Comparison of Moloney Murine Leukemia Virus Mutation Rate with the Fidelity of Its Reverse Transcriptase \textit{in Vitro}

(Received for publication, May 14, 1992)

Alfredo Varela-Echavarria$\S$, Nancy Garvey$\D$, Bradley D. Preston$\D$, and Joseph P. Dougherty$\D$**

From the $\D$Department of Molecular Genetics and Microbiology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, and the $\S$Laboratory for Cancer Research, College of Pharmacy, Rutgers University, Piscataway, New Jersey 08854

The role of Moloney murine leukemia virus (MoMLV) reverse transcriptase (RT) in the generation of base substitution mutations during retroviral replication was analyzed. To that effect, the \textit{in vitro} fidelity of the MoMLV RT was compared to the rate of base substitution mutations occurring during the replication of an MoMLV-based retroviral vector. Using the vector in an amber reversion assay, the base substitution mutation rate at a single locus was found to be $2 \times 10^{-4}$/base pair in one cycle of vector virus replication. Analysis of the fidelity of the purified RT using the same template sequence revealed that, of the two mismatches (A-C and T-G) that would lead to reversion of the amber codon during replication, A-C occurs at a rate of $4.0 \times 10^{-8}$, and T-G occurs at a rate of $0.7 \times 10^{-4}$. While the rate of formation of A-C is very similar to the vector mutation rate, the rate of formation of T-G is more than 30 times higher. This discrepancy in rates suggests that there are other elements in the infected cells that contribute to the fidelity of viral replication.

Retroviruses are RNA viruses which replicate through a DNA intermediate (1, 2). Upon infection, the viral RNA serves as template for minus strand DNA synthesis by the virus-encoded reverse transcriptase (RT) followed by plus strand DNA synthesis by RT using the minus strand DNA as template. The viral DNA is then integrated into the genome of the host cell where it is stably maintained as a provirus. Transcription of the provirus by RNA polymerase II (pol II) and packaging of the newly synthesized RNA into progeny virions completes the retroviral replication cycle.

Retroviral genomes evolve at relatively rapid rates (3, 4). Mutations can be introduced during transcription by RNA pol II, minus or plus strand DNA synthesis by RT, or provirus replication by cellular DNA polymerases. Due to the high fidelity of cellular DNA replication ($10^{-9}$ to $10^{-11}$ substitutions/base pair) (5), it is unlikely that mutations occurring during cellular replication of the provirus contribute significantly to the high retroviral mutation rates (6). Therefore, most mutations are probably introduced during provirus transcription or reverse transcription.

Retroviral mutation rates or the mutation frequency in a single cycle of retroviral replication have been quantitated for some retroviruses (7-9). These determinations, which measure the combined contributions of RT and RNA pol II to the mutation rate, have been done for two retroviruses of avian origin, spleen necrosis virus (SNV) and Rous sarcoma virus. For SNV, retroviral vectors and helper cells were employed and the base substitution mutation rate at a defined locus was found to be $2 \times 10^{-4}$/base pair/replication cycle (7). Using a different SNV-based vector, the base substitution mutation rate over a longer sequence was found to be $0.7 \times 10^{-4}$/base pair/replication cycle (9). Using a different experimental system, the rate of base pair substitution mutations of Rous sarcoma virus was calculated as $1.4 \times 10^{-5}$ mutations/base pair/replication cycle (8). Thus, the average base substitution mutation rate of these avian retroviruses is $6 \times 10^{-6}$ or about 0.5 base substitutions/retroviral genome/replication cycle.

The fidelity of different purified RTs has been determined using cell-free systems. Nucleotide misincorporation rates were found to be sequence- and polymerase-dependent. Rates, in general, varied from $2.5 \times 10^{-5}$ to $6 \times 10^{-4}$ (10-17). Based on these studies, it has been proposed that the high mutation rate during retroviral replication is, at least in part, a consequence of errors made by RT (10). However, RNA pol II could also contribute to retroviral mutation rates.

To date, experiments have not been performed to directly compare the fidelity of purified RT to retroviral mutation rates at identical genetic loci. This is particularly important in view of the established effects of template sequence on RT fidelity. To gain insight into the relative contribution of RT to retroviral mutation rates, we compared the base substitution mutation rate at a defined locus during replication of a Moloney murine leukemia virus (MoMLV)-based vector in cultured cells with the fidelity of the purified MoMLV RT measured \textit{in vitro} at the same locus. We found that the fidelity of plus strand synthesis by MoMLV RT in \textit{vitro} was about 30 times lower than the overall mutation rate of the virus. These results suggest that other factors increase the fidelity of viral replication in the cell.

**MATERIALS AND METHODS**

\textbf{Enzymes}

The MoMLV RT was obtained from Pharmacia LKB Biotecnology Inc. and is a recombinant product (18) purified from \textit{Escherichia coli}.
coli to a specific activity of 47,000 units/mg. T4 polynucleotide kinase was obtained from New England BioLabs (Beverly, MA).

**DNA Constructions**

**Plasmids**—The vector pAV was constructed in two sequential steps as follows. The SNV promoter contained in the EcoRI-Aval fragment of pJD214Hy (19) was cloned by blunt-end ligation into the SalI site of pJD220Hy (19) resulting in the plasmid pJD220SNVHy. From this vector, the fragment containing the SNV promoter linked to the hygromycin phosphotransferase gene (hygro) was excised with XbaI and ClaI and cloned into the Xhol site of pN2 (20) by blunt-end ligation yielding pAV. The vector pAVneoAm differs from pAV by a single base pair which creates an amber codon at the 15th amino acid of the neomycin phosphotransferase gene (neo). pAVneoAm was made by substituting the BclI-BstBI fragment of pAV with the corresponding fragment from pJD216NeoAmHy (7) which contains the 5' coding region of neo with the amber codon. Vector viruses derived from the plasmid pAV and pAVneoAm are referred to as AV and AVneoAm, respectively.

**M13 Constructions**—The 1.3-kilobase pair HindIII fragment of pJD215Am (map available upon request), containing the same neo gene with the amber codon (neoAm) that was used in the construction of pAVneoAm, was cloned into the HindIII site of M13 mp18 in both orientations to create M13neoAm+ and M13neoAm-. In M13neoAm-, the neo coding sequence is inserted in the same orientation as that of the phage transcription; in M13neoAm+ it is in the opposite orientation. These constructs yield single-stranded phage DNA (21) containing neoAm template sequences corresponding to the plus and minus strands, respectively, of the MoMLV-based pAVneoAm vector.

**Cells**

NIH/3T3 is a murine fibroblast cell line permissive for infection by MoMLV and MoMLV-based vectors. GP+E-86 is an NIH/3T3-derived helper cell line which provides the MoMLV trans-acting functions required for propagation of replication-defective MoMLV vectors without producing replication competent virus (22). The MoMLV provirus clone used to establish the helper cell line was made by substituting the BclI-BstBI fragment of pAV with the corresponding fragment from pJD216NeoAmHy (7) which contains the 5' coding region of neo with the amber codon. Vector viruses derived from the plasmid pAV and pAVneoAm are referred to as AV and AVneoAm, respectively.

**Transfections and Infections**

Vector plasmid DNA was transfected by the polybrene-dimethyl sulfoxide method (23) into GP+E-86 helper cells. Vector virus harvested from these cells was in turn used to infect fresh GP+E-86 cells (see below). Infected helper cell clones harboring a single provirus were isolated and used as source of vector virus for the determination of mutation rates. Vector virus titers of these clones were determined by infecting 2 x 10^5 NIH/3T3 cells in 60-mm dishes with 0.2 ml of 10-fold serial dilutions of virus stock and allowing the cells to grow to confluence. The single-stranded DNA was sequenced with a wild-type hygro gene (expressed from an internal SNV promoter) by selection with hygromycin B (see below). Infected helper cell clones harboring a single provirus were isolated and used as source of vector virus for the determination of mutation rates. Vector virus titers of these clones were determined by infecting 2 x 10^5 NIH/3T3 cells in 60-mm dishes with 0.2 ml of 10-fold serial dilutions of virus stock and allowing the cells to grow to confluence. The single-stranded DNA was sequenced with a wild-type hygro gene (expressed from an internal SNV promoter) by selection with hygromycin B (see below). Infected helper cell clones harboring a single provirus were isolated and used as source of vector virus for the determination of mutation rates. Vector virus titers of these clones were determined by infecting 2 x 10^5 NIH/3T3 cells in 60-mm dishes with 0.2 ml of 10-fold serial dilutions of virus stock and allowing the cells to grow to confluence.

**DNA Amplification and Sequencing**

**Asymmetric Enzymatic Amplification Reaction**—The primer extension assay used for kinetic analysis is basically that described by Boosalis et al. (28). The reactions contained MoMLV RT at 0.002-0.15 unit/µl (units as defined by the manufacturer). Reaction components, in addition to enzyme, were 20 mM Tris-HCl (pH 8.0), 2 mM dithiothreitol, 10 mM MgCl2, 0.1 mg/ml bovine serum albumin, 2.2 mM of hybridized DNA template, and dNTPs as indicated in a total volume of 10 µl. Reactions were initiated by the addition of dNTP, incubated at 50 °C for varying times, then terminated by the addition of NaOEtDA, and transferred to an ice-water bath. Each reaction was eluted through a 0.5-ml Sephadex G-100 column to remove non-hybridized primer, electrophoresed in 20% polyacrylamide-urea gels, visualized by autoradiography, and quantitated by densitometric scanning (Hoeffer GS-300 Scanning Densitometer, Ref. 11). Reaction times and enzyme levels were experimentally predetermined to ensure steady-state conditions during the measurement of both mispair and correct pair formation (<20% of total primer extension).

**Mispair Formation and Extension Assays**

**Assay for MoMLV Mutation Rates during a Single Cycle of Replication**—To determine the base substitution mutation rate of MoMLV, a MoMLV-based vector system was developed that scored reversion mutations during a single cycle of viral replication (Fig. 1). This system is an extension of an assay previously used to measure SNV mutation rate (7). A replication-defective MoMLV-based vector pAVneoAm (Fig. 1A) was constructed carrying two marker genes: a wild-type hygro gene (expressed from an internal SNV promoter)
arise during vector virus replication. Resistance to hygromycin B (hygro') reflects the total virus titer, while resistance to were then selected for drug resistance (Fig. 1B). The presence of two marker genes in the AVneoAm vector permits the mutant titer divided by the total titer represents the mutation frequency. The spread of the vector virus among the cells which provide the viral proteins without producing replication-competent viruses (see Transfections and Infections under "Materials and Methods"). To measure the mutation frequency, vector virus from helper cell clones harboring a single vector provirus was used to infect NIH/3T3 cells that provide the viral proteins. Vector virus was produced using GP+E-86 helper cell clones (Table I, clones A-H) were used to generate vector virus and to score for reversion frequencies following infection of target cells. All clones yielded total virus titers ranging from $9 \times 10^2$ to $13 \times 10^4$ colony forming units (hygro' colonies)/0.2 ml of virus stock thus giving a high sensitivity for the detection of reversion mutants. Of the 30.4

**FIG. 1. Vector virus construct and protocol for retrovirus mutation rates.** A, diagram of the two vectors used in these studies. The sequence shown above the provirus represents codons 14, 15, and 16 of the neo-coding sequence in vector AV. The sequence below represents the sequence of the vector AVneoAm at the same site showing the base substitution which converted the tryptophan codon TGG into the amber codon TAG, thereby generating a DdeI cleavage site (CTNAG). LTR, MoMLV long terminal repeat; neo, neomycin phosphotransferase gene; hygro, hygromycin phosphotransferase gene; *, the site of the base change in the neo gene; horizontal lines, MoMLV sequences containing the cis-acting elements required for vector replication. B, schematic representation of the protocol for mutation rate determination which involves a single cycle of viral replication. Inside the cells a vector provirus is depicted. The two open squares represent long terminal repeats; the two rectangles represent exogenous neo and hygro marker genes; the jagged lines represent genomic DNA; GP+E-86 is an MoMLV-derived helper cell line; NIH/3T3 is a cell line permissive for infection by MoMLV; G418' and hygro' denote G418-resistant and hygromycin B-resistant colonies, respectively.

and a mutant neo gene (expressed from the 5'-MoMLV LTR promoter) containing an amber codon in the 5'-coding sequence. The wild-type marker genes confer resistance to hygromycin B and G418, respectively (29, 30). The vector contains the cis-acting MoMLV sequences required for replication but not the trans-acting sequences encoding the viral proteins. Vector virus was produced using GP+ E-86 helper cells which provide the viral proteins without producing replication-competent viruses (see Transfections and Infections under "Materials and Methods"). To measure the mutation frequency, vector virus from helper cell clones harboring a single vector provirus was used to infect NIH/3T3 cells that were then selected for drug resistance (Fig. 1B). The presence of two marker genes in the AVneoAm vector permits the highly sensitive detection of neoAm reversion mutations that arise during vector virus replication. Resistance to hygromycin B (hygro') reflects the total virus titer, while resistance to G418 (G418') reflects the titer of neoAm reversion mutants.

The mutant titer divided by the total titer represents the mutation frequency. The spread of the vector virus among the helper cells is effectively blocked due to superinfection immunity (31), and in the target NIH/3T3 cells due to the absence of viral proteins required for replication. Thus, this system limits retroviral replication to a single cycle which starts with a provirus in a GP+E-86 helper cell and ends with a provirus in the target NIH/3T3 cell. Since the measurement of the mutation frequency is confined to a single cycle, it constitutes a measure of the mutation rate. It should be noted, however, that sequence analysis of the mutants is required before the final computation of the mutation rate (see below).

The polymerization steps during replication of this vector system are identical to those in MoMLV replication involving one step of genomic RNA synthesis by transcription (catalyzed by cellular RNA pol II in the helper cells), and two steps of DNA polymerization (minus and plus strand synthesis catalyzed by MoMLV RT in the target cells).

**Mutation Rate during Replication of MoMLV Vector Virus—**Measurements of mutant frequency following a single cycle of AVneoAm vector virus replication revealed that neoAm reversion mutations occurred at a relatively low rate (Tables I and II). In one experiment, eight independent GP+E-86 helper cell clones (Table I, clones A-H) were used to generate vector virus and to score for reversion frequencies following infection of target cells. All clones yielded total virus titers ranging from $9 \times 10^2$ to $13 \times 10^4$ colony forming units (hygro' colonies)/0.2 ml of virus stock thus giving a high sensitivity for the detection of reversion mutants. Of the 30.4

| Titers (cfu/0.2 ml) | Titer$^a$ |
|---------------------|----------|
| Hygro' | 4.0 $\times 10^4$ |

$^a$To obtain virus titers, vector virus was harvested from the helper cell clones (GP+E-86) harboring an AVneoAm provirus and used to infect NIH/3T3 cells. The infected cells were in turn selected for G418-resistance (G418') or hygromycin B-resistance (hygro').

| Helper cell clones | G418' | Hygro' |
|-------------------|-------|-------|
| A                 | 5.0 $\times 10^2$ | 1.4 $\times 10^4$ |
| B                 | 1.3 $\times 10^3$ | 6.6 $\times 10^3$ |
| C                 | 2.1 $\times 10^5$ | 6.8 $\times 10^4$ |
| D                 | 2.5 $\times 10^5$ | 4.0 $\times 10^5$ |
| E                 | 2.4 $\times 10^5$ | 2.6 $\times 10^5$ |
| F                 | 0.9 $\times 10^5$ | 4.0 $\times 10^5$ |
| G                 | 1.0 $\times 10^5$ | 4.0 $\times 10^5$ |
| H                 | 3.5 $\times 10^5$ | 4.0 $\times 10^5$ |

Total 12, 30.4 $\times 10^4$

Mutation rate = G418'/Hygro' = 4 $\times 10^{-6}$

**TABLE I**

**Summary of the mutation rates**

The abbreviation used is: cfu, colony forming unit.

| Experiment | No. of helper cell clones $^*$ | Titers (cfu/0.2 ml) | Mutation rate (G418'/Hygro') |
|------------|--------------------------------|---------------------|-----------------------------|
| 1          | 2                              | 15 $\times 10^3$    | 1.4 $\times 10^4$           |
| 2          | 3                              | 6.1 $\times 10^4$   | 6.6 $\times 10^3$           |
| 3          | 5                              | 43 $\times 10^3$    | 6.8 $\times 10^4$           |
| 4          | 8                              | 30 $\times 10^4$    | 4.0 $\times 10^5$           |
| 5          | 10                             | 34 $\times 10^5$    | 2.6 $\times 10^5$           |

Average 4.0 $\times 10^5$

$^*$A total of 16 helper cell clones (GP+E-86) harboring an AVneoAm provirus were used as source of virus. Five clones were used in three different experiments, two clones in two experiments and the other nine in only one experiment.

$^a$Titers represent the sum of the individual titers obtained with the helper cell clones involved in each experiment.
×10⁶ total virus screened in this experiment, 12 G418' NIH-3T3-infected clones were detected, corresponding to a mutation rate of 4×10⁻⁶/cycle.

To obtain significant numbers of G418' reversion mutants and to examine the possible effects of different helper cell clones on mutation rate, four additional experiments were conducted (Table II). All experiments yielded similar mutation rates (1.4×10⁻⁶ to 6.8×10⁻⁶/cycle) with an average mutation rate from 16 independent proviral helper cell clones of 4×10⁻⁶/cycle.

For each experiment a control was performed using a mass population of helper cells established by infection with the wild-type neo/hygro provirus AV which yielded very similar titers for both marker genes (the ratio of the G418' titers to the hygro' titers ranged from 1.3 to 5.7 with an average value of 3.2; data not shown). Thus, the conditions for selection were such that allowed the wild-type neo and hygro genes to be scored with similar efficiencies. Therefore, the ratio of G418'/hygro' titers obtained with AVneoAm virus should accurately reflect the mutation rate.

**Southern Blot Analysis of Proviral DNA from G418' Colonies**—During the construction of the vector pAVneoAm, a Ddel site (CTTAG) was generated in the neo gene by the substitution of the amber codon TAG for the tryptophan codon TGG (Fig. 1A). Any base change in the second or third position of the amber codon should result in the loss of the Ddel site. To determine whether the Ddel site was lost in the G418' mutant clones obtained during AVneoAm vector replication, their genomic DNA was digested with Ddel and analyzed by Southern blotting. Of 45 clones analyzed, 31 had multiple proviral copies integrated in the genome since the effective multiplicity of infection was usually greater than one. Of the 14 cell clones with one provirus/cell, seven lost the Ddel site and seven retained it. Two cases in which the Ddel site was lost and two in which it was retained are shown in Fig. 2.

**Sequence Analysis of G418' Mutant Proviral DNA**—To determine the exact nature of the mutations in the G418' clones, the region of the provirus containing the amber codon was sequenced using asymmetric polymerase chain reaction and the dyeoxy-chain termination method (26). Analysis of the 14 singly infected G418' mutant clones revealed that in those in which the Ddel site was lost had an A·T to G·C base pair substitution in the second position of the amber codon (data not shown). This substitution restored the wild-type neo sequence coding for tryptophan (TGG). This A·T to G·C transition was the only mutation seen at the amber codon in the Ddel site. If the provirus has an intact amber site (represented by the asterisk), it will generate a 760-bp fragment. Any base substitution at the amber codon which disrupts the Ddel site will generate a 940-bp fragment. The smaller Ddel fragments with homology to the probe were too small to detect under the conditions employed.

from two sequential polymerization events at the target site. The first is incorporation of a noncomplementary nucleotide to generate a mispair at the 3' terminus of the nascent strand. The second step is extension of the terminal mispair, thereby permitting complete genome synthesis. Thus, mutagenesis depends upon the frequency of mispair formation as well as the efficiency of mispair extension. In order to explore the contribution of RT in the generation of the vector amber reversion mutations, we have measured the relative efficiencies of both mispair formation and extension at the amber reversion site by purified MoMLV RT using the kinetic assay depicted in Fig. 3 (28). Briefly, a radiolabeled oligonucleotide (15 nucleotides in length, 16-mer) is hybridized to a single-stranded neoAm template so that the 3' end of the primer is immediately adjacent to the second position of the amber codon. Purified MoMLV RT is incubated in the presence of excess template-primer and increasing levels of a single dNTP species, either complementary or noncomplementary to the template base. Reaction products are separated by polyacrylamide gel electrophoresis, and the relative amounts of product formed at each dNTP concentration are quantitated to obtain apparent Vₘₐₓ/Kₘ values for incorporation of both correct and incorrect dNTPs at the amber site. The efficiency of base pair formation is described by the term Vₘₐₓ/Kₘ (28, 32). Thus, the relative frequency for the formation of each mispair is given by Vₘₐₓ/Kₘ for that mispair divided by Vₘₐₓ/Kₘ for formation of the correct pair (28). Efficiencies of mispair extension are measured basically in the same manner, using template-primers with preformed 3'-terminal mispairs at the amber site.

**Mispair Formation by MoMLV RT**—All of the reversion mutations observed at the amber codon during vector replication in culture resulted from A·T to G·C transitions at the central base pair of the TAG codon. These mutations could
MoMLV Mutation Rate and RT Fidelity

Fig. 3. Primer extension assay for reverse transcriptase fidelity.
The M13neoAm templates are shown. Panel A depicts mispair formation and extension using M13neoAm+ as template. This is analogous to the generation of mutations during minus strand DNA synthesis of vector virus replication. Panel B depicts the same process on the M13neoAm- template, which corresponds to plus strand DNA synthesis during vector virus replication. The amber reversion site is the second base pair of the amber codon.

As a model for plus strand DNA synthesis, we examined the reaction products made by MoMLV RT on M13neoAm- templates which contain the minus strand neo sequence found in the vector AVneoAm (Figs. 3B, C, and D). Incorporation was measured opposite the T in the 5'-CTA-3' sequence which is complementary to the amber codon. The correct substrate, dATP, readily supported polymerization (Fig. 4D) at μM concentrations, while mispaired product was only detectable at mM concentrations of dGTP (Fig. 4D). The product containing the T, G mispair was 19 nucleotides in length due to correct incorporation of dGTP opposite the two template C residues downstream from the target site. To confirm that the observed 19-mer resulted from dGMP incorporation at the template T, a correctly paired 19-mer was made by extension of 16-mer in the presence of 50 μM each of dATP and dGTP. Under these conditions, the product 19-mer will contain the correct base, A, at the target site, since mispairs are not formed detectably at micromolar concentrations of dGTP. It is clear that this 19-mer marker migrates differently than the 19-mer formed in the presence of millimolar concentrations of dGTP (Fig. 4D, marker lane indicated with an *). From this we infer that the 19-mer formed in the kinetic fidelity assay contained an incorrectly incorporated G residue opposite the T of the 5'-CTA-3' sequence, and extension was not due to correct incorporation of contaminant dATP in the dGTP stocks. These data also provide evidence...
that the RT is able to extend the T-G mispair because only the 19-mer product was detectable on the gel.

Apparent \( V_{\text{max}}/K_m \) values for formation of each nucleotide pair were determined, and the frequencies of mispair formation \( (f_{\text{err}}) \) were calculated as described (Table III, Ref. 28). The \( f_{\text{err}} \) for A-C is \( 3.5 \pm 0.6 \times 10^{-6} \) and the \( f_{\text{err}} \) for T-G is \( 6.7 \pm 3.2 \) \( \times 10^{-6} \). Thus, the error rate of T-G mispair formation in vitro is over 30 times higher than the mutation rate at this site during AVneoAm vector virus replication.

**Extension of Preformed Mispaired Termini by MoMLV RT**—Since extension of mispairs might influence mutagenesis during viral replication, we quantitated the propensity of purified MoMLV RT to extend preformed A-C and T-G mispairs. For these experiments, primers 17 bases in length were constructed. The primers contained C or G terminal bases so that upon hybridization, A-C or T-G mispaired termini were formed. For assays of correct pair extension, A-T or T-A, 17-mers terminating in bases complementary to the target site were used (Fig. 5). Inspection of Fig. 5, A and C, shows that the 17-mers containing the correctly paired termini were extended with \( \mu M \) dNTPs to yield primarily 19mers representing the addition of two template-directed residues. Extension of both A-C and T-G mispaired termini were detected at millimolar nucleotide concentrations of the next nucleotide (Fig. 5, B and D).

Apparent \( V_{\text{max}}/K_m \) values and relative efficiencies of mispair extension \( (f_{\text{err}}) \) are listed in Table IV. There was no significant difference between the efficiency of extension of the A-C and T-G mispairs \( (f_{\text{err}} = 1.5 \times 10^{-6} \text{ for A-C}; f_{\text{err}} = 0.8 \times 10^{-5} \text{ for T-G}) \).

**DISCUSSION**

To gain insight into the role of RT in the generation of retroviral mutations, we compared the fidelity of MoMLV RT during polymerization in vitro of an amber codon in the neo gene to the mutation rate of an MoMLV-based viral vector at the same locus during replication in the cell. After a single cycle of MoMLV vector replication, G418-resistant revertants arose at a rate of \( 4 \times 10^{-6} \). Sequence analysis of the mutant clones revealed that seven of 14 had undergone an A-T to G-C transition at the second base pair of the amber codon; the other seven clones showed no sequence changes in the amber codon and thus were not true revertants. Therefore, the base substitution mutation rate at the second base pair of the neo amber codon is \( 2 \times 10^{-6} \) substitutions/base pair/virus replication cycle.

The A-T to G-C transitions observed in the revertant proviruses could arise as an A-C mispair formed during minus strand DNA synthesis by RT or as a T-G mispair formed during plus strand DNA synthesis by RT or during provirus transcription by RNA pol II. In an attempt to assess which of those steps contributed most significantly to the mutation rate, we examined the fidelity of MoMLV RT during polymerization of both strands of the amber codon in vitro. Our results showed that the A-C mispair is formed at an average rate of \( 4 \times 10^{-6} \) which is comparable to the \( 2 \times 10^{-6} \) rate observed during MoMLV vector replication. However, the T-G mispair is formed at a rate of \( 7 \times 10^{-6} \) in vitro which is over 30 times higher than the vector virus mutation rate in the cell. This difference might be slightly greater since the G418" tilters obtained with the control vector AV, which contains both wild-type neo and hygro genes, were on average three times higher than the corresponding hygro tilters, thus resulting in a small overestimation of the mutation rate during virus replication.

During replication, viral RNA serves as the template for minus strand DNA synthesis, and minus strand DNA serves as template for plus strand DNA synthesis. In the in vitro system employed, DNA templates were used as models for both minus and plus strand DNA synthesis during vector replication. Several studies show that the average fidelity of MoMLV RT is comparable on both RNA and DNA templates (16, 33, 34), including ribo- and deoxyribo-templates of identical sequence (18). Thus, DNA appears to be a reasonable model for RNA in MoMLV RT fidelity assays. Furthermore, if the rate of misincorporation for the in vitro model corresponding to minus strand synthesis was significantly higher than the mutation rate obtained during replication, it could be argued that the discrepancy is the result of the use of a DNA template instead of RNA, the normal template during replication. However, a significant difference was found only between the in vitro model corresponding to DNA-directed plus strand synthesis and the mutation rate measured during vector virus replication. Since DNA serves as template for plus strand synthesis both during virus replication and in our in vitro model, it is difficult to argue that the nature of the templates accounts for that difference.

There are at least four possible reasons for the discrepancy in rates. First, conditions in the cellular environment may contribute to fidelity by decreasing the rate of mispair formation. During reverse transcription in the cell, RT is initially found in an RNA-protein complex, and by the end of reverse transcription, the product viral DNA is still associated with subviral particles that contain RT, gag proteins, and viral integrase (35–37). Such nucleoprotein complexes could enhance the fidelity of RT by affecting processivity, template-primer recognition, or dNTP discrimination. Unidentified cellular factors might also contribute to RT fidelity.

Second, the rate of mispair formation could be the same both in vitro and during viral replication, but its contribution

**TABLE III**

Relative rates of mispair formation by purified MoMLV reverse transcriptase

| Base pair | Experiment | \( V_{\text{max}}/K_m \) | \( f_{\text{err}} \) |
|-----------|-------------|----------------|-------------|
| A-T       | 1           | \( 1.3 \times 10^6 \) | 1           |
|           | 2           | \( 1.5 \times 10^6 \) | 1           |
|           | 3           | \( 8.4 \times 10^5 \) | 1           |
| A-C       | 1           | \( 5.4 \times 10^5 \) | 4.2 \times 10^{-6} |
|           | 2           | \( 5.0 \times 10^4 \) | 3.5 \times 10^{-6} |
|           | 3           | \( 2.5 \times 10^3 \) | \( \sim 3.1 \times 10^{-5} \) |
| T-A       | 1           | \( 1.5 \times 10^4 \) | 1           |
|           | 2           | \( 1.9 \times 10^4 \) | 1           |
|           | 3           | \( 4.3 \times 10^3 \) | 1           |
| T-G       | 1           | \( 5.0 \times 10^4 \) | 3.3 \times 10^{-6} |
|           | 2           | \( 1.8 \times 10^3 \) | 3.5 \times 10^{-6} |
|           | 3           | \( 3.2 \times 10^2 \) | 7.4 \times 10^{-5} |

* Template base shown first.

† Data are from three independent experiments.

‡ \( f_{\text{err}} \) = relative incorporation frequency \( = \frac{V_{\text{max}}}{K_m} \) for mispair formation \( / \frac{V_{\text{max}}}{K_m} \) for correct pair formation.

§ Estimated from the initial linear slopes of plots of [dNTP] \( \text{versus} \) initial velocity (V) under conditions where [dNTP] \( \ll K_m \) and V/ [dNTP] approximates the rate constant \( V_{\text{max}}/K_m \) (32). It was necessary to quantitate mispair rate constants by this method because saturation kinetics did not occur below 2 mM dNTP and substrate inhibition at higher concentrations precluded reliable determinations of individual \( V_{\text{max}} \) and \( K_m \) values. Rate constants for correct base pair reactions were similarly estimated and were consistent with those calculated from individual apparent \( K_m \) and \( V_{\text{max}} \) values (e.g. \( K_m \) for A-T = 4.8 \( \mu M \) and \( V_{\text{max}} \) = 8.5 \% min/mM). All \( V_{\text{max}}/K_m \) values were corrected for the 5-fold greater RT concentrations used in the mispair reactions (0.15 unit/\( \mu M \)) compared to correct pair reactions (0.03 unit/\( \mu M \)).

∥ Initial velocities were directly proportional to RT concentration (0.15 unit/pl) compared to correct pair reactions (0.03 unit/pl), corrected for the 5-fold greater RT concentrations used in the mispair reactions.

(30) MoMLV Mutation Rate and RT Fidelity
to the final rate of viral mutation might be reduced due to inefficient mispair extension in the infected cell. Our data on T1-G mispair extension in vitro (Table IV) demonstrate that the efficiency of extension at the amber codon is almost exclusively a function of the concentration of the next correct nucleotide dGTP (i.e. a $K_a$ effect, Ref. 28). Thus, if dGTP concentrations are sufficiently high in the cell during viral replication and there are no other cellular factors affecting mispair extension, then the T1-G mispairs will present no impediment to DNA polymerization. However, limiting dGTP concentrations will result in slowed extension and possibly loss of mutants due to DNA turnover. Studies on whole cell extracts from log-phase cells demonstrate that total cellular dNTP concentrations range from about 4 to 50 µM for each dNTP with dGTP present in the lowest concentrations (4-5 µM, Refs. 38-42). If the same concentrations are present at the site of retroviral DNA synthesis, our data predict that T1-G mispairs will slow polymerization by some 200-fold. Assuming normal RT polymerization rates of about 0.5-1 nucleotides incorporated/s during retroviral replication (43, 44), this slowed extension of a single mispair will result in a pause of only 200-400 s or about 3-7 min. Since unintegrated dNTP concentrations range from about 4 to 2 mM for mispair extension (10-min reactions). Panel A, T1-A terminus extension on M13neoAm*; panel B, A-C mispair extension on M13neoAm*; panel C, T1-A terminus extension on M13neoAm*; panel D, T1-G mispair extension on M13neoAm*.

TABLE IV

Relative rates of mispair extension by purified MoMLV reverse transcriptase

| Base pair at 3' terminus (template-primer) | $K_a(\text{app})$ | $V_{\text{max(app)}}$ | $V_{\text{max(app)}}/K_a(\text{app})$ | fext* |
|------------------------------------------|------------------|-----------------------|--------------------------------------|-------|
| A-T                                      | ND               | ND                    | 2.6 x 10^-6                          | 1     |
| A-C                                      | ND               | ND                    | 3.9 x 10^-6                          | 1     |
| T1-A                                     | 0.9              | 8.8                   | 9.7 x 10^-6                          | 1     |
| T1-G                                     | 780              | 6.1                   | 7.8 x 10^-3                          | 0.8 x 10^-3 |

*fext = relative extension efficiency = $V_{\text{max(app)}}/K_a(\text{app})$ for mispair extension

*ND, not determined.

*Estimated as described in Table III.

* Determined from $V_{\text{max(app)}}$ and $K_a(\text{app})$ values.

MoMLV Mutation Rate and RT Fidelity

of pyrophosphate on the fidelity of avian myeloblastosis virus RT (46).

Third, cellular DNA mismatch repair mechanisms might correct mispairs formed by RT. This would require a repair system that gains access to the replicative complex or integrated proviral DNA, discriminates between the template and nascent DNA strands, and selectively repairs incorrectly inserted bases on the nascent strand. DNA repair systems capable of strand-specific mismatch repair have been detected in nuclear and cytoplasmic extracts of HeLa cells (47-49) and nuclear extracts of Drosophila melanogaster cells (47). These systems might be capable of correcting replication errors during or after retroviral plus strand DNA synthesis. Studies of retroviral recombination suggest that mismatch repair might occur during replication sometime after reverse transcription and before cellular replication of integrated proviral DNA (50). Although strand discrimination for mismatch repair in retroviruses has not been examined, selective removal of RT errors in the nascent plus strand could be signaled by discontinuities often present in this strand (44). It should be noted that repair by the mammalian G-T to G-C short patch repair system cannot account for the low rate of AVneoAm virus mutation in our experiments, since this mismatch repair system exhibits a strand bias that would actually increase the reversion rate during replication (51).

The fourth possibility is that differences between the MoMLV RT used by the vector virus during replication and the recombinant MoMLV RT used in vitro resulted in the observed discrepancies. The purified RT utilized in our in vitro studies is a fusion protein produced in E. coli which contains some amino acid residues of non-viral origin at both termini and lacks 7 residues normally present at the carboxyl end of the mature enzyme (18). Similar experiments to those described herein revealed that the recombinant RT used in our studies and another recombinant MoMLV RT (Bethesda Research Laboratories, Ref. 52) incorporate nucleotides with similar fidelity opposite a G on a ΦX174 single-stranded DNA template. The second enzyme has all the amino acids present in the authentic RT, but it also contains an extra methionine at the amino terminus as well as 6 extra residues present in the polyprotein (not found in the normally processed MoMLV RT) plus 6 residues encoded by a terminator linker at the carboxyl terminus. Since the accuracies of both recombinant proteins are the same, it suggests that the differences

* M. L. Kuo, unpublished observations.

* N. Garvey and B. D. Preston, unpublished results.
at their termini do not affect fidelity.

Of the four possible mechanisms discussed above, we favor the first, that accessory viral or cellular factors increase RT fidelity during plus strand synthesis in the cell. Replicative polymerases do not generally function as isolated enzymes and often require accessory factors to conduct efficient, faithful DNA synthesis (53). Several retroviral proteins exhibit biochemical properties that suggest accessory roles during reverse transcription (54), but contribution of these and other cellular factors to RT fidelity has not been addressed experimentally. Analyses of the fidelity of reverse transcription complexes in vitro involving other virus-associated proteins should provide valuable information concerning the possible roles of these proteins in retroviral DNA synthesis. Similar analyses including cellular extracts might prove useful in elucidating the role of mismatch repair systems in retroviral replication fidelity. In light of these results, caution should be exercised when extrapolating in vitro measurements of RT fidelity to the corresponding mutation rates during retroviral replication.

Our results also provide information about the fidelity of RNA pol II. RNA pol II transcribes genomic RNA from proviral DNA and thus can contribute to retroviral mutagenesis. Establishment of proviral DNA and thus can contribute to retroviral mutagenesis. Our data indicate that RNA pol II forms T-G mispairs in the neo amber codon at a rate no greater than 2 x 10^-4 (i.e., the overall mutation rate of the virus). This relative high fidelity is not predicted from teleological arguments nor from the overall mutation rate of the virus). This relative high fidelity is not predicted from teleological arguments nor from the overall mutation rate of the virus. This relative high fidelity is not predicted from teleological arguments nor from the overall mutation rate of the virus.

Acknowledgments—We thank Christine M. Prorock for technical assistance and Ralph Dornburg, Han-Mo Koo, Paul Olen, Shobha Panhatarami, and Monica J. Roth for discussion and helpful comments on the manuscript.

Note Added in Proof—The fidelity of MoMLV RT was determined during minus strand DNA synthesis in vitro on an RNA neoAm template generated by SP6 RNA polymerase-catalyzed run-off transcription of the neoAm gene cloned into the vector pGEM-3zf (+) (Promega Corporation, Madison, WI). The frequency of A-G mispair formation (fAG) at the neoAm reversion site was 1.2 x 10^-4. Thus, the rate of A-G mispair formation in vitro is about 10 times lower on RNA compared to DNA templates of the same neoAm sequence. These data do not change the conclusions of this paper. However, they do suggest an asymmetry in the contribution of DNA- and RNA-directed errors to retroviral mutagenesis.

REFERENCES

1. Baltimore, D. (1970) Nature 226, 1209-1211
2. Temin, H. M., and Mizutani, S. (1970) Nature 226, 1211-1213
3. Coffin, J. M. (1989) in Applied Virology Research (Kurstak, E., Maruyuk, R. G., Murphy, F. A., and Van Regenmortel, M. H. V., eds) Vol. 2, pp. 11-33, Plenum Press, New York.
4. Katz, R. A., and Skalka, A. M. (1990) Annu. Rev. Genet. 24, 409-445
5. Wahi, M., Burrows, P. D., Gabain, A. V., and Steinberg, C. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 479-492