MOLECULAR AND CELLULAR BIOLOGY, Feb. 1989, p. 469–476
0270-7306/89/020469-08$02.00/0

Vol. 9, No. 2

Fidelity of Two Retroviral Reverse Transcriptases during DNA-Dependent DNA Synthesis In Vitro

JOHN D. ROBERTS,[1] BRADLEY D. PRESTON,[2] LAURA A. JOHNSTON,[2] ARUNA SONI,[1] LAWRENCE A. LOEB,[2] AND THOMAS A. KUNKEL[1]*

Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709,[1] and Department of Pathology, University of Washington, Seattle, Washington 98195[2]

Received 18 July 1988/Accepted 13 October 1988

We determined the fidelity of avian myeloblastosis virus and Moloney murine leukemia virus reverse transcriptases (RTs) during DNA synthesis in vitro using the M13mp2 *lacZα* gene as a mutational target. Both RTs commit an error approximately once for every 30,000 nucleotides polymerized. DNA sequence analysis of mutants generated in a forward mutation assay capable of detecting many types of errors demonstrated that avian myeloblastosis virus RT produced a variety of different mutations. The majority (58%) were single-base substitutions, all of which resulted from the misincorporation of either dAMP or dGMP. Minus-one frameshifts were also common, composing about 30% of the mutations. In addition to single-base events, eight mutants contained sequence changes involving from 2 to 59 bases. The frequency of these mutants suggests that, at least during DNA synthesis in vitro, RTs also commit errors by mechanisms other than classical base miscoding and misalignment. We examined the ability of RTs to synthesize DNA from a mismatched primer terminus at a sequence where the mismatched base was complementary to the next base in the template. Unlike cellular DNA polymerases which polymerize from the mismatched template-primer, RTs preferred to polymerize from a rearranged template-primer containing a matched terminal base pair and an unpaired base in the template strand. The unusual preference for this substrate suggests that the interactions between RTs and the template-primer are different from those of cellular DNA polymerases. The overall error rate of RT in vitro is sufficient to account for the estimated mutation rate of these viruses.

Retroviral reverse transcriptases (RTs) (EC 2.7.7.7) are DNA polymerases that transcribe retroviral genomic RNA into double-stranded DNA. RT was first identified in virions of murine leukemia virus (MuLV) by Baltimore (1) and Rous sarcoma virus by Temin and Mizutani (54). These enzymes perform multiple enzymatic functions, including the copying of RNA into DNA, the hydrolysis of the RNA from an RNA-DNA hybrid (RNase H), and the copying of single-stranded DNA into double-stranded DNA; thus, RT plays a central role in the life cycle of retroviruses (55, 56).

The physical structure of purified RT from avian and murine retroviruses has been studied extensively (for a review, see reference 58). Avian retrovirus RT is a heterodimeric molecule containing two related subunits, one with a molecular size of ~63 kilodaltons (the α subunit) and the other ~95 kilodaltons (the β subunit) (18). These two proteins, along with a pp32^{pol} phosphoprotein endonuclease, are derived by differential processing of a *pol* polyprotein precursor (45). The endonuclease specifically nicks DNA at the junction of adjacent retroviral long terminal repeat sequences (14). The α subunit of avian RT appears to carry both polymerase and RNase H activities.

The murine retroviral RT differs from the avian form in that it is a single 84-kilodalton polypeptide (57). Fine-structure mutational analysis of the murine enzyme indicated that both the polymerase and the RNase H activities are contained in this polypeptide, with the polymerase activity mapping to the N terminus and the RNase H activity at the C terminus (52). Thus, the murine RT appears to be analogous to the α subunit of the avian RT.

Studies of the variability and evolution of RNA tumor viruses have suggested that, on average, these viruses have higher mutation rates than do the genomes of eucaryotic cells (7–9, 11–13, 17, 40, 46, 50, 53). It has been suggested that much of this variability could arise from the infidelity of reverse transcription (e.g., see reference 7). RTs, which lack 3′→5′ exonuclease activity and therefore cannot proofread errors made during synthesis (3), have high base substitution error rates as measured on synthetic polynucleotide templates (2, 38, 39, 48, 49, 51, 59). During synthesis on natural DNA templates, avian myeloblastosis virus (AMV) RT has low fidelity for certain base substitution errors at several different nonsense codons in φX174 DNA (16, 29, 33, 34, 38, 44).

Since these polymerases are likely to commit a variety of other types of errors (24, 25), we set out to define the complete mutagenic potential of RT during gap-filling DNA synthesis in vitro. For this we used an M13mp2-based forward mutation assay that allows the detection of all types of base substitution errors at a number of sites as well as frameshifts, deletions, insertions, and complex errors. We report here the mutation rate and error specificity of retroviral RT for a single round of DNA synthesis.

MATERIALS AND METHODS

Viral and bacterial strains. All the bacterial and viral strains used in this study have been described previously (23, 24, 32, 35).

Enzymes. AMV and Moloney MuLV (Mo-MuLV) RTs were obtained from Boehringer Mannheim Biochemicals (Indianapolis, Ind.) and Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), respectively. HeLa cell apurinic endonuclease (fraction VII; specific activity, 2×10^5 U/mg [21]) was a gift of Dale W. Mosbaugh (Department of Chemistry, University of Texas, Austin).

Construction of gapped M13mp2 molecules. M13mp2 DNA

* Corresponding author.

molecules containing a single-stranded gap in the region of the *lacZα* gene were constructed as described previously (24, 35) with either wild-type single-stranded DNA for the forward assay (24) or M13mp2A89 single-stranded DNA for the opal codon reversion assay (32).

Conditions and analysis of RT gap-filling synthesis. RT polymerase reactions (50 μl) contained 20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid)-KOH (pH 7.8); 10 mM $MgCl_2$; 500 μM dATP, dTTP, dGTP, and dCTP; 2 mM dithiothreitol; 300 ng of gapped M13mp2 or M13 mp2A89 DNA; and 20 to 50 U of RT. One unit is that amount of enzyme that incorporates 1 nmol of dTMP into an acid-insoluble product in 10 min at 37°C with a poly(rA · $dT)_{12-18}$ template-primer. Reactions were incubated at 37°C for 60 min and stopped by adding disodium EDTA to a final concentration of 15 mM. DNA synthesis products were analyzed by agarose gel electrophoresis as described previously (24) and were assayed for plaque phenotype by transfection of competent *Escherichia coli* MC1061 (35). Mutant or revertant bacteriophage were genetically purified as described previously (23). DNA sequence analysis of mutant phage was performed by the chain-termination method (43) with oligonucleotides as described previously (23).

Assay for DNA synthesis from mismatched primer terminus. M13mp2 gapped DNA containing a T · C mispair at the primer terminus was constructed as described previously (35). This DNA was incubated with RT as described above except that the reaction volume was 30 μl and 12 U of RT was added. DNA synthesis products were examined for complete gap-filling synthesis by agarose gel electrophoresis (24) and were used to transfect competent *E. coli* MC1061. Plaque phenotypes were scored as described previously (35).

Reversion and forward mutation assays. Single-base substitution errors that occurred during DNA synthesis in vitro were measured by the opal codon reversion assay as previously described (32). An M13mp2 DNA molecule was constructed that had a 361-nucleotide gap containing a single base change (G→A in the viral [plus] template strand at position +89 of the *lacZα* coding sequence). This change creates an opal (TGA) codon, resulting in a colorless plaque phenotype under the appropriate plating conditions. The gap was filled by a single cycle of DNA synthesis in vitro with the chosen DNA polymerase and appropriate reaction conditions. A sample of the product was analyzed to confirm complete synthesis (24), and the remainder was used to transfect competent α-complementation *E. coli* host cells to determine the colors of the resulting M13mp2 plaques (23). Base substitution errors at the opal codon during gap-filling DNA synthesis were detected as blue plaques. The reversion frequency (the proportion of blue to total plaques) reflects the error rate for the single round of gap-filling DNA synthesis. Eight of nine possible base substitution errors at the TGA codon yield a detectable blue plaque phenotype.

A large variety of errors produced during DNA synthesis in vitro can be quantitated and subsequently recovered for DNA sequence analysis by the M13mp2 forward mutation assay (24). As described above, a gapped DNA substrate was used (in this case, the single-stranded gap was 390 nucleotides) but the target for mutations was the 250-nucleotide wild-type M13mp2 DNA sequence. The assay scores the loss of α-complementation of β-galactosidase activity, i.e., mutants are identified as light blue or colorless plaques. Since α-complementation activity is not essential for M13mp2 plaque production, over 200 different base substitution errors at ~110 different sites can be scored within the 250-nucleotide *lacZα* sequence (28). In addition, frameshift mutations (26), deletions (23, 25), and more complex errors (24, 25) can be detected. The mutant frequency, i.e., the number of light blue and white plaques relative to the total number of plaques scored, reflects the error rate, which can be precisely calculated as described previously (26, 28). Following confirmation of the altered plaque phenotype, the precise nature of the mutation can be determined by DNA sequence analysis (23).

TABLE 1. Base substitution fidelity of AMV RT in nonsense codon reversion assays

Reversion target	Plaques scored		Revertant frequency (10^6)	Error rate[a]
	Total	Revertant		
M13mp2A89 opal codon				
Uncopied[b]	2,900,000	4	1.4	
Copied, 20 μM dNTPs[b]	230,000	23	100	1/18,000
φX174 am3				
Uncopied[c]			2.5	
Copied, 10 μM dNTPs[d]			13.9	1/17,000
Copied, 5 μM dNTPs[e]			17.8	1/13,000

[a] The error rate was calculated by subtracting the background mutant frequency, dividing by the average probability of an error being expressed (0.6 in the M13mp2 assay, 0.2 in the φX174 assays [this includes a correction for 50% of the input φX174 molecules being copied]), and then dividing by the number of detectable sites (three for M13mp2A89, one for φX174 am3).

[b] RT copying reactions were performed as described in Materials and Methods. See reference 24 for a discussion of the uncopied background determination. dNTPs, Deoxynucleoside triphosphates.

[c] Data from reference 34.

[d] Data from reference 29.

[e] Data from reference 33.

RESULTS

Base substitution fidelity of AMV RT in opal codon reversion assay. To examine the frequency of base substitutions by RT during DNA synthesis in vitro, we used an opal codon reversion assay (32, 35). Gap-filling polymerization reactions were performed with AMV RT and the opal codon-containing M13mp2 DNA template. When a sample of the reaction mixture was analyzed by agarose gel electrophoresis, the DNA product appeared as a single homogeneous band that migrated coincident with replicative-form II DNA standard (data not shown; see reference 24), demonstrating that the vast majority of the gapped molecules had been copied. The remaining product DNA was then used to transfect competent cells (24) to score mutant and total plaques (Table 1). The frequency of revertants in the RT-copied DNA is 70-fold above the background, demonstrating that AMV RT produced base substitution errors during gap-filling synthesis. Using the revertant frequency to calculate the error rate (28, 35) (Table 1, footnote *a*) yields a value of one base substitution error for every 18,000 nucleotides incorporated. This error rate is similar to that found in previous studies of AMV RT with the φX174 am3 reversion assay (Table 1) (29, 33, 34). The error rates estimated for two other amber sites in φX174 (34) vary from one per 2,000 to one per 31,000 nucleotides incorporated, depending on the template position being scored. These data indicate that in reversion assays that focus on only a few sites, the base substitution fidelity of AMV RT is low.

Fidelity of RTs in M13mp2 forward mutation assay. We next used the M13mp2 forward mutation assay to determine the frequency of forward mutations (*lacZ*$^+$→*lacZ*) during DNA synthesis by several preparations of AMV and Mo-

TABLE 2. Forward mutant frequency of M13mp2 DNA copied by RTs[a]

RT	Plaques scored		Mutant frequency (10^4)
	Total	Mutant	
None[b]	199,655	128	6.4
AMV			
Expt 1	39,647	144	36
Expt 2	3,866	26	67
Expt 3	38,070	68	18
Expt 4	10,200	48	47
Mo-MuLV	50,356	113	22

[a] Reactions, product analysis, and transfections were performed as described in Materials and Methods and reference 35. Experiments 1 through 4 with AMV RT were done with different preparations of enzyme. Mutants for DNA sequence analysis were selected from experiment 1.
[b] Taken from reference 24.

MuLV RTs. DNA synthesis and product analyses were performed as before except that wild-type M13mp2 gapped DNA was used. In each case, the enzyme filled the gap (data not shown), and reaction products were used to transfect competent *E. coli*. The frequency of mutant (i.e., light blue or colorless) plaques for DNA copied by these enzymes (Table 2) was significantly (3- to 10-fold) above the background.

Mutational specificity of AMV RT. To examine the types of mutations most frequently produced by RT, we determined the DNA sequence of 66 *lacZ*α mutants generated by AMV RT. The results are shown in Fig. 1 and Table 3. The collection contained single-base substitutions, −1 frameshifts, and several mutations involving two or more template nucleotides. These data, in combination with existing information on the total number of detectable sites in the target (28), permit the calculation of the base substitution and −1 frameshift error rates. Thus, AMV RT commits base substitutions at a rate of one error for every 37,000 nucleotides incorporated. This reflects the average fidelity throughout the target (i.e., ~110 sites [28]). AMV RT appears to be more accurate for frameshifts, making −1 errors at a rate of one for every 87,000 nucleotides incorporated. This is an average value calculated for the entire target sequence containing ~150 template positions at which −1 errors can be scored (26).

Among the base substitutions were two striking asymmetries in specificity. First, all 38 base substitutions resulted from the misinsertion of purines, with a 2:1 preference for dAMP (26 misinsertions) over dGMP (12 misinsertions). This observation is consistent with previous results from a study of AMV RT by the φX174 am3 reversion assay in which 9 of 10 mutations identified by DNA sequencing resulted from the misinsertion of purine deoxynucleotides opposite the template A at position 587 of the φX174 sequence (34). The preferential incorporation of dAMP and dGMP opposite template purines is reminiscent of the specificity for insertion opposite abasic sites (44) and suggests that copying past cryptic sites of DNA damage is a mechanism for the specificity of base substitutions by AMV RT. To examine this possibility, we treated the M13mp2 gapped duplex with an apurinic endonuclease prior to copying the *lacZ*α region and then used the copied DNA to transfect *E. coli* as before. Treatment with the apurinic endonuclease had no significant effect on the mutation frequency in the AMV RT-copied DNA. Furthermore, when seven base substitution mutations from apurinic endonuclease-treated AMV RT-copied DNA were identified by DNA sequencing, all were found to be due to the misinsertion of purine residues. Thus, the potential presence of apurinic sites in the M13mp2 DNA does not explain the specificity of misinsertion of purines by AMV RT. However, other types of cryptic damage, e.g., that caused by oxygen free radicals, are known to induce the preferential misincorporation of purines in SOS-induced *E. coli* (37); our controls do not eliminate the possible influence of this type of damage.

The second asymmetry in specificity involves the sites of base substitutions within the target. Even though the *lacZ*α sequence contains about the same number of detectable sites for errors at template A, G, T, and C positions, 22 of 38 base substitutions are at template G sites (7 are at A, 5 at C, and 4 at T). These lead to substantial and potentially informative differences in fidelity for specific mispairs. For example, mutants derived from the misinsertion of dAMP opposite G were recovered 15 times, while those resulting from the reciprocal mispair, dGMP opposite template A, were seen only once.

Similarly, asymmetries for the frameshifts are apparent. For example, all 20 single-base frameshifts were −1 events; no +1 errors were observed. Additionally, 16 of the 20 −1 errors were the loss of a base in a non-run or two-base run template sequence, and 15 of these 16 were the loss of a template purine.

In addition to the base substitutions and −1 errors committed by AMV RT, eight mutants contained other changes. The frequency of these changes is significantly greater than their background frequency (24). These mutants include two examples of the loss of two adjacent bases, one mutation involving the loss of two adjacent bases along with a G-to-T transversion 42 bases away, two examples of double-base substitutions at adjacent sites, a deletion of 40 bases in the regulatory region, a single-base insertion four bases away from a single-base deletion, and a complex mutant involving the replacement of 59 bases with 54 bases of new sequence, the origin of which is unclear.

TABLE 3. Classes of errors made by AMV RT

Class	No. of mutations	Frequency (10^4)[a]	Error rate[b]
Base substitutions	38	18	1/37,000
Minus-one-base frameshifts	20	10.3	1/87,000
Other[c]	8	2.4	

[a] Background mutant frequencies of 3×10^{-4}, 0.6×10^{-4}, and 2×10^{-4} for base substitutions, −1 frameshifts, and other mutations, respectively, were subtracted.
[b] The average fidelity (error rate) is calculated as described in Table 1, footnote *a*. The number of detectable sites in this assay is 110 for base substitutions and 150 for −1 errors.
[c] Other mutations include two-base losses, multiple base substitutions, and complex insertion-deletion events.

DNA synthesis by RT on a DNA template containing a mismatched 3′ primer terminus. The error rate of a polymerase will depend in part on the efficiency with which the enzyme continues synthesis of the nascent DNA strand from the mismatched primer terminus created by a misinsertion event. To test the ability of RT to extend DNA from a mispaired terminus, as well as to confirm the reported absence of 3′→5′ exonuclease activity in AMV RT (3), we constructed a gapped M13mp2 template that contained a T · C mismatch at the 3′-OH end of the primer strand (Fig. 2A). The position and mismatch were chosen so that expression of the T-containing template strand will produce a dark blue plaque, whereas expression of the C-containing primer

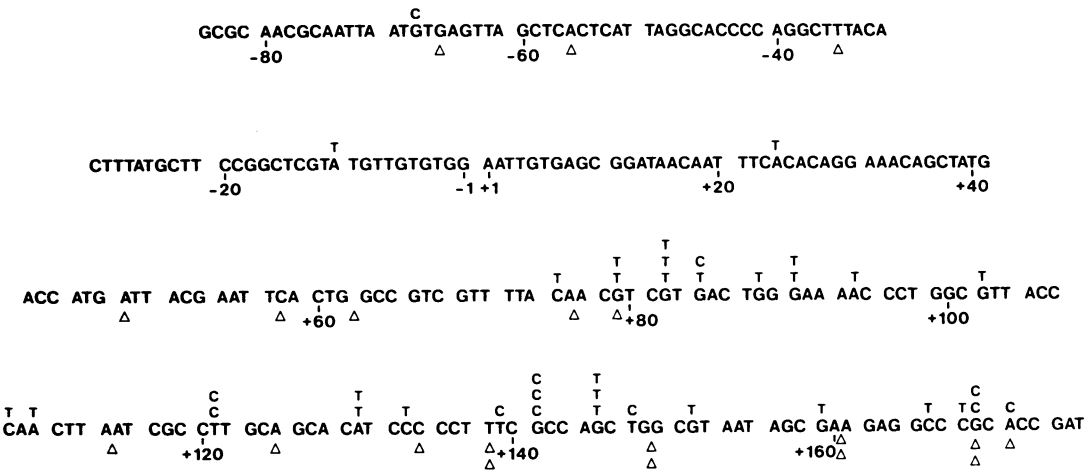

FIG. 1. Spectrum of single-base mutations produced by AMV RT. Four lines of primary wild-type viral strand DNA sequence are shown. The upper two lines are the regulatory regions for the *lacZ*α gene; the lower two lines indicate the protein-coding region of the gene. This figure presents only single-base mutations, with base substitutions shown above each line of wild-type sequence and frameshifts shown below each line. The letters used for the base substitutions (shown directly above the wild-type base) indicate the new base found in the viral template strand DNA sequence. For −1 errors, the loss of a base is indicated by Δ directly below the base. When a frameshift occurs in a sequence with two or more adjacent identical bases, it is not possible to assign the event to an individual base, and the symbol is centered under the run. The spectrum shown here is a composite of independent mutations.

strand will yield a light blue plaque (35). A polymerase with an active proofreading function should remove the mismatched terminal C residue prior to nucleotide addition and will produce predominantly dark blue plaques upon transfection of competent *E. coli* cells with the DNA product. On the other hand, an enzyme that polymerizes from the mispair without excising the mismatched nucleotide should produce a heteroduplex DNA that will yield a mixture of light blue and dark blue plaques. Figure 2B describes both of these possibilities, including the pertinent codons, amino acids, and plaque colors.

We examined several DNA polymerases for their ability to extend or excise the terminal mismatched base in this assay (Table 4). Polymerization reactions were performed and a sample of each reaction mixture was analyzed. In all cases, gap-filled DNA synthesis had occurred (data not shown). As expected, the Klenow fragment of *E. coli* polymerase I excised almost all the mismatched C at the primer terminus prior to polymerization, since >99% of the resulting plaques were dark blue (i.e., 4+). In contrast, AMV and Mo-MuLV RTs, known to lack associated exonuclease activity (for example, see reference 3), failed to remove the mispaired C and yielded a significant number of light blue plaques. This result confirms the absence of a proofreading exonuclease activity in these two RTs. Surprisingly, an even larger proportion of colorless plaques was generated by the RTs. Since the colorless plaque frequency was far in excess of what would be expected based on the infidelity of gap-filling DNA synthesis (24), several RT-generated colorless mutants were examined by DNA sequence analysis. In every case the mutant was missing one of the two template T residues at the template-primer terminus (position 103 or 104). A model for how this may have occurred includes a template-primer rearrangement such that the mispaired terminal C in the primer pairs with the next template base, a G. This creates a template-primer with a correct G · C base pair at the terminus but an unpaired T residue in the template strand. Because the DNA sequence at this site consists of two adjacent T residues in the template strand, the unpaired T could reside two base pairs upstream from the primer terminus. The product of this rearrangement is shown as the slippage event in Fig. 2B.

The results with another exonuclease-deficient polymerase, eucaryotic polymerase β (see reference 15 and references therein), are also shown in Table 4. As with the RTs, these enzymes were unable to remove the mismatched C, thus producing a substantial number of light blue plaques, a result consistent with the lack of a 3′→5′ exonuclease in this enzyme. However, polymerase β yielded a much smaller proportion of colorless plaques, with a colorless-to-light blue ratio 10- to 40-fold lower than that generated by the RTs. This result suggests that the slippage event, proposed above to explain the RT result, is not occurring as frequently with polymerase β and may reflect differences in how these polymerases interact with the template-primer.

DISCUSSION

We examined retroviral RTs from AMV and Mo-MuLV in fidelity assays and found that these enzymes are error prone, as manifested by the frequencies of base substitutions, −1 frameshifts, and complex errors in the polymerization products. These error frequencies are high relative to those of DNA polymerases that are capable of proofreading, such as *E. coli* DNA polymerase I (4), *E. coli* DNA polymerase III (36), T4 DNA polymerase (19), chicken polymerase γ (35), and calf polymerase δ (5, 32). The base substitution fidelity of exonuclease-deficient cellular DNA polymerases varies over a 6,000-fold range and depends not only on the polymerase but also on the template position and type of error (for a review, see reference 28a). In the φX174 DNA fidelity assay, AMV RT is much less accurate than the DNA polymerase α-primase complex from *Drosophila melanogaster* (20) and some preparations of polymerase α from calf thymus (41; but see reference 31). On the other hand, in the M13mp2 assay, RT preparations are approximately as accurate as human (25, 42) and calf thymus (28b) DNA polymerase α-primase and are much more accurate than preparations of DNA polymerase β (24). This difference could be a manifestation of how these enzymes interact at the am3 site

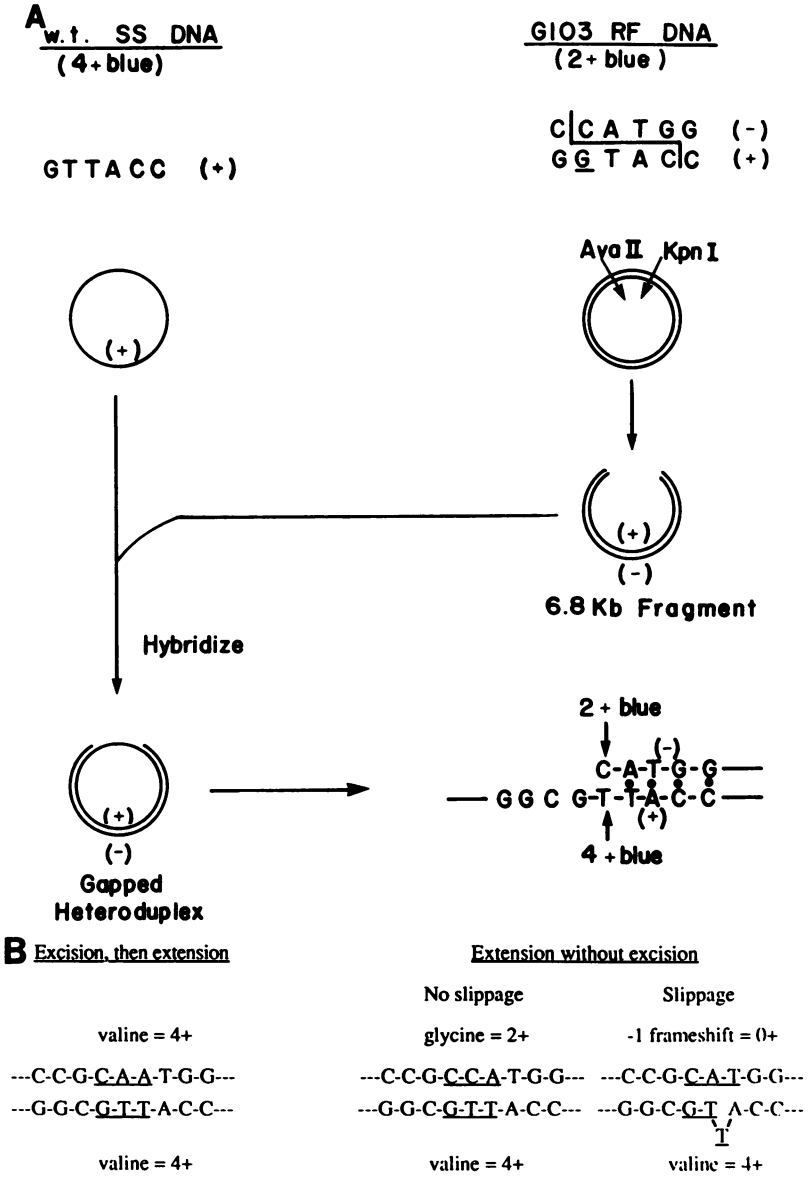

FIG. 2. (A) Construction of M13mp2 gapped molecule containing a terminal mismatch at the 3'-OH end of the primer. M13mp2G103 DNA containing unique sites for both restriction endonucleases *Ava*II and *Kpn*I (the latter site is not present in M13mp2 but was created by the T-to-G base change at position 103) was digested to produce fragments ~6,800 and ~360 base pairs long. These fragments were separated by precipitation with polyethylene glycol as described previously (35). The large fragment was then hybridized to single-stranded circular viral (plus) DNA as described previously (35). Using a 1/1 ratio of fragment to viral DNA, about one-half of the single-stranded DNA was converted to a heteroduplex molecule containing a 363-base gap from positions −261 to +102, with a terminal mispair at position +10. w.t., Wild type; RF, replicative form. (B) Results of DNA synthesis from a terminal mismatch in M13mp2. The upper line in each case represents the minus-strand sequence; the lower line, the plus (viral)-strand sequence. If a proofreading exonuclease removes the C from the primer strand of the terminal mismatch, extension yields the wild-type sequence in both strands, i.e., coding for the amino acid valine at codon 21 of the *lacZ* sequence. If the terminal C is not removed, two outcomes are possible. Extension from the mispair (No slippage) produces a heteroduplex molecule in which the minus-strand sequence codes for glycine at codon 21, producing a light blue (2+) phenotype. Slippage of the template-primer to create a G · C base pair at the primer terminus yields, upon extension, a −1 frameshift in the minus strand, generating, upon replication of the heteroduplex in vivo, the loss of a T at position +103 to +104, producing a colorless phenotype.

on φX174 DNA or could result from differences in the preparations of DNA polymerases.

Studies of the fidelity of RT in vitro have produced estimates of error rates that vary from one error per 300 nucleotides incorporated (16) to one error per 31,000 nucleotides incorporated (34). There are at least three possible reasons for these variations in error rates. First, many of the early studies utilized homopolymeric or simple repeating heteropolymeric templates that may not reflect accurately the structure of natural DNA sequences. Second, purification procedures may have evolved to yield more intact or different forms of RT. Third, previous studies with the φX174 amber reversion assay and the present study both demonstrate that error rates depend on the nature and position of the error as well as on the polymerase itself (34).

Each of the asymmetries noted here for the AMV-induced

TABLE 4. Terminal mismatch utilization by DNA polymerases[a]

DNA polymerase used	Plaques scored			Colorless/ light blue ratio
	Total[b]	Colorless (0+)	Light blue (2+)	
AMV RT	284	63	27	2/1
Mo-MuLV RT	1,162	216	30	7/1
Polymerase β	2,151	136	839	1/6
E. coli polymerase I (Klenow fragment)	1,521	0	11	

[a] Reactions (30 μl) was performed as described in Material and Methods, using 150 ng of gapped M13mp2 DNA with a T · C mismatch at the primer terminus. RT reaction mixtures (containing 12 U of enzyme) were incubated for 60 min; polymerase β reaction mixtures (containing 0.2 U of enzyme) were incubated for 60 min; and *E. coli* polymerase I (Klenow fragment) reaction mixtures (containing 0.5 U of enzyme) were incubated for 20 min.

[b] Total includes dark blue (4+), light blue (2+), and colorless (0+) plaques.

base substitution and frameshift errors has been observed before with other exonuclease-deficient polymerases, and potential mechanisms have been discussed (26, 28; see reference 28b for a review). For example, the A · G versus G · A mispair symmetry difference observed with AMV RT may have a simple structural basis. The crystal structure of the G · A mispair in the B-type helix suggests that the orientation of the base about the glycosyl linkage is *anti* for G but *syn* for A, allowing two hydrogen bonds (22). It is possible that a template A cannot assume the *syn* conformation required to mispair with an incoming G because stacking interactions with neighboring bases prevent this rotation, whereas an incoming A can easily rotate from *anti* to *syn* to pair with a template G. With regard to the −1 errors, we have proposed that the frequent production of frameshifts at non-run sites by some DNA polymerases may occur by misincorporation of a single base followed by rearrangement of the template-primer before the next base is added (27). In this situation, the misinserted base in the primer strand forms a single correct base pair with a template base at some other position. The demonstration that RTs produce frameshifts at non-run template sites (9 of the 20 −1 errors; Fig. 1) and use template-primers that contain an unpaired template base close to the primer terminus lends support to this idea. Furthermore, the preference of RT for committing both base substitutions and −1 errors at template purines is consistent with this model. These results also suggest that, given the choice of extending from a template-primer containing a T · C terminal mispair or a correctly paired terminus with an unpaired template base two base pairs upstream of the primer terminus, cellular DNA polymerases usually make a very different decision than do RTs.

Several complex errors were also found in the collection of forward mutants from the AMV RT polymerization reaction. This indicates that RT also creates errors by mechanisms that involve larger misalignments. Previous examples of mutations created during DNA synthesis in vitro that involve more than one nucleotide position (10, 24, 27, 30) have led to models that involve the unpairing of bases at the primer terminus followed by hybridization at a new position to provide a functional template-primer for continued synthesis. Errors that fit this model are simple deletions between direct repeats. A more complex derivative of this model, involving transient misalignment, has been described to explain a complex deletion (27) and simple two-base deletions (10). The more complicated errors seen here may reflect such mechanisms and, perhaps more importantly, properties of retroviral RTs that are relevant to their ability to jump from one DNA molecule to another in vivo (56). It is possible, and in some cases appears likely (46, 47), that recombination, perhaps involving copy choice reverse transcription (6), is involved in retrovirus variation.

There are multiple steps in the life cycle of a retrovirus that could contribute to mutation rates in the virus, including reverse transcription, replication by host cell enzymes during maintenance of the provirus, and transcription of the genomic RNA by cellular RNA polymerases. Most of the viral studies to date have provided data that only allow estimates of cumulative mutations per genome per year and do not allow an estimate of the error rate per round of replication. However, recent advances in retrovirus expression vectors have led to estimates of error rates (2×10^{-5} base substitutions per base pair per replication cycle) in viral sequences during the progression from the proviral state in one cell through a single round of transcription, infection, and integration into another cell type (11, 12). The error rate of AMV RT (2.7×10^{-5} for base substitutions), as determined in the M13mp2 forward mutation assay in vitro, is sufficient to account for the viral mutation rates reported in these studies. It should prove interesting to learn what role, if any, other steps in the retroviral life cycle play in determining the rate and specificity of mutation in these genomes.

ACKNOWLEDGMENTS

We thank L. Boone and R. Schaaper (National Institute of Environmental Health Services) for their helpful suggestions and critical evaluation of the manuscript.

LITERATURE CITED

1. **Baltimore, D.** 1970. RNA-dependent DNA polymerase in virions of RNA tumor viruses. Nature (London) **226**:1209–1211.
2. **Battula, N., and L. A. Loeb.** 1974. The infidelity of avian myeloblastosis virus deoxyribonucleic acid polymerase in polynucleotide replication. J. Biol. Chem. **249**:4086–4093.
3. **Battula, N., and L. A. Loeb.** 1976. On the fidelity of DNA replication. Lack of exodeoxyribonuclease activity and error-correcting function in avian myeloblastosis virus DNA polymerase. J. Biol. Chem. **251**:982–986.
4. **Brutlag, D., and A. Kornberg.** 1972. Enzymatic synthesis of deoxyribonucleic acid. XXXVI. A proofreading function for the 3′-5′ exonuclease activity in deoxyribonucleic acid polymerases. J. Biol. Chem. **247**:241–248.
5. **Byrnes, J. J., K. M. Downey, V. L. Black, and A. G. So.** 1976. A new mammalian DNA polymerase with 3′ to 5′ exonuclease activity: DNA polymerase δ. Biochemistry **15**:2817–2823.
6. **Coffin, J.** 1979. Structure, replication and recombination of retrovirus genomes: some unifying hypotheses. J. Gen. Virol. **42**:1–26.
7. **Coffin, J. M.** 1986. Genetic variation in AIDS viruses. Cell **46**:1–4.
8. **Coffin, J. M., P. N. Tsichlis, C. S. Barker, S. Voynew, and H. L. Robinson.** 1980. Variation in retrovirus genomes. Ann. N.Y. Acad. Sci. **354**:410–425.
9. **Darlix, J.-L., and P.-F. Spahr.** 1983. High spontaneous mutation rate of Rous sarcoma virus demonstrated by direct sequencing of the RNA genome. Nucleic Acids Res. **11**:5953–5967.
10. **de Boer, J. G., and L. S. Ripley.** 1988. An *in vitro* assay for frameshift mutations: hotspots for deletions of 1 bp by Klenow-fragment polymerase share a consensus DNA sequence. Genetics **118**:181–191.
11. **Dougherty, J. P., and H. M. Temin.** 1987. High mutation rate of a spleen necrosis virus-based retrovirus vector. Mol. Cell. Biol. **168**:4387–4395.
12. **Dougherty, J. P., and H. M. Temin.** 1988. Determination of the

rate of base-pair substitution and insertion mutations in retrovirus replication. J. Virol. 62:2817–2822.
13. **Duran-Reynals, F.** 1942. The reciprocal infection of ducks and chickens with tumor-inducing viruses. Cancer Res. 2:343–369.
14. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. deHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. J. Virol. 56:589–599.
15. **Fry, M., and L. A. Loeb.** 1986. Animal cell DNA polymerases, p. 77–78. CRC Press, Inc., Boca Raton, Fla.
16. **Gopinathan, K. P., L. A. Weymouth, T. A. Kunkel, and L. A. Loeb.** 1979. Mutagenesis *in vitro* by DNA polymerase from an RNA tumor virus. Nature (London) 278:857–859.
17. **Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science 232:1548–1553.
18. **Hizi, A., and W. K. Joklik.** 1977. RNA-dependent DNA polymerase of avian sarcoma virus B77. I. Isolation and partial characterization of the α, β2 and αβ forms of the enzyme. J. Biol. Chem. 252:2281–2289.
19. **Huang, W. M., and I. R. Lehman.** 1972. On the exonuclease activity of phage T4 deoxyribonucleic acid polymerase. J. Biol. Chem. 247:3139–3146.
20. **Kaguni, L. S., R. A. DiFrancesco, and I. R. Lehman.** 1984. The DNA polymerase-primase from *Drosophila melanogaster* embryos. Rate and fidelity of polymerization on single-stranded DNA templates. J. Biol. Chem. 259:9314–9319.
21. **Kane, C. M., and S. Linn.** 1981. Purification and characterization of an apurinic/apyrimidinic endonuclease from HeLa cells. J. Biol. Chem. 256:3405–3414.
22. **Kennard, O.** 1987. The molecular structure of base pair mismatches, p. 25–52. *In* F. Eckstein and D. M. J. Lilley (ed.), Nucleic Acids and Molecular Biology, vol. 1. Springer-Verlag KG, Berlin.
23. **Kunkel, T. A.** 1984. Mutational specificity of depurination. Proc. Natl. Acad. Sci. USA 81:1494–1498.
24. **Kunkel, T. A.** 1985. The mutational specificity of DNA polymerase-β during *in vitro* DNA synthesis. Production of frameshift, base substitution, and deletion mutants. J. Biol. Chem. 260:5787–5796.
25. **Kunkel, T. A.** 1985. The mutational specificity of DNA polymerases-α and -γ during *in vitro* DNA synthesis. J. Biol. Chem. 260:12,866–12,874.
26. **Kunkel, T. A.** 1986. Frameshift mutagenesis by eucaryotic DNA polymerases *in vitro*. J. Biol. Chem. 261:13,581–13,587.
27. **Kunkel, T. A.** 1988. Mutagenesis by transient misalignment. J. Biol. Chem. 263:14,784–14,789.
28. **Kunkel, T. A., and P. S. Alexander.** 1986. The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. J. Biol. Chem. 261:160–166.
28a. **Kunkel, T. A., R. A. Bambara, K. Bebenek, J. D. Roberts, R. D. Sabatino, M. P. Smith, and A. Soni.** 1988. Analysis of mutational mechanisms with eukaryotic DNA polymerases. ICN-UCLA Symp. Mol. Cell. Biol. New Ser. 83:521–528.
28b. **Kunkel, T. A., and K. Bebenek.** 1988. Recent studies of the fidelity of DNA synthesis. Biochim. Biophys. Acta 951:1–15.
29. **Kunkel, T. A., F. Eckstein, A. S. Mildvan, R. M. Koplitz, and L. A. Loeb.** 1981. Deoxynucleoside [1-thio]triphosphates prevent proofreading during *in vitro* DNA synthesis. Proc. Natl. Acad. Sci. USA 78:6734–6738.
30. **Kunkel, T. A., K. P. Gopinathan, D. K. Dube, E. T. Snow, and L. A. Loeb.** 1986. Rearrangements of DNA mediated by terminal transferase. Proc. Natl. Acad. Sci. USA 83:1867–1871.
31. **Kunkel, T. A., and L. A. Loeb.** 1981. Fidelity of mammalian DNA polymerases. Science 213:765–767.
32. **Kunkel, T. A., R. D. Sabatino, and R. A. Bambara.** 1987. Exonucleolytic proofreading by calf thymus DNA polymerase δ. Proc. Natl. Acad. Sci. USA 84:4865–4869.
33. **Kunkel, T. A., R. M. Schaaper, R. Beckman, and L. A. Loeb.** 1981. On the fidelity of DNA replication. Effect of the next nucleotide on proofreading. J. Biol. Chem. 256:9883–9889.
34. **Kunkel, T. A., R. M. Schaaper, and L. A. Loeb.** 1983. Depurination-induced infidelity of deoxyribonucleic acid synthesis with purified deoxyribonucleic acid replication proteins *in vitro*. Biochemistry 22:2378–2384.
35. **Kunkel, T. A., and A. Soni.** 1988. Exonucleolytic proofreading enhances the fidelity of DNA synthesis by chick embryo DNA polymerase-γ. J. Biol. Chem. 263:4450–4459.
36. **Livingston, D. M., and C. C. Richardson.** 1975. Deoxyribonucleic acid polymerase III of *Escherichia coli*. Characterization of associated exonuclease activities. J. Biol. Chem. 250:470–478.
37. **Loeb, L. A., E. A. James, A. M. Waltersdorph, and S. J. Klebanoff.** 1988. Mutagenesis by the autoxidation of iron with isolated DNA. Proc. Natl. Acad. Sci. USA 85:3918–3922.
38. **Loeb, L. A., and T. A. Kunkel.** 1982. Fidelity of DNA synthesis. Annu. Rev. Biochem. 52:429–457.
39. **Mizutani, S., and H. M. Temin.** 1976. Incorporation of noncomplementary nucleotides at high frequencies by ribodeoxyvirus DNA polymerases and *Escherichia coli* DNA polymerase I. Biochemistry 15:1510–1516.
40. **O'Rear, J. J., and H. M. Temin.** 1982. Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus and avian retrovirus. Proc. Natl. Acad. Sci. USA 79:1230–1234.
41. **Reyland, M. E., and L. A. Loeb.** 1987. On the fidelity of DNA replication. Isolation of high fidelity DNA polymerase-primase complexes by immunoaffinity chromatography. J. Biol. Chem. 262:10824–10830.
42. **Roberts, J. D., and T. A. Kunkel.** 1988. Fidelity of a human cell DNA replication complex. Proc. Natl. Acad. Sci. USA 85:7064–7068.
43. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
44. **Schaaper, R. M., T. A. Kunkel, and L. A. Loeb.** 1983. Infidelity of DNA synthesis associated with bypass of apurinic sites. Proc. Natl. Acad. Sci. USA 80:487–491.
45. **Schiff, R. D., and D. P. Grandgenett.** 1978. Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the β polypeptide of the avian retrovirus DNA polymerase. J. Virol. 28:279–292.
46. **Shields, A., O. N. Witte, E. Rothenberg, and D. Baltimore.** 1978. High frequency of aberrant expression of Moloney murine leukemia virus in clonal infections. Cell 14:601–609.
47. **Shimotohno, K., and H. M. Temin.** 1982. Spontaneous variation and synthesis in the U3 region of the long terminal repeat of an avian retrovirus. J. Virol. 41:163–171.
48. **Sirover, M. A., and L. A. Loeb.** 1974. Infidelity of DNA synthesis: a general property of RNA tumor viruses. Biochem. Biophys. Res. Commun. 61:410–414.
49. **Sirover, M. A., and L. A. Loeb.** 1977. On the fidelity of DNA replication. Effect of metal activators during synthesis with avian myeloblastosis virus DNA polymerase. J. Biol. Chem. 252:3605–3610.
50. **Smith, D. B., and S. C. Inglis.** 1987. The mutation rate and variability of eukaryotic viruses: an analytical review. J. Gen. Virol. 68:2729–2740.
51. **Springgate, C. F., N. Battula, and L. A. Loeb.** 1973. Infidelity of DNA synthesis by reverse transcriptase. Biochem. Biophys. Res. Commun. 52:401–406.
52. **Tanese, N., and S. P. Goff.** 1988. Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. Proc. Natl. Acad. Sci. USA 85:1777–1781.
53. **Temin, H. M.** 1961. Mixed infection with two types of Rous sarcoma virus. Virology 13:158–163.
54. **Temin, H. M., and S. Mizutani.** 1970. RNA-directed DNA polymerase in virions of Rous sarcoma virus. Nature (London)

226:1211–1213.
55. **Varmus, H.** 1988. Retroviruses. Science **240**:1427–1435.
56. **Varmus, H., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
57. **Verma, I. M.** 1975. Studies on reverse transcriptase of RNA tumor viruses. III. Properties of purified Moloney murine leukemia virus DNA polymerase and associated RNase H. J. Virol. **15**:843–854.
58. **Verma, I. M.** 1981. Reverse transcriptase, p. 87–103. *In* P. D. Boyer (ed.), The enzymes, vol. 14. Academic Press, Inc., Orlando, Fla.
59. **Weymouth, L. A., and L. A. Loeb.** 1977. Infidelity of DNA synthesis by temperature-sensitive DNA polymerases from RNA tumor viruses. Biochim. Biophys. Acta **478**:305–315.