MOLECULAR AND CELLULAR BIOLOGY, May 1997, p. 2844–2850 Vol. 17, No. 5
0270-7306/97/$04.00+0

Frameshift Intermediates in Homopolymer Runs Are Removed Efficiently by Yeast Mismatch Repair Proteins

CHRISTOPHER N. GREENE[1] AND SUE JINKS-ROBERTSON[1,2]*

Graduate Program in Genetics and Molecular Biology[1] and Department of Biology,[2] Emory University, Atlanta, Georgia 30322

Received 23 December 1996/Returned for modification 28 January 1997/Accepted 13 February 1997

A change in the number of base pairs within a coding sequence can result in a frameshift mutation, which almost invariably eliminates the function of the encoded protein. A frameshift reversion assay with *Saccharomyces cerevisiae* that can be used to examine the types of insertions and deletions that are generated during DNA replication, as well as the editing functions that remove such replication errors, has been developed. Reversion spectra have been obtained in a wild-type strain and in strains defective for defined components of the postreplicative mismatch repair system (*msh2*, *msh3*, *msh6*, *msh3 msh6*, *pms1*, and *mlh1* mutants). Comparison of the spectra reveals that yeast mismatch repair proteins preferentially remove frameshift intermediates that arise in homopolymer tracts and indicates that some of the proteins have distinct substrate or context specificities.

The integrity of the genetic material during genome duplication requires that the process of DNA replication be highly accurate. The overall fidelity of replication is determined at three steps: (i) the error rate of nucleotide incorporation during polymerization, (ii) the efficiency with which the exonucleolytic proofreading activity of DNA polymerase removes terminal nucleotides that are incorrectly base paired, and (iii) the efficiency with which postreplicative mismatch repair (MMR) systems remove any remaining errors. Mutations resulting from persistent replication errors generally can be classified as either insertion or deletion events involving a small number of nucleotides or base substitution events. Frameshift mutations are specifically those nucleotide insertion or deletion events that do not occur in multiples of three and consequently alter the reading frame of the translated mRNA. Whereas base substitution events often are phenotypically silent, frameshift mutations almost invariably compromise protein function. Given the very deleterious nature of frameshift mutations, it is important to understand the mechanisms for generating insertions and deletions during DNA replication as well as the editing functions that prevent fixation of such replication errors.

Two major approaches have been utilized to study frameshift mutagenesis: analysis of spontaneous or mutagen-induced frameshift spectra generated in vivo and analysis of frameshift events generated during in vitro replication of defined DNA templates. In vivo studies done primarily with prokaryotic organisms have demonstrated the existence of multiple mechanisms of frameshift mutagenesis, all of which are facilitated by local DNA sequence context (36). As first noted by Streisinger et al. in studies with phage T4 (48), deletion or addition of one or more units of a tandemly repeated sequence (e.g., a monotonic run of a single nucleotide or a dinucleotide repeat) can be explained by DNA polymerase slippage. Transient dissociation of the nascent 3′ end from the template strand, followed by misalignment between repeats and subsequent polymerase extension, will lead to a deletion event if the extrahelical base(s) is on the template strand or an insertion event if the extrahelical base(s) is on the nascent strand. In addition to events involving tandem repeats, polymerase slippage is likewise invoked to explain larger deletions and insertions having their endpoints in short direct repeats (37, 41, 44, 49). More complex frameshift events in which there are changes in both the number and the sequence of bases have been attributed to mispairing between quasipalindromic sequences (11) or imperfect direct repeats (13). In addition to the frameshift events that are templated by various types of DNA misalignments, in vitro work has provided strong evidence that base substitution events can initiate frameshift mutagenesis through a process known as dislocation, in which slippage following base misincorporation leads to correct pairing at the 3′ terminus (2). Dislocation may account for many of the in vivo frameshifts that do not involve iterated DNA sequences.

DNA synthesis errors that escape the proofreading activity of DNA polymerase are subject to postreplicative MMR systems, the best understood of which is the methyl-directed MMR system of *Escherichia coli* (for a review, see reference 29). Three key components of this system are MutS, MutL, and MutH. A MutS homodimer binds directly to mismatched bases, MutH nicks the unmethylated strand at hemimethylated *dam* (GATC) sites, thus marking the newly replicated strand for removal, and a MutL homodimer promotes interaction between the MutS-DNA complex and MutH, thereby activating the latent MutH endonuclease activity. Frameshift as well as base substitution intermediates are efficiently recognized and repaired by the methyl-directed MMR system (40).

Eukaryotic homologs of the bacterial MutS and MutL proteins have recently gained much attention because of their association with human hereditary nonpolyposis colon cancer (for reviews, see references 8 and 52). In the yeast *Saccharomyces cerevisiae*, multiple homologs of both MutS (Msh1p-6p) and MutL (Pms1p and Mlh1p-3p) have been identified (for reviews, see references 5 and 21). Msh2p, Msh3p, Msh6p, Pms1p, and Mlh1p are important for correcting mismatches arising during nuclear DNA replication and recombination (6, 19, 22, 28, 30, 32, 35, 42, 45, 46), whereas Msh1p corrects mismatches that arise in mitochondrial DNA (35). In contrast, Msh4p and Msh5p are meiosis-specific proteins that have no apparent effect on mismatch repair but rather affect crossing-

* Corresponding author. Mailing address: Department of Biology, 1510 Clifton Rd., Emory University, Atlanta, GA 30322. Phone: (404) 727-6312. Fax: (404) 727-2880. E-mail: jinks@biology.emory.edu.

over (16, 38). The roles of Mlh2p and Mlh3p have not been established. Mlh1p and Pms1p can form a heterodimer (33), and because mutations in either gene completely eliminate replication-associated MMR (32), the heterodimer is presumed to be the active form. Msh2p is required for all replication and recombination-associated MMR and can form a heterodimer with either Msh3p or Msh6p (12, 18, 28). The phenotypes of *msh2*, *msh3*, *msh6*, and *msh3 msh6* mutants suggest that the Msh2p-Msh6p heterodimer recognizes both base substitution and insertion or deletion mismatches, whereas the Msh2p-Msh3p heterodimer appears to be specific for insertion or deletion mismatches (19, 28). An association between yeast MMR proteins and proliferating cell nuclear antigen has been reported (18a, 51), suggesting that the MMR machinery may track with the replication apparatus in eukaryotic organisms.

In order to better understand the process of frameshift mutagenesis in eukaryotes, we have developed an in vivo system that identifies frameshift events occurring within a defined 150-bp region of the yeast *LYS2* gene. This system has been used to obtain a frameshift spectrum in wild-type yeast cells as well as in cells deficient in various components of the MMR machinery. These studies indicate that the yeast MMR proteins have distinct in vivo specificities in terms of recognizing and removing frameshift intermediates.

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 2% glycerol and 2% ethanol (YEPGE) or 2% dextrose (YEPD). Synthetic complete (SC) medium (43) containing 2% dextrose and lacking the appropriate amino acid was used for selective growth. 5-Fluoroorotic acid (5-FOA) medium containing 1 g of 5-FOA per liter was used to select for Ura^- yeast segregants (3). LB medium (1% yeast extract, 0.5% Bacto Tryptone, 1% NaCl [1.5% agar for plates]) was used for growth of *E. coli* strains. Ampicillin (100 μg/ml) was added to LB medium for growth of plasmid-containing strains. Yeast and bacterial strains were grown at 30 and 37°C, respectively.

Strain construction. All strains used in this study are isogenic derivatives of SJR357 (*MAT*α *ade2-101 his3*Δ*200 ura3*Δ*Nco lys2*Δ*Bgl*) and were constructed by standard transformation procedures (9). The *lys2*Δ*Bgl* allele was introduced into SJR357 with plasmid pSR125 as described previously (7). The *pms1*Δ strain was constructed by a two-step transplacement procedure, whereas all other mutants were constructed by gene disruption (39). Ura^+ colonies were selected in all transformations; subsequent loss of the *URA3* marker (either by plasmid loss from the chromosome in two-step transplacement experiments or recombination between the flanking *hisG* repeats when the gene in question was disrupted with the *hisG-URA3-hisG* cassette [1]) was detected on 5-FOA medium. All disruptions were confirmed by PCR.

*msh2*Δ*::hisG* was introduced by transformation with *Aat*II/*Xba*I-digested GC1914 (see below), *msh3*Δ*::hisG* was introduced by transformation with *Eco*RI-digested pEN33 (6), *msh6*Δ*::hisG* was introduced by transformation with *Sac*I/*Eco*RI-digested pBUH-MSH6-NT/CT (see below), *pms1*Δ was introduced by transformation with *Bst*XI-digested pSR211 (6), and *mlh1*Δ*::URA3* was introduced by transformation with *Sac*I/*Bam*HI-digested *mlh1*Δ*::URA3* (32). Plasmid GC1914 (obtained from G. F. Crouse) is a pBluescript derivative in which the *MSH2*-internal *Eco*NI/*Hpa*I fragment was replaced with the *hisG-URA3-hisG* cassette. pBUH-MSH6-NT/CT (obtained from W. Kramer) is a pBluescript derivative in which the *hisG-URA3-hisG* cassette is flanked by upstream and downstream *MSH6* sequences.

Measurement of reversion rates and isolation of revertants. Two-day old colonies were excised from YEPD plates, inoculated into 5 ml of YEPGE liquid medium (growth in YEPGE prevents accumulation of petite mutants), and grown for 2 days on a roller drum. Cells were harvested by centrifugation, washed once with sterile H_2O, and resuspended in 1 ml of H_2O. Aliquots (100 μl) were plated onto SC-Lys to select Lys^+ revertants, and appropriate dilutions were plated on YEPD to determine viable cell numbers. Lys^+ cells were counted on day 3 after selective plating. Reversion rates were determined by the method of the median (24), with data from 10 to 15 cultures of each strain.

To isolate independent Lys^+ revertants for DNA sequence analysis, 1-ml YEPGE cultures were grown as described above and a single aliquot was plated on SC-Lys. One revertant from each culture was purified for subsequent molecular analysis.

DNA sequencing. Total genomic DNA was prepared from 2 ml of cells by glass bead lysis (15). DNA was resuspended in a final volume of 100 μl, and 0.5 μl was used a template in a 50-μl PCR. A 900-bp region containing the *lys2*Δ*Bgl* reversion window was amplified by using primers 5′-GTAACCGGTGACGAT GAT-3′ (forward) and 5′-CCAATTGTCCAGCAGCTC-3′ (reverse). PCR was performed as follows: 3 min at 94°C followed by 30 cycles of 30 s at 94°C, 30 s at 52°C, and 60 s at 72°C. The PCR mixture (5 μl) was treated with 1 U each of exonuclease I and shrimp alkaline phosphatase to remove the amplification primers. Standard dideoxy sequencing was performed directly, without further purification, with [^{35}S]dCTP, Sequenase (U.S. Biochemical), and an internal primer (5′-CGCAACAATGGTTACTCT-3′).

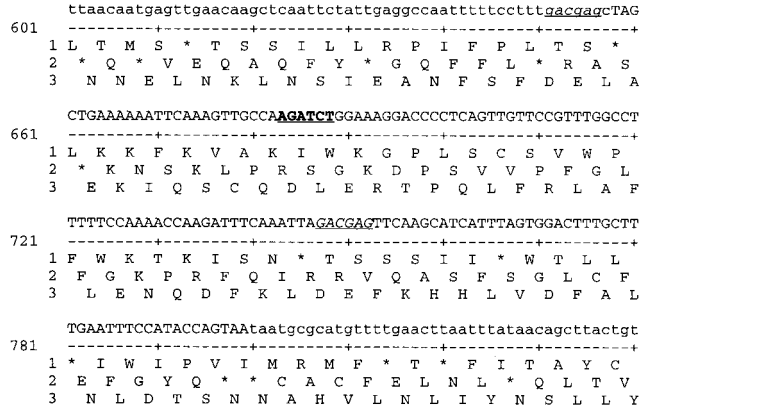

FIG. 1. Partial sequence of the *LYS2* gene. Nucleotides are numbered beginning at the upstream *Xba*I site, and the three possible reading frames are shown. The *Bgl*II site at position 683 is underlined and in boldface type; the ATG start site (not shown) is at position 297. The nucleotides outside the reversion window are in lowercase letters, whereas those in the reversion window are in uppercase letters. The 6-bp repeats at the endpoints of the 100-bp deletion are italicized and underlined.

RESULTS

A system for detecting frameshift events. The 4.2-kb *LYS2* locus contains a single *Bgl*II site (5′-A↓GATCT, where the arrow indicates the restriction site) at position 683, approximately 400 bp downstream of the ATG start codon. The *lys2*Δ*Bgl* frameshift allele was constructed in vitro by filling in the *Bgl*II-generated 5′ overhangs with the Klenow fragment of DNA polymerase I, resulting in direct duplication of the sequence GATC. The addition of 4 bp to the coding sequence is the equivalent of a +1 frameshift and causes a shift from the normal reading frame (reading frame 3 [Fig. 1]) to an alternative reading frame (reading frame 2) when encountered by a translating ribosome. The *lys2*Δ*Bgl* allele reverts spontaneously at a detectable rate, and early genetic analyses indicated that the vast majority, if not all, of the reversion events are second-site, intragenic events (our unpublished studies). Inspection of the coding sequence reveals an approximately 150-bp region where such a compensatory $3N - 1$ (where N is any nucleotide) frameshift must occur. As shown in Fig. 1, the reversion window is defined by the first stop codon in reading frame 2 downstream of the *Bgl*II fill-in (position 797), and the first stop codon in reading frame 1 upstream of the fill-in (position 658).

Reversion spectrum in a wild-type strain. The molecular natures of spontaneous compensatory $3N - 1$ frameshift events were determined by DNA sequence analysis of 145 independent *lys2*Δ*Bgl* revertants (Fig. 2A). The simplest way to revert the *lys2*Δ*Bgl* +1 frameshift is by a compensatory 1-bp deletion, so it is not surprising that the majority (136 of 145, or 94%) of the reversion events were 1-bp deletions. The remaining nine reversion events in the wild-type strain were comprised of seven 2-bp insertions (one of these is complex), one 5-bp insertion, and one 100-bp deletion. One of the +2 events was found three times within the same run of three thymines (3T run), whereas all other +2 events were unique. The end-

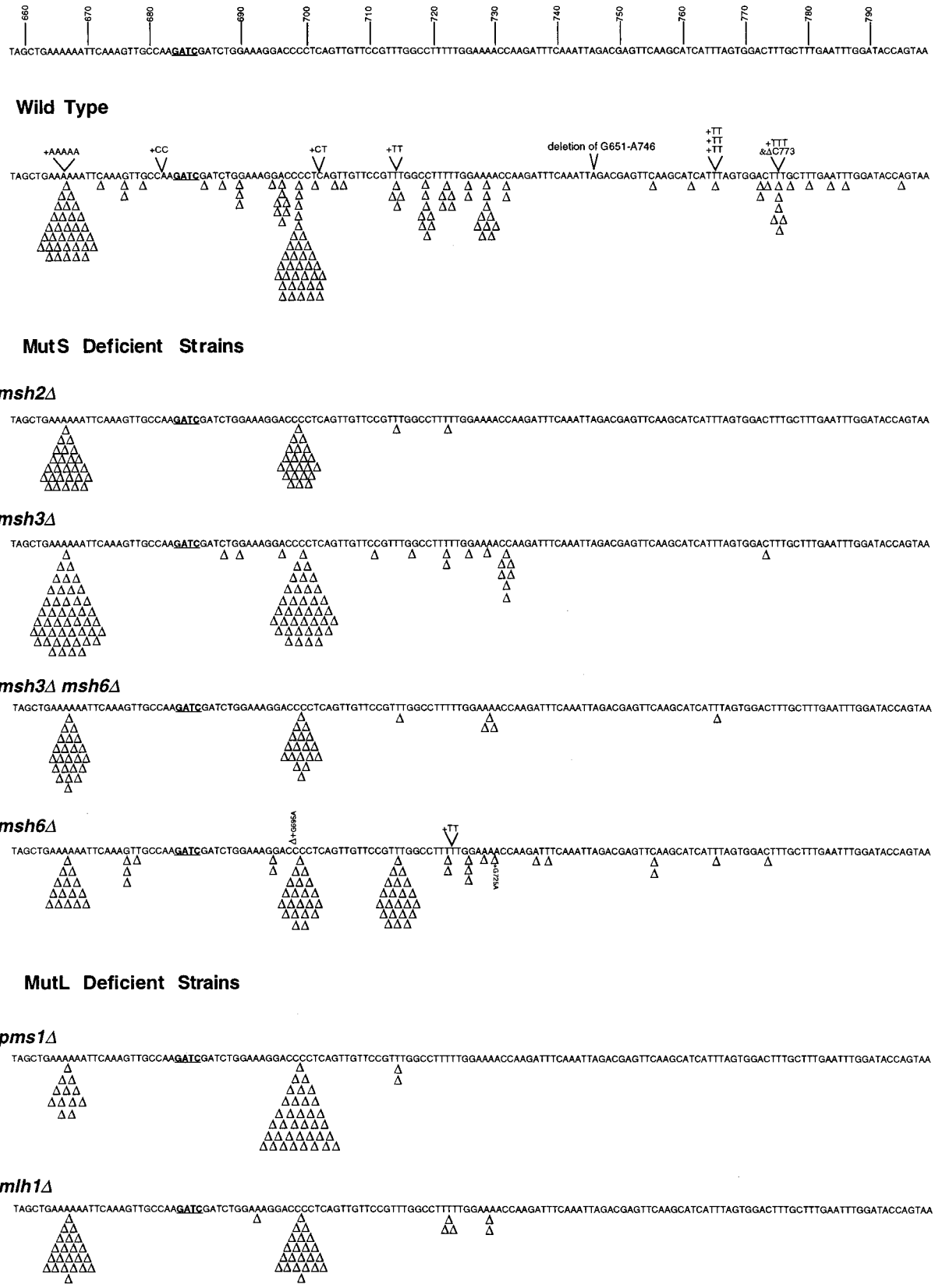

FIG. 2. Reversion spectra in wild-type and MMR-defective yeast strains. The +4 insertion that created the *lys2ΔBgl* allele is underlined and in boldface type. Single-base deletions are indicated below the sequences, and insertions are indicated above the sequences. Each independent −1 event identified is indicated (Δ). (A) Wild-type spectrum; (B) spectra in MutS homolog-deficient strains; (C) spectra in MutL homolog-deficient strains.

TABLE 1. Locations of spontaneous frameshift mutations in a wild-type strain[a]

Type of run	Proportion of window	No. of expected events	No. of observed events	No. of observed events/no. of expected events
Noniterated	56/146	55	23*	0.42
2*N* (22)	44/146	43	20*	0.47
3*N* (9)	27/146	27	18	0.67
4*N* (2)	8/146	8	44*	5.5
5*N* (1)	5/146	5	5	1.0
6*N* (1)	6/146	6	34*	5.7
Total		144	144	

[a] All runs are monotonic. The number in parentheses following a sequence type indicates the number of times it occurs in the 150-bp reversion window. The 100-bp deletion has one endpoint outside of the reversion window and was excluded from this analysis. The difference between the observed and expected distributions is highly significant, as shown by χ^2 analysis ($P < 0.001$). *, the difference between the observed and expected number of events is highly significant, as shown by χ^2 analysis.

points of the 100-bp deletion were within two 6-bp direct repeats (GACGAG at positions 651 to 656 and 747 to 722). The recovery of compensatory frameshift events that change over 30 amino acids within the window, as well as recovery of a deletion removing 32 amino acids, suggests few, if any, functional limitations on the types of events that can be identified with this system. The assay used in this study thus should provide an essentially unbiased sample of all possible frameshift events that occur within the defined reversion window.

Table 1 classifies reversion events according to their location in noniterated sequences versus mononucleotide runs of 2*N*, 3*N*, 4*N*, 5*N*, or 6*N*. The experimental distribution was then compared to the distribution expected if each site within the reversion window were equally mutable. This analysis clearly demonstrates that the reversion events are not distributed randomly throughout the target sequence but rather cluster in mononucleotide runs of more than 3*N*. There were approximately sixfold more events in the 6*N* and 4*N* runs than expected, and approximately twofold fewer events than expected in the 3*N*, 2*N*, and noniterated sequences. One-half of the −1 frameshift events localized to two notable mononucleotide runs: a run of six adenines (6A run) at nucleotides 664 to 669 and a run of four cytosines (4C run) at nucleotides 697 to 700. Although the longer mononucleotide runs are clearly hot spots for −1 events, the relative frequencies of events in different monotonic runs do not appear to be related in any simple way to run length or sequence. There were, for example, 3.5 times more events in the 4C run than in a 4A run (nucleotides 727 to 730), and there were as many events in the 4C run (35 events) as in the longer 6A run (33 events). Additionally, there is a 5T run (nucleotides 720 to 724) where only five events occurred, making the 5T run mutationally "cold" relative to the 6A and the 4C runs.

It appears that, at least in the system used here, 3*N* runs are no better targets for frameshift mutagenesis in yeast than are 2*N* runs and noniterated sequences. The 3*N* runs do not appear to be equally susceptible to −1 mutational events; there are nine 3*N* runs (three 3A runs and six 3T runs), and yet 10 of the 13 deletions clustered in only two of the 3T runs. A similar nonrandom distribution of −1 deletions is seen with the 2*N* runs and noniterated sequences. With the 2*N* repeats, χ^2 analysis indicates that the tracts are not equally mutable. There are approximately equal numbers of AA, CC, GG, and TT repeats in the reversion window, but more than one-half (10 of 19) of the −1 events were in CC tracts. This bias is likely explained by a single CC hot spot; among the five CC repeats in the reversion window, 8 of the 10 C deletions occurred at one of these tracts (nucleotides 718 to 719). For the noniterated bases, a classification of the −1 events by the nucleotide deleted (A, T, C, or G) indicates that the deletions mirror the overall distribution of the bases in the reversion window. Sixteen out of the 56 noniterated bases are adenine, for example, so out of 22 noniterated bases deleted in our sample, one would expect for 6 of these to be adenine residues. We actually observed the deletion of an adenine residue 10 times, which is not significantly different from the expected number of 6. A closer look at the data, however, reveals a deletion hot spot; of the 16 noniterated adenines that could be deleted, 7 out of the 10 examples of an A deletion were at the same position (nucleotide 696).

Reversion spectra in strains defective in MutS homologs. Strains containing individual deletions of the *MSH2*, *MSH3*, and *MSH6* genes, which encode the three MutS homologs implicated in spontaneous nuclear mutagenesis, were constructed in order to investigate the roles of the encoded proteins in preventing the accumulation of frameshift mutations. In addition, an *msh3*Δ *msh6*Δ double-mutant strain was constructed. The reversion rates of the *lys2*Δ*Bgl* allele in the wild-type and mutant backgrounds were determined by fluctuation analysis and are presented in Table 2. The *msh2*Δ and *msh3*Δ *msh6*Δ strains were strong mutators, with the reversion rate of

TABLE 2. Rates and distributions of spontaneous frameshift mutations in MMR-defective strains[a]

Strain	Reversion rate (fold increase)	% of reversion events in run type[b]:					
		6*N*	5*N*	4*N*	3*N*	2*N*	Noniterated
Wild type	2.8×10^{-9} (1)	24 (34/144)	3 (5/144)	31 (44/144)	13 (18/144)	14 (20/144)	16 (23/144)
*msh2*Δ	5.4×10^{-7} (190)	52 (26/50)	2 (1/50)	44 (22/50)	2 (1/50)	0 (0/50)	0 (0/50)
*msh3*Δ	1.1×10^{-8} (3.8)	46 (47/102)	2 (2/102)	38 (39/102)	0 (0/102)	11 (11/102)	3 (3/102)
*msh6*Δ	4.4×10^{-9} (1.6)	19 (15/79)	4 (3/79)	30 (24/79)	30 (24/79)	8 (6/79)	9 (7/79)
*msh3*Δ *msh6*Δ	6.9×10^{-7} (250)	50 (23/46)	0 (0/46)	46 (21/46)	4 (2/46)	0 (0/46)	0 (0/46)
*pms1*Δ	6.8×10^{-7} (240)	24 (12/50)	0 (0/50)	72 (36/50)	4 (2/50)	0 (0/50)	0 (0/50)
*mlh1*Δ	6.0×10^{-7} (210)	44 (22/50)	6 (3/50)	48 (24/50)	2 (1/50)	0 (0/50)	0 (0/50)

[a] Numbers in parentheses following the reversion rates are the fold increases relative to the rate in the wild-type strain. The distributions of reversion events within runs of different lengths were compared in a pairwise fashion by the contingency χ^2 test. Pairwise comparisons of the total spectra of reversion events were done using the C++ version of the Adams and Skopek algorithm (4). The differences between the spectrum of each mutant strain and that of the wild-type strain were highly significant by both tests. The *msh2*Δ spectrum was not significantly different from either the *msh3*Δ spectrum or the *msh3*Δ *msh6*Δ spectrum by either test, the difference between the *msh6*Δ spectrum and either the *msh2*Δ spectrum or the *msh3*Δ spectrum was highly significant by both tests, and the difference between the *pms1*Δ and *mlh1*Δ spectra was significant by the χ^2 test and highly significant by the Adams and Skopek algorithm.

[b] Ratios in parentheses are the ratios of the number of events observed to the total number of events analyzed.

the *lys2ΔBgl* allele increasing approximately 200- and 250-fold in these strains, respectively. The *msh3*Δ and *msh6*Δ strains showed very weak mutator phenotypes for reversion of *lys2ΔBgl*, with the reversion rates increasing only four- and twofold, respectively. The rates of *lys2ΔBgl* reversion reported here are similar to those reported in previous studies (28, 30).

The spectra of reversion events in the mutant strains are shown in Fig. 2B, and these data are summarized in Table 2. The *msh2*Δ and *msh3*Δ *msh6*Δ strains had a dramatic increase in the 1-bp deletion events that occurred in monotonic runs, specifically the 6A and 4C runs. Approximately 95% of the −1 events localized to the 6A and 4C runs in the *msh2*Δ and *msh3*Δ *msh6*Δ strains, compared to only 50% of the events in the wild-type control strain. The reversion rates at these two homopolymeric runs thus were increased at least several hundredfold; the rate increases for events outside of the 6A and 4C tracts are difficult to assess, given the relative rarity of these events. Although the *msh3*Δ strain had only a weak mutator phenotype, its mutation spectrum was very similar to those of the *msh2*Δ and *msh3*Δ *msh6*Δ strains, with 84% of the −1 events occurring in the 6A and 4C runs. Statistical analyses confirm that the *msh2*Δ, *msh3*Δ, and *msh3*Δ *msh6*Δ spectra are significantly different from the wild-type spectrum (see Table 2 footnote).

The reversion spectrum for the *msh6*Δ strain is quite distinct from those of the wild type and the other *msh*Δ strains. Most notably, there was a dramatic increase in reversion events that occur within a specific 3T tract (nucleotides 713 to 715). This particular run only accounts for 3% of the total events in a wild-type background (four −1 events and one +2 event) but represents 28% of the events in the *msh6*Δ strain. Only one event at this 3T tract was seen in each of the *msh2*Δ and *msh3*Δ *msh6*Δ strains; none were observed in the *msh3*Δ strain.

Reversion spectra in strains defective in MutL homologs. The *PMS1* and *MLH1* genes encoding MutL homologs were deleted individually in order to assess the roles of these proteins in the removal of potential frameshift intermediates. Both the *pms1*Δ and the *mlh1*Δ strains were strong mutators, with each showing an approximately 200-fold increase in the *lys2ΔBgl* reversion rate (Table 2). The reversion spectra for the *pms1*Δ and the *mlh1*Δ strains showed a specific increase in 1-bp deletions in the 6A and 4C runs (Fig. 2C), which is similar to the spectra in all of the *msh*Δ strains except *msh6*Δ. If one examines the distribution of events in the 6A and 4C tracts, however, the reversion spectrum for *pms1*Δ differs subtly from those of the *mlh1*Δ strain and all of the *msh*Δ strains. In the *pms1*Δ strain, there was a 3:1 bias for deletions in the 4C run relative to the 6A run; the ratio of −1 events in the 6A run to those in the 4C run was approximately 1:1 for all other strains, including the wild type. This difference between the *pms1*Δ strain and all other strains is statistically significant (see Table 2 footnote).

DISCUSSION

Frameshift events account for only 10% of the mutations identified in forward mutation assays (10, 25), which limits the usefulness of these systems for studying aspects of frameshift mutagenesis in yeast. The specificity necessary for studying frameshift mutagenesis can be achieved by selecting for reversion of a frameshift mutation, and such frameshift reversion assays have been utilized extensively in prokaryotes (for a review, see reference 36). Yeast frameshift-specific assays also have been described but generally are limited either by the use of defined substrates to target the frameshift events (e.g., dinucleotide repeats) or involve a relatively small target that restricts the total types of events that can be detected (14, 20, 50). The current study utilizes phenotypic reversion of a +4 frameshift allele (*lys2ΔBgl*) to isolate compensatory frameshift events occurring within a defined 150-bp region of the yeast *LYS2* gene and focuses on the role of the yeast MMR proteins in eliminating frameshift intermediates. The identification of compensatory frameshift mutations throughout the reversion window (Fig. 2A) indicates that the *lys2ΔBgl* allele provides a suitable system for obtaining an unbiased set of frameshift mutations within a relatively large target.

The specificity of individual MutS and MutL homologs for mutational intermediates has been inferred by comparing *lys2ΔBgl* reversion spectra in wild-type and MMR-defective yeast strains. It is noteworthy that in the DNA sequence analysis of over 500 independent revertants reported here, we did not observe a true reversion event. The majority of *lys2ΔBgl* revertants arising in a wild-type, DNA repair-proficient background were −1 deletion events (Fig. 2A). One-half of the −1 events occurred at two notable monotonic runs (a 6A run and a 4C run) and presumably arose via slippage between the nascent and template strands during DNA replication. In addition to the two mutationally hot runs, there are a 5T run and a 4A run where relatively few frameshift events occurred. Although in vitro and in vivo studies have demonstrated convincingly that the frequency of frameshift events within a given run is directly correlated with the length of the run (23, 50a), our data indicate that factors other than run length may influence the accumulation of frameshifts (see also reference 31). The sequence context surrounding a particular monotonic run may be important as well as the nucleotide composition of a run or the identity of the nucleotides on the leading- versus lagging-strand template during DNA synthesis.

In bacteriophage systems, frameshift events occur more often in 3*N* runs than expected by chance alone, whereas 2*N* runs contain no more −1 deletions than do noniterated sequences (37). In the yeast system used here, 3*N* runs appear to be no better targets for frameshift mutagenesis than are 2*N* runs or noniterated bases. The reason for the discrepancy between the bacterial and yeast studies is not clear, but such a discrepancy could reflect different proofreading capacities of the relevant polymerases in 3*N* runs. In considering the 3*N*, 2*N*, and noniterated sequences in the *lys2ΔBgl* reversion window, there are clearly mutational hot spots for each type of sequence. For example, 8 of 10 C deletions clustered within one of five CC repeats. Again, the most likely explanation for such hot spots is some type of sequence context effect.

The roles of the yeast MutS homologs in eliminating frameshift intermediates were ascertained using strains with individual deletions of *MSH2*, *MSH3*, and *MSH6* as well as a strain harboring a deletion of both *MSH3* and *MSH6* (Fig. 2B). In general, our results agree with those of Marsischky et al. (28), who sequenced a limited number of *lys2ΔBgl* revertants arising in various MMR-defective strains. In the *msh2*Δ strain, the rate of frameshift mutagenesis increased approximately 200-fold; of 50 reversion events sequenced, all were −1 events and >95% of these were in the 6A and 4C runs. Although we cannot eliminate a role for Msh2p in recognizing other types of −1 frameshift intermediates, there clearly is a strong preference for events in the longer mononucleotide runs. A similar bias has been reported for frameshift mutations in MMR-defective *E. coli* cells (26, 41). An interesting issue that emerges concerns the targeted repair of an extrahelical base in a mononucleotide run versus an extrahelical base in noniterated sequences. One possibility is that the frameshift intermediates in reiterated versus noniterated sequences are generated by different mechanisms. Events in mononucleotide runs, for exam-

ple, likely result from DNA polymerase slippage, whereas those in noniterated sequences may result from dislocation initiated by base misincorporation. If one assumes that both slippage and dislocation occur during genome duplication, however, then the specificity of Msh2p is still difficult to explain. A more attractive hypothesis is that the events in mononucleotide runs originate predominantly during genome duplication, whereas those in noniterated sequences might originate from other DNA metabolic processes. Studies in yeast, for example, have demonstrated that genes required for induced mutagenesis also are important for spontaneous mutagenesis, thus implicating DNA damage and repair as a major source of spontaneous mutation events (34). Additionally, it has been inferred that the error rate of DNA synthesis during recombination in yeast is much higher than the error rate of normal replicative DNA synthesis (47). It recently has been reported that the yeast MMR machinery interacts with proliferating cell nuclear antigen, which suggests that the MMR machinery may actually track with the replication apparatus (18a, 51). Such tracking could explain not only how strand specificity is achieved in those organisms that do not differentially mark the newly replicated and template strands but also why frameshift intermediates that arise out of the context of genome duplication might be refractory to repair by the MMR machinery.

Studies in yeast have demonstrated that strains deleted individually for *MSH3* and *MSH6* have weak mutator phenotypes, whereas simultaneous deletion of *MSH3* and *MSH6* yields a strong mutator phenotype indistinguishable from that of an *msh2*Δ strain (19, 28, 43a). These in vivo data are the basis of a model in which Msh2p forms a heterodimer with either Msh3p or Msh6p; Msh2p-Msh3p is thought to recognize only insertion or deletion mismatches, whereas Msh2p-Msh6p exhibits a preference for base substitution mismatches (19, 28). In vitro data have supported this basic model (12, 18). As expected, both the mutator phenotype and the frameshift mutation spectrum of the *msh3*Δ *msh6*Δ strain were indistinguishable from those of the *msh2*Δ strain in the present work. Although both the *msh3*Δ and *msh6*Δ strains showed only marginal increases in the rate of *lys2*Δ*Bgl* reversion, these strains had distinct reversion spectra. The *msh3*Δ strain exhibited an increased bias (relative to the wild-type strain) in −1 events in the 6A and 4C runs. The bias was similar to that in the *msh2*Δ strain but, as would be predicted by the reversion rate differences in these strains, was not nearly as extreme. All of the reversion rate increase in the *msh3*Δ strain can be accounted for by additional mutations at the 6A and 4C runs.

Examination of the frameshift spectrum in the *msh6*Δ strain revealed a distinctive −1 hot spot within a specific 3T run (note that there are six 3T runs within the *lys2*Δ*Bgl* reversion window) that was not evident in any of the other MMR-defective strains examined. The slight increase in the *lys2*Δ*Bgl* reversion rate in the *msh6*Δ strain can be accounted for solely by −1 events in this 3T run. Although this result generally agrees with other studies demonstrating a role of Msh6p in recognition of one-base loops, it also suggests a specificity for Msh6p that has not previously been noted. The basis for this striking specificity is unclear, but it may be related to the surrounding sequence context. Sequence context could influence the frequency with which a specific frameshift intermediate arises, or the efficiency with which an intermediate is repaired. It also is possible that this unique hot spot reflects a role of Msh6p in leading versus lagging strand replication or in repairing frameshift intermediates that arise out of the context of normal DNA replication.

In addition to examining the roles of individual MutS homologs in removing frameshift intermediates, we also examined the roles of two of the yeast MutL homologs: Pms1p and Mlh1p. Although in vivo and in vitro data support the notion that these two proteins function as a heterodimer (32, 33), the *lys2*Δ*Bgl* reversion spectra suggest a subtle difference between a *pms1*Δ strain and an *mlh1*Δ strain. As with the *msh2*Δ strain, the mutation rates were elevated approximately 200-fold in the *pms1*Δ and *mlh1*Δ strains, and more than 95% of the −1 events were in homopolymer tracts of more than 3*N*. As in the *msh2*Δ strain, the *mlh1*Δ strain had reversion events equally distributed between the 6A and 4C tracts. In contrast, there was a 3:1 bias for events in the 4C tract relative to those in the 6A tract in the *pms1*Δ strain. We have no explanation for this bias but suggest that it may reflect a functional difference between the two MutL homologs.

The present study compares frameshift spectra in wild-type versus MMR-defective yeast strains and demonstrates that the MMR machinery very effectively eliminates those frameshift intermediates that arise in mononucleotide runs. Frameshift intermediates arising in noniterated sequences appear to be relatively ineffective targets of the MMR machinery. This result has implications for the types of frameshift mutations that are likely to arise in replication error tumor cells (17, 27). Those genes containing long homopolymer tracts should be particularly susceptible to loss-of-function frameshift mutations, which may in part explain the apparent tissue specificity of tumor development in MMR-defective cells.

ACKNOWLEDGMENTS

We thank Kimberly Plafcan for excellent technical assistance and members of our lab for constructive feedback throughout the course of this work. We thank T. Petes and D. Gordenin for communicating results prior to publication and for helpful comments on the manuscript.

This work was supported by a grant from the National Science Foundation to S.J.-R. C.N.G. was supported in part by the Graduate Division of Biological and Biomedical Sciences at Emory University.

REFERENCES

1. **Alani, E., L. Cao, and N. Kleckner.** 1987. A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. Genetics **116:**541–545.
2. **Bebenek, K., and T. A. Kunkel.** 1990. Frameshift errors initiated by nucleotide misincorporation. Proc. Natl. Acad. Sci. USA **87:**4946–4950.
3. **Boeke, J. D., J. Trueheart, G. Natsoulis, and G. R. Fink.** 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. **154:**164–175.
4. **Cariello, N. F., W. W. Piegorsch, W. T. Adams, and T. R. Skopek.** 1994. Computer program for the analysis of mutational spectra: application to *p53* mutations. Carcinogenesis **15:**2281–2285.
5. **Crouse, G. F.** Mismatch repair systems in *Saccharomyces cerevisiae. In* J. A. Nickoloff and M. F. Hoekstra (ed.), DNA damage and repair, vol. 1. DNA repair in prokaryotes and lower eukaryotes, in press. Humana Press, Totowa, N.J.
6. **Datta, A., A. Adjiri, L. New, G. F. Crouse, and S. Jinks-Robertson.** 1996. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **16:**1085–1093.
7. **Datta, A., and S. Jinks-Robertson.** 1995. Association of increased spontaneous mutation rates with high levels of transcription in yeast. Science **268:** 1616–1619.
8. **Fishel, R., and R. D. Kolodner.** 1995. Identification of mismatch repair genes and their role in the development of cancer. Curr. Opin. Genet. Dev. **5:**382–395.
9. **Geitz, D., A. St. Jean, R. A. Woods, and R. H. Schiestl.** 1992. Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. **20:**1425.
10. **Giroux, C. N., J. R. A. Mis, M. K. Pierce, S. E. Kohalmi, and B. A. Kunz.** 1988. DNA sequence analysis of spontaneous mutations in the *SUP4-o* gene of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **8:**978–981.
11. **Glickman, B. W., and L. S. Ripley.** 1984. Structural intermediates of deletion mutagenesis: a role for palindromic DNA. Proc. Natl. Acad. Sci. USA **81:** 512–516.
12. **Habraken, Y., P. Sung, L. Prakash, and S. Prakash.** 1996. Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. Curr. Biol. **6:**1185–1187.

13. **Hampsey, D. M., J. F. Ernst, J. W. Stewart, and F. Sherman.** 1988. Multiple base-pair mutations in yeast. J. Mol. Biol. **201:**471–486.
14. **Henderson, S. T., and T. D. Petes.** 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **12:**2749–2757.
15. **Hoffman, C. S., and F. Winston.** 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene **57:**267–272.
16. **Hollingsworth, N. M., L. Ponte, and C. Halsey.** 1995. *MSH5*, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. Genes Dev. **9:**1728–1739.
17. **Huang, J., N. Papadopoulos, A. J. McKinley, S. M. Farrington, L. J. Curtis, A. H. Wyllie, S. Zheng, J. K. V. Willson, S. D. Markowitz, P. Morin, K. W. Kinzler, B. Vogelstein, and M. G. Dunlop.** 1996. *APC* mutations in colorectal tumors with mismatch repair deficiency. Proc. Natl. Acad. Sci. USA **93:**9049–9054.
18. **Iaccarino, I., F. Palombo, J. Drummond, N. F. Totty, J. J. Hsuan, P. Modrich, and J. Jiricny.** 1996. MSH6, a *Saccharomyces cerevisiae* protein that binds to mismatches as a heterodimer with MSH2. Curr. Biol. **6:**484–486.

18a. **Johnson, R. E., G. K. Kovvali, S. N. Guzder, N. S. Amin, C. Holm, Y. Habraken, P. Sung, L. Prakash, and S. Prakash.** 1996. Evidence for involvement of yeast proliferating cell nuclear antigen in DNA mismatch repair. J. Biol. Chem. **271:**27987–27990.

19. **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.
20. **Kalinowski, D. P., F. W. Larimer, and M. J. Plewa.** 1995. Analysis of spontaneous frameshift mutations in *REV1* and *rev1-1* strains of *Saccharomyces cerevisiae*. Mutat. Res. **331:**149–159.
21. **Kolodner, R.** 1996. Biochemistry and genetics of eukaryotic mismatch repair. Genes Dev. **10:**1433–1442.
22. **Kramer, W., B. Kramer, M. S. Williamson, and S. Fogel.** 1989. Cloning and nucleotide sequence of DNA mismatch repair gene *PMS1* from *Saccharomyces cerevisiae*: homology of PMS1 to procaryotic MutL and HexB. J. Bacteriol. **171:**5339–5346.
23. **Kroutil, L. C., K. Register, K. Bebenek, and T. A. Kunkel.** 1996. Exonucleolytic proofreading during replication of repetitive DNA. Biochemistry **35:** 1046–1053.
24. **Lea, D. E., and C. A. Coulson.** 1949. The distribution of the numbers of mutants in bacterial populations. J. Genet. **49:**264–285.
25. **Lee, G. S.-F., E. A. Savage, R. G. Ritzel, and R. C. Von Borstel.** 1988. The base-alteration spectrum of spontaneous and ultraviolet radiation-induced forward mutations in the *URA3* locus of *Saccharomyces cerevisiae*. Mol. Gen. Genet. **214:**396–404.
26. **Longerich, S., A. M. Galloway, R. S. Harris, C. Wong, and S. M. Rosenberg.** 1995. Adaptive mutation sequences reproduced by mismatch repair deficiency. Proc. Natl. Acad. Sci. USA **92:**12017–12020.
27. **Markowitz, S., J. Wang, L. Myeroff, R. Parsons, L. Sun, J. Lutterbaugh, R. S. Fan, E. Zborowska, K. W. Kinzler, B. Vogelstein, M. Brattain, and J. K. V. Willson.** 1995. Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability. Science **268:**1336–1338.
28. **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.
29. **Modrich, P., and R. Lahue.** 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem. **65:**101–133.
30. **New, L., K. Liu, and G. F. Crouse.** 1993. The yeast gene *MSH3* defines a new class of eukaryotic MutS homologues. Mol. Gen. Genet. **239:**97–108.
31. **Pribnow, D., C. Sigurdson, L. Gold, B. S. Singer, C. Napoli, J. Brosius, T. J. Dull, and H. F. Noller.** 1981. rII cistrons of bacteriophage T4: DNA sequence around the intercistronic divide and position of genetic landmarks. J. Mol. Biol. **149:**337–376.
32. **Prolla, T. A., D.-M. Christie, and R. M. Liskay.** 1994. Dual requirement in yeast DNA mismatch repair for *MLH1* and *PMS1*, two homologs of the bacterial *mutL* gene. Mol. Cell. Biol. **14:**407–415.
33. **Prolla, T. A., Q. Pang, E. Alani, R. D. Kolodner, and R. M. Liskay.** 1994. MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science **265:**1091–1093.
34. **Quah, S.-K., R. C. Von Borstel, and P. J. Hastings.** 1980. The origin of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics **96:**819–839.
35. **Reenan, R. A. G., and R. D. Kolodner.** 1992. Characterization of insertion mutations in the *Saccharomyces cerevisiae MSH1* and *MSH2* genes: evidence for separate mitochrondrial and nuclear functions. Genetics **132:**975–985.
36. **Ripley, L. S.** 1990. Frameshift mutation: determinants of specificity. Annu. Rev. Genet. **24:**189–213.
37. **Ripley, L. S., A. Clark, and J. G. deBoer.** 1986. Spectrum of spontaneous frameshift mutations: sequences of bacteriophage T4 *rII* gene frameshifts. J. Mol. Biol. **191:**601–613.
38. **Ross-Macdonald, P., and G. S. Roeder.** 1994. Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction. Cell **79:**1069–1080.
39. **Rothstein, R.** 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. **194:**281–301.
40. **Schaaper, R. M., and R. L. Dunn.** 1987. Spectra of spontaneous mutations in *Escherichia coli* strains defective in mismatch correction: the nature of *in vivo* DNA replication errors. Proc. Natl. Acad. Sci. USA **84:**6220–6224.
41. **Schaaper, R. M., and R. L. Dunn.** 1991. Spontaneous mutation in the *Escherichia coli lacI* gene. Genetics **129:**317–326.
42. **Selva, E. M., L. New, G. F. Crouse, and R. S. Lahue.** 1995. Mismatch correction acts as a barrier to homeologous recombination in *Saccharomyces cerevisiae*. Genetics **139:**1175–1188.
43. **Sherman, F.** 1991. Getting started with yeast. Methods Enzymol. **194:**3–20.

43a. **Sia, E. A., R. J. 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.

44. **Singer, B. S., and J. Westlye.** 1988. Deletion formation in bacteriophage T4. J. Mol. Biol. **202:**233–243.
45. **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.
46. **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.
47. **Strathern, J. N., B. Shafer, and C. B. McGill.** 1995. DNA synthesis errors associated with double-strand-break repair. Genetics **140:**965–972.
48. **Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye.** 1966. Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:**77–84.
49. **Tran, H. T., N. P. Degtyareva, N. N. Koloteva, A. Sugino, H. Masumoto, D. A. Gordenin, and M. A. Resnick.** 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.
50. **Tran, H. T., D. A. Gordenin, and M. A. Resnick.** 1996. The prevention of repeat-associated deletions in *Saccharomyces cerevisiae* by mismatch repair depends on size and origin of deletions. Genetics **143:**1579–1587.

50a. **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.

51. **Umar, A., A. B. Buermeyer, J. A. Simon, D. C. Thomas, A. B. Clark, R. M. Liskay, and T. A. Kunkel.** 1996. Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. Cell **87:**65–73.
52. **Umar, A., and T. A. Kunkel.** 1996. DNA-replication fidelity, mismatch repair and genome instability in cancer cells. Eur. J. Biochem. **238:**297–307.