Exonucleolytic Proofreading during Replication of Repetitive DNA

Lisa C. Kroutil, Kelly Register, Katarzyna Bebenek, and Thomas A. Kunkel*

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Received September 12, 1995; Revised Manuscript Received November 22, 1995[⊗]

ABSTRACT: We are attempting to understand the processes required to accurately replicate the repetitive DNA sequences whose instability is associated with several human diseases. Here we test the hypothesis that the contribution of exonucleolytic proofreading to frameshift fidelity during replication of repetitive DNA sequences diminishes as the number of repeats in the sequence increases. The error rates of proofreading-proficient T7, T4, and *Pyrococcus furiosis* DNA polymerases are compared to their exonuclease-deficient derivatives, for +1 and −1 base errors in homopolymeric repeat sequences of three to eight base pairs. All three exonuclease-deficient polymerases produce frameshift errors during synthesis at rates that increase as a function of run length, suggesting the involvement of misaligned intermediates. Their wild-type counterparts are all much more accurate, suggesting that the majority of the intermediates are corrected by proofreading. However, the contribution of the exonuclease to fidelity decreases substantially as the length of the homopolymeric run increases. For example, the exonuclease enhances the frameshift fidelity of T7 DNA polymerase in a run of three A·T base pairs by 160-fold, similar to its contribution to base substitution fidelity. However, in a run of eight consecutive A·T base pairs, the exonuclease only enhances frameshift fidelity by 7-fold. A similar pattern was observed with T4 and *Pfu* DNA polymerases. Thus, both polymerase selectivity and exonucleolytic proofreading efficiency are diminished during replication of repetitive sequences. This may place an increased relative burden on post-replication repair processes to reduce rates of addition and deletion mutations in organisms whose genome contains abundant simple repeat DNA sequences.

The fidelity of a DNA polymerization reaction depends on the selectivity of the DNA polymerase and on the ability of its associated 3′→5′ exonuclease to proofread errors. DNA polymerases favor correct over incorrect dNTP[1] incorporation by factors of 10^4-10^5, depending on the polymerase, the mispair, and its location [for a review, see Roberts and Kunkel (1996)]. Proofreading enhances the base substitution fidelity of DNA polymerases up to several hundredfold (Eger *et al.*, 1991; Donlin *et al.*, 1991; West Frey *et al.*, 1993; Schaaper, 1993; Cai *et al.*, 1995). Editing efficiency primarily depends on kinetic partitioning of the terminal base pair between the polymerase and exonuclease active sites [reviewed in Johnson (1993)]. For a correctly paired template-primer terminus, polymerization is greatly favored over exonucleolytic cleavage (Figure 1, top). However, polymerization from a terminal mismatch is much slower (Figure 1, second line), providing additional time for exonucleolytic proofreading. Moreover, the DNA polymerase active site is designed to accommodate duplex (i.e., paired) DNA, whereas the exonuclease active site preferentially binds single-stranded DNA [reviewed in Joyce and Steitz (1994)]. A terminal mispair is thought to induce fraying (Brutlag & Kornberg, 1972) to yield a single-stranded 3′-OH terminus that preferentially binds to the exonuclease site, allowing hydrolysis of the misincorporated nucleotide.

These concepts on the roles of template-primer breathing and kinetic partitioning have been derived from 25 years of studies of base substitution fidelity. However, even before this, Streisinger *et al.* (1966) proposed that base addition and deletion errors could arise via a different process, strand slippage at repetitive DNA sequences to generate misaligned premutational intermediates (e.g., last line, Figure 1). Interest in this hypothesis has increased greatly in the last few years with the observation that several hereditary diseases are associated with expansion of specific triplet repeat sequences [reviewed in Willems (1994)] and that repetitive sequences are highly unstable in the genome of cells from a variety of human tumors [reviewed in Marra and Boland (1995) and in Eshleman and Markowitz (1995)]. Understanding the origins of these disease-associated instabilities may be

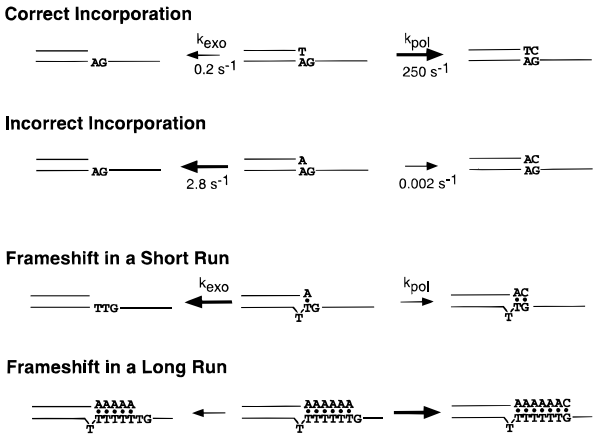

FIGURE 1: Kinetic partitioning of misinsertion and misalignment errors. The kinetic constants shown are for T7 DNA polymerase, taken from Wong *et al.*, (1991). Boldface arrows indicate a fast rate of either polymerization or hydrolysis relative to the opposite reaction.

* To whom correspondence should be addressed. Phone: (919)-541-2644. Fax: (919)-541-7613. Email: kunkel@niehs.nih.gov.

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1996.

[1] Abbreviations: dNTP, deoxyribonucleoside triphosphates; DTT, dithiothreitol; *Pfu*, *Pyrococcus furiosis*.

facilitated by studies of the basic processes operating in cells to control the rate of frameshift[2] mutations.

The idea that slippage occurs during DNA synthesis to generate misaligned intermediates is supported by studies of the fidelity of DNA synthesis by DNA polymerase *in vitro* [reviewed in Kunkel (1990)]. DNA polymerases generate single-base deletion and addition errors and do so preferentially at homopolymeric repeat sequences and at a rate that increases with the number of repeats in the sequence (Kunkel, 1990; Kunkel et al., 1994; Bebenek et al., 1995). Moreover, polymerase frameshift error hot spots at homopolymeric runs are eliminated by interruptions with nonidentical template nucleotides (Kunkel, 1986; Bebenek et al., 1993). These observations suggest that the intermediates for a deletion or an addition arising in a polymerization reaction contain paired termini with unpaired nucleotides in the template or the primer strand, respectively. The relationship between run length and error rate further implies that the unpaired nucleotide may be located as far away as possible from the 3′-OH primer terminus, allowing the maximum number of correct base pairs in the repeated sequence to stabilize the intermediate and also removing any distortion resulting from the presence of the unpaired nucleotide the greatest distance from the polymerase active site.

This logic suggests that an unpaired nucleotide in a short run may be somewhat similar to a terminal mismatch in terms of fraying and kinetic partitioning (Figure 1, third line), and thus may be edited efficiently. However, for misalignment in a longer run, the primer terminus may fray less and partition more in favor of the polymerase rather than the exonuclease (Figure 1, line 4), resulting in decreased proofreading efficiency. In the present study, we test this hypothesis by comparing the frameshift error rates of three DNA polymerases having inherent 3′→5′ exonuclease activities to their respective exonuclease-deficient derivatives. The contribution of exonuclease activity to frameshift fidelity is estimated during DNA synthesis templated by homopolymeric tracts containing varying numbers of nucleotides.

MATERIALS AND METHODS

Strains and Reagents. The *Escherichia coli* strains (MC1061, CSH50, and NR9099), reagents, and preparation of substrates have been described (Bebenek & Kunkel, 1995). T7 DNA polymerase and its exonuclease-deficient D5A,-E7A derivative were kind gifts of Smita S. Patel and Kenneth A. Johnson (Patel *et al.*, 1991) . These proteins were isolated in the absence of thioredoxin. T4 DNA polymerase and its exonuclease-deficient D219A derivative were generously provided by Michelle West Frey and Stephen J. Benkovic (West Frey et al., 1993). Exonuclease-proficient and -deficient *Pyrococcus furiosus* (*Pfu*) DNA polymerases were provided by Eric Mathur of Stratagene.

DNA Synthesis Reactions. T7 and T4 DNA polymerase reactions were performed as previously described (Kunkel *et al.*, 1994; Pavlov *et al.*, 1994). *Pfu* polymerase reactions (25 μL) contained 20 mM Tris-HCl (pH 8.7), 10 mM $MgCl_2$, 2 mM DTT, 200 μM dNTPs, 45 fmol of the designated gapped M13mp2 DNA substrate, and either 370 fmol of wild-type or 710 fmol of mutant enzyme. Reactions were incubated at 72 °C for 5 min and quenched by adding EDTA to a final concentration of 15 mM. All reactions completely filled the gap, as determined by agarose gel electrophoresis (Bebenek & Kunkel, 1995).

Fidelity Assays. The forward assay scores mutations that result from polymerization to fill a single-stranded gap containing the *LacZ*α gene in an M13mp2 DNA substrate (Kunkel, 1985). Correct gap-filling produces DNA that yields blue plaques upon transfection into MC1061 and plating on CSH50. Errors are scored as lighter blue or colorless plaques. Reversion assays use gapped substrates encoding a colorless plaque phenotype due to frameshift mutations [constructed as described in Kunkel *et al.* (1991)] in runs of 3−8 consecutive A·T base pairs. Polymerase errors that restore the correct reading frame ($3n-1$ or $3n+1$) yield blue revertants. Frameshifts within the run produce dark blue revertant plaques, as confirmed for all assays by DNA sequence analysis of revertants. Light blue revertants due to mutations at sites other than the run (also confirmed by sequence analysis) were not used to calculate error rates. Procedures for these assays have been described in detail (Bebenek & Kunkel, 1995).

RESULTS

Frameshift Fidelity of T7 DNA Polymerase. For the initial test of the hypothesis, we studied the frameshift fidelity of T7 DNA polymerase. This enzyme was chosen for four reasons. First, it is a replicative DNA polymerase with an inherent 3′→5′ exonuclease activity that strongly contributes to base substitution fidelity (Donlin *et al.*, 1991). Second, the fidelity of the wild-type enzyme can be directly compared to that of a mutant protein that retains DNA polymerase activity but lacks exonuclease activity due to changes in two amino acid residues (D5A,E7A) that are highly conserved among all proofreading-proficient DNA polymerases. These amino acids are thought to reside in the exonuclease active site. This site is physically distant from the polymerase active site, and the characteristics of the DNA polymerase of the mutant protein are relatively normal. Thus, any differences observed in the fidelity of the wild-type and mutant protein are inferred to reflect the contribution of exonucleolytic proofreading. Third, in T7-infected *E. coli* cells, the polymerase forms a tight, 1:1 complex with thioredoxin, a 12 kDa host protein that confers stability and high processivity to the polymerase. We have previously shown that the exonuclease-deficient DNA polymerase lacking thioredoxin has low frameshift fidelity (Kunkel *et al.*, 1994) particularly for single-nucleotide additions in homopolymeric runs. Because other DNA polymerases readily generate one-nucleotide deletions but rarely generate addition errors (Kunkel, 1990), the use of this polymerase permits the contribution of the exonuclease to frameshift fidelity to be examined for both types of frameshift errors. Finally, the frameshift error rate of the polymerase increases as a function of increasing homopolymeric run length (Kunkel *et al.*, 1994), a key observation on which the current hypothesis is based.

We first compared the fidelity of the wild-type[3] and mutant T7 polymerases in the forward mutation assay where

[2] Although the term frameshift mutation usually refers to changes in the number of base pairs in a protein-coding sequence that are not multiples of three, for convenience it refers here to mutations resulting from any difference in the number of base pairs and regardless of location.

[3] In this study, wild-type T7 DNA polymerase refers to the T7 gene 5 protein in the absence of thioredoxin, and mutant refers to the exonuclease-deficient derivative of the gene 5 protein.

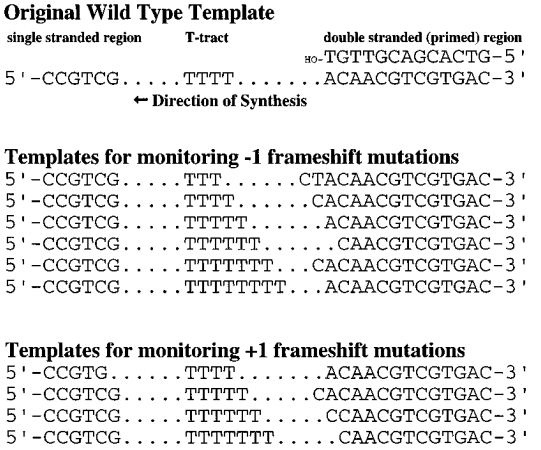

FIGURE 2: Template DNA sequences of reversion substrates. The original wild-type (i.e., blue) sequence of the LacZ α-complementation gene in M13mp2 is shown at the top. The 5′-most T is nucleotide 70, where +1 is the first transcribed nucleotide. The double-stranded primer stem is shown here merely to orient the reader. Synthesis actually initiates from a 3′-OH provided at nucleotide +145 in the the gapped substrates. Frameshifts within the T runs produce dark blue revertant plaques, as confirmed for all assays by DNA sequence analysis of revertants. Light blue revertants result from mutations at sites other than the run, as also confirmed by sequence analysis. Light blue revertants were not used to calculate error rates.

Table 1: Mutation Frequencies of T7 DNA Polymerases by Class, in the Forward Assay[a]

	wild type		exonuclease deficient	
type of mutation	mutants	mut freq ($\times 10^{-4}$)	mutants	mut freq ($\times 10^{-4}$)
base substitutions	9	2.5	18	44
C→T	37	10	3	7.3
minus-one	2	0.55	20	48
minus-two	0	≤0.27	11	27
plus-one	21	5.8	21	51
317-base deletion	9	2.5	0	≤2.4
others	6	1.6	22	53

[a] A total of 84 and 95 mutants were sequenced, respectively, for the wild-type and mutant enzymes, from 45 min reactions containing 1 mM dNTPs and 1.8 pmol of polymerase. "Others" include mutants containing more than one change, deletions, or complex sequence changes. The background frequency of the uncopied control for each mutant class was as follows: base substitutions other than C→T, 1.9×10^{-4}; C→T, 1.6×10^{-4}; minus-one, 0.4×10^{-4}; minus-two, $\leq 0.05 \times 10^{-4}$; plus-one, 0.15×10^{-4}; 317 base deletion, $\leq 0.05 \times 10^{-4}$; others, 2.5×10^{-4}.

synthesis is templated by the wild-type *LacZ*α gene in M13mp2 (original wild-type template in Figure 2). The gap contains 199 nucleotides that are sensitive to frameshift mutagenesis (Bebenek & Kunkel, 1995). Gap-filling synthesis followed by transfection of *E. coli* cells with the products of these reactions yielded *LacZ* mutant frequencies of 23×10^{-4} and 230×10^{-4} for the wild-type and mutant enzymes, respectively. Both values are above the mutation frequency of uncopied DNA (6×10^{-4}), indicating that errors were made by both DNA polymerases during synthesis *in vitro*.

To define the error specificities of the enzymes, DNA sequence analysis of independent mutants was performed. A variety of mutations were observed in both mutant collections (Table 1). The mutation frequencies for base substitutions and one- and two-nucleotide deletions recovered from the wild-type polymerase reaction were similar to those obtained with uncopied DNA (legend to Table 1), suggesting that the proofreading-proficient polymerase was highly accurate for errors that would result in these types of mutations. The frequencies for plus-one nucleotide frameshift errors and a specific 317-nucleotide deletion flanked by a pentanucleotide direct repeat are 120-fold and 50-fold higher, respectively, than those observed with uncopied DNA, suggesting that the wild-type polymerase is error-prone for these mistakes despite the exonuclease activity.

For the exonuclease-deficient polymerase, the mutation frequency is higher than that obtained with the wild-type polymerase for most classes of mutations, providing an estimate of the contribution of the exonuclease to fidelity for each class. Thus, the 87-fold, 100-fold, and 9-fold differences for minus-one, minus-two and plus-one nucleotide errors suggest that, on average, the frameshift intermediates that lead to these mutations are efficiently proofread. For base substitutions, the 18-fold difference (Table 1, 2.5×10^{-4} versus 44×10^{-4}) provides a minimum estimate of the contribution of the exonuclease to fidelity, since the former value does not reflect polymerase errors because it is not above the mutation frequency for uncopied DNA. The frequency of C→T transitions is similar for both polymerase reactions, indicating that this "error" is not proofread. We have suggested earlier (Cai *et al.*, 1995) that these mutations result from correct incorporation of dAMP opposite uracil in the gapped substrate resulting from cytosine deamination. This intermediate would not be subject to proofreading. Likewise, the comparative frequency data for the 317-nucleotide deletion suggest that the intermediate responsible for this mutation (suggested to contain five correct terminal base pairs, see below) is also not efficiently proofread.

To estimate the contribution of the exonuclease to frameshift fidelity as a function of run length, we calculated the average frameshift error rate per nucleotide polymerized at non-run sequences and at the homopolymeric runs in the template sequences that are 2, 3, 4, and 5 nucleotides in length. We reported earlier that, in the absence of thioredoxin, exonuclease-deficient T7 DNA polymerase is error-prone for one-nucleotide additions in runs, and that the error rate increases with increasing run length (Kunkel *et al.*, 1994). The present study confirms that observation (Figure 3B) and then compares the error rates of the exonuclease-proficient polymerase under identical reaction conditions. The latter polymerase is more accurate at all run lengths (Figure 3A). However, the difference in error rate between the two enzymes diminishes as the run length increases (Figure 3C), consistent with the hypothesis we set out to test.

Eleven of the 20 one-nucleotide deletions recovered from the exonuclease-deficient polymerase reaction were at noniterated nucleotides, while no such mutants were obtained from the wild-type polymerase reaction. The difference in error rate between the wild-type and mutant enzymes for loss of a noniterated nucleotide is thus ≥100-fold (error rates of $\leq 4.6 \times 10^{-7}$ and 4.6×10^{-5}, respectively). However, a similar calculation based on the difference at the one five-nucleotide run in the target is only 11-fold (error rates of 8.1×10^{-6} and 9.1×10^{-5}, respectively). Thus, as for addition errors, the difference in error rate between the two enzymes for one-nucleotide deletions diminishes as the run length increases.

Frameshift Fidelity in Reversion Assays. The above analysis involves small numbers of frameshift mutants and relatively short homopolymeric runs, with none longer than

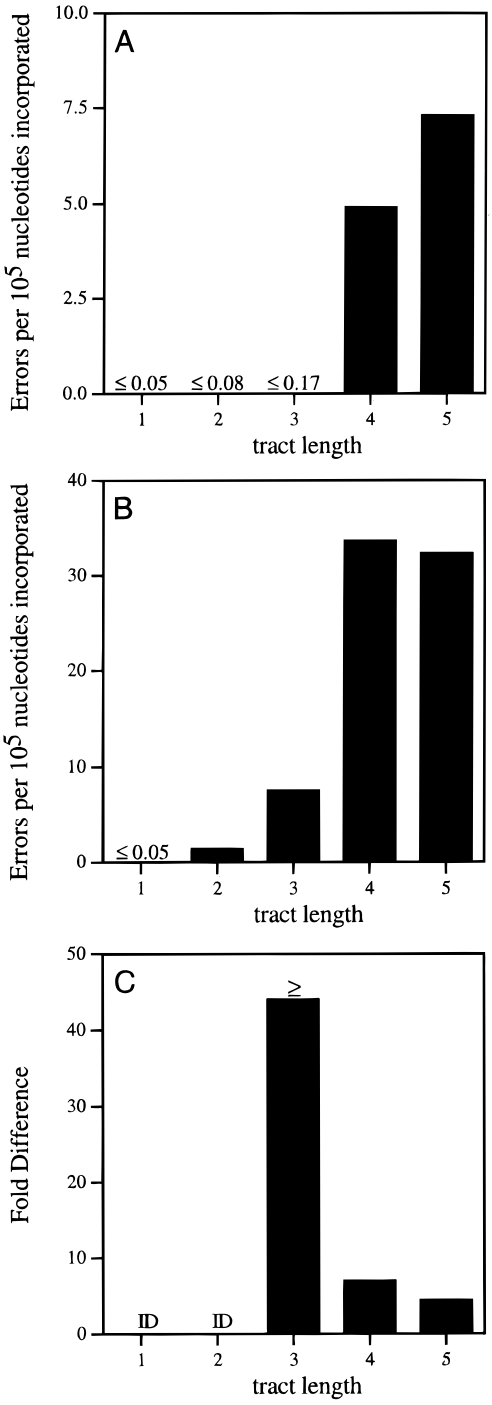

FIGURE 3: One-nucleotide addition error rates of T7 DNA polymerases in homopolymeric tracts of increasing length. In the template sequence copied, the target sizes for scoring frameshift errors within runs of 1, 2, 3, 4, and 5 nucleotides are 97, 58, 27, 12, and 5 nucleotides, respectively (Bebenek & Kunkel, 1995). The number of addition mutants for each of these run lengths, respectively, were as follows: wild-type polymerase, 0, 0, 0, 13, 8 mutants; exo-deficient polymerase, 0, 2, 5, 11, 4 mutants. Error rates were calculated by dividing the number of mutants observed for any run length by the total number of mutants sequenced. This value was multiplied by the total mutation frequency and then divided by 0.6, the probability of expressing an error in *E. coli* (Bebenek & Kunkel, 1995). This value was then divided by the number of detectable nucleotides for scoring that class of error, to correct for different target sizes and expression rates "per detectable nucleotide polymerized", these values were then plotted against run length for the exonuclease-proficient polymerase (panel A) and the exonuclease-deficient polymerase (panel B). Fold effects were obtained as the ratio of the error rates and were indeterminate (ID) for runs of three or less since no mutations were scored for the wild-type enzyme; these values are plotted against run length in panel C.

Table 2: One-Nucleotide Deletion Fidelity of Wild-Type and Mutant T7 DNA Polymerases[a]

	plaques scored		
substrate	total	dark blue	rev freq ($\times 10^{-5}$)
Exonuclease-Proficient T7 DNA Polymerase			
T3	670 000	0	≤0.15
T4	360 000	2	0.6
T5	370 000	28	7.6
T6	160 000	41	26
T7	43 000	41	95
T8	44 000	79	180
Exonuclease-Deficient T7 DNA Polymerase			
T3	220 000	53	24
T4	280 000	238	84
T5	630 000	845	130
T6	89 000	109	120
T7	33 000	140	430
T8	66 000	705	1100

[a] See text for description. The reversion frequencies of uncopied DNAs were as follows: T3, $\leq 0.08 \times 10^{-4}$; T4, $\leq 0.02 \times 10^{-4}$; T5, 0.13×10^{-4}; T6, 0.11×10^{-4}; T7, 0.23×10^{-4}; T8, 2×10^{-4}.

five base pairs and only one of this length. Moreover, use of a forward assay necessarily involves comparison of runs that differ not only in length but also in flanking nucleotide composition, a parameter that can strongly influence frameshift fidelity (Bebenek *et al.*, 1993). A more rigorous test of the hypothesis was thus performed using reversion assays that score either addition or deletion errors in runs of A·T base pairs. The run was constructed to contain from three to eight T(template)·A(primer) base pairs with minimal variation in the flanking sequence (Figure 2).

Synthesis by the exonuclease-proficient enzyme, first using the substrates that score one-nucleotide deletions, generated reversion frequencies (Table 2) from which error rates were calculated (Figure 4). The deletion rate per template, T, copied increased from $\leq 0.08 \times 10^{-5}$ for the T3 substrate to 33×10^{-5} for the T8 substrate (Figure 4A), a difference of ≥410-fold. The fidelity of the exonuclease-deficient polymerase also decreased as the run length increased, and the error rate was significantly higher relative to the wild-type enzyme at all run lengths (Table 2 and Figure 4B). The difference in error rate between the wild-type and mutant enzyme decreased from ≥160-fold with the T3 substrate to only 7-fold with the T8 substrate (Figure 4C). The greatest increment of decrease was observed as the run increased from four to five A·T base pairs.

Synthesis using the substrates with the homopolymeric runs in the other reading frame yielded reversion frequencies well above background with all substrates and both enzymes (Figure 4D,E), consistent with the high error rate of the thioredoxin-deficient polymerase for one-nucleotide addition errors (Kunkel *et al.*, 1994; Table 1). The error rate for the wild-type enzyme increased from 1.1×10^{-5} for the T4 substrate to 12×10^{-5} for the T7 substrate. This 11-fold difference is substantially less than for the equivalent runs in the other reading frame, which yielded a 160-fold increase when comparing runs of 4 versus 7 Ts. The same trend was observed with the exonuclease-deficient polymerase, where the error rate increased with increasing tract length, but only 2-fold as the run increased from 4 to 7 Ts (75×10^{-5} for the run of 4 Ts versus 150×10^{-5} for the run of 7 Ts). Despite the lesser dependence of error rate on length, the apparent contribution of the exonuclease to addition frame-

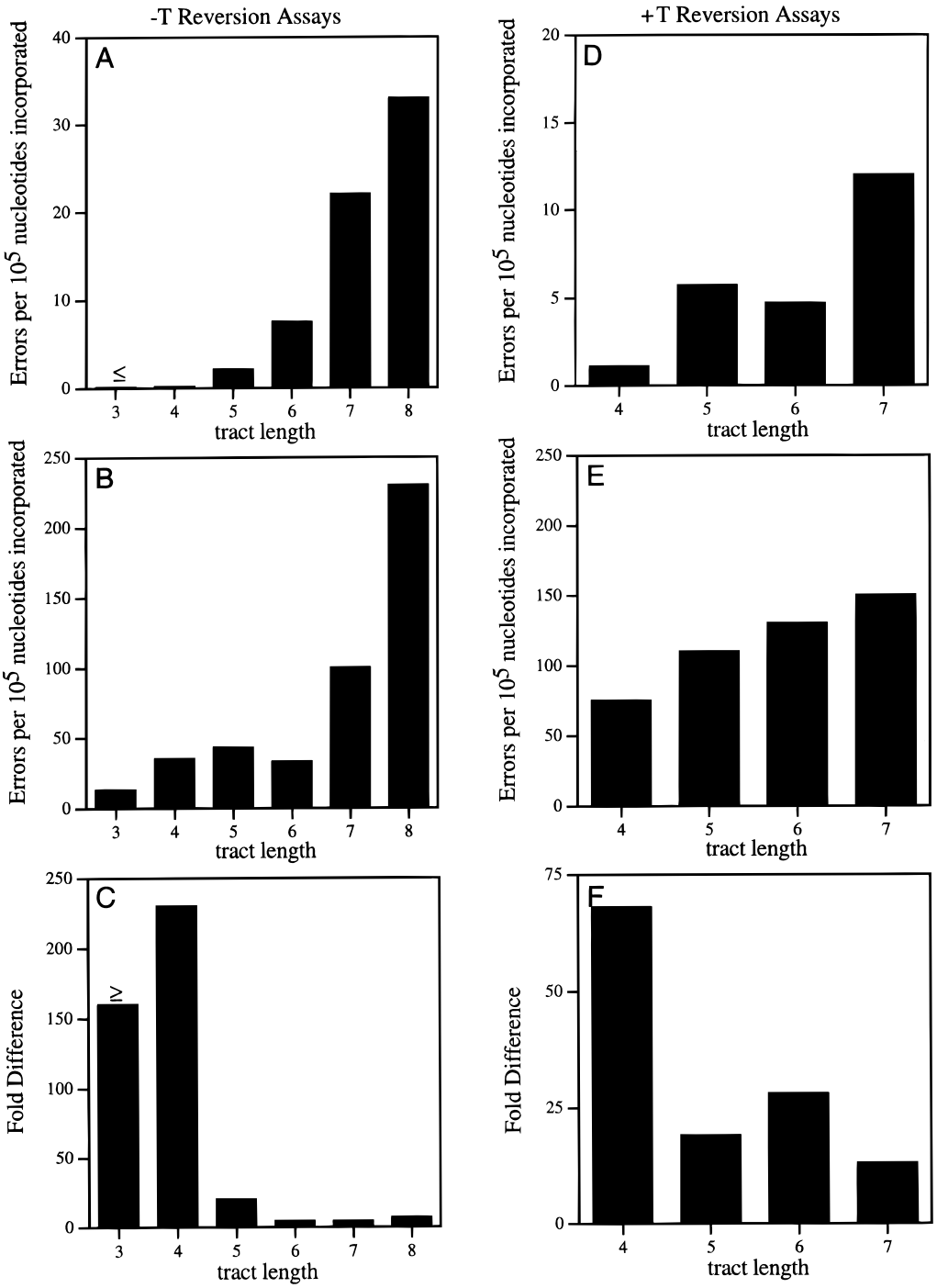

FIGURE 4: T7 DNA polymerase error rates for one-nucleotide deletions and additions as a function of run length. Error rates were calculated as described in the legend to Figure 3, using as the target size the number of T residues in each run and the reversion rates in Table 2. Panel A: deletion error rates for wild-type polymerase. Panel B: deletion error rates for exonuclease-deficient polymerase. Panel C: mutant: wild-type polymerase deletion error rate ratios. Panel D: addition error rates for wild-type polymerase. Panel E: addition error rates for exonuclease-deficient polymerase. Panel F: mutant:wild-type polymerase addition error rate ratios. The reversion frequencies of uncopied DNA for the one-nucleotide addition substrates were as follows: T4, 0.13×10^{-4}; T5, $\leq 0.08 \times 10^{-4}$; T6, 0.07×10^{-4}; T7, 0.83×10^{-4}. Following polymerization, the revertant and total plaque counts and reversion frequencies at each run length for the one-nucleotide addition substrates were as follows: wild-type polymerase: T4, 11/280 000, 0.39×10^{-4}; T5, 27/150 000, 1.8×10^{-4}; T6, 56/320 000, 1.8×10^{-4}; T7, 21/37 000, 5.7×10^{-4}; exonuclease-deficient polymerase: T4, 526/300 000, 18×10^{-4}; T5, 254/75 000, 34×10^{-4}; T6, 2556/561 000, 46×10^{-4}; T7, 296/46 000, 64×10^{-4}.

shift fidelity did decrease with increasing run length. For example, comparison of the error rates of the two polymerases with the 4T run substrate reveals a 68-fold, difference (Figure 4F), whereas a 13-fold decrease in fidelity was observed with the T7 substrate.

Proofreading of One-Base Deletions by T4 DNA Polymerase. To test the generality of the relationship between proofreading efficiency and increasing tract length, we examined the fidelity of wild-type and exonuclease-deficient mutants of T4 DNA polymerase. As with the T7 DNA polymerase, the T4 mutant protein contains substitutions in conserved amino acids thought to be in the exonuclease active site and known not to impair polymerase properties. Studies of the overall fidelity of the wild-type and mutant T4 polymerases using the forward mutation assay[4] revealed that, like most DNA polymerases, they rarely generate one-

[4] T. A. Kunkel, unpublished observations.

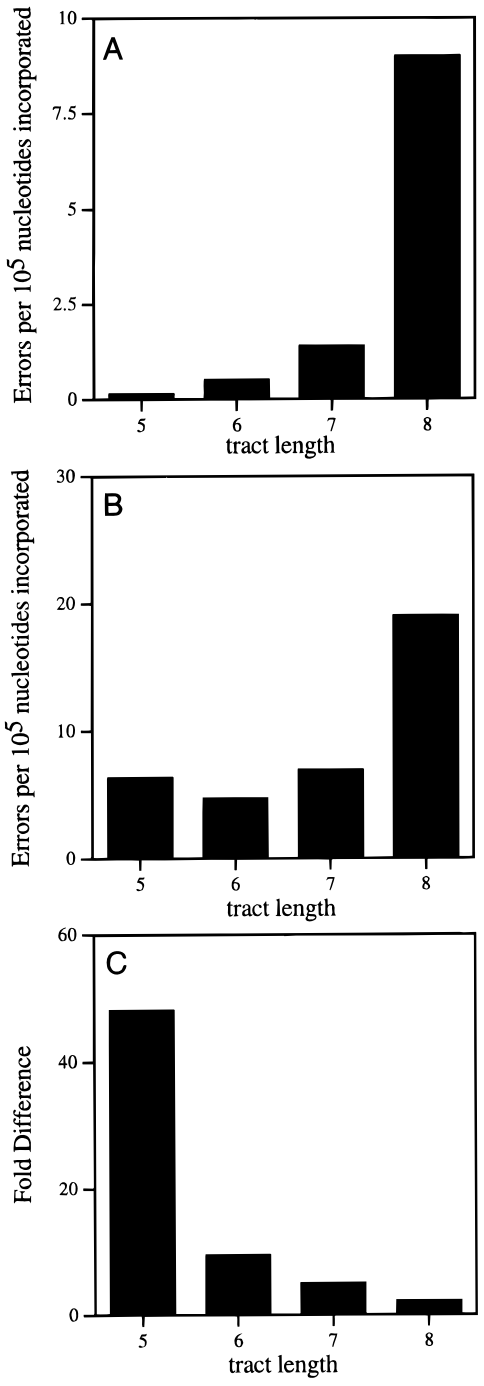

FIGURE 5: T4 DNA polymerase error rates for one-nucleotide deletions as a function of run length. Error rates were calculated as described in the legend to Figure 3, using as the target size the number of T residues in each run and the following reversion data: wild-type T4 polymerase: T5, 9/530 000, 0.17×10^{-4}; T6, 9/310 000, 0.29×10^{-4}; T7, 23/280 000, 0.82×10^{-4}; T8, 63/100 000, 6.3×10^{-4}; exonuclease-deficient T4 polymerase: T5, 42/210 000, 2.0×10^{-4}; T6, 39/220 000, 1.8×10^{-4}; T7, 55/180 000, 3.1×10^{-4}; T8, 125/110 000, 11×10^{-4}. Panel A: deletion error rates for wild-type polymerase. Panel B: deletion error rates for exonuclease-deficient polymerase. Panel C: mutant: wild-type polymerase deletion error rate ratios.

nucleotide addition errors. Thus, we examined fidelity using substrates that score one-nucleotide deletions. The error rates of the wild-type T4 DNA polymerase increased with tract length, with values ranging from 0.13×10^{-5} for the T5 substrate to 9×10^{-5} for the T8 substrate (Figure 5A). The mutant polymerase was less accurate, with rates ranging from 6.3×10^{-5} for T5 to 19×10^{-5} for T8 (Figure 5B). As for the T7 DNA polymerase, the difference in error rate between the wild-type and mutant T4 DNA polymerases decreased as the tract length increased (Figure 5C).

Proofreading of One-Base Deletions at High Temperature by Pfu DNA Polymerase. The efficiency of proofreading has been related to the stability of the template-primer; i.e., its tendency to "fray" favors exonuclease over polymerase activity. We therefore examined the contribution of exonucleolytic proofreading to frameshift fidelity during polymerization at an elevated reaction temperature that could enhance fraying, by comparing the fidelity of the wild-type and exonuclease-deficient mutant of the thermostable DNA polymerase from *Pyrococcus furiosis*. Error rates were determined in the forward mutation assay in reactions catalyzed by *Pfu* DNA polymerase at 72 °C. The complete analysis of error specificity will be presented elsewhere,[5] but the results for one-nucleotide deletion errors are shown in Figure 6. The wild-type polymerase, which contains an inherent and highly active proofreading exonuclease activity (Lundberg *et al.*, 1991), is highly accurate for one-nucleotide deletions, but its error rate increases with increasing tract length (Figure 6A). The exonuclease-deficient mutant is less accurate, consistent with a role for the exonuclease in correcting frameshift errors. There is an obvious trend toward lower fidelity in longer runs (Figure 6B). Thus, the comparative fidelity of the two polymerases suggests that, in reactions incubated at 72 °C, frameshift errors are proofread, but less efficiently as the run length increases (Figure 6C).

DISCUSSION

This study was undertaken to test the hypothesis that the contribution of exonucleolytic proofreading to frameshift fidelity during replication of repetitive DNA sequences diminishes as the number of repeats in the sequence increases. This idea is partly based on Streisinger's original hypothesis that frameshift errors are initiated by strand slippage to generate misaligned template-primers stabilized by the correct base pairs that are possible if slippage occurs within repetitive DNA sequences. The model predicts that the frameshift error rate during synthesis should increase as the number of consecutive repeats increases because as the repeat number goes up, so does the potential number of stabilizing base pairs and misaligned intermediates and the distance between the polymerase active site and the unpaired nucleotide(s). Previous studies with human DNA polymerase α (Kunkel, 1986, 1990), HIV-1 reverse transcriptase (Bebenek *et al.*, 1995), and T7 DNA polymerase (Kunkel *et al.*, 1994) have demonstrated such frequency versus length correlations. This study has extended these observations to additional enzymes and substrates, offering compelling evidence for the involvement of misaligned intermediates in the production of frameshift mutations.

Our starting hypothesis was also suggested by the concepts of template-primer fraying and kinetic partitioning that were derived from studies of DNA polymerase fidelity for misincorporation errors. The decreased stability of the duplex template-primer stem conferred by a terminal mispair, the preference of the polymerase active site for binding duplex DNA, and the preference of the exonuclease active site for binding to and digesting single-stranded DNA

[5] K. Register, K. Bebenek, and T. A. Kunkel, manuscript in preparation.

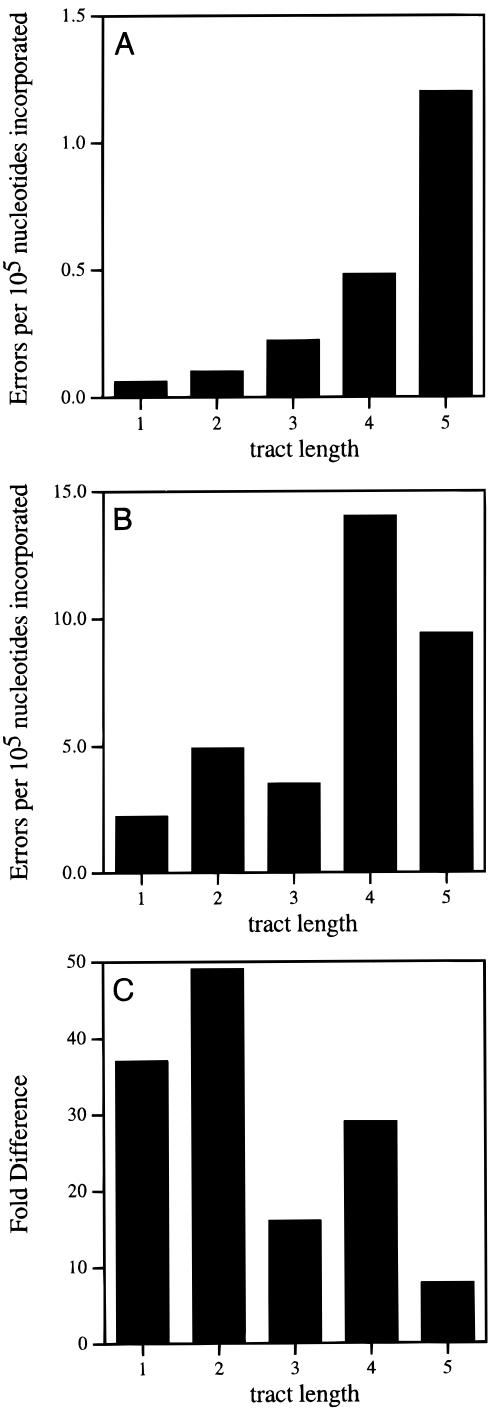

FIGURE 6: One-nucleotide deletion error rates of *Pfu* DNA polymerases in homopolymeric tracts of increasing length. Error rates were calculated as described in the legend to Figure 3. The number of deletion mutants for runs of 1, 2, 3, 4, or 5 nucleotides were as follows: wild-type polymerase, 1, 1, 0, 1, and 1 mutants of 35 sequenced; exonuclease-deficient polymerase: 9, 12, 4, 7, and 2 mutants of 113 sequenced. Mutation frequencies were 12×10^{-4} for the wild-type enzyme and 160×10^{-4} for the exonuclease-deficient enzyme. Panel A: deletion error rates for wild-type polymerase. Panel B: deletion error rates for exonuclease-deficient polymerase. Panel C: mutant:wild-type polymerase deletion error rate ratios.

[reviewed in Joyce and Steitz (1994)] are realized in the preferential rate of polymerization from correct terminal pairs and the selective hydrolysis of a misincorporated nucleotide [Figure 1, and reviewed in Johnson (1993)]. Our results strongly suggest that these same principles are relevant to premutational intermediates involving misaligned template-primers. The data in Figures 3C, 4C, 4F, 5C, and 6C all suggest that the contribution of exonucleolytic proofreading to the frameshift fidelity of a DNA polymerization reaction diminishes as the homopolymeric tract length increases. For deletions of noniterated nucleotides, the contribution of the exonuclease to the error rate is similar to the efficiency with which misincorporation errors are proofread. This observation is consistent with our previous suggestion that some frameshift errors are initiated not by misalignment, but by misincorporation followed by primer relocation (Kunkel & Soni, 1988; Bebenek & Kunkel, 1990; Bebenek *et al.*, 1993). However, despite the fact that T7 and T4 DNA polymerases are among the most accurate DNA polymerases known and have robust proofreading exonucleases, the exonuclease contributes only 2−13-fold to frameshift fidelity in tracts of seven or eight T·A base pairs. Thus, proofreading contributes much less to frameshift fidelity in longer repeats, consistent with misaligned intermediates that are sufficiently stable to be seen more often by the polymerase as correct substrates and partitioned in favor of polymerization rather than hydrolysis (Figure 1, bottom).

Decreasing frameshift fidelity with increasing tract length is observed for reactions catalyzed by several DNA polymerases, suggesting that frameshift error rates are strongly determined by the properties of the DNA substrate. Nonetheless, frameshift error rates vary widely among different polymerases when the same substrate is used [for review, see Kunkel (1990)], illustrating the important influence of the protein. As examples in this study, note that T7 polymerase is consistently less accurate for one-nucleotide deletions than is T4 DNA polymerase (compare Figure 4A,B to Figure 5A,B) and that the increment of increase as a function of tract length by one nucleotide is not constant in all situations. Such effects may result from any of several differences in how polymerases interact with their substrates. The increment of change with tract length may vary depending on which strand contains the unpaired nucleotide(s), its position relative to the polymerase active site, and the spatial relationship and communication between the polymerase and exonuclease active sites. Since it may also depend on the number and types of unpaired nucleotides, it will be interesting to examine frameshift fidelity and the contribution of proofreading during replication of other simple repetitive sequences, such as di- and trinucleotide repeats. Eukaryotic cells, which contain many such microsatellites, may have evolved proofreading processes to deal with these sequences, that are not apparent from this study with homopolymeric sequences and prokaryotic DNA polymerases.

An unexpected and interesting result from the present study was the frequent generation by the exonuclease-proficient T7 DNA polymerase of errors involving loss of 317 nucleotides flanked by the directly repeat sequence 5′-CCCGC-3′ (Table 1). A simple model to account for this error involves synthesis through the first copy of the repeat, disruption of five terminal base pairs, formation of base pairs involving the newly-made DNA and the downstream direct repeat, and continued synthesis. Deletions consistent with this model have been found in a previous study of the fidelity of exonuclease-deficient DNA polymerase β (Kunkel, 1985; Kunkel & Soni, 1988). However, here they were recovered from the exonuclease-proficient T7 polymerase reaction at a frequency well above the background mutation frequency. This suggests that, if the degree of homology (here five G·C base pairs) is sufficient, deletions between direct repeats can

be generated during polymerization reactions despite the presence of a highly active exonuclease that successfully proofreads substitutions and many frameshifts. We have recently observed similar large deletion errors in a collection of mutants generated by exonuclease-proficient *E. coli* DNA polymerase II (Cai *et al.*, 1995). Thus, inefficient proofreading is not limited to only tandem repeats, but can include widely separated repeats.

Streisinger and Owen (1985) hypothesized that addition mutations would be preferentially edited in comparison to deletion mutations. Their view was that correction would occur when the bulge migrated to the 3′-OH end of the primer. Since exonucleolytic hydrolysis occurs only on the primer strand and a primer strand bulge is required for an addition while a template strand bulge is required for a deletion, editing of addition intermediates would be more efficient. The comparative data in Figure 4C (deletions) and 4F (additions) suggest that, for a run of 4 T·A base pairs, deletions are proofread more efficiently than additions, whereas the opposite is true for tracts of 5, 6, and 7 T·A base pairs. Further tests of this idea will depend on sensitive assays to detect addition errors, which, at least in homopolymeric runs, are generally produced at much lower rates than are deletion errors.

That exonucleolytic proofreading contributes less to frameshift fidelity as the length of a homopolymeric repeat increases may explain the fact that mutations in the proofreading function of replicative DNA polymerases had little effect on the stability of long dinucleotide repeat sequences in yeast (Strand *et al.*, 1993). The fact that both polymerase selectivity and editing efficiency are run length-dependent has implications for the stability of eukaryotic genomes that contain abundant simple repetitive sequences (microsatellites) in which the number of consecutive repeats is equal to or larger than those used in this study. For frameshift mutagenesis, only the third major step for determining replication fidelity, postreplication mismatch, is not sensitive to repeat number. Thus, unlike the situation in *E. coli* where DNA polymerase selectivity, exonucleolytic proofreading, and postreplication mismatch repair contribute about 10^{-5}, 10^{-2}, and 10^{-3}, respectively, to *in vivo* rates for base substitution (Schaaper, 1993), the burden of maintaining the stability of long microsatellites may be borne more by the last step. This is consistent with the possibility that eukaryotic cells may contain novel proteins and/or processes for repairing frameshift intermediates (Umar *et al.*, 1994), and may explain why eukaryotic mismatch repair may be more differentiated than prokaryotic repair with respect to substrate specificity [for a recent review, see Kunkel (1995)]. It may also explain why microsatellite instability has been such an informative molecular biomarker for defects in postreplication mismatch repair; i.e., human cells that exhibit microsatellite instability are uniformly deficient in mismatch repair.

ACKNOWLEDGMENT

We thank Cheng-I Wang and Hiep Tran for critical evaluation of the manuscript.

REFERENCES

Bebenek, K., & Kunkel, T. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4946−4950.
Bebenek, K., & Kunkel, T. A. (1995) *Methods Enzymol. 262*, 217−232.
Bebenek, K., Abbotts, J., Wilson, S. H., & Kunkel, T. A. (1993) *J. Biol. Chem. 268*, 10324−10334.
Bebenek, K., Beard, W. A., Casas-Finet, J. R., Kim, H.-R., Darden, T. A., Wilson, S.H., & Kunkel,T. A. (1995) *J. Biol. Chem. 270*, 19516−19523.
Brutlag, D., & Kornberg, A. (1972) *J. Biol. Chem. 247*, 241−248.
Cai, H., Yu, H., McEntee, K., Kunkel, T. A., & Goodman, M. F. (1995) *J. Biol. Chem. 270*, 15327−15335.
Donlin, M. J., Patel, S. S., & Johnson, K. A. (1991) *Biochemistry 30*, 538−546.
Eger, B. T., Kuchta, R. D., Carroll, S. S., Benkovic, P. A., Dahlberg, M. E., Joyce, C.M., & Benkovic, S. J. (1991) *Biochemistry 30*, 1441−1448.
Eshleman, J. R., & Markowitz, S. D. (1995) *Curr. Opin. Oncol. 7*, 83−89.
Johnson, K. A. (1993) *Annu. Rev. Biochem. 62*, 685−713.
Joyce, C. M., & Steitz, T. A. (1994) *Annu. Rev. Biochem. 63*, 777−882.
Kunkel, T. A. (1985) *J. Biol. Chem. 260*, 5787−5796.
Kunkel, T. A. (1986) *J. Biol. Chem. 261*, 13581−13587.
Kunkel, T. A. (1990) *Biochemistry 29*, 8003−8011.
Kunkel, T. A. (1995) *Curr. Biol. 5*, 1091−1094.
Kunkel, T. A., & Soni, A. (1988) *J. Biol. Chem. 263*, 14784−14789.
Kunkel, T. A., Bebenek, K., & McClary, J. (1991) *Methods Enzymol. 204*, 125−139.
Kunkel, T. A., Patel, S. S., & Johnson, K. A. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6830−6834.
Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A., & Mathur, E. J. (1991) *Gene 108*, 1−6.
Marra, G., & Boland, C. R. (1995) *J. Natl. Cancer Inst. 87*, 1114−1125.
Patel, S. S., Wong, I., & Johnson, K. A. (1991) *Biochemistry 30*, 511−525.
Pavlov, Y. I., Minnick, D. T., Izuta, S., & Kunkel, T. A. (1994) *Biochemistry 33*, 4695−4701.
Roberts, J. D., & Kunkel, T. A. (1996) in *DNA Replication in Eukaryotic Cells: Concepts, Enzymes and Systems* (De Pamphilis, M., Ed.) Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (in press).
Schaaper, R. M. (1993) *J. Biol. Chem. 268*, 23762−23765.
Strand, M., Prolla, T. A., Liskay, R. M., & Petes, T. D. (1993) *Nature 365*, 274−276.
Streisinger, G., & Owen, J. (1985) *Genetics 109*, 633−659.
Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77−84.
Umar, A., Boyer, J. C., & Kunkel, T. A. (1994) *Science 266*, 814−816.
West Frey, M., Nossal, N. G., Capson, T. L., & Benkovic, S. J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2579−2583.
Willems, P. J. (1994) *Nat. Genet. 8*, 213−215.
Wong, I., Patel, S. S., & Johnson, K. A. (1991) *Biochemistry 30*, 526−537.

BI952178H