 1997). As run length increases, it is thought that the extrahelical loop formed when the nascent and template strands reanneal out of register can migrate far enough away from the elongating polymerase to escape detection. In contrast to the inverse length dependence exhibited by proofreading, correction of DNA slippage events by the yeast MMR system appears to become more efficient as run length increases (Tran *et al.* 1997).

The yeast MMR system can be completely inactivated by eliminating Msh2p, the central homolog of the *E. coli* MutS mismatch-binding protein. Msh2p functions in heterodimeric complexes with the MutS homologs Msh3p and Msh6p (reviewed in Crouse 1998; Kolodner and Marsischky 1999), and both complexes are

Corresponding author: Sue Jinks-Robertson, Department of Biology, 1510 Clifton Rd., Emory University, Atlanta, GA 30322.
E-mail: jinks@biology.emory.edu

capable of initiating the repair of +1 and −1 frameshift intermediates (Marsischky *et al.* 1996; Greene and Jinks-Robertson 1997; Sia *et al.* 1997; Harfe and Jinks-Robertson 1999). Although examination of frameshift spectra derived for nonrepetitive DNA indicates that the Msh2p-Msh3p and Msh2p-Msh6p complexes have distinct recognition specificities (Greene and Jinks-Robertson 1997; Flores-Rozas and Kolodner 1998; Harfe and Jinks-Robertson 1999), the recognition specificities of these complexes for frameshift intermediates in mononucleotide runs of different composition have not been examined. For an 18G run, elimination of either Msh3p or Msh6p produced a modest increase in frameshift events (130-fold and 30-fold, respectively), whereas there was a strong, 5000-fold increase when both were eliminated (Sia *et al.* 1997). Such synergism indicates competing repair specificities for the Msh2p-Msh3p and Msh2p-Msh6p mismatch-binding complexes, with one complex being able to compensate for loss of the other complex. In addition to the MutS homologs, the repair of frameshift intermediates also requires the yeast Pms1p and Mlh1p proteins, which form a heterodimeric complex and are homologs of the *E. coli* MutL protein (Prolla *et al.* 1994a,b). Mlh1p also can form heterodimeric complexes with Mlh2p and Mlh3p (Wang *et al.* 1999), but these complexes have only minor roles in correcting frameshift intermediates (Flores-Rozas and Kolodner 1998; Harfe *et al.* 2000).

In addition to considering the composition of a mononucleotide run, the sequences flanking a run can potentially affect the frequency of polymerase slippage. This would be the case, for example, if a base misincorporation preceded the slippage event, with slippage restoring proper base pairing of the 3′ end of the nascent strand with the template (see Bebenek and Kunkel 1990). In addition to context effects on polymerase slippage, sequence context might also impact mismatch recognition by the Msh2p-Msh3p or Msh2p-Msh6p complex. For the human MSH2-MSH6 complex, binding to a frameshift intermediate *in vitro* was influenced by the immediate flanking sequence (Macpherson *et al.* 1998). Similarly, *in vitro* binding of the yeast Msh2p-Msh6p complex to base substitution intermediates was impacted by sequence context (Marsischky and Kolodner 1999). To date, *in vivo* studies have not addressed the impact of sequence context either on polymerase slippage or on the repair of frameshift intermediates in mononucleotide runs.

To address the role of sequence composition and context on mononucleotide run stability, we constructed yeast strains in which runs of 10 guanines, 10 cytosines, 10 thymines, or 10 adenines (10G, 10C, 10T, or 10A runs, respectively) were engineered at a defined position in the *LYS2* gene. The resulting +1 frameshift alleles reverted predominantly by loss of a single nucleotide from the 10N runs, thus allowing tract stability to be assessed by measuring the rate of Lys^+ revertants. In addition to varying the run composition, the contexts of the runs were altered either by changing the identities of flanking nucleotides or by changing the position of the runs. The impact of mononucleotide run composition and context on the rate of DNA polymerase slippage and the efficiency of MMR were determined by measuring reversion rates in wild-type cells and in cells defective in various components of the yeast MMR machinery.

MATERIALS AND METHODS

Media and growth conditions: Yeast strains were grown nonselectively in YEP medium (1% yeast extract, 2% Bacto-peptone; 2.5% agar for plates) supplemented with either 2% dextrose (YEPD) or 2% glycerol/2% ethanol (YEPGE). Synthetic complete (SC) medium (Sherman 1991) containing 2% dextrose and lacking the appropriate amino acid was used for selective growth. Lys^- segregants were selected on minimal medium containing α-aminoadipate as the primary nitrogen source (Chattoo *et al.* 1979). All yeast strains were grown at 30°.

Yeast strain constructions: All strains used were isogenic derivatives of SJR195 (*MAT*α *ade2-101 his3*Δ*200 ura3*Δ*Nco*; Harfe and Jinks-Robertson 1999). *MSH2*, *MSH3*, *MSH6*, *MLH1*, and *PMS1* genes were disrupted as described previously (Greene and Jinks-Robertson 1997). 10N runs were introduced into the *LYS2* locus using a two-step allele replacement procedure (Rothstein 1991). In the first step, the relevant plasmid was linearized within the *lys2* sequence, and cells that had integrated the plasmid were selected as His^+ transformants. Plasmid excision events subsequently were identified by selecting Lys^- segregants on α-aminoadipate medium and then screening for simultaneous acquisition of a His^- phenotype.

Creation of mononucleotide runs in the *LYS2* gene: The 6A tract beginning at nucleotide 664 of the *LYS2* coding strand (nucleotides are numbered from the upstream *Xba*I site) was replaced with 10N mononucleotide runs using the Chameleon mutagenesis system (Stratagene, La Jolla, CA) and plasmid pSR531 (Harfe and Jinks-Robertson 1999) as the template. pSR531 is a nonreplicative plasmid containing *HIS3* as a selectable marker and an internal fragment of *LYS2*. The 10N runs at position 664 are referred to as "position 1" mutations and the resulting alleles will be referred to as *lys2-10N-1* alleles. The following oligonucleotides were used for the mutagenesis:

10C-1 run: 5′-GACGAGCTAGCTGCCCCCCCCCCTTCAAA GTTGCC

10G-1 run: 5′-GACGAGCTAGCTCGGGGGGGGGGTTCAA AGTTGCC

10T-1 run: 5′-GACGAGCTAGCTGTTTTTTTTTTGCCAAA GTTGCC

10A-1 run: 5′-GACGAGCTAGCTGAAAAAAAAAATTCAAA GTTGCC.

Plasmids containing *lys2-10N-1* mutations are as follows: pSR607 (*lys2-10A-1*), pSR608 (*lys2-10G-1*), pSR609 (*lys2-10T-1*), and pSR619 (*lys2-10C-1*). The mutations were introduced into strain SJR195 to create yeast strains SJR938 (*lys2-10G-1*), SJR939 (*lys2-10A-1*), SJR980 (*lys2-10T-1*), and SJR981 (*lys2-10C-1*).

To place 10N runs at "position 2" (a unique *Pfl*MI site located at nucleotide 686), two complementary oligonucleotides containing *Pfl*MI compatible ends were annealed and directionally cloned into the *Pfl*MI site of pSR531. The re-

sulting alleles are referred to as *lys2-10N-2x* alleles, with "x" indicating flanking sequence differences between different position 2 alleles (see Figure 1). The following pairs of oligonucleotides were annealed for the mutagenesis:

10C-2a run: 5′-ACCCCCCCCCCACGAT and 5′-GTGGGGGGGGGGTATC
10G-2a run: 5′-AGGGGGGGGGGACGAT and 5′-GTCCCCCCCCCCTATC
10T-2a run: 5′-ATTTTTTTTTTACGAT and 5′-GTAAAAAAAAAATATC
10C-2b run: 5′-ACCCCCCCCCCTTGAT and 5′-AAGGGGGGGGGGTATC
10G-2b run: 5′-AGGGGGGGGGGTTGAT and 5′-AACCCCCCCCCCTATC
10C-2c run: 5′-GCCCCCCCCCCTTGAT and 5′-AAGGGGGGGGGGCATC
10G-2c run: 5′-CGGGGGGGGGGTTGAT and 5′-AACCCCCCCCCCGATC.

Plasmids containing *lys2-10N-2x* mutations are as follows: pSR646 (*lys2-10G-2a*), pSR647 (*lys2-10T-2a*), pSR648 (*lys2-10C-2a*), pSR649 (*lys2-10C-2b*), pSR650 (*lys2-10G-2b*), pSR651 (*lys2-10C-2c*), and pSR652 (*lys2-10G-2c*). The mutations were introduced into strain SJR195 by two-step replacement to create yeast strains SJR1047 (*lys2-10G-2a*), SJR1045 (*lys2-10T-2a*), SJR1046 (*lys2-10C-2a*), SJR1093 (*lys2-10G-2b*), SJR1092 (*lys2-10C-2b*), SJR1095 (*lys2-10G-2c*), and SJR1094 (*lys2-10C-2c*). The *lys2-10N-2a* alleles were used in a previous study to demonstrate MMR activity of the yeast Mlh2 and Mlh3 proteins (Harfe *et al.* 2000).

Mutation rates and spectra: Reversion rates to lysine prototrophy were determined by the method of the median (Lea and Coulson 1949) using data derived from 10–24 cultures of each strain. It should be noted that two independent isolates of each strain were used to generate these data. For the rate measurements, 5 ml of YEPGE medium were inoculated with single colonies from YEPD plates and the cultures were incubated for 2 days on a roller drum. Cells were harvested by centrifugation, washed with sterile H_2O, and resuspended in 1 ml of H_2O. A total of 100 μl aliquots of appropriate dilutions were plated onto SC-lys to identify Lys^+ revertants and on YEPD to determine viable cell numbers. Lys^+ colonies were counted 2 days after selective plating. The 95% confidence interval (CI) for each reversion rate was calculated using the 95% CI for the corresponding experimentally determined median (Dixon and Massey 1969).

To isolate independent Lys^+ revertants for DNA sequence analysis, YEPGE cultures were grown as above and plated on SC-lys. To ensure independence, only one revertant from each culture was purified for subsequent molecular analysis. Automated DNA sequence analysis of PCR-amplified genomic fragments was performed as described previously (Chen and Jinks-Robertson 1998) using primer 5′-CGCAACAATGGTTACTCT.

RESULTS

We and others have used the *LYS2* locus as an assay system for studying frameshift mutagenesis in yeast (Greene and Jinks-Robertson 1997; Tran *et al.* 1997; Harfe and Jinks-Robertson 1999). In general, these assays have monitored the reversion of *lys2* alleles that contain net +1 or −1 frameshift mutations. Compensatory frameshift mutations that restore the correct reading frame can, in principle, occur throughout a reversion window defined by stop codons in the alternative, incorrect reading frames. Our previous studies have focused on the characterization of net +1 and −1 frameshift events that revert the *lys2ΔA746* and *lys2ΔBgl* frameshift alleles, respectively (Greene and Jinks-Robertson 1997; Harfe and Jinks-Robertson 1999). These alleles have overlapping reversion windows near the 5′ end of the *LYS2* coding sequence that correspond to a functionally dispensable portion of Lys2p, thus allowing a wide variety of different types of frameshift events to be identified. To study the generation and repair of frameshift intermediates that occur in repetitive sequence, Tran *et al.* (1997) introduced out-of-frame poly(A) runs of varying length into a different, centrally located region of *LYS2*. This latter analysis demonstrated that most compensatory frameshift events occurred within the run itself when run length was greater than eight nucleotides.

To examine the effects of nucleotide composition on mononucleotide run stability, we engineered 10N runs into the region of *LYS2* that corresponds to the reversion window for the *lys2ΔA746* and *lys2ΔBgl* frameshift alleles. The 10N runs generate net +1 frameshift mutations, which are most easily "reverted" by deletion of a single base pair. Furthermore, the runs are sufficiently long that most, if not all, reversion events should occur in the runs. DNA sequence analysis confirmed this prediction for all runs used (data not shown), thus validating the assumption that Lys^+ reversion rates accurately reflect the −1 mutation rates in the 10N runs. To examine the impact of sequence context on mononucleotide run stability, runs of varying composition were introduced at two different positions within the reversion window (see Figure 1). First, a 6A run was replaced with 10N runs, resulting in frameshift alleles referred to as "position 1" *lys2-10N-1* alleles. Second, 10N runs were introduced 18 bp downstream of the 6A sequence, thus changing the local context of the runs. The resulting frameshift alleles are referred to as "position 2" *lys2-10N-2x* alleles. Finally, to examine the effects of flanking sequences on run stabilities, the nucleotides immediately flanking the position 2 runs also were varied (*e.g.*, *lys2-10N-2a* and *lys2-10N-2b*).

The yeast MMR system corrects the majority of polymerase slippage events that occur at long mononucleotide runs (Tran *et al.* 1997). To examine the contributions of individual MMR proteins in the correction of frameshift intermediates in the 10N runs, we individually eliminated yeast MutS (Msh2p, Msh3p, and Msh6p) and MutL (Pms1p and Mlh1p) homologs. Reversion rates of the *lys2-10N* alleles were measured in both wild-type strains and in *msh2*, *msh3*, *msh6*, *pms1*, and *mlh1* mutant strains. All reversion rate data are presented in Table 1.

Reversion rates of *lys2* alleles containing 10N runs at position 1: In a wild-type background, the reversion rates of the *lys2-10N-1* alleles differed over an ~20-fold

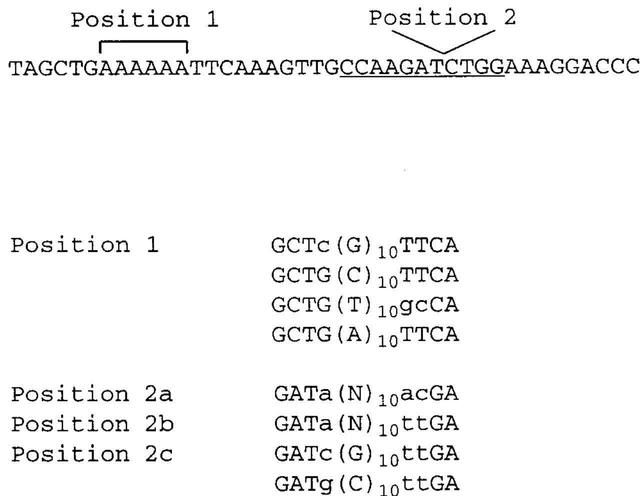

Figure 1.—Position of mononucleotide runs within the *LYS2* gene and sequence of nucleotides flanking the runs. The *LYS2* coding sequence (6A run) replaced by the 10N runs at position 1 is shown at the top. The position 2 10N runs were introduced at the indicated *Pfl*MI site (5′-CCA NNNN↓NTGG; underlined) using appropriate oligonucleotides. When annealed, the oligonucleotides yielded duplex molecules with 3′ overhangs complementary to those produced by *Pfl*MI digestion. In addition, each 10N run was flanked by one additional nt on the 5′ side and two additional nt on the 3′ side so that the immediately flanking sequences could be altered. Insertion of the 10N runs at site 2 thus was accompanied by an insertion of three additional nt, and a duplication of the GAT sequence of the *Pfl*MI site. At the bottom are shown the sequences of the *lys2-10N* alleles. Lowercase letters denote nt outside of the 10N runs that were altered. "N" denotes G, C, or T, depending on the allele.

range. The *lys2-10A-1* and *lys2-10T-1* alleles exhibited the lowest reversion rates (0.9×10^{-8} and 1.4×10^{-8}, respectively) and the *lys2-10G-1* allele the highest reversion rate (1.7×10^{-7}). In the *msh2*, *pms1*, and *mlh1* strains, the reversion rates were elevated 4,500- to 66,000-fold. Although the reversion rates of any given allele were similar in the three mutant backgrounds, the alleles containing the 10G or 10C runs reverted at a 5- to 10-fold higher rate than the alleles containing the 10A or 10T runs.

Individual disruption of either *MSH3* or *MSH6* increased the reversion rate of each *lys2-10N-1* allele examined, but the reversion rates of different alleles varied over a very broad range. In an *msh3* strain, for example, the 10G and 10C runs were much more unstable than the 10A and 10T runs, with the reversion rate of the *lys2-10G-1* allele being 340-fold higher than that of the *lys2-10T-1* allele (1×10^{-4} *vs.* 2.9×10^{-7}, respectively). Furthermore, the 10G run was 10-fold more unstable than the 10C run in an *msh3* background. In *msh6* strains, the 10G and 10C runs were even more dissimilar, with the reversion rates of the respective *lys2-10N-1* alleles differing by more than 200-fold (6.6×10^{-5} and 2.8×10^{-7} for the *lys2-10G-1* and *lys2-10C-1* alleles, respectively). The stability of the 10T run was similar to that of the 10C run in the *msh6* strains and these runs each were 10-fold more unstable than the 10A run. Simultaneous disruption of *MSH3* and *MSH6* resulted in a synergistic increase in mutation rates relative to the single disruptions. As expected, the *msh3 msh6* double mutants exhibited mutation rates comparable to those seen in the *msh2* single mutants (data not shown).

Reversion rates of *lys2* alleles containing 10N runs at position 2: In the position 1 10N stability assays, the DNA sequences flanking the 10N runs were not identical (Figure 1). Specifically, during the construction of the 10G run, the nucleotide immediately preceding the run was changed from G to C to maintain a constant run length while varying the run composition. Similarly, for the 10T run, the TT dinucleotide immediately following the run was changed to GC. To investigate the contribution of flanking sequence to polymerase slippage, three of the runs (10G, 10C, and 10T) were flanked by adenines and placed 18 bp downstream of position 1 (position 2a; see Figure 1). Placing the runs 18 bp away from position 1 allowed us to assay whether the sequence context of the runs affected polymerase slippage and/or MMR efficiency.

In a wild-type background, the *lys2-10N-2a* runs had stabilities that differed by approximately fourfold, and it should be noted that the *lys2-10C-2a* allele was slightly more unstable than the *lys2-10G-2a*, a reverse of the pattern seen at position 1. As at position 1, the reversion rate of the *lys2-10T-2a* allele was lower than that of the *lys2-10G-2a* and *lys2-10C-2a* alleles in the *msh2* strains. Although there were subtle differences between the position 1 and position 2a reversion rates for a given 10N run in wild-type strains, sequence context had no significant effect on reversion rates in *msh2* mutants (this also was true for the 2b and 2c runs; see below). In stark contrast, some runs of the same composition behaved very differently in an *msh3* mutant and/or an *msh6* mutant when the context was altered. Most notably, the 10C run was 50-fold more unstable at position 2a than at position 1 in either an *msh3* or an *msh6* mutant. The reverse pattern was seen for the corresponding 10G alleles in *msh3* and *msh6* mutants, with the position 1 allele being two- to eightfold more unstable than the position 2a allele. The *lys2-10T-1* and *lys2-10T-2a* alleles had similar reversion rates in the *msh3* and *msh6* single mutants.

The very different stabilities of the 10G and 10C runs at positions 1 and 2a in the *msh3* and *msh6* single mutants could be attributed to the general locations of the runs and/or to the identities of the immediately flanking nucleotides. To address these possibilities, we modified the sequences flanking the position 2a 10G and 10C runs so that they were identical to the position 1 flanking sequences (position 2c runs; see Figure 1). This allowed us to compare the stabilities of 10C and 10G runs having identical 5′ and 3′ flanking sequences but slightly differ-

TABLE 1

Reversion rates ($\times 10^{-8}$) of *lys2* alleles containing 10N runs

Run position	Genotype	Run composition			
		10G	10C	10T	10A
1	Wild type	17	3.9	1.4	0.9
		(16–21)[a]	(2.6–8.0)	(0.7–1.4)	(0.4–1.0)
	mlh1	75,000	75,000	22,000	6,000
		(70,000–140,000)	(63,000–87,000)	(15,000–32,000)	(3,400–8,200)
	pms1	110,000	77,000	19,000	4,100
		(30,000–160,000)	(77,000–140,000)	(6,400–23,000)	(3,300–5,700)
	msh2	66,000	68,000	19,000	7,000
		(52,000–70,000)	(42,000–82,000)	(14,000–22,000)	(5,700–8,600)
	msh3	10,000	1,000	29	42
		(8,300–13,000)	(630–1,700)	(23–37)	(37–65)
	msh6	6,600	28	26	2.5
		(5,700–13,000)	(14–28)	(21–39)	(2.1–2.9)
2a	Wild type	5.3	8.5	2.1	ND
		(3.8–5.4)	(7.2–11)	(1.6–3.9)	
	msh2	77,000	98,000	20,000	ND
		(48,000–88,000)	(82,000–156,000)	(6,500–29,000)	ND
	msh3	5,200	51,000	48	ND
		(4,100–8,100)	(39,000–64,000)	(28–62)	
	msh6	800	1,800	16	ND
		(530–950)	(1,400–2,100)	(15–24)	
2b	Wild type	4.4	3.4	ND	ND
		(4.2–5.4)	(2.3–4.9)		
	msh2	79,000	72,000	ND	ND
		(53,000–110,000)	(18,000–110,000)		
	msh3	3,100	800	ND	ND
		(2,200–4,200)	(130–1,000)		
	msh6	3,500	3,800	ND	ND
		(2,400–5,900)	(3,200–4,500)		
2c	Wild type	12	3.3	ND	ND
		(10–20)	(2.6–7.2)		
	msh2	73,000	80,000	ND	ND
		(60,000–85,000)	(66,000–110,000)		
	msh3	14,000	900	ND	ND
		(13,000–18,000)	(840–1,600)		
	msh6	3,200	1,800	ND	ND
		(2,200–6,300)	(1,400–2,600)		

[a] 95% confidence interval.
ND, not determined.

ent positions within *LYS2*. Finally, strains were made in which the 10G and 10C runs at position 2a were modified so that 3′ flanks were identical to those at position 1 (TT dinucleotide) but the 5′ flanks were different (position 2b runs; see Figure 1).

As noted previously, context appeared to have no effect on run stabilities in *msh2* strains and only small effects on the stabilities of the 10G and 10C runs in wild-type strains. In wild-type strains, the *lys2-10G-1* and *lys2-10G-2c* alleles, which have identical flanking sequences, were more unstable than the *lys2-10G-2a* and *lys2-10G-2b* alleles. For the 10C runs in wild-type strains, the stabilities of the *lys2-10C-1*, *lys2-10C-2b*, and *lys2-10C-2c* alleles, all of which have a TT dinucleotide immediately 3′ to the run, were greater than that of the *lys2-10C-2a* allele. In the *msh3* and *msh6* strains, there were very dramatic context-dependent differences in the *lys2-10N* reversion rates. First, in an *msh3* background the *lys2-10C-2a* allele was 50-fold more unstable than the *lys2-10C-1*, *lys2-10C-2b*, or *lys2-10C-2c* allele, each of which had a comparable stability. As note previously, the 10C runs in contexts 1, 2b, and 2c have in common a 3′ flanking TT dinucleotide, which is changed to AC in context 2a. Second, in an *msh6* mutant background the *lys2-10C-1* allele was 50-fold more stable than the *lys2-10C-2a*, *lys2-10C-2b*, or *lys2-10C-2c* alleles. The stabilities of the *lys2-10G* alleles also exhibited context effects in the *msh3* and *msh6* single mutants, but the mutation rate variations generally were not as dramatic as those observed with the *lys2-10C* alleles.

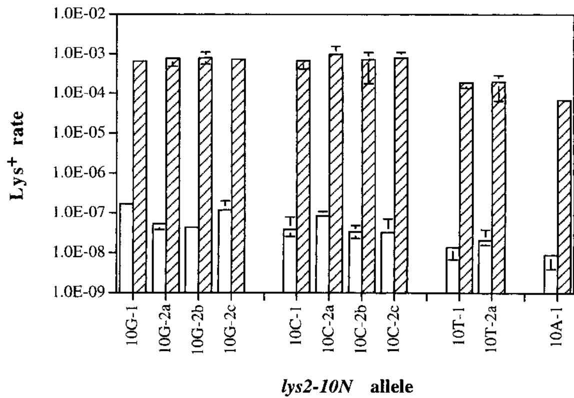

FIGURE 2.—Reversion rates of *lys2-10N* alleles in wild-type (open bars) and *msh2* (hatched bars) strains. The 95% CI for each rate is indicated.

DISCUSSION

Mononucleotide runs are found frequently within the coding regions of eukaryotic genes. In the yeast genome, for example, ~25% of genes contain a poly(A) or poly (T) tract that is at least eight nucleotides in length (Tran *et al.* 1997). Because polymerase slippage events occur at an increased rate in mononucleotide runs, it is important to understand how the nucleotide composition and sequence context of a run affect the rate of DNA polymerase slippage. In addition to affecting the rate of polymerase slippage, run composition and context may also affect the efficiency with which the slippage errors are removed by the MMR machinery. To address these issues, we developed an *in vivo* reversion assay that allows us to measure the rate of −1 frameshift events within defined 10N mononucleotide runs located in the yeast *LYS2* gene (*lys2-10N* alleles). Both the composition of the runs and the sequences flanking the runs were altered, allowing us to examine the effects of these changes on mononucleotide run stability. To facilitate comparisons relevant to the points discussed below, the reversion rates of the *lys2-10N* alleles in wild-type and in completely MMR-defective (*msh2*) strains are presented in Figure 2. Figure 3 presents the reversion rates of the *lys2-10N* alleles in *msh2*, *msh3*, and *msh6* strains normalized to the corresponding rates obtained in wild-type strains.

As shown in Figure 2, sequence context had only small (at most fourfold) effects on the stabilities of the 10G, 10C, or 10T runs in MMR competent cells. Because these subtle effects were not evident in completely MMR-defective (*msh2*) cells, it can be concluded that the overall efficiency of the MMR system is influenced by the sequence context of the mutational intermediate. The context of the 10A run was not altered, but we would expect to obtain similar results. The data from the *msh2* strains suggest that most of the −1 frameshift mutations resulted from simple slippage within the runs rather than slippage initiated by base misincorporation events adjacent to the runs. In the latter case, one would expect for slippage rates to exhibit context dependence. We consistently observed a higher mutation rate in the 10G and 10C runs than in the 10T and 10A runs in both wild-type and *msh2* strains. One interpretation of this observation is that G and C runs are inherently more unstable than A and T runs, which would be consistent with observations made in *E. coli* (Sagher *et al.* 1999). Although one might expect that slippage would occur less often in G/C tracts than in A/T tracts because of the number of hydrogen bonds that must be disrupted, a consideration of base stacking interactions suggests that a slippage intermediate in a G/C run should be more stable than one in an A/T run (see Sagher *et al.* 1999). In addition, slippage intermediates in G or C runs would be expected to be less efficiently proofread than those in A or T runs because of the relative difficulty of disrupting GC base pairs (Bessman and Reha-Krantz 1977; Goodman and Fygenson 1998). It should be noted that the exonucleolytic proofreading activities of the yeast replicative polymerases were present in our studies, so the stability differences between the G/C tracts *vs.* the A/T tracts could reflect different proofreading efficiencies of −1 frameshift intermediates in A/T *vs.* G/C runs.

The MMR proteins Msh3p and Msh6p interact with Msh2p to form Msh2p-Msh3p and Msh2p-Msh6p mismatch-recognition complexes (Kolodner and Marsischky 1999). These complexes have overlapping recognition specificities for +1 and −1 frameshift intermediates (Marsischky *et al.* 1996; Greene and Jinks-Robertson 1997; Sia *et al.* 1997; Harfe and Jinks-Robertson 1999), and these overlaps are manifested as strong synergistic effects on mutation rates when both complexes are eliminated (*i.e.*, an *msh3 msh6* double mutant or an *msh2* single mutant). It should be noted that the reversion rate in an *msh3* mutant reflects the repair capacity of the remaining Msh2p-Msh6p com-

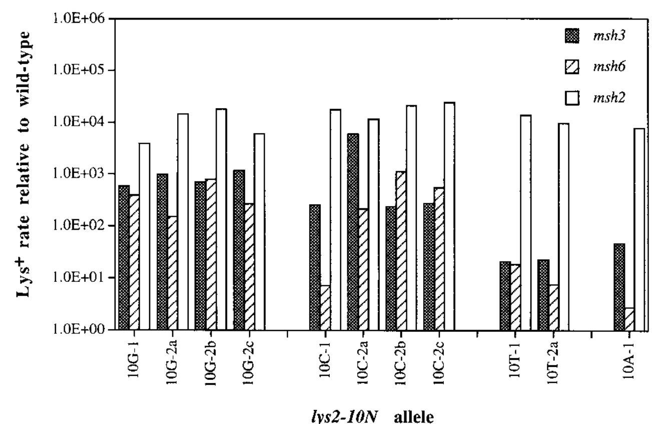

FIGURE 3.—Lys+ reversion rates in *msh2*, *msh3*, and *msh6* mutants. For each *lys2-10N* allele, the mutation rates have been normalized to the corresponding rate obtained in a wild-type strain.

plex, while the reversion rate in an *msh6* mutant indicates the repair capacity of the Msh2p-Msh3p complex. The relative abilities of Msh2p-Msh3p and Msh2p-Msh6p complexes to initiate repair of DNA polymerase slippage errors thus can be assessed by comparing reversion rates of a given allele in *msh6* and *msh3* mutants, respectively. If the reversion rate of a given allele is similar in the *msh3* and *msh6* single mutants, then one could argue that the Msh2p-Msh3p and Msh2p-Msh6p complexes have similar affinities for the corresponding frameshift intermediate. If the reversion rate of a given *lys2-10N* allele differs dramatically in the two single mutants, however, then one would conclude that the complexes have different recognition specificities for the corresponding frameshift intermediate. As shown in Figure 3, mutation rates for the *lys2-10C* and *lys2-10G* alleles generally were elevated to a much greater extent in the single mutants than were the reversion rates for the *lys2-10A* and *lys2-10T* alleles. It thus appears that the Msh2p-Msh3p and Msh2p-Msh6p complexes may compete more effectively for the repair of frameshift intermediates in A and T runs than in G and C runs. There did not seem to be general specificity of a given complex for an A or T *vs.* a G or C run.

The Msh2p-Msh3p and Msh2p-Msh6p complexes did not exhibit any striking repair biases in terms of run composition, but we found that the context of a run could dramatically affect the repair of frameshift intermediates. This was particularly evident when examining the reversion rates of the *lys2-10C* alleles in the mutant strains relative to the corresponding wild-type strains (Figure 3). In an *msh6* background, for example, the mutation rates of the three *lys2-10C-2x* alleles were 50-fold higher than that of the *lys2-10C-1* allele. Thus the Msh2p-Msh3p complex appears to recognize frameshift intermediates much better when the 10C run is at position 1 (*lys2-10C-1* allele) than when the run is at position 2 (*lys2-10C-2x* alleles). In addition to this position-dependent context effect, an effect of the immediately flanking nucleotide (nt) on the repair efficiency of the Msh2p-Msh6p complex was evident when the stabilities of the 10C runs were examined in an *msh3* background. In this case, the *lys2-10C-2a* allele had a much higher reversion rate than the *lys2-10C-1, lys2-10C-2b*, or *lys2-10C-2c* allele. The *lys2-10C-2a* allele differs from the other three alleles in that it is immediately flanked on the 3′ side by an AC dinucleotide instead of a TT dinucleotide. It should be noted that the reversion rate of the *lys2-10C-2a* allele in an *msh3* background was almost as high as in an *msh2* background, indicating that the Msh2p-Msh6p complex has very little repair activity against the relevant frameshift intermediate.

Although context effects for the binding of Msh2p-Msh6p to defined mismatches *in vitro* have been reported (Macpherson *et al.* 1998; Marsischky and Kolodner 1999), the results presented here provide the first clear example of strong context effects on the repair efficiencies of the Msh2p-Msh3p and Msh2p-Msh6p complexes *in vivo*. Thus, the repair of a frameshift intermediate is not simply related to the identity of the extrahelical nucleotide(s), but also can be profoundly impacted by the nucleotides that flank the mutational intermediate, even when the intermediate is embedded in a highly repetitive sequence. Such context effects may not only be determined by the specificities of the Msh2p-Msh3p and Msh2p-Msh6p mismatch-binding complexes, but also could be related to whether the frameshift intermediate arises during leading strand *vs.* lagging strand DNA synthesis. In the case of the *LYS2* gene, a DNA replication fork would be expected to enter from the 5′ end (Tran *et al.* 1995; Freudenreich *et al.* 1997), so that the coding strand would be the template for lagging strand synthesis. In addition, the context effects reported here could be related to the identity of the polymerase that generates the frameshift intermediate or even to the transcriptional state of the sequence (Morey *et al.* 2000). Additional studies will be required to determine whether these additional factors influence MMR efficiency.

We gratefully acknowledge the contributions of R. McVoy in constructing some of the plasmids used in this work. We thank K. Hill and R. Spell for critically reading the manuscript. This work was funded by a grant from the National Science Foundation.

LITERATURE CITED

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

Bessman, M. J., and J. Reha-Krantz, 1977 Studies on the biochemical basis of spontaneous mutation. J. Mol. Biol. **116:** 115–123.

Chattoo, F. F., F. Sherman, D. A. Azubalis, T. A. Fjellstedt, D. Mehnert *et al.*, 1979 Selection of *lys2* mutants of the yeast *Saccharomyces cerevisiae* by the utilization of alpha-aminoadipate. Genetics **93:** 51–65.

Chen, W., and S. Jinks-Robertson, 1998 Mismatch repair proteins regulate heteroduplex formation during mitotic recombination in yeast. Mol. Cell. Biol. **18:** 6525–6537.

Crouse, G. F., 1998 Mismatch repair systems in *Saccharomyces cerevisiae*, pp. 411–448 in *DNA Damage and Repair, Volume 1: DNA Repair in Prokaryotes and Lower Eukaryotes*, edited by J. A. Nickoloff and M. F. Hoekstra. Humana Press, Totowa, NJ.

Dixon, W. J., and F. J. Massey, Jr., 1969 *Introduction to Statistical Analysis*. McGraw-Hill, New York.

Flores-Rozas, H., and R. D. Kolodner, 1998 The *Saccharomyces cerevisiae MLH3* gene functions in MSH3-dependent suppression of frameshift mutations. Proc. Natl. Acad. Sci. USA **95:** 12404–12409.

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

Goodman, M. F., and D. K. Fygenson, 1998 DNA polymerase fidelity: from genetics toward a biochemical understanding. Genetics **148:** 1475–1482.

Greene, C. N., and S. Jinks-Robertson, 1997 Frameshift intermediates in homopolymer runs are removed efficiently by yeast mismatch repair proteins. Mol. Cell. Biol. **17:** 2844–2850.

Harfe, B. D., and S. Jinks-Robertson, 1999 Removal of frameshift intermediates by mismatch repair proteins in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **19:** 4766–4773.

Harfe, B. D., B. K. Minesinger and S. Jinks-Robertson, 2000 Discrete in vivo roles for the MutL homologs Mlh2p and Mlh3p in

the removal of frameshift intermediates in budding yeast. Curr. Biol. **10:** 145–148.

Kolodner, R. D., and G. T. Marsischky, 1999 Eukaryotic DNA mismatch repair. Curr. Opin. Genet. Dev. **9:** 89–96.

Kroutil, L. C., K. Register, K. Bebenek and T. A. Kunkel, 1996 Exonucleolytic proofreading during replication of repetitive DNA. Biochemistry **35:** 1046–1053.

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

Macpherson, P., O. Humbert and P. Karran, 1998 Frameshift mismatch recognition by the human MutSα complex. Mutat. Res. **408:** 55–66.

Marsischky, G. T., and R. D. Kolodner, 1999 Biochemical characterization of the interaction between the *Saccharomyces cerevisiae* MSH2-MSH6 complex and mispaired bases in DNA. J. Biol. Chem. **274:** 26668–26682.

Marsischky, G. T., N. Filosi, M. F. Kane and R. Kolodner, 1996 Redundancy of *Saccharomyces cerevisiae MSH3* and *MSH6* in *MSH2*-dependent mismatch repair. Genes Dev. **10:** 407–420.

Morey, N. J., C. N. Greene and S. Jinks-Robertson, 2000 Genetic analysis of transcription-associated mutation in *Saccharomyces cerevisiae*. Genetics **154:** 109–120.

Prolla, T. A., D.-M. Christie and R. M. Liskay, 1994a Dual requirement in yeast DNA mismatch repair for *MLH1* and *PMS1*, two homologs of the bacterial *mutL* gene. Mol. Cell. Biol. **14:** 407–415.

Prolla, T. A., Q. Pang, E. Alani, R. D. Kolodner and R. M. Liskay, 1994b MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science **265:** 1091–1093.

Rothstein, R., 1991 Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. **194:** 281–301.

Sagher, D., A. Hsu and B. Strauss, 1999 Stabilization of the intermediate in frameshift mutation. Mutat. Res. **423:** 73–77.

Sherman, F., 1991 Getting started with yeast. Methods Enzymol. **194:** 3–20.

Sia, E. A., C. A. Butler, M. Dominska, P. Greenwell, T. D. Fox *et al.*, 1997 Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell. Biol. **17:** 2851–2858.

Streisinger, G., and J. Owen, 1985 Mechanisms of spontaneous and induced frameshift mutation 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.

Wang, T.-F., N. Kleckner and N. Hunter, 1999 Functional specificity of MutL homologs in yeast: evidence for three Mlh1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. Proc. Natl. Acad. Sci. USA **96:** 13914–13919.

Communicating editor: M. Lichten