Frameshift Fidelity during Replication of Double-Stranded DNA in HeLa Cell Extracts

John D. Roberts, Dinh Nguyen, and Thomas A. Kunkel*

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

Received November 16, 1992; Revised Manuscript Received January 4, 1993

ABSTRACT: The processes by which minus-one frameshifts arise during replication of double-stranded DNA by a human replication apparatus were examined. Using M13mp2 DNA containing the simian virus 40 (SV40) origin of replication and a plus-one frameshift mutation in the *lacZα* reporter gene, we performed replication reactions using a HeLa cell extract and the SV40 large T antigen. Frameshifts that restore the reading frame to give a blue-plaque phenotype include the loss of one of five consecutive A·T base pairs or any one of 36 non-reiterated base pairs. Although both types of deletions were generated at rates substantially above the background mutant frequency of unreplicated DNA, the rate was highest at the A·T run, suggesting the involvement of a misaligned replication intermediate at this homopolymeric sequence. The error rate for both types of deletions increased as the concentration of dNTPs was increased. A small increase in error rate at the run of A·T base pairs was also observed when a dNMP was added to the replication reaction. These results are consistent with the correction of frameshift intermediates during replication by exonucleolytic proofreading. To examine frameshift error rates on the leading and lagging strands, we compared reversion frequencies for two vectors containing the origin of replication close to, but on opposite sides of, the mutational target. To generate strand-specific errors, nucleotide substrate imbalances were used in replication reactions with these vectors. The results suggest that there is less than a 2-fold difference in the fidelity of leading- and lagging-strand synthesis for deletions at the run of A·T base pairs. However, site-specific differences of more than 10-fold were observed for deletion of certain non-reiterated base pairs. Furthermore, the pattern of minus-one frameshifts over the 41-base-pair target was different for the two vectors, suggesting that error specificity is different on the leading and lagging strands. This could reflect differences in nucleotide selectivity and/or proofreading during replication or a difference in the efficiency of postreplication repair of misaligned heteroduplexes on the two strands.

The loss or gain of one or more base pairs of DNA comprises a significant portion of spontaneous and induced mutations in many organisms. The scheme first proposed by Streisinger et al. (1966) provided an initial framework for considering how such mutations arise. This model suggests that at repetitive sequences slippage of the two DNA strands generates a misaligned intermediate containing one or more unpaired nucleotides stabilized by correct base-pairing. Subsequent processing of this intermediate by reapir or replication could yield a frameshift mutation. Information obtained in subsequent years has supported this model and suggested additional mechanisms by which frameshifts may be generated (Kunkel, 1990; Ripley, 1990).

Our attempts to understand the processes that generate frameshifts are directed toward those that operate during DNA polymerization. We began with reactions catalyzed in vitro by purified DNA polymerases [reviewed in Kunkel (1990)]. In the work described here, we extend studies of frameshift fidelity to the semidiscontinuous replication of double-stranded DNA by a multiprotein replication apparatus in extracts of human HeLa cells. DNA replication provides a substantial potential for formation of misaligned intermediates because eukaryotic genomes contain large numbers of repetitive nucleotides and because their replication involves the formation of many template–primer termini at which misalignments could occur.

Studies of replication fidelity in vitro were made possible by the development of the simian virus 40 (SV40) in vitro replication system. This system has proven an excellent model for human chromosomal replication (Kelly, 1988; Stillman, 1989; Hurwitz et al., 1990). SV40 origin-dependent DNA replication involves a series of complex reactions catalyzed by host cell proteins, including two or more DNA polymerases, DNA primase, topoisomerases, single-standed DNA binding proteins, ligase, and other accessory factors, and a single viral protein, the SV40 large T antigen (Tag). We have used this replication activity in HeLa cell extracts to establish that replication is highly accurate for single base substitution errors (Roberts & Kunkel, 1988; Thomas et al., 1991b) and for single base pair frameshift errors (Thomas et al., 1991b). Furthermore, for both types of errors, replication fidelity can be modified by providing unequal concentrations of deoxyribonucleoside triphosphates (Roberts et al., 1991; Bebenek et al., 1992).

These observations led to development of a strategy to examine base substitution fidelity on the leading and lagging strands during replication (Roberts et al., 1991). In the present study, we have applied this approach (described below) to a reversion assay for measuring single base deletion error rates in a target sequence containing five consecutive A·T base pairs and 36 non-reiterated base pairs. We wished to address three issues relevant to the frameshift accuracy of the human replication machinery. First, we wanted to determine if, as predicted by the slippage model, single base deletion (minus-one frameshift) error rates during replication were higher at the homopolymeric sequence than at the non-reiterated sequences. Second, we wanted to examine the possibility that frameshift errors were subject to exonucleolytic proofreading.

* Author to whom correspondence should be addressed.

Third, we wanted to examine the possibility that frameshift fidelity on the leading and lagging strands might be different. This last issue was prompted by several observations. Current models [reviewed in Linn (1991) and Kunkel (1992)] suggest that different polymerases may replicate the two strands, and eukaryotic DNA polymerases α, δ, and ϵ generate single base pair frameshift errors at distinctly different rates (Thomas et al., 1991b). Also, concomitant replication of both strands by 5′→3′ polymerization requires that the lagging strand be replicated as a series of fragments that are subsequently ligated. This could provide greater opportunity for slippage at termini than for a continuously replicated leading strand, especially if more than one DNA polymerase were to participate in lagging strand replication, requiring switching from one enzyme to the next to complete each fragment.

EXPERIMENTAL PROCEDURES

Materials

Bacterial Strains and Reagents. Escherichia coli strains CSH50, NR9099, and NR9162 have been described (Kunkel, 1985; Roberts & Kunkel, 1988). SV40 Tag was purchased from Molecular Biology Resources (Milwaukee, WI). Restriction enzymes were from New England Biolabs. [α-^{32}P]-dCTP and [γ-^{32}P]ATP were purchased from Amersham Corp.

Methods

Preparation of DNA Substrates. Templates were based on M13mp2SV ori-left and ori-right vectors (Roberts & Kunkel, 1988; Roberts et al., 1991), constructed by site-directed mutagenesis (Kunkel et al., 1991a), and prepared as described (Roberts & Kunkel, in press).

Frameshift Reversion Assay. The fidelity assay scores minus-one frameshifts in a M13mp2 DNA substrate at the following sequence:

```
3′-AC TAA TGC TTA AGT GAC CGG
   •• ••• ••• ••• ••• ••• •••
5′-TG ATT ACG AAT TCA CTG GCC
       •                 •
      50                60
           CAG CAA AAA TGT TGC AGC ACT-5′
           ••• ••• ••• ••• ••• ••• •••
           GTC GTT TTT ACA ACG TCG TGA-3′
               •                 •
              70                80
```

The sequence is that of codons 2–15 of the α-complementation domain of the *lacZ* gene, but contains an extra A·T base pair in the homopolymeric sequence at positions 70–73 (underlined). The resulting M13 plaques are therefore colorless. Blue plaques result from frameshift errors that restore the reading frame. Minus-one frameshifts within the TTTTT run restore the wild-type coding sequence and yield dark blue plaques [also confirmed earlier by DNA sequence analysis (Bebenek et al., 1990)]. Minus-one frameshifts at any of the 36 other nucleotide positions also yield blue plaques, but these plaques are lighter blue because, although the reading frame is restored, one or more amino acids are no longer wild-type. DNA sequence analysis of dark and light blue revertants was performed to confirm that these colors do reflect one-nucleotide deletions at the predicted sites. We have now identified minus-one frameshifts at 35 of the 36 nonrun positions in the target; all had a light blue phenotype. Both light and dark blue revertants can be quantitatively scored on plates containing as many as 10 000 colorless plaques (Frederico et al., 1990), a density that was not exceeded in the experiments reported here. An unequivocal distinction between wild-type dark blue and pseudo-wild-type lighter blue revertants, and hence between frameshifts at the TTTTT run versus other sites, can be made by comparing, on the same plate, candidate revertants with wild-type plaques. Except for the mutational spectra, all assignments of frameshifts as run or nonrun were made by a compsrison of plaque color. For the spectra, dark blue plaques were assigned as resulting from minus-one frameshifts in the A·T run, and light blue plaques were subjected to DNA sequence analysis to identify the location of the nonrun minus-one frameshift.

SV40 Origin-Based Replication in Human Cell Extracts. Replication reactions (25–250 μL) were carried out as described (Roberts & Kunkel, 1988) using M13mp2SV RFI DNA. Unreplicated molecules were inactivated by treatment with the restriction endonuclease *Dpn*I. A portion (1–3 μL) of the replicated DNA was used for product analysis by agarose gel electrophoresis as described (Roberts & Kunkel, 1988). Another aliquot (1–3 μL in deionized H_2O) was used to transfect competent *E. coli* host strain NR9162 (*mutS*) in a volume of 50 μL of cells ($\sim 2 \times 10^{10}$ cells/mL in 10% glycerol) by electroporation in a 0.2-cm cuvette with a Bio-Rad gene pulser. The instrument settings were 2000 V, 400 W, and 25 μF, and the resulting time constants were 9.0–9.3 ms. Appropriate dilutions of the transfected cells were plated such that plates contained 2000–10 000 pfu (Kunkel, 1984; Bebenek et al., 1990). Estimates of the total number of plaques per plate were made from dilutions of the electroporated samples, plated separately to yield 100–1000 plaques per plate. Blue plaques were scored, recovered, and replated along with wild-type phase to determine the blue color intensity.

RESULTS

Frameshift Error Rates during Replication. We had previously determined that minus-one frameshift error rates per mutationally detectable nucleotide during replication in HeLa cell extracts are $\leq 3.4 \times 10^{-6}$ and $\leq 1.6 \times 10^{-7}$, respectively, at reiterated and nonreiterated sites (Thomas et al., 1991b). Both values are "less than or equal to" rates because frameshift errors were not detected above the background frequency of unreplicated DNA. In an attempt to measure frameshift error rates in the present study, we used a reversion assay that focuses exclusively on minus-one nucleotide frameshift errors (or the genetic equivalent, i.e., −4, +2, etc.) and has a low background reversion frequency for unreplicated DNA ($\sim 0.5 \times 10^{-5}$).

Replication of SV40 origin-containing M13mp2SV double-stranded circular DNA was performed using an extract of HeLa cells and SV40 Tag. As described previously (Roberts & Kunkel, 1988), replication depended onthe presence of Tag (Table I), the SV40 origin sequence, and the HeLa extract. Analysis of an aliquot of the reaction mixture (and all subsequent reaction mixtures) by agarose gel electrophoresis [(not shown, but see Roberts and Kunkel (1988)] demonstrated that the replication products were primarily monomer circular molecules, with some high molecular weight DNA present. Diagnostic restriction endonuclease digestion demonstrated that the monomers were primarily the result of a single round of replication.

The remaining products were digested with *Dpn*I to linearize unreplicated molecules, and the newly replicated DNA was used to transfect competent host cells to score reversion frequencies. Replication using 100 μM dNTPs (Table I) yielded dark blue plaque reversion frequencies 3–5-fold above that of unreplicated DNA. These experiments were carried out multiple times to ensure that this difference was repro-

Table I: Effect of dNTP and dGMP Concentration on Error Rates for Minus-One Frameshifts[a]

dNTP concn (μM)	dGMP (mM)	incorporation[c] (%)	no. of plaques			error rate (×10⁻⁶)[b]	
			total (×10⁵)	dark blue	light blue	A·T run	non-reiterated sites
oriL Vector, Increasing dNTP Concentration							
100, no-Tag control[d]		4	166	121	9	2.9	0.03
20		110	17	39	3	9.2	0.10
100[e]		100	53	139	17	10.5	0.18
1000		55	4.4	59	21	54	2.7
oriL Vector, Addition of dGMP							
110		100	23	58	7	10	0.17
100	1	60	17	64	4	15	0.13
100	3	14	0.9	16	0	71	≤0.62
oriR Vector, Increasing dNTP Concentration							
100, no-Tag control[f]		5	13	9	4	2.8	0.17
100[g]		100	17	36	16	8.4	0.52
1000		51	4.8	51	17	43	2.0

[a] Reactions were performed as described in Methods except for changes in nucleotide concentrations and addition of dGMP. [b] Error rates are number of errors per mutationally detectable base pair per round of replication, calculated by multiplying the reversion frequency (blue plaques/total plaques) by 2 to correct for the 50% recovery of errors upon transfection (Roberts & Kunkel, 1988) and then dividing by 5 (for the A·T run) or 36 (for the non-reiterated sites). [c] The 100% values for incorporation were 110 pmol in the ori-left pool-bias experiment, 139 pmol in the ori-left dGMP experiment, and 98 pmol in the ori-right pool-bias experiment. [d] Sum of eight experiments. The no-Tag control samples were not digested with *Dpn*I. The mean error rate for the eight determinations was (2.8 ± 1.1) × 10^{-6} for the minus-one base errors in the run (dark blue). The very low frequency of light blue plaques in the eight experiments yields "less than or equal to" values for the error rate in many experiments, making a calculation of the mean and standard deviation impossible. [e] Sum of six experiments. The mean error rate for the six determinations was (10.6 ± 2.9) × 10^{-6} for the minus-one base errors in the run (dark blue) and (0.19 ± 0.2) × 10^{-6} for the nonrun frameshifts (light blue plaques). [f] Sum of four experiments. The mean error rate for the four determinations was (3.0 ± 1.2) × 10^{-6} for the minus-one base errors in the run (dark blue). As above, the absence of light blue plaques in some experiments precludes a calculation of the mean and standard deviation for error rates at nonrun positions. [g] Sum of three experiments. The mean error rate for the four determinations was (10 ± 5) × 10^{-6} for the minus-one base errors in the run (dark blue). Again, the absence of light blue plaques in some experiments precludes a calculation of the mean and standard deviation for error rates at nonrun positions.

ducible. The means and standard deviations for these measurements, when possible, are given in the legend to Table I and were used to calculate the range of error rates for minus-one frameshifts. After subtracting the background frequency of dark blue plaques (obtained from the no-Tag control), the minus-one frameshift error rates per nucleotide (see legend to Table I for calculation) were 5.6–12 × 10^{-6} for the loss of an A·T pair at the homopolymeric sequence. In the same 100 μM reactions, light blue mutants were also generated at frequencies 2–4-fold above that obtained with unreplicated DNA. After subtracting the background frequency of light blue plaques, the minus-one frameshift error rates per nucleotide polymerized were 4–35 × 10^{-8} for the loss of non-reiterated base pairs. These rates are much lower than at the A·T run.

Evidence for Exonucleolytic Proofreading of Frameshift Errors. For base substitution errors, the competition between polymerization from mispaired termini and exonucleolytic removal of misinsertions is critical for determining the final error rate of a polymerization reaction (Ninio, 1975). The polymerase:exonuclease ratio can be modulated by increasing the concentration of the next correct nucleotide to be incorporated after a misinsertion or by inhibiting the exonuclease activity by adding a nucleoside monophosphate, the end product of exonuclease action. In both cases, the polymerase:exonuclease ratio is increased; such increases in error rate have been interpreted to reflect the contribution of proofreading to base substitution fidelity. We took this approach to examine the possibility that proofreading contributes to frameshift fidelity during replication.

Minus-one frameshift error rates were determined in reactions containing either a 5-fold lower or 10-fold higher concentration of dNTPs, as well as in reactions containing dGMP. As shown in Table I, a 50-fold variation in dNTP concentration had only a 2-fold effect on the efficiency of replication. For errors at the A·T run, increasing the dNTP concentration from 20 to 1000 μM increased the error rate 8-fold (after subtracting the background). A similar (7-fold) effect was seen with a second (ori right) vector, comparing results at 100 and 1000 μM. In these same reactions, the minus-one frameshift error rate at non-reiterated sites increased to an even greater extent. For example, after subtracting the background, the increase in error rate resulting from the 50-fold increase in dNTP concentration was 38-fold (ori-left vector, 1000 vs 20 μM, 2.67 × 10^{-6}/7 × 10^{-8}). These data are consistent with the suggestion that exonucleolytic proofreading enhances frameshift fidelity during DNA replication.

In reactions containing 100 μM dNTPs, the addition of 1 mM dGMP had little effect on either replication efficiency or error rates (Table I). Increasing the dGMP concentration to 3 mM inhibited replication by 86%, yielding little product for transfection. For the plaques that were obtained, the data demonstrate a 10-fold increase in minus-one frameshift error rate at the A·T run (68 × 10^{-6}/7 × 10^{-6}). No light blue plaques were obtained, so that the effect on frameshifts at non-reiterated sites could not be calculated. Nonetheless, at least for minus-one frameshifts occurring in the A·T run, the data are again suggestive of proofreading of frameshift errors.

Strategy To Examine Replication Fidelity on the Leading and Lagging Strands. In an attempt to determine the frameshift fidelity on the leading and lagging strands, we constructed two vectors for the frameshift reversion assay. One contains the origin on the left of the frameshift target sequence, approximately 500 nucleotides from the A·T run. This distance is small relative to the size of the vector (~7400 base pairs). Previous studies (Edenberg & Huberman, 1975; Li & Kelly, 1985) have indicated that the rate of replication fork movement is similar in both directions from the origin. Thus, with this vector the strand containing the run of T residues at the five-base-pair A·T run is likely to be replicated as the lagging strand (Figure 1A). The second vector contains the origin to the right of the target, approximately 200 nucleotides from the A·T run. In this vector, the strand containing the run of T residues is likely to be replicated as the leading strand (Figure 1B).

Fidelity measurements with these two vectors allow determination of the fidelity of replication of the same sequence by either the leading- or the lagging-strand apparatus, provided that the strand on which the error was made can be assigned. In an attempt to do this, we took advantage of the observation that frameshift error rates with a replicative DNA polymerase can be affected in a predictable manner by selective dNTP substrate imbalances. In reactions catalyzed by DNA polymerase α (pol-α), decreasing the concentration of one dNTP increased the error rate for deletions of homopolymeric template nucleotides complementary to that dNTP (Bebenek

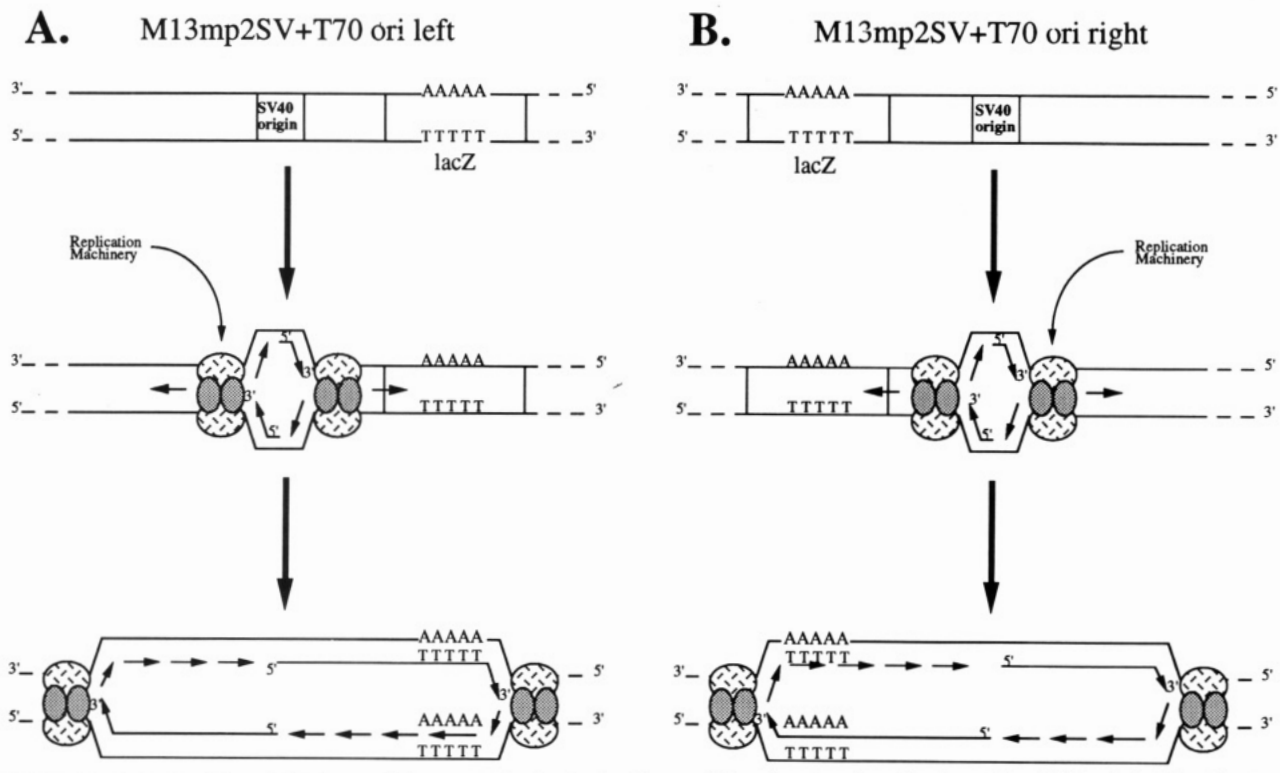

FIGURE 1: Strategy for determining frameshift error rates during leading- and lagging-strand replication. The SV40 origin of replication is placed on either the left (A) or the right (B) of the *lacZα* gene. Semiconservative DNA synthesis at a replication fork requires that one strand be copied by the leading-strand apparatus and one strand by the lagging-strand apparatus. The polarity of the DNA strands at the minus-one frameshift target remain the same regardless of the position of the origin of replication. Thus, the assignment of leading and lagging strands during replication at the *lacZα* target depends only on the position of the origin and the assumption that the replication forks move at approximately the same rate in both directions from the origin. The distance from the A·T run to the site of the first incorporated nucleotide during initiation of DNA replication is 494 nucleotides in the ori-left vector and 249 nucleotides in the ori-right construct. The size of the entire molecule is 7398 base pairs.

et al., 1992). Thus, an increased error rate at the run of A·T base pairs resulting from a decrease in the dATP concentration would suggest that the error occurred during replication of the TTTTT-containing strand. Conversely, an increased error rate at the run of A·T base pairs resulting from a decrease in the dTTP concentration suggests that the error occurred during replication of the AAAAA-containing strand.

Frameshift Error Rates at the Run of A·T Base Pairs on the Leading and Lagging Strands. The results for replication reactions performed with unbalanced dNTP pools are shown in Table II. A 50-fold decrease in the dTTP concentration led to minus-one frameshift error rates in the A·T base pair run of 160×10^{-5} and 200×10^{-5}, respectively, with the ori-left and ori-right vectors. Both rates are substantially higher than for reactions at equal dNTP concentrations (54×10^{-5} and 43×10^{-5}, ori left and ori right, respectively, at 1000 μM equal pools; Table I) and thus result primarily from the substrate imbalance. The fact that the rates are similar for the two vectors implies that the fidelity of leading- and lagging-strand replication is similar for this error under this pool-bias condition.

With a low dATP concentration, the error rates with both vectors (62×10^{-5} and 87×10^{-5}) are only slightly higher than those obtained with 1000 μM equal dNTPs (Table I, 54×10^{-5} and 43×10^{-5}). By themselves, these small increases preclude any conclusion about the fidelity of leading- and lagging-strand replication when only dATP is lowered. However, given the sequence at the homopolymeric site,

```
        (-) strand template:  3'-CAGC AAAAA TGTT-5'
synthesis of daughter strand:  5'-GTCG TTTTT A

synthesis of daughter strand:        C AAAAA TGTT-5'
        (+) strand template:  5'-GTCG TTTTT ACAA-3'
```

Table II: Effect of dNTP Pool Imbalances on Minus-One Frameshift Error Rates[a]

dNTP concn (μM)	incorporation (%)	no. of plaques			error rate ($\times 10^{-6}$)	
		total ($\times 10^5$)	dark blue	light blue	A·T run	non-reiterated sites
oriL Vector						
equal pools 1000	55	4.4	59	21	54	2.7
20 dTTP/1000 others	150	6.7	268	14	160	1.2
20 dATP/1000 others	130	20	307	36	61	1.0
20 dATP/20 dCTP/1000 others	44	20	88	10	18	0.28
oriR Vector						
equal pools, 1000	51	4.8	51	17	43	2.0
20 dTTP/1000 others	150	6.6	328	34	200	2.9
20 dATP/1000 others	110	7.7	167	57	87	4.1
20 dATP/20 dCTP/1000 others	37	11	48	14	17	0.71

[a] Reactions were performed as described in Methods except for changes in the dNTP concentrations as described above. Error rates were calculated as described in the legend to Table I. The 100% values for incorporation at equal dNTPs were 110 pmol (ori-left vector) and 139 pmol (ori-right vector).

dCMP (underlined) is the next correct nucleotide to be incorporated following replication of the TTTTT-run in the (+)-strand template, but not following replication of the AAAAA-run in the (–)-strand template. If errors that appeared in the reactions with the dATP bias occurred during replication of the (+)-strand, then the results in Table I, which suggest that the concentration of the next nucleotide modulates frameshift fidelity, imply that the error rate here should be affected by the dCTP concentration. Thus, lowering the concentration of both dATP and dCTP would be predicted

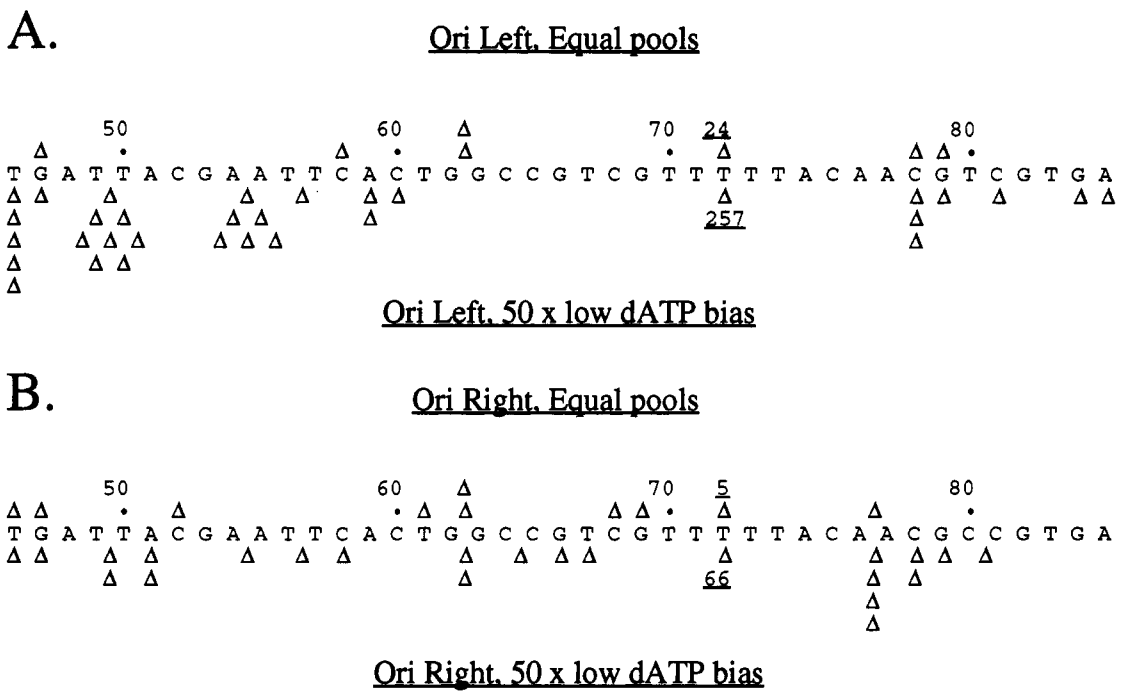

FIGURE 2: Spectra of minus-one frameshift errors at non-reiterated sites within the frameshift reversion target of M13mp2SV. Reactions were performed with either the ori-left (A) or the ori-right (B) construct as the template. Dark and light blue revertants were identified and isolated, and light blue plaques were used to prepare ssDNA for sequencing. Deletions listed above the line of sequence are from reactions done with equal dNTP pools. Deletions listed below the line of sequence are from reactions done with a 50 × pool bias for dATP (20 μM dATP and 1 mM dGTP, dCTP, and dTTP). The underlined number at the run of 5 Ts indicates the number of reversion events within the run. The total number of plaques examined for each variable were ori left, equal pools, 6×10^5; ori left, dATP bias, 17×10^5; ori right, equal pools, 4×10^5; and ori right, dATP bias, 2.9×10^5. The C at position +80 in the ori-right construct is a silent change that was identified in some of the light blue revertants.

to lead to a decrease in error rate over that obtained when only dATP concentration is lowered. This effect was in fact observed with both vectors (Table II). These results are consistent with the following interpretations for errors at the run of A·T base pairs: (i) minus-one frameshift errors with the low dATP concentration alone are occurring during replication of the TTTTT-containing strand, (ii) frameshift fidelity on both strands is similar when either dTTP or dATP concentration is lowered, and (iii) the minus-one frameshift error rate on both strands is influenced similarly by the concentration of the next correct nucleotide after the reiterated sequence.

Differences in Frameshift Error Rates at Non-Reiterated Sites on the Leading and Lagging Strands. Comparison of the average error rates for loss of non-reiterated base pairs (Tabel II) revealed 3–4-fold differences between the ori-left and ori-right vectors with all three pool-bias conditions. To quantitate error rates at specific positions, we analyzed the DNA sequences of collections of light blue mutants. Since the highest error rate was for the dATP pool bias with the ori-right vector (Table II, 4.1×10^{-6}), we compared the sequences of 31 mutants from this reaction to those of 22 mutants from the corresponding reaction with the ori-left vector. In addition, we sequenced a number of light blue mutants from the 100 μM equal-pools reactions. In the spectra of errors from reactions with equimolar dNTPs (Figure 2), 12 of 15 mutants (80%) had lost a G·C base pair (Figure 2). In contrast, when the concentration of dATP was 50-fold lower than that of the other three dNTPs, 35 of 53 mutants (66%) had lost a non-reiterated A·T base pair. Thus, the pool bias did affect the error specificity.

The data in Table II and Figure 2 can also be used to calculate site-specific error rates (Table III). At some, but not all, positions in the 36-base-pair target, error rates with the ori-left and ori-right vectors differ by more than 10-fold.

Table III: Site-Specific Rates for Non-Reiterated Base Deletions with Ori-Left and Ori-Right Vectors

		oriL vector[a]		oriR vector[b]		
nucleotide sequence	position no.	mutants	error rate ($\times 10^{-6}$)	mutants	error rate ($\times 10^{-6}$)	ratio (R/L)
TAC AA CGC	76,77	0	≤0.6	4	14	≥23
ATT A CGA	51	0	≤1.2	2	14	≥12
ACT GG CCG	62,63	0	≤0.6	2	7.0	≥12
ACG AA TTC	54,55	6	3.5	1	3.5	1.0
CCA T GAT	46	5	5.8	1	7.0	1.2
TGA TT ACG	49,50	8	4.6	2	7.0	1.5

[a] The total number of plaques from the experiment with the ori-left vector and a dATP pool bias was 1.7×10^6. Analysis of the DNA sequence of the 33 light blue revertants in this collection indicated that 31 revertants had lost a single nucleotide in the regions flanking the run of A·T base pairs. [b] The total number of plaques from the experiment with the ori-right vector and a dATP pool bias was 6.5×10^5. Of the 50 light blue revertants in this collection, 22 DNA sequences were determined, and all showed a loss of a single base pair in the non-reiterated sequences.

This is also consistent with the possibility that replication frameshift fidelity on the leading and lagging strands differs at some sites.

To further address this issue, a statistical analysis was performed to compare the overall spectra generated with the two vectors with a low dATP concentration. Zelterman's method, as described by Piegorsch and Bailer (1992), was used to examine the complete spectra, including light and dark blue revertants. The analysis indicated that the patterns of mutations in the two collections (Figure 2) are significantly different ($P < 0.001$). To ask which sites are responsible for the observed heterogeneity between the two spectra, we then used an approach described by Perli (1985). For the mutational data here, the AA site at positions 76–77 exhibits the most heterogeneity ($P < 0.006$), followed by the run of Ts at positions 70–73 ($P < 0.02$). (Note that this analysis

does not compare the error rates at individual sites, but is rather a measure of the distribution of mutations throughout the target. Thus, while the error rate for minus-one frameshifts at the run of A·T base pairs is similar in the two vectors, the percentage of total errors occurring at this site in the two vectors is different.) Although not seen as significant by this conservative statistical analysis, the next most heterogeneous sites were positions A_{51} and GG_{62-63} ($P < 0.2$). Overall, the results in Table III and from both statistical analyses indicate that, for replication reactions performed at a low dATP concentration, significant differences exist in the specificity of minus-one frameshift errors in the two vectors, suggesting that there are differences in error rates on the leading and lagging strands.

DISUCSSION

Although a wealth of information exists on the frameshift fidelity of DNA synthesis catalyzed by DNA polymerases (Kunkel, 1990), much less is known about frameshift error rates during replication of double-stranded DNA except for the fact that they are low. In fact, previous measurements with equimolar substrate concentrations (Thomas et al., 1991b) provided only "less than or equal to" error rates. The present study was an attempt to (i) determine whether the human replication apparatus does indeed generate frameshift errors in reactions containing equimolar substrate concentrations and, if so, obtain quantitative error rates, (ii) examine whether exonucleolytic proofreading contributes to high frameshift fidelity during replication, and (iii) examine possible mechanisms for their production on the leading and lagging strands. In this study, we focused on single-base deletions because these are the most common frameshift errors generated by eukaryotic replicative DNA polymerases (Thomas et al., 1991b).

We have considered the results in Table I from several perspectives. First, the data with the ori-left vector at 100 μM dNTPs (Table I), representing the average of eight independent determinations, demonstrate that replication at equimolar dNTPs in human cell extracts does indeed generate single base pair deletions at detectable rates. This directly supports the possibility that replication errors are one source of the frameshift mutations known to result in human genetic diseases (Cooper & Krawczak, 1990).

Second, these error rates can be compared directly to those of replicative DNA polymerases plus their accessory proteins in reactions that fill single-stranded gaps. The replication error rate for the loss of one A·T base pair in the five-base-pair run (7.6×10^{-6}) is 130-fold lower than the error rate for DNA polymerase α-primase plus replication factor A (i.e., single-stranded DNA binding protein) when copying the same sequence (TTTTT-containing strand) (Kunkel et al., 1991b). Replication by HeLa cell extracts across the five-base run is even 9-fold more accurate than for DNA synthesis by the exonuclease-proficient DNA polymerase δ plus PCNA copying this sequence as a four-base TTTT run [calculated from the data in Thomas et al. (1991b)]. Similarly, replication by HeLa cell extracts is much more accurate than synthesis by isolated DNA polymerases for deletions of non-reiterated nucleotides [Table I versus data in Thomas et al. (1991b)]. These differences may reflect the contribution of additional accessory proteins to fidelity or perhaps synthesis by DNA polymerase ϵ, which has high frameshift fidelity even without accessory proteins (Thomas et al., 1991b).

A third notable aspect of the data in Table I is that the error rate per detectable base pair is almost 100-fold higher at the A·T run than at the non-reiterated sites (e.g., ori-left vector, 100 μM dNTPs). This suggests that the premutagenic replication intermediate at the homopolymeric sequence is a misaligned substrate containing an unpaired template base (A or T) stabilized by one or more correct A·T base pairs. Such an intermediate could arise by strand slippage, as first proposed by Streisinger et al. (1966). Alternatively, it is possible that the event that initiates some frameshifts is the misinsertion of a nucleotide (Kunkel & Soni, 1988; Bebenek & Kunkel, 1990). If this nucleotide is complementary to the next correct template base, a rearrangement could generate a misaligned intermediate stabilized by one or more correct terminal base pairs, ultimately yielding a frameshift mutation. This offers one possible explanation for the origin of the frameshift errors at non-reiterated sites (Table III). The misinsertion model is especially relevant to those sites that have no common adjacent neighbor (e.g., positions +46 and +51 in Table III), because initiation of a frameshift at such sites by strand slippage would not allow formation of a misaligned intermediate stabilized by a correct terminal base pair. Of course, misinsertion could initiate frameshift errors at repetitive sites as well.

A fourth observation from Table I is that frameshift error rates are influenced by the absolute concentration of dNTPs and by the presence of dGMP. These effects are similar to those obtained earlier for base-substitution errors during replication and are hallmarks of exonucleolytic proofreading [for review, see Kunkel (1988)]. The data are in agreement with several experiments suggesting that single base deletion errors at repetitive and nonrepetitive sites are proofread during polymerase reactions (Bebenek et al., 1990; Bebenek & Kunkel, 1990; Thomas et al., 1991b) and suggest that frameshift intermediates generated during replication may be proofread as well.

The experiments in Tables II and III represent our initial attempt to examine frameshift fidelity on the leading and lagging strands. There are several reasons to believe that frameshift error rates on the two strands could be different. Models for eukaryotic replication (Linn, 1991; Kunkel, 1992) posit that different polymerases replicate the two strands. DNA polymerases α, β, δ, and ϵ generate single base pair frameshift errors at distinctly different rates (Kunkel & Alexander, 1986; Thomas et al., 1991b). Also, the lagging strand is replicated as a series of fragments that are subsequently ligated, potentially providing a greater opportunity for slippage at termini than in the case of a continuously replicated leading strand. Finally, participation of more than one DNA polymerase in lagging-strand replication requires switching from one enzyme to another to complete each fragment, which could provide an additional opportunity for generating errors.

The observation that error rates for delections at the run of A·T base pairs are similar for the two different vectors (ori left and ori right, Figure 1) in reactions containing unbalanced dNTP concentrations (Table II) argues that the frameshift rate for this error is similar on both strands. However, the data for deletion of non-reiterated base pairs (Tables II and III) indicate differences between the two vectors in error rates at some sites, but not at others. Furthermore, a statistical analysis comparing the two spectra indicates that the pattern of minus-one frameshifts is different and that this difference is strongest in the same sites at which the error rates are the most different. We cannot rule out, of course, possible long-range effects of the SV40 origin sequence, located 200–500 nucleotides away. However, since identical sequences were copied under the same conditions in both reactions, the observed differences are not due to a local DNA sequence

context effect or to differences in reaction conditions.

On the basis of the following logic, we speculate that the differences in Table III and Figure 2 result from an inherent difference in replication fidelity on the leading and lagging strands. The observed specificity, i.e., mostly loss of A·T base pairs, is consistent with these non-reiterated-site errors having arisen as a result of the lowered dATP pool. They could have been initiated via misinsertion opposite template T residues because the dATP concentration was low, followed by rearrangement to generate a misalignment. There is substantial evidence with purified DNA polymerases that misinsertions followed by rearrangement can generate frameshift errors (Bebenek & Kunkel, 1990; Bebenek et al., 1992). If this were the case during double-strand replication, the template strand for the generation of the error would be identified as the T-containing strand. Thus, for the two positions (+51 and +76–77, Table III) where the largest differences were found for loss of an A·T base pair, the higher error rates would occur during lagging-strand replication. By this same logic, error rates are the same on both strands for the last three positions listed in Table III.

The complexity inherent in replication reactions catalyzed by a cell extract provides several potential explanations for site-specific differences in fidelity on the two strands. These differences could reflect synthesis by different DNA polymerases on the leading and lagging strands. The DNA polymerases that could be involved (pol α, β, δ, and/or ϵ) differ in both base selectivity and proofreading capacity. It is also possible that post-replication mismatch repair capacity (Holmes et al., 1990; Thomas et al., 1991a) might differ for errors generated on the leading and lagging strands. Future studies will attempt to distinguish among these possibilities.

ACKNOWLEDGMENT

We thank Walter Piegorsch of NIEHS for performing the statistical analysis, David Thomas of NIEHS for many helpful discussions, and Roel Schaaper and Craig Bennett, both of NIEHS, for their thoughtful critiques of the manuscript.

REFERENCES

Bebenek, K., & Kunkel, T. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4946–4950.

Bebenek, K., Joyce, C. M., Fitzgerald, M. P., & Kunkel, T. A. (1990) *J. Biol. Chem. 265*, 13878–13887.

Bebenek, K., Roberts, J. D., & Kunkel, T. A. (1992) *J. Biol. Chem. 267*, 3589–3596.

Cooper, D. N., & Krawczak, M. (1990) *Hum. Genet. 85*, 55–74.

Edenberg, H. J., & Huberman, J. A. (1975) *Annu. Rev. Genet. 9*, 245–284.

Frederico, L. A., Kunkel, T. A., & Shaw, B. R. (1990) *Biochemistry 29*, 2532–2537.

Holmes, J., Jr., Clark, S., & Modrich, P. (1990) *Proc. Natl. Acad. Sci. U.S.A 87*, 5837–5841.

Hurwitz, J., Dean, F. B., Kwong, A. D., & Lee, S.-H. (1990) *J. Biol. Chem. 265*, 18043–18046.

Kelly, T. J. (1988) *J. Biol. Chem. 263*, 17889–17892.

Kunkel, T. A. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1494–1498.

Kunkel, T. A. (1985) *J. Biol. Chem. 260*, 5787–5796.

Kunkel, T. A. (1988) *Cell 53*, 837–840.

Kunkel, T. A. (1990) *Biochemistry 29*, 8003–8011.

Kunkel, T. A. (1992) *BioEssays 14*, 303–308.

Kunkel, T. A., & Alexander, P. S. (1986) *J. Biol. Chem. 261*, 160–166.

Kunkel, T. A., & Soni, A. (1988) *J. Biol. Chem. 263*, 14784–14789.

Kunkel, T. A., Bebenek, K., & McClary, J. (1991a) *Methods Enzymol. 204*, 125–139.

Kunkel, T. A., Roberts, J. D., & Sugino, A. (1991b) *Mutat. Res. 250*, 175–182.

Li, J. J., & Kelly, T. J. (1985) *Mol. Cell. Biol. 5*, 1238–1246.

Linn, S. (1991) *Cell 66*, 185-187.

Ninio, J. (1975) *Biochimie 57*, 587–595.

Perli, H.-G., Hummel, G., & Lehmacher, W. (1985) *Biom. J. 27*, 885–893.

Piegorsch, W. W., & Bailer, A. J. (1992) Report No. 8270, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Ripley, L. S. (1990) *Annu. Rev. Genet. 24*, 189–213.

Roberts, J. D., & Kunkel, T. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7064–7068.

Roberts, J. D., & Kunkel, T. A. (1993) *Methods Mol. Genet.* (in press).

Roberts, J. D., Thomas, D. C., & Kunkel, T. A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3465–3469.

Stillman, B. (1989) *Annu. Rev. Cell Biol. 5*, 197–245.

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77–84.

Thomas, D. C., Roberts, J. D., & Kunkel, T. A. (1991a) *J. Biol. Chem. 266*, 3744–3751.

Thomas, D. C., Roberts, J. D., Sabatino, R. D., Myers, T. W., Tan, C. K., Downey, K. M., So, A. G., Bambara, R. A., & Kunkel, T. A. (1991b) *Biochemistry 30*, 11751–11759.