JOURNAL OF VIROLOGY, Nov. 1996, p. 7594–7602 Vol. 70, No. 11
0022-538X/96/$04.00+0

Retroviral Mutation Rates and A-to-G Hypermutations during Different Stages of Retroviral Replication

THEODORE KIM, RONALD A. MUDRY, JR., CHARLES A. REXRODE II, AND VINAY K. PATHAK*

Department of Biochemistry and Mary Babb Randolph Cancer Center, West Virginia University, Morgantown, West Virginia 26506

Received 29 April 1996/Accepted 17 July 1996

Retroviruses mutate at a high rate in vivo during viral replication. Mutations may occur during proviral transcription by RNA polymerase II, during minus-strand DNA synthesis (RNA template) by viral reverse transcriptase, or during plus-strand DNA synthesis (DNA template) by reverse transcriptase. To determine the contributions of different stages of replication to the retroviral mutation rates, we developed a spleen necrosis virus-based in vivo system to selectively identify mutations occurring during the early stage (RNA transcription plus minus-strand synthesis) and the late stage (plus-strand synthesis plus DNA repair). A *lacZ*α reporter gene was inserted into the long terminal repeat (LTR) of a spleen necrosis virus shuttle vector, and proviruses were recovered from infected cells as plasmids containing either one or both LTRs. Plasmids containing both LTRs generated a mutant phenotype only if the *lacZ*α genes in both LTRs were mutated, which is most likely to occur during the early stage. Mutant phenotypes were identified from plasmids containing one LTR regardless of the stage at which the mutations occurred. Thus, mutant frequencies obtained after recovery of plasmids containing both LTRs or one LTR provided early-stage and total mutation rates, respectively. Analysis of 56,409 proviruses suggested that the retroviral mutation rates during the early and late stages of replication were equal or within twofold of each other. In addition, two mutants with A-to-G hypermutations were discovered, suggesting a role for mammalian double-stranded RNA adenosine deaminase enzyme in retroviral mutations. These experiments provide a system to selectively identify mutations in the early stage of retroviral replication and to provide upper and lower limits to the in vivo mutation rates during minus-strand and plus-strand synthesis, respectively.

Retrovirus populations exhibit high variation and evolutionary potential (5, 38). The rates of genetic variation in retroviral populations depend upon the mutation and recombination rates per replication cycle, the replication rate (replication cycles/time), and the selective forces that act on the population (5, 8, 16, 23, 24, 28, 29). Recent studies of human immunodeficiency virus type 1 (HIV-1) in vivo dynamics suggest that approximately 10^4 to 10^5 mutations arise at each nucleotide position every day in an HIV-1-infected patient (7, 14, 43). High variation allows retroviruses to evade the immune response (26), acquire drug resistance (37), and frustrate efforts to develop effective vaccines (9). One model for HIV-1 pathogenesis has suggested that antigenic diversity is a cause and not a consequence of HIV-1-induced immunodeficiency (26).

One mechanism for generating mutations in retroviral genomes is likely to involve the virally encoded reverse transcriptase (RT), which lacks exonucleolytic proofreading activity and carries out error-prone polymerization. Two other polymerases could potentially contribute to the retroviral mutation rate. First, the host cell DNA polymerases that replicate the provirus through each cell division may introduce mutations in the viral genomes, but the mutation rates of higher eukaryotes is estimated to be at least 10^5- to 10^6-fold lower than retroviral mutation rates, and their contributions to the retroviral mutation rates are negligible (11). Second, the host cell RNA polymerase II that transcribes the provirus may generate retroviral mutations (6). The error rate of RNA polymerase II has not been measured, and its contributions to the retroviral mutation rates remain unknown.

The in vivo forward mutation rate for a single replication cycle was first determined for the spleen necrosis virus (SNV) (28, 29). Subsequently, the in vivo forward mutation rates have been measured for the bovine leukemia virus (23), HIV-1 (24), and Moloney murine leukemia virus (27). In these studies, the retroviral mutation rates for a single replication cycle have ranged from 0.48×10^{-5} to 3.4×10^{-5}/bp/replication cycle.

A single cycle of retroviral replication is defined as the steps in going from a provirus in a helper cell to a provirus in a target cell (8) (Fig. 1A). This replication cycle includes RNA transcription by RNA polymerase II, RNA-dependent DNA synthesis by RT, and DNA-dependent DNA synthesis by RT. Mutations occurring during DNA-dependent DNA synthesis result in DNA heteroduplexes, which may be acted upon by the host cell DNA repair enzymes. It is unknown whether the viral DNA is amenable to DNA repair before integration and cell division.

It would be desirable to know the contributions of the various nucleic acid polymerization steps in retroviral replication to the in vivo retroviral mutation rates. Several in vitro mutation studies have compared mutation rates during RNA- and DNA-dependent DNA synthesis (4, 17, 18). These studies suggest a higher mutation rate during RNA-dependent DNA synthesis (17), a higher mutation rate during DNA-dependent DNA synthesis (4), or equal mutation rates during RNA- and DNA-dependent DNA synthesis (18). An additional complicating factor in interpretation of these studies is that recent comparisons of in vitro and in vivo mutation rates for HIV-1 and murine leukemia virus RTs have indicated significant differences, suggesting that some elements affecting fidelity in vivo are absent in in vitro assays (24, 32, 34, 35, 40).

* Corresponding author. Mailing address: Department of Biochemistry, Mary Babb Randolph Cancer Center, West Virginia University, Morgantown, WV 26506. Phone: (304) 293-5949. Fax: (304) 293-4667. Electronic mail address: VPATHAK@WVUMBRCC1.hsc.wvu.edu.

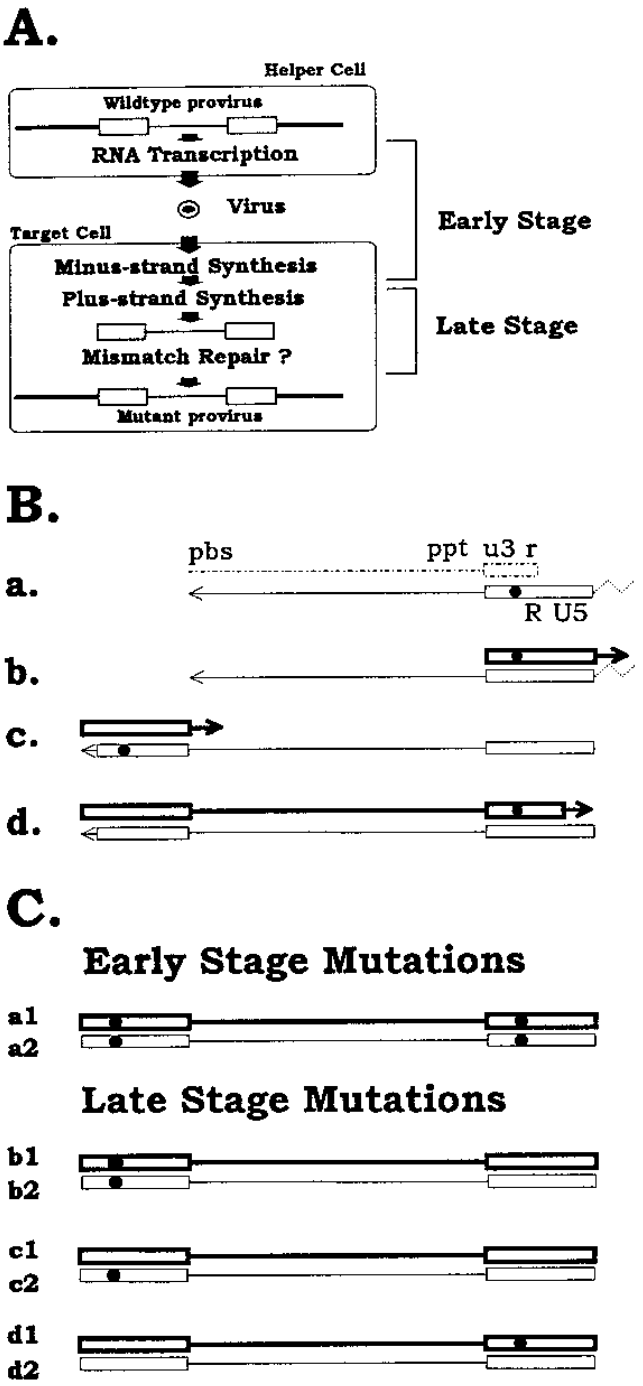

FIG. 1. (A) Retroviral mutations in different stages of retroviral replication. Different steps in retroviral replication that may affect the retroviral mutation rates are illustrated. Helper Cell, a virus-producing cell; Target Cell, a cell infected by retrovirus. A single round of retroviral replication is defined as the steps in going from a provirus in the helper cell to a provirus in the target cell. The replication cycle includes one cycle of RNA transcription in the helper cell, one cycle of minus-strand synthesis, and one cycle of plus-strand synthesis. Some mutations occurring during plus-strand synthesis result in heteroduplex formation, which may theoretically be acted upon by target cell DNA repair machinery. For the purposes of this experiment, the RNA transcription plus minus-strand synthesis is defined as the early stage, and plus-strand synthesis plus DNA repair is defined as the late stage. (B) Mutations during reverse transcription in the LTR regions of the retroviral genome. Four different stages of reverse transcription are shown. Stage a represents a mutation occurring in the early stage of replication, and stages b, c, and d represent mutations occurring in late stages of replication. a, a mutation occurs during minus-strand synthesis; b, a mutation occurs during the synthesis of plus-strand strong-stop DNA; c, a mutation occurs during the copying of the plus-strand strong-stop DNA; d, a mutation occurs during extension of the plus-strand strong-stop DNA after strand transfer. Dotted lines, RNA being degraded by RNaseH; thin lines, minus-strand DNA; thick lines, plus-strand DNA; boxes, LTR sequences; solid circles, mutations. (C) Viral genotypes resulting from mutations occurring in the LTR region. Each set of boxes connected by a line represents one DNA strand of a viral DNA. The designations a1 through d2 refer to each of the eight DNA strands shown. The set a1 and a2 refers to viral double-stranded DNA that resulted from the mutational event shown in panel B stage a. The sets b1 and b2, c1 and c2, and d1 and d2 also refer to double-stranded viral DNAs resulting from events depicted in panel B. Thick lines, plus-strand DNA; thin lines, minus-strand DNA; solid circles, mutations. The a1-a2 genotype resulted from mutations occurring during the early stage of retroviral replication (A). The sets b1-b2, c1-c2, and d1-d2 are genotypes that result from mutations occurring during the late stage of replication. The sets c1-c2 and d1-d2 form heteroduplexes.

TABLE 1. Recovery of BK-2 proviruses after *Bam*HI and *Hin*dIII digestions

Helper cell clone no.	Proviruses recovered after *Bam*HI digestion		Proviruses recovered after *Hin*dIII digestion		Mutant frequency ratio (%)[a]
	No. of mutants/ total no. recovered	Mutant frequency (10^{-3})	No. of mutants/ total no. recovered	Mutant frequency (10^{-3})	
P1C2	28/8,844	3.17	20/10,488	1.91	60
P2C1	37/9,586	3.86	14/5,259	2.66	69
P7C3	38/12,218	3.11	25/10,014	2.50	80
Total	103/30,648	3.36	59/25,761	2.29	68

[a] A mutant frequency ratio was calculated by dividing the *Hin*dIII recovery mutant frequency by the *Bam*HI recovery mutant frequency, and multiplying by 100%. Statistical analysis was performed with SAS-JMP version 3. Analysis of variance with the helper cell clone and digest as the independent variables and mutation frequency as the dependent variable indicated that there was no significant difference ($P > 0.99$) between the helper cell clones. Analysis of variance revealed that the *Bam*HI recovery mutation rate was higher than the *Hin*dIII recovery mutation rate ($P = 0.0290$).

In an effort to begin dissecting the retroviral mutation process, we have designed a retroviral vector and a strategy to distinguish in vivo retroviral mutations in the early and late stages of retroviral replication. We have defined the early stage to include transcription by RNA polymerase II and RNA-dependent DNA synthesis by RT and the late stage to include DNA-dependent DNA synthesis by RT and possible DNA repair of mismatches in the viral DNA prior to integration and cell division (Fig. 1A).

The strategy to distinguish in vivo mutations is based on the fact that during retroviral replication, mutations that occur in the long terminal repeat (LTR) regions result in proviruses with different genotypes (Fig. 1B and C). A similar strategy was previously used to determine whether a stem-loop structure was deleted during minus-strand synthesis or plus-strand synthesis (30). First, a mutation that occurs during transcription by RNA polymerase II is present in the template RNA, and reverse transcription of this template will result in proviruses that have the same mutation in both DNA strands of the 5′ and the 3′ LTRs (Fig. 1C, a1 and a2). Similarly, a mutation that occurs during minus-strand synthesis will result in proviruses that also have the same mutation in both DNA strands of the 5′ and the 3′ LTRs (Fig. 1B, stage a; Fig. 1C, a1 and a2). Second, a mutation that occurs during the synthesis of plus-strand strong-stop DNA will have the mutation on both strands of the 5′ LTR, and the 3′ LTR will be wild type (Fig. 1B, stage b; Fig. 1C, b1 and b2). Third, a mutation that occurs during the copying of the plus-strand strong-stop DNA after strand transfer will result in a mutation in one of the DNA strands of the

5′ LTR, and the 3′ LTR will be wild type (Fig. 1B, stage c; Fig. 1C, c1 and c2). Fourth, a mutation that occurs during extension of plus-strand DNA after strand transfer will result in a mutation in one DNA strand of the 3′ LTR, and the 5′ LTR will be wild type (Fig. 1B, stage d; Fig. 1C, d1 and d2). Thus, mutations that occur during the early stage are present in both LTRs whereas mutations that occur during the late stage are present in either the 5′ LTR or the 3′ LTR (Fig. 1C). It is theoretically possible for two independent mutations to occur during plus-strand synthesis. However, two independent mutations in a single provirus are not expected to be detected in this experiment, since the frequency of such an event is expected to be extremely low. The frequency of a single mutation occurring in a provirus is 3.36×10^{-3} (Table 1); therefore, the frequency of two independent mutations occurring in a single provirus is approximately 1×10^{-5} (3.36×10^{-3} multiplied by 3.36×10^{-3}).

The viral DNAs shown in Fig. 1C will integrate into the host cell DNA to form proviruses. During cell division, each DNA strand will be replicated and separated into a daughter cell. Therefore, eight different proviral genotypes can be generated from viral DNAs that contain mutations. Two of eight proviral genotypes (Fig. 1C, a1 and a2) will contain mutations in both the 5′ and 3′ LTRs and will be identified as mutations that occurred during the early stage of replication. Assuming no strand-specific DNA repair, three of eight proviral genotypes will contain mutations in the 3′ LTR (Fig. 1C, a1, a2, and d1). In addition, five of eight proviral genotypes will contain mutations in the 5′ LTR (Fig. 1C, a1, a2, b1, b2, and c2). Mutations that occur during two stages of plus-strand synthesis (Fig. 1B, stages c and d) result in the formation of heteroduplexes that may be amenable to DNA repair by host cell enzymes. It is not clear, however, whether the heteroduplexes are corrected by DNA repair and, if they are repaired, whether the correction exhibits strand specificity. Assuming no DNA repair or repair without strand specificity, 50% of the mutated strands will be corrected to wild type and 50% of the wild-type strands will be corrected to the mutants. As a result, the frequency of mutant proviruses present after cell division will not be affected by the efficiency of DNA repair. (The potential impact of hypothetical strand-specific DNA repair on the retroviral mutation rates will be discussed more fully below.)

We designed a retroviral shuttle vector that contained a *lacZ*α mutation reporter gene in the U3 portion of the 3′ LTR. After one cycle of retroviral replication, we recovered proviruses containing only the 3′ *lacZ*α gene or both *lacZ*α genes. Proviral plasmids were screened in *Escherichia coli* for inactivation of the *lacZ*α genes by using α-complementation and white or light-blue colony phenotypes (28). Mutations in recovered proviral plasmids containing both LTRs are identified only if both *lacZ*α genes are inactivated, which is most likely to occur during the early stage of replication. Mutations in proviral plasmids containing only the 3′ *lacZ*α gene were identified regardless of the stage of replication in which the mutation occurred.

Using this vector and strategy, we analyzed 56,409 proviruses for mutations in a *lacZ*α gene after a single replication cycle. The results suggest that mutations occur at nearly equal rates during the early and late stages of retroviral replication.

MATERIALS AND METHODS

Plasmid construction. SNV-based retroviral vector plasmid BK-2 was derived from the previously described retroviral vector VP254 (30). Standard molecular cloning procedures were used (36). Briefly, the stem-loop structure consisting of four copies of an 8-bp *Aat*II linker (5′-GGACGTCC-3′) was removed and replaced with a single copy of the *Aat*II linker to generate VP258. VP258 contains a 1.33-kb *Hin*dIII fragment encoding the neomycin resistance gene (*neo*) from Tn*5* (19). VP258 was digested with *Hin*dIII and treated with Klenow enzyme to generate blunt ends, and the blunt-ended *neo* fragment was reinserted into the vector to generate BK-1. BK-1 was identical to VP258, except that the *Hin*dIII restriction sites flanking the *neo* fragment were destroyed. Next, BK-1 was digested with *Xba*I, which cuts between the encapsidation sequence and *neo*, and was treated with Klenow enzyme to generate blunt ends. A 1-kb fragment encoding the hygromycin phosphotransferase B gene (12) was obtained from the vector pTZHy (kindly provided by W.-S. Hu) and ligated to the *Xba*I-digested BK-1 to generate BK-2. A detailed description of all cloning steps is available upon request.

Cells, transfections, and infections. D17 cells and C3A2 cells (obtained from the American Type Culture Collection) were maintained in Dulbecco's modified Eagle's medium (ICN) supplemented with 6% bovine calf serum (Hyclone Laboratory), penicillin (50 U/ml; Gibco), and streptomycin (50 μg/ml; Gibco). D17 is a canine osteosarcoma cell line that can be infected with SNV. C3A2 is a D17-derived reticuloendotheliosis virus-based helper line that can be used to package SNV (42). The selective agents hygromycin B (Calbiochem) and ouabain (Sigma) were present at final concentrations of 120 μg/ml (0.23 mM) and 73 μg/ml (10^{-7} M), respectively.

Cells were transfected by the previously described dimethyl sulfoxide-Polybrene method (20). D17 cells were infected by plating 2×10^5 cells on 60-mm-diameter dishes or 1×10^6 cells on 100-mm-diameter dishes. At 24 h later, cells were infected with either 0.2 ml of virus (60-mm-diameter dishes) or 1 ml of virus (100-mm-diameter dishes) by using Polybrene (50 μg/ml [final concentration]) as previously described (16). Transfected or infected cells were subjected to hygromycin selection 24 h later or ouabain selection 48 h later.

Recovery, restriction mapping, and DNA sequence analysis of proviruses. Proviruses derived from infection of D17 cells with BK-2 were recovered by the previously described Lac repressor protein-mediated recovery method (28). Briefly, 500 μg of genomic DNA from the infected D17 cells was digested with either *Bam*HI or *Hin*dIII. The digestion reaction mixtures were adjusted to 150 mM NaCl, 10 mM EDTA, 50 μg of bovine serum albumin per ml, and 10% (vol/vol) glycerol at a final volume of 600 μl. Digested DNA was incubated with 1 μg of Lac repressor fusion protein (Promega) at room temperature for 20 min. The reaction mixture was filtered through nitrocellulose (Gelman), which was pretreated with 1.5 ml of wash buffer (150 mM NaCl, 10 mM EDTA). The nitrocellulose was washed again with 1.5 ml of wash buffer after the filtration and then eluted with 1.0 ml of elution buffer (10 mM Tris-HCl [pH 7.5], 10 mM EDTA, 10 mM isopropyl-β-D-thiogalactoside [IPTG]) at 37°C for 1 h. The nitrocellulose filter was removed, and the remaining elution buffer was extracted with an equal volume of phenol-chloroform. The sample was ethanol precipitated, centrifuged, and lyophilized. The dried sample was resuspended in ligation buffer (20 U/400 μl [Boehringer Mannheim Biochemicals] and ligated at 16°C for 16 h. The ligated sample was treated with 10 U of *Dpn*I at 37°C for 2 h, extracted with phenol-chloroform, ethanol precipitated, and lyophilized. The DNA was resuspended in 4.0 μl of H_2O and used to transform *E. coli* DH10B competent bacterial cells (Bethesda Research Laboratories/Life Sciences). The transformed bacterial cells were allowed to recover for 1 h in a 37°C shaking bath and then plated onto Luria-Bertani plates containing 400 μg of IPTG per liter, 50 mg of kanamycin per liter, and 1 ml of 2% 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) per liter.

Transformed DH10B bacterial cells were screened for light-blue or white mutant colony phenotypes. Recovered mutant proviruses were analyzed by restriction digestion and gel electrophoresis and further analyzed by DNA sequencing. All plasmid DNA sequencing was performed by using the Sequenase kit and protocols (United States Biochemical).

RESULTS

Strategy for identifying mutations during early and late stages of replication. To identify mutations that occur in the early and late stages of retroviral replication, we constructed an SNV-based retroviral vector, BK-2 (Fig. 2A). BK-2 contains all of the SNV *cis*-acting sequences needed for viral replication. In addition, it contains a hygromycin phosphotransferase B gene that is expressed from the 5′ LTR promoter, a neomycin phosphotransferase gene that is expressed only in prokaryotic cells from a prokaryotic promoter in the vector, a pBR322 origin of replication for propagation of proviral plasmids in *E. coli*, and a *lacZ*α gene inserted in the U3 region of the 3′ LTR. During reverse transcription, the *lacZ*α gene in the 3′ LTR is duplicated and is present in both LTRs of the resulting provirus. The *lacZ*α gene is driven by the prokaryotic promoter and is expressed only in *E. coli.*

The protocol used to generate pools of BK-2-infected cells and to recover proviral plasmids is outlined in Fig. 2B. The

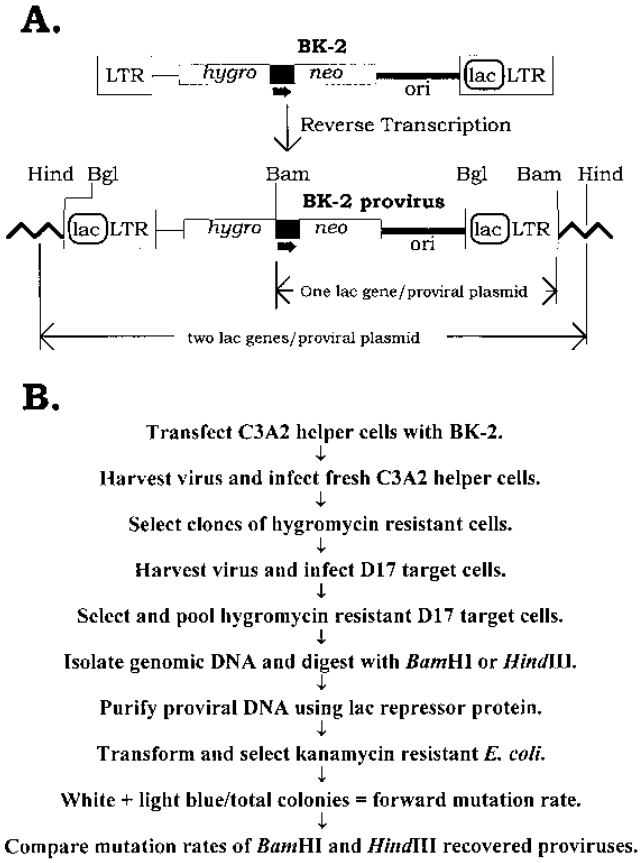

FIG. 2. (A) Structures of the SNV-based retroviral vector BK-2 and of the provirus derived from BK-2 and experimental strategy. BK-2 is designed to distinguish retroviral mutations occurring during early and late stages of retroviral replication. *hygro*, hygromycin phosphotransferase gene; *neo*, bacterial aminoglycoside 3′-phosphotransferase gene; black square and solid black arrow, bacterial promoter that expresses *neo* in *E. coli* and direction of transcription; ori, pBR322 bacterial origin of replication; lac, bacterial *lacZ*α peptide-encoding gene; thick zigzag lines, host chromosomal DNA adjacent to the integrated provirus; Hind, *Hin*dIII restriction sites; Bam, *Bam*HI restriction sites; Bgl, *Bgl*I restriction sites. The BK-2 viral genome has a *lacZ*α gene inserted at the 5′ end of the U3 region of the 3′ LTR. After a single round of replication, the BK-2 provirus contains a *lacZ*α in both 5′ and 3′ LTRs. (B) Outline of the experimental approach for determination of retroviral mutation rates during early and late stages of retroviral replication. The experimental approach taken to generate pools of target cells infected with BK-2 viruses is also described in Results. Genomic DNA of infected D17 target cells was purified and digested with *Bam*HI or *Hin*dIII. Proviral plasmids were recovered with Lac repressor protein (see Materials and Methods). The forward mutation rate was calculated by dividing the number of white plus light-blue colonies by the total number of recovered colonies. Proviral plasmids recovered after *Bam*HI digestion contain one *lac* gene from the 3′ LTR, whereas proviral plasmids recovered after *Hin*dIII digestion contain two *lac* genes (A). Mutant colonies are identified in the *Bam*HI pool regardless of the stage of retroviral replication at which the mutations occurred; mutant colonies are identified in the *Hin*dIII pool only if the mutations occurred during the early stage of retroviral replication (see Fig. 1 and Results).

BK-2 DNA was transfected into the C3A2 helper cell line, and pools of hygromycin-resistant helper cells were selected; virus was harvested and used to infect fresh C3A2 cells. Again, hygromycin-resistant cells were selected and individual cell clones were isolated and expanded. Thus, the BK-2 retroviral vector was introduced into the C3A2 cells by infection to avoid any mutations that could occur during transfection. Virus was harvested from the expanded C3A2 helper cell clones and used to generate pools of infected D17 target cells.

Genomic DNA was isolated from infected D17 cells and digested with *Bam*HI or *Hin*dIII (Fig. 2). The proviral DNA was purified by using the Lac repressor protein and its high affinity to the *lac* operator sequences (present in the BK-2 vector) as previously described (28, 29). The purified proviral DNA was ligated and transformed into *E. coli* by electroporation. The resulting transformants were plated on kanamycin plates in the presence of X-Gal, a color indicator substrate for the bacterial β-galactosidase gene. Mutations in the *lacZ*α gene were identified by the white or light-blue colony phenotype. A ratio of the mutant colonies to the total colonies recovered provided an in vivo forward mutation rate.

The strategy for distinguishing mutations in different stages of retroviral replication is based on the fact that mutations in the LTR regions during early and late stages of retroviral replication result in integrated proviruses with different genotypes (Fig. 1B and C). As shown in Fig. 2A, *Bam*HI digestion of genomic DNA results in recovery of proviral plasmids containing only one *lacZ*α gene present in the 3′ LTR of the integrated provirus. Mutations in this *lacZ*α gene result in a mutant colony phenotype without regard to the stage of retroviral replication at which the mutation occurred. In contrast, digestion of genomic DNA with *Hin*dIII, which does not cut in BK-2, results in recovery of proviral plasmids containing two *lacZ*α genes present in both LTRs of the integrated provirus. *Hin*dIII-recovered proviral plasmids must have a mutation in both copies of the *lacZ*α genes to provide a mutant colony phenotype, which is most likely if the mutation occurred in the early stage of retroviral replication (Fig. 1B and C). A mutation in only one *lacZ*α gene provides a wild-type colony phenotype, since the bacterial cells contain one functional copy of the *lacZ*α gene. Any mutations occurring in the late stage of retroviral replication, which mutate only one copy of the *lacZ*α gene, are not identified in *Hin*dIII-recovered proviral plasmids. Therefore, *Hin*dIII-recovered proviruses provide a mutation rate for only the early stage of replication whereas *Bam*HI-recovered proviruses provide the total mutation rate. Subtraction of the *Hin*dIII recovery mutation rate from the *Bam*HI recovery mutation rate provides an estimation of the mutation rate for the late stage of replication.

In vivo mutation rates for early and late stages of replication. BK2 vector virus was harvested from three C3A2 cell clones, and proviruses from pools of infected D17 cells were recovered and analyzed for mutations in the *lacZ*α reporter gene (Table 1). Each pool of colonies contained at least 5×10^6 independent colonies. No replication-competent viruses were found in cell supernatants from the pools of infected cells (data not shown). Proviruses were recovered from each pool of D17 target cells after *Bam*HI or *Hin*dIII digestions of the genomic DNAs and colony phenotype of the bacterial colonies was determined in the presence of X-Gal. A total of 30,648 proviruses were recovered after *Bam*HI digestion. Of these, 103 colonies had a white or light-blue mutant-colony phenotype, indicating a mutant frequency of 3.36×10^{-3} (1 in 298). In addition, a total of 25,761 colonies were recovered after *Hin*dIII digestion. Of these, 59 colonies had a mutant-colony phenotype, providing a lower mutant frequency of 2.29×10^{-3} (1 in 437). A comparison of the mutant frequencies after *Bam*HI and *Hin*dIII digestions revealed that the mutant frequency after *Hin*dIII digestions was approximately 68% of the mutant frequency after *Bam*HI digestion. This result indicated that approximately 68% of the retroviral mutations occur in the early stage and 32% of the mutations occur in the late stage of replication. No significant differences were noted in the pools generated from each of the three C3A2 virus-producing clones ($P > 0.99$; factorial analysis of variance). The *Bam*HI recovery mutant frequencies ranged from 3.11×10^{-3} (1 in

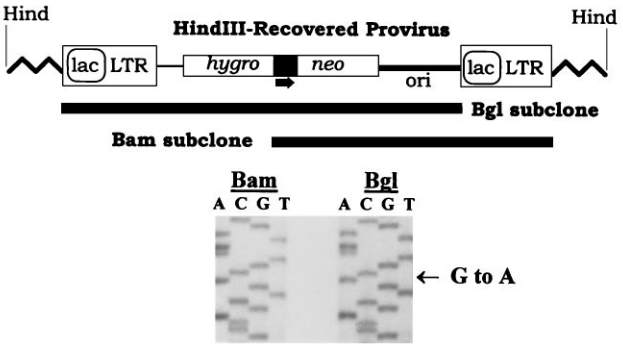

FIG. 3. Analysis of mutations present in both *lacZ*α genes of *Hin*dIII-recovered proviruses. The structure of a *Hin*dIII-recovered provirus is shown. All abbreviations are defined in the legend to Fig. 2. *Hin*dIII-recovered proviruses were digested with either *Bgl*I or *Bam*HI. The digested DNAs were self-ligated to form plasmids labelled Bgl subclone or Bam subclone (black boxes). The Bgl subclone contains the 5′ *lacZ*α gene, 5′ LTR, *hygro*, *neo*, and pBR origin of replication. The Bam subclone contains *neo*, pBR origin of replication, 3′ *lacZ*α gene, and 3′ LTR. DNA sequencing of the subclones derived from one *Hin*dIII-recovered provirus is shown. Both the 5′ and 3′ *lacZ*α genes contained the same G-to-A substitution.

322) to 3.86×10^{-3} (1 in 259). The *Hin*dIII recovery mutant frequencies ranged from 2.66×10^{-3} (1 in 376) to 1.91×10^{-3} (1 in 524). The ratios of *Hin*dIII to *Bam*HI mutant frequencies ranged from 60 to 80%. Statistical analysis of these results indicated a high level of significance ($P = 0.029$).

***Hin*dIII-recovered proviruses have two mutant *lacZ*α genes.** We experimentally verified the expectation that *Hin*dIII-recovered proviral plasmids contain two *lacZ*α genes and that both *lacZ*α genes must have the same mutation for generating a mutant colony phenotype (Fig. 3). Five *Hin*dIII-recovered mutant plasmids from three independent pools of infected cells were digested with either *Bam*HI or *Bgl*I. *Bam*HI digestion released a fragment containing *neo*, pBRori, and the 3′ *lacZ*α gene and LTR, whereas *Bgl*I digestion released a fragment containing the 5′ *lacZ*α gene and LTR, *hygro*, *neo*, and pBRori. The digested DNAs were ligated to form circular kanamycin-resistant plasmids containing either the 5′ *lacZ*α gene or the 3′ *lacZ*α gene. The ligated DNAs were transformed into an appropriate *E. coli* strain and plated on kanamycin plates containing X-Gal and IPTG. The phenotypes of colonies obtained after *Bam*HI and *Bgl*I digestions of *Hin*dIII-recovered proviral mutants were compared. All of the colonies obtained after *Bam*HI and *Bgl*I digestions had mutant colony phenotypes. This indicated that both the 5′ and 3′ copies of the *lacZ*α genes in *Hin*dIII-recovered mutants contained mutations that affected β-galactosidase activity. DNA sequencing of the 5′ and 3′ *lacZ*α genes verified that they contained the same mutations (Fig. 3).

Characterization of mutant proviruses recovered after *Bam*HI and *Hin*dIII digestion. The results obtained from DNA sequencing analysis of the mutant proviruses are summarized in Table 2. A total of 37 of 103 mutant clones of proviruses recovered after *Bam*HI digestion and 35 of 59 mutant proviruses recovered after *Hin*dIII digestion were characterized by DNA sequencing. The total mutation rate after *Bam*HI digestion was 2.2×10^{-5}/bp/cycle. The overall early stage mutation rate after *Hin*dIII digestion was 1.5×10^{-5}/bp/cycle. At least 27 different mutations were recovered after *Bam*HI digestion, and 30 different mutations were recovered after *Hin*dIII digestion. The spectrum of mutations and the frequencies of specific types of mutations after *Bam*HI and *Hin*dIII recoveries were compared. All of the previously described classes of mutations found in SNV were also identified in *Bam*HI- and *Hin*dIII-recovered proviruses (28, 29). These classes of mutations included substitutions, frameshifts, simple deletions, deletions with insertions, and duplications. In addition, two proviruses with uncommon A-to-G hypermutations and two duplication mutations were identified.

*Hin*dIII-recovered proviral plasmids had frameshift mutations, substitutions, simple deletions, deletions with insertions, and duplications, indicating that these classes of mutations can occur in the early stage of replication. The *Bam*HI recovery substitution rate (17.7×10^{-6} substitution/bp/cycle) was 53% higher than the *Hin*dIII recovery substitution rate (11.6×10^{-6} substitution/bp/cycle). The higher *Bam*HI recovery substitution mutation rate indicated that substitution mutations also occur in the late stage of replication. Similarly, the *Bam*HI recovery frameshift mutation rate (2.6×10^{-6} frameshift/bp/cycle) was 24% higher than the *Hin*dIII recovery frameshift mutation rate (2.1×10^{-6} frameshift/bp/cycle), suggesting that

TABLE 2. Mutation rates for different classes of mutations after *Bam*HI and *Hin*dIII digestions

Mutation class	*Bam*HI digestion		*Hin*dIII digestion	
	No. of different mutations/ total no. of mutations[a]	Mutation rate (10^{-6})[b]	No. of different mutations/ total no. of mutations	Mutation rate (10^{-6})
Substitutions	14/22	17.7	16/20	11.6
Frameshifts	7/8	2.6	8/9	2.1
Simple deletions	2/2	0.6	3/3	0.7
Deletions with insertions	1/2	0.6	2/2	0.5
Duplications	1/1	0.3	1/1	0.2
Hypermutations	2/2	0.6[c]	0	0
Total	27/37	22.4	30/35	15.1

[a] The first number indicates the number of different mutants identified. The second number indicates the number of mutants sequenced containing each mutation type.

[b] The mutation rates represents the number of mutations per base pair per replication cycle. Mutation rates were calculated as follows:

$$\frac{\text{no. of mutants identified}}{\text{total no. of mutants sequenced}} \times \frac{\text{total no. of mutants}}{\text{total no. of colonies}} \div T \text{ (target nucleotides)}$$

The total number of mutants and total number of colonies recovered are listed in Table 1. For substitutions, $T = 113$ target nucleotides of the *lacZ*α gene (2). For example, the *Bam*HI recovery substitution mutation rate equals $22/37 \times 103/30{,}648 \div 113 = 17.7 \times 10^{-6}$ substitutions per bp per replication cycle. For frameshifts, simple deletions, deletions with insertions, duplications and hypermutants, $T = 280$ target nucleotides (*lacZ*α gene plus promoter region).

[c] The mutation rate within the hypermutated regions was approximately 1.0×10^{-1}/bp (21 substitutions/180 target nucleotides for P1C2B:73, and 8 substitutions/76 nucleotides for P7C3B:8).

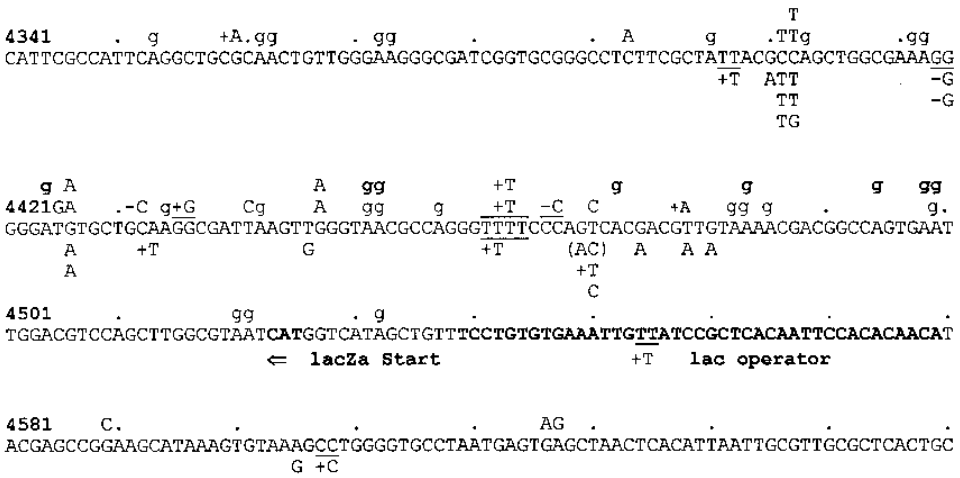

FIG. 4. Plus-strand nucleotide sequence of the *lacZ*α gene and location of mutations identified by DNA sequencing. Nucleotide numbers above each line refer to number beginning at the start of the 5′ LTR. Mutations identified in *Bam*HI-recovered proviruses are shown above the sequence; mutations identified in *Hin*dIII-recovered proviruses are shown below the sequence. All mutations shown are single-base substitutions, insertions or deletions, with the exception of a double substitution shown in parentheses. Mutations preceded by a plus or minus sign indicate +1 and −1 frameshifts, respectively. In cases when the frameshift occurred in runs of identical nucleotides, the run is identified by a line above or below the sequence. Two separate A-to-G hypermutant proviruses were identified. The A-to-G substitutions in the two separate hypermutants are shown as lowercase g: A-to-G substitutions in the P7C3B:8 hypermutant are shown in boldface type, and A-to-G substitutions in the P1C2B:73 hypermutant are shown in regular type.

frameshift mutations also occur in the late stage of replication. Simple deletions, deletions with insertions, and duplications occurred at similar rates in the *Bam*HI- and *Hin*dIII-recovered proviruses.

Analysis of substitutions and frameshift mutations. The locations of substitutions and frameshift mutations are shown with the *lacZ*α sequence (Fig. 4). The *Bam*HI-recovered mutations are shown above the sequence, and the *Hin*dIII-recovered mutations are shown below the sequence. At least 14 different substitution mutations were identified from the *Bam*HI-recovered pools and 16 substitutions were identified from the *Hin*dIII-recovered pools. Approximately 80% of the substitutions were transitions, and G-to-A and C-to-T substitutions predominated (62%). The ratios of transitions to transversions and the mutational specificities were similar in *Bam*HI recovery and *Hin*dIII recovery pools. We analyzed whether the substitutions observed here could have occurred by the dislocation mutagenesis model by analyzing the next nucleotide effect. In this model, a transient misalignment of the template strand and the nascent strand occurs, so that the nucleotide adjacent to the substitution site is copied. Only 6 of 30 substitutions could be explained by the dislocation mutagenesis model, indicating that most if not all of the in vivo substitution mutations occurred by misincorporation.

Fifteen frameshift mutations were also identified in *Bam*HI-recovered and *Hin*dIII-recovered proviruses. No differences were noted in the types of frameshifts or their locations with respect to the enzyme used for proviral recovery. Of 15 frameshifts, 11 were +1 frameshifts and the other 4 were −1 frameshifts. In addition, 11 of 15 frameshifts occurred in runs of identical nucleotides whereas the other 4 did not occur in runs of identical nucleotides.

A-to-G hypermutation. Two *Bam*HI-recovered mutants exhibited multiple A-to-G substitutions, constituting A-to-G hypermutation (Fig. 4). The A-to-G substitutions were located primarily within the *lacZ*α reading frame. The hypermutant P2C1H:73 had a region of 180 bp hypermutated, and 21 of 39 (54%) adenines in the *lacZ*α reading frame were mutated to guanines. Similarly, the hypermutant P7C3H:8 had a region of 76 bp hypermutated, and 8 of 20 (40%) adenines were mutated to guanines. DNA sequencing of more than 90% of the hypermutant plasmid P2C1H:73 revealed only one additional A-to-G substitution located 7 nucleotides 5′ of the *lacZ*α gene (data not shown). Additional DNA sequencing of the P7C3H:8 hypermutant plasmid was also performed, and no additional A-to-G substitutions were found 400 bp upstream or downstream of the *lacZ*α gene (data not shown).

Deletions, deletions with insertions, and duplications. Five independent simple deletions with characteristic short direct repeats at the deletion junctions were identified (Fig. 5A). The sizes of the direct repeats ranged from 2 to 8 nucleotides, and the sizes of the deletions ranged from 18 to 275 nucleotides. No differences were noted with respect to the size of the direct repeat or the length of deletion between *Bam*HI- and *Hin*dIII-recovered mutants. Similarly, three independent deletions with insertions containing characteristic short insertions at the site of deletions were identified (Fig. 5B). No differences were noted with respect to the size of deletion or insertion between *Bam*HI- and *Hin*dIII-recovered mutants. Two short duplications of 10 and 11 nucleotides were also identified (Fig. 5C). Both involved a short 3-nucleotide direct repeat and transient misalignments of the template and nascent strand.

DISCUSSION

Retroviral mutation rates in early and late stages of replication. The experiments reported here are the first in vivo studies to determine retroviral mutation rates in different stages of the viral replication. Our analysis of the retroviral mutations in the *lacZ*α reporter gene indicates that 68% of the mutations occur during the early stage and 32% of the mutations occur during the late stage of replication. These results suggest that the mutation rate during the early stage of replication is either equal to or twofold higher than the mutation rate during the late stage.

These slightly different interpretations are based on assumptions made about the role of DNA repair in modulating retroviral mutations. If it is assumed that DNA repair of viral DNA does not occur before cell division or if the DNA repair occurs without strand specificity, then the early- and late-stage

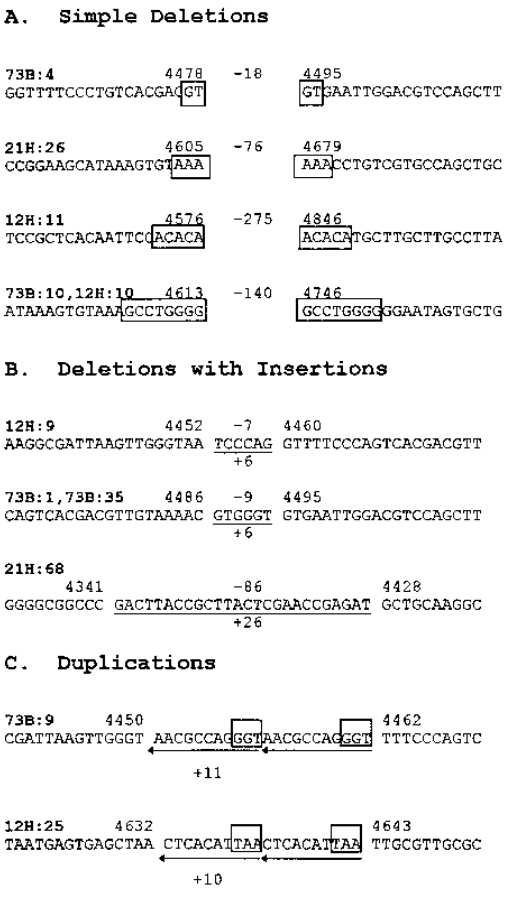

FIG. 5. Sequence analysis of deletions, deletions with insertions, and duplications. Clone numbers are shown above the sequence in boldface type. (A) Simple deletions. Short direct repeats at the deletion junctions are shown in boxes. The number of nucleotides deleted is indicated between the deletion junctions preceded by a minus sign. (B) Deletions with insertions. The number of nucleotides deleted is indicated between deletion junctions preceded by a minus sign. Inserted nucleotides are underlined, with the length of insertion shown below their sequence. (C) Duplications. Nucleotides within the boxes again represent direct repeats. The arrows indicate the duplicated nucleotides. The exact number of duplicated sequences is shown below the bold arrow.

mutation rates are equal. Little is known about whether DNA repair of viral DNA occurs. On the basis of identification of two genotypically different proviruses at the same chromosomal locations, it was concluded that correction of mismatches in murine leukemia virus is not 100% (3). As shown in Fig. 1C, two of the eight proviral genotypes generated during replication will have mutations in both *lacZ*α genes (Fig. 1C, a1 and a2) whereas three of the eight proviral genotypes will have a mutation in the 3′ *lacZ*α gene (Fig. 1C, a1, a2, and d1). Therefore, if the mutation rates during the early and late stages are equal, we would expect the *Hin*dIII recovery mutant frequency to be 66.6% (2 of 3) of the *Bam*HI recovery mutant frequency. The observed result of 68% is in close agreement with this expectation.

If it is assumed that DNA repair occurs efficiently in a strand-specific manner, the results can be interpreted in two ways, depending on the strand that is presumed to be specifically repaired. First, if it is assumed that the minus strand of the 3′ LTR is specifically corrected to the plus strand, the early-stage mutation rate would be twofold higher than the late-stage mutation rate. Pulsinelli and Temin hypothesized that single-strand gaps formed at the sites of proviral integration may be utilized for strand-specific DNA repair, with the gapped DNA strand being the strand that is repaired (33). This strand-specific DNA repair system is limited to about 1 kb from the gap in the DNA (15, 39). If this system affects the viral DNA, it would be expected to correct the minus-strand DNA of the heteroduplex to the mutant strand in the 3′ LTR (Fig. 1C, d1 and d2). Therefore, assuming that this strand-specific repair is operating, four viral genotypes would have mutations in the 3′ LTR (Fig. 1C, a1, a2, d1, and d2), and two genotypes would have mutations in both LTRs (Fig. 1C, a1 and a2). If the mutation rates during the early and late stages were equal, we would expect the *Hin*dIII recovery mutation rate to be 50% (2 of 4) of the *Bam*HI recovery mutation rate. If this strand-specific repair occurs efficiently, the observed results would suggest that the early-stage mutation rate is approximately twofold higher than the late-stage mutation rate (68% of the total rate for early stage divided by 32% of the total rate for the late stage). Second, it is possible that the plus strand of the 3′ LTR is specifically corrected to the minus strand. If this were the case, the mutation rate during the late stage would be more than 32% of the total mutation rate. It can be concluded that 32% of the mutations occurred during the late stage and "survived" (not corrected to the wild type) the DNA repair system.

The same assumptions for DNA repair can be applied to estimate the mutation rate during plus-strand synthesis. If no DNA repair occurs or if DNA repair occurs without strand specificity, the mutation rate during plus-strand synthesis is equal to the late-stage mutation rate (32% of the total mutation rate). If strand-specific DNA repair of the minus strand occurs, the plus-strand synthesis mutation rate would be half of the late-stage mutation rate (16% of the total mutation rate). If strand-specific DNA repair of the plus strand occurs, the plus-strand mutation rate is any number greater than 32% of the total mutation rate.

The early stage includes mutations occurring during RNA transcription by RNA polymerase II and minus-strand synthesis catalyzed by RT. These experiments do not determine the relative contributions of RNA transcription and minus-strand synthesis to the retroviral mutation rate. RT is believed to be less processive than RNA polymerase II, which suggests that the minus-strand synthesis mutation rate is at least equal to the RNA polymerase II mutation rate (27). If so, the minus-strand synthesis mutation rate would be 34 to 68% of the total mutation rate.

Mutational spectra in different stages of retroviral replication. Analysis of the types of mutations and their frequencies suggests that substitutions, frameshifts, and deletions occur in both the early and late stages of viral replication. The rate of substitutions in *Hin*dIII recovery pools is approximately 68% of the rate in *Bam*HI recovery pools, suggesting that these mutations occur at essentially equivalent rates during minus-strand and plus-strand synthesis.

Few template-switching mutations involving simple deletions, deletions with insertions, and duplications were identified in this study. The rate of template-switching mutations in this study was approximately threefold lower than in our previous studies with the same *lacZ*α reporter gene. Because we identified very few template-switching mutations, their relative rates during early and late stages of replication could not be accurately determined. Previously, a deletion with insertion mutation that deleted the viral polypurine tract was characterized, indicating that the mutation occurred during plus-strand synthesis (29). The lower rate of template-switching mutants may reflect the different location of the *lacZ*α genes. In the BK-2 vector, *lacZ*α is located in the LTR, whereas in the

previous VP212 vector, it is located between the LTRs. In VP212, the *lacZ*α coding region was adjacent to a 1-kb region of the F1 origin of replication, which was not essential for viral replication or proviral DNA isolation. Consequently, deletion mutations extending into the F1 region could be isolated. In the BK-2 vector, the *lacZ*α coding region is flanked by the polypurine tract and the *lac* operator sequence. Thus, deletions that extend beyond the coding region are not expected to complete reverse transcription and survive the selection process. The different location of the *lacZ*α reporter gene could explain the lower rate of template-switching mutations in the present study.

A-to-G hypermutation. Two *Bam*HI-recovered clones had undergone A-to-G hypermutation of their *lacZ*α coding regions. G-to-A hypermutation was discovered in an SNV retroviral vector, and DNA synthesis by an error-prone RT was hypothesized to result in hypermutation (28). G-to-A hypermutation has since been found in HIV-1 genomes, indicating that hypermutation occurs often and in different retroviruses (22, 25, 41). A-to-G hypermutation has recently been described in an oncogene-transducing Rous-associated virus type 1 derivative, an avian retrovirus (10); in this virus, 48% of the A residues in the U3 region of the LTR were mutated to G residues. Similarly, A-to-G hypermutation also occurs in a region of an avian retrovirus likely to form a hairpin structure (13). In the latter study, a double-stranded RNA adenosine deaminase (dsRAD) was postulated to cause retroviral hypermutation. The dsRAD utilizes complementary sequences within the RNA strand and deaminates adenosine into inosine, which in turn is converted to guanosine during reverse transcription (1, 31).

It is not clear whether the two A-to-G hypermutants identified in this study were generated by the dsRAD activity, by an error-prone RT, or by another mechanism. Two observations strongly suggest that the mechanism of A-to-G hypermutation is different from the mechanism of G-to-A hypermutation. First, G-to-A hypermutation in SNV occurred primarily at GpA (13 of 15 G residues), suggesting a reverse transcriptase-mediated dislocation mutagenesis mechanism of mutation (2, 21, 28). In contrast, only 10 of 29 A-to-G hypermutations have a G residue as a 3′ neighbor. Second, G-to-A hypermutation resulted in mutation of only 6 to 7% of the G nucleotides in the hypermutated region (15 of 225 G nucleotides). In contrast, A-to-G hypermutation resulted in a much higher proportion of mutations, i.e., 40 to 50% of the A nucleotides.

Two observations support the hypothesis that the A-to-G hypermutation occurred from the activity of mammalian dsRAD enzyme. First, the proportion of mutated A nucleotides, 40 to 50%, is consistent with the previous reports of dsRAD hypermutation activity (1, 13). Second, dsRAD enzyme has a 5′ neighbor preference of A = U > C > G (31). Analysis of the 5′ neighbors of the hypermutated A nucleotides indicates identical preferences. The proportions of each of the four dinucleotides that were hypermutated are as follows: ApA, 87% (13 of 15 in the hypermutated region); TpA, 86% (6 of 7); CpA, 55% (6 of 11); and GpA, 33% (3 of 9). These 5′ neighbor frequencies are in close agreement with those predicted by dsRAD preferences.

A secondary-structure analysis of the hypermutated region (DNAsis software) indicated that 16 of 29 hypermutated A residues were predicted to be in double-stranded regions while the remaining 13 residues were predicted to be single stranded (data not shown). These analyses argue against intrastrand secondary structure playing a role in this dsRAD-mediated hypermutation. On the other hand, an antisense transcript initiating in cellular sequences could have annealed to the proviral RNA transcript, forming interstrand secondary structure that led to dsRAD-mediated hypermutation. If this hypothesis is correct, mammalian dsRAD may regularly play a role in retroviral hypermutagenesis.

In summary, we have developed an in vivo approach to determine retroviral mutation rates at different stages of viral replication. The results strongly suggest that the retroviral mutation rates during minus-strand and plus-strand synthesis are either equal to or within twofold of each other. The results also provide upper and lower limits for retroviral mutation rates during minus-strand and plus-strand synthesis. The in vivo approach to selectively identify retroviral mutations occurring in the early stage should be useful in determining the effect of antiretroviral drugs, nucleotide pools, and other environmental conditions that affect retroviral mutation rates.

ACKNOWLEDGMENTS

We thank David Hash and Balween Kaur for technical assistance and Wei-Shau Hu, Jeffery Anderson, Gregory Arnold, Ella Bowman, Robert Bowman, Jeanine Certo, Krista Delviks, Elias Halvas, John Julias, and Philip Yin for helpful comments on the manuscript and for valuable discussions of results. We especially thank Keith Garbutt for statistical analysis of the results, John Coffin for helpful discussion of the results, and Wei-Shau Hu for valuable intellectual input and discussions throughout the project.

This work was supported by Public Health Service grant CA58875 from National Institutes of Health.

REFERENCES

1. **Bass, B. L., and H. Weintraub.** 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell **55:**1089–1098.
2. **Bebenek, K., J. Abbots, J. D. Roberts, S. H. Wilson, and T. A. Kunkel.** 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. J. Biol. Chem. **264:**16948–16956.
3. **Berwin, B., and E. Barklis.** 1993. Retrovirus-mediated insertion of expressed and non-expressed genes at identical chromosomal locations. Nucleic Acids Res. **21:**2399–2407.
4. **Boyer, J. C., K. Bebenek, and T. A. Kunkel.** 1992. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. Proc. Natl. Acad. Sci. USA **89:**6919–6923.
5. **Coffin, J. M.** 1990. Genetic variation in retroviruses, p. 11–33. *In* E. Kurstak, R. G. Marusyk, F. A. Murphy, and M. H. V. Van Regenmortel (ed.), Applied virology research, virus variability, epidemiology and control. Plenum Press, New York.
6. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. *In* B. Fields (ed.), Virology, 2nd ed. Raven Press, New York.
7. **Coffin, J. M.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science **267:**483–489.
8. **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.
9. **Essex, M.** 1995. Strategies of research for a vaccine against AIDS. Pubbl. Stn. Zool. Napoli Sect. II **17:**141–149.
10. **Felder, M.-P., D. Laughier, B. Yatsula, P. Dezelee, G. Calothy, and M. Marx.** 1994. Functional and biological properties of an avian variant long terminal repeat containing multiple A-to-G conversions in the U3 sequence. J. Virol. **68:**4759–4767.
11. **Glickman, B. W., V. A. Saddi, and J. Curry.** 1994. Spontaneous mutations in mammalian cells. Mutat. Res. **304:**19–32.
12. **Gritz, L., and J. Davies.** 1979. Plasmid encoded hygromycin-b resistance: the sequence of hygromycin b phosphotransferase and its expression. Gene **25:**179–188.
13. **Hajjar, A. M., and M. L. Linial.** 1995. Modification of retroviral RNA by double-stranded RNA adenosine deaminase. J. Virol. **69:**5878–5882.
14. **Ho, D. D., A. U. Neumann, A. S. Perelson, W. Chen, J. M. Leonard, and M. Markowitz.** 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature (London) **373:**123–126.
15. **Holmes, J., Jr., S. Clark, and P. Modrich.** 1990. Strand-specific mismatch correction in nuclear extracts of human and drosophila melanogaster cell lines. Proc. Natl. Acad. Sci. USA **87:**5837–5841.
16. **Hu, W.-S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and a high rate of genetic recombination. Proc. Natl. Acad. Sci. USA **87:**1556–1560.
17. **Hubner, A., M. Kruhoffer, F. Grosse, and G. Krauss.** 1992. Fidelity of human immunodeficiency virus type I reverse transcriptase in copying natural RNA. J. Mol. Biol. **223:**595–600.

18. **Ji, J., and L. A. Loeb.** 1992. Fidelity of HIV-1 reverse transcriptase copying RNA *in vitro*. Biochemistry **31:**954–958.
19. **Jorgensen, R. A., S. J. Rothstein, and W. J. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. Mol. Gen. Genet. **177:**65–72.
20. **Kawai, S., and M. Nishizawa.** 1984. New procedures for DNA transfection with polycation and dimethyl sulfoxide. Mol. Cell. Biol. **4:**1172–1174.
21. **Kunkel, T. A., and P. S. Alexander.** 1986. The base substitution fidelity of eucaryotic DNA polymerases-mispairin frequencies, site preferences, insertion preferences, and base substitution by dislocation. J. Biol. Chem. **261:** 160–166.
22. **Li, Y., J. C. Kappes, J. A. Conway, R. W. Price, G. M. Shaw, and B. H. Hahn.** Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication-competent and -defective viral genomes. J. Virol. **65:**3973–3985.
23. **Mansky, L. M., and H. M. Temin.** 1994. Lower mutation rate of bovine leukemia virus relative to that of spleen necrosis virus. J. Virol. **68:**494–499.
24. **Mansky, L. M., and H. M. Temin.** 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. **69:**5087–5094.
25. **Martinez, M. A., M. Sala, J. P. Vartanian, and S. Wain-Hobson.** 1995. Reverse transcriptase and substrate dependence of the RNA hypermutagenesis reaction. Nucleic Acids Res. **23:**2573–2578.
26. **Nowak, M. A., R. M. Anderson, A. R. McLean, T. F. W. Wolfes, J. Goudsmit, and R. M. May.** 1991. Antigenic diversity thresholds and the development of AIDS. Science **254:**963–969.
27. **Parthasarathi, S., A. Varela-Echavarria, Y. Ron, B. D. Preston, and J. P. Dougherty.** 1995. Genetic rearrangements occurring during a single cycle of murine leukemia virus vector replication: characterization and implications. J. Virol. **69:**7991–8000.
28. **Pathak, V. K., and H. M. Temin.** 1990. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. Proc. Natl. Acad. Sci. USA **87:**6019–6023.
29. **Pathak, V. K., and H. M. Temin.** 1990. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: deletions and deletions with insertions. Proc. Natl. Acad. Sci. USA **87:**6024–6028.
30. **Pathak, V. K., and H. M. Temin.** 1992. 5-Azacytidine and RNA secondary structure increase the retrovirus mutation rate. J. Virol. **66:**3093–3100.
31. **Polson, A. G., and B. L. Bass.** 1994. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. EMBO J. **13:** 5701–5711.
32. **Preston, B. D., B. J. Poiesz, and L. A. Loeb.** 1988. Fidelity of HIV-1 reverse transcriptase. Science **242:**1168–1171.
33. **Pulsinelli, G. A., and H. M. Temin.** 1994. High rate of mismatch extension during reverse transcription in a single round of retrovirus replication. Proc. Natl. Acad. Sci. USA **91:**9490–9494.
34. **Roberts, J. D., K. Bebenek, and T. A. Kunkel.** 1988. The accuracy of reverse transcriptase from HIV-1. Science **242:**1171–1173.
35. **Roberts, J. D., B. D. Preston, L. A. Johnson, A. Soni, L. A. Loeb, and T. A. Kunkel.** 1989. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. Mol. Cell. Biol. **9:**469–476.
36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37. **Tantillo, C., J. Ding, A. Jacobo-Molina, R. G. Nanni, P. L. Boyer, S. H. Hughes, R. Pauwels, K. Andries, P. A. J. Janssen, and E. Arnold.** 1994. Location of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. J. Mol. Biol. **243:**369–387.
38. **Temin, H. M.** 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc. Natl. Acad. Sci. USA **90:**6900–6903.
39. **Thomas, D. C., J. D. Roberts, and T. A. Kunkel.** 1991. Heteroduplex repair in extracts of human HeLa cells. J. Biol. Chem. **266:**3744–3751.
40. **Varela-Echavarria, A., N. Garvey, B. D. Preston, and J. P. Dougherty.** 1992. Comparison of Moloney murine leukemia virus mutation rate with the fidelity of its reverse transcriptase in vitro. J. Biol. Chem. **267:**24681–24688.
41. **Vartanian, J. P., A. Meyerhans, B. Asjo, and S. Wain-Hobson.** 1991. Selection, recombination, and G-A hypermutation of human immunodeficiency virus genomes. J. Virol. **65:**1179–1788.
42. **Watanabe, S., and H. M. Temin.** 1983. Construction of a helper cell line for reticuloendotheliosis virus cloning vectors. Mol. Cell. Biol. **3:**2241–2249.
43. **Wei, X., S. K. Ghosh, M. E. Taylor, V. A. Johnson, E. A. Emini, P. Deutsch, J. D. Lifson, S. Bonhoeffer, M. A. Nowak, B. H. Hahn, M. S. Sag, and G. M. Shaw.** 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature (London) **373:**117–122.