Determination of the Mutation Rate of a Retrovirus

JASON M. LEIDER, PETER PALESE, AND FRANCES I. SMITH*

Department of Microbiology, Mount Sinai School of Medicine, New York, New York 10029

Received 5 February 1988/Accepted 25 May 1988

The mutation rate of Rous sarcoma virus (RSV) was measured. Progeny descended from a single virion were collected after one replication cycle, and seven regions of the genome were analyzed for mutations by denaturing-gradient gel electrophoresis. In all, 65,250 nucleotides were screened, yielding nine mutations, and the RSV mutation rate was calculated as 1.4×10^{-4} mutations per nucleotide per replication cycle. These results indicate that RSV is an extremely mutable virus. We speculate that the mutation rate of a virus may correlate inversely with the effectiveness of vaccination against a given virus and suggest that prevention of retrovirus-mediated disease via vaccination may prove difficult.

Viral variation may be the most important factor which limits the effectiveness of vaccine development against viral infection. Recent data have suggested that the amount of variation of a virus in nature can be correlated with its mutation rate (26). Because of the emerging importance of retroviral infections in humans, it was of interest to determine the mutation rate of a representative retrovirus and to compare this value with that of other human RNA viruses.

Previous workers have studied variation of retroviruses (4, 6) and, although a precise mutation rate at the nucleotide level could not be calculated from these studies, the results obtained suggested that extremely high variability could arise in a population derived from a cloned virus over a relatively short period. In this study, we attempted to measure directly the mutation rate of the retrovirus Rous sarcoma virus (RSV) by screening for mutations arising over a single replication cycle. Denaturing-gradient gel electrophoresis, originally developed by Lerman et al. (21) and subsequently modified to detect point mutations in RNA molecules by Smith et al. (32), was used to screen for mutations. The ease and rapidity of this new technique allowed us to analyze several thousand nucleotides distributed over seven portions of the RSV genome. The results indicate that retroviruses are among the fastest-mutating viruses.

MATERIALS AND METHODS

Cells. Chicken embryo fibroblasts (CEFs) were prepared from 10-day-old fertilized chicken eggs according to published methods (15). CEFs susceptible to the Prague C strain of RSV (RSV Pr-C) were derived from eggs that were virus negative, chicken helper factor negative, and avian leukosis virus group-specific antigen negative (v^- chf^- gs^- ; SPAFAS, Inc., Norwich, Conn.). CEFs genetically resistant to RSV Pr-C infection were derived from fertilized chicken eggs that were the kind gift of W. E. Briles, Department of Biological Sciences, Northern Illinois University. The culture medium employed consisted of Hams F-10 medium (Hazelton Research Products, Inc.) supplemented with 10% tryptose phosphate broth, 5% calf serum, 6.6 mM sodium bicarbonate, 100 U of penicillin per ml, and 100 μ g of streptomycin per ml.

Plasmids. pATV-8, a recombinant plasmid containing the RSV Pr-C genome cloned into pBR322 (19), kindly provided

by J. Coffin of the Tufts University School of Medicine, was the source of RSV Pr-C sequences. Various regions of the RSV Pr-C genome were subcloned into plasmids containing RNA polymerase promoters for the purpose of the production of riboprobes. pGEM-4 (Promega Biotec, Madison, Wis.) and pIBI 31 (IBI) were used as cloning vectors. The following RSV Pr-C genomic fragments were subcloned from pATV-8; 256 to 520 into the *SmaI-SphI* site of pGEM-4 (pgag1); 520 to 1006 into the *SmaI-SphI* site of pGEM-4 (pgag3); 4068 to 4997 into the *AccI-KpnI* site of pGEM-4 (ppol1); 5257 to 5563 into the *PstI-SmaI* site of pGEM-4 (penv4); 5564 to 5837 into the *PstI-SmaI* site of pGEM-4 (penv1); 6440 to 6865 into the *SacI-PstI* site of pGEM-4 (penv3); 6983 to 7631 into the *SalI-AvaI* site of pGEM-4 (psrc1); 9238 to 9291 and 1 to 2318 into the *EcoRI* site of pIBI 31 (pGG-AMT3); 6285 to 9238 into the *EcoRI-HindIII* site of pIBI 31 (pE3S-MT7); and 2870 to 6085 into the *HindIII* site of pIBI 31 (pPE14-AMT7). All plasmids were made by standard techniques, transfected into *Escherichia coli* HB101 or DH5- α , selected by hybridization with radiolabeled nick-translated restriction fragments of pATV-8, and expanded, and the plasmid DNA was purified by the alkaline lysis method (23).

Preparation of RNA probes. Plasmid DNA was digested with *EcoRI* or *HindIII* restriction enzymes, depending on the desired sense of the transcripts. Linear templates (100 μ g/ml) were then transcribed by SP6 RNA polymerase (7.5 U/ μ g of template DNA; Promega Biotec), T7 RNA polymerase (7.5 U/ μ g of template DNA; Promega Biotec), or T3 RNA polymerase (10 U/ μ g of template DNA; IBI) in the presence of [α - 32 P]CTP as described by Melton et al. (24). Typically, 3 to 5 μ g of single-stranded RNA was produced per microgram of input plasmid DNA.

Preparation of double-stranded (ds) RNA hybrids. 32 P-labeled minus-sense RNA was suspended in 30 μ l of hybridization buffer (80% deionized formamide, 40 mM PIPES [piperazine-*N,N'*-bis(2-ethanesulfonic acid)], pH 6.7, 0.4 M NaCl, 1 mM EDTA), mixed with 1 μ g of viral RNA, and incubated for 12 to 16 h at 45°C. To remove unhybridized RNA, 300 μ l of S1 digestion buffer (30 mM sodium acetate [pH 4.4], 280 mM NaCl, 5 mM ZnCl₂, containing 1,000 U of nuclease S1 per ml; Sigma Chemical Co., St. Louis, Mo.) was then added, and the reaction was incubated at 37°C for 1 h. This step was followed by a phenol-chloroform extraction, and the [32 P]RNA was ethanol precipitated, dissolved in 50 μ l of sterile water, and stored at -20°C.

The wild-type control duplexes were made by hybridizing

* Corresponding author.

³²P-labeled SP6- or T7-derived transcripts from the various pGEM-4 plasmids with T7- or T3-derived transcripts of the RSV Pr-C genomic fragments in the pIBI 31 plasmids. The latter plasmids contained regions complementary to the SP6 or T7 transcripts as well as additional retroviral sequences at the 5' and the 3' ends. It was necessary to employ such a strategy to make the control heteroduplexes in order to avoid the complication of hybridization due to polylinker sequence. Adjacent hybridized polylinker would alter the length and perhaps the melting properties of the hybrids. Conditions for hybridization and nuclease S1 digestion were the same as mentioned above.

Electrophoresis. The submerged gel apparatus used was similar to that described by Fischer and Lerman (8). Gels were 6.5% polyacrylamide containing a denaturant gradient either perpendicular or parallel to the direction of electrophoresis (100% denaturant = 7 M urea and 60% formamide). Parallel gradients had a short-stacking gel (3.8% acrylamide, no denaturant). The gels were run at 200 V submerged in an aquarium containing TAE buffer (40 mM Tris acetate [pH 7.4], 20 mM sodium acetate, 1 mM EDTA) that was maintained at a constant temperature of 65°C by a thermomix (B. Braun Melsungen, AG) heater and stirrer. After electrophoresis, the gels were fixed in a 7.8% acetic acid-26.5% methanol solution, dried, and exposed to Cronex 4 film (E. I. du Pont de Nemours & Co., Inc., Wilmington, Del.) for 12 to 48 h. Nondenaturing 6.5% polyacrylamide gels run at room temperature were also used to check the integrity of all heteroduplexes prior to loading on denaturing-gradient gels.

Transfection of CEFs. CEFs were transfected by the method of Potter et al. (27) with RSV Pr-C cDNA purified from a partial *Hind*III digestion of pATV-8. A bank of capacitors (effective capacity, 14 μ F) charged to 1,500 V was discharged via an electronic switch (model ZA 1000; PDS, Madison, Wis.) through the sample suspended in 0.5 ml of 1 $\times$ ZAP buffer (140 mM NaCl, 0.75 mM Na₂HPO₄, 25 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [pH 7.2]) by using a Cell Chamber (PDS) 5 mm in length and with a cross-sectional area of 1 cm² at 0°C. Cells were then allowed to adhere to tissue culture dishes at 37°C. Medium was changed at 3-day intervals, and complete transformation was seen in 14 days. Individual transformed cells were grown at a limiting dilution of 0.2 transformed cells per 2.5-cm dish containing a feeder layer of cells genetically resistant to infection by RSV Pr-C to minimize virus spread across feeder layers. Growth of transformed cells was scored by light microscopy. These clones represented parental candidates (see Fig. 1).

Preparation of RSV and purification of viral RNA. To generate sufficient RNA for analysis of the viral genotype, supernatant was collected from dishes containing a single colony, and the virus was amplified by addition to a 5-cm dish of susceptible CEFs followed by subsequent expansion. Supernatant containing virus liberated from these confluent transformed cells was then harvested until a total volume of 200 ml had been collected. The supernatant was cleared by centrifugation at 8,000 $\times g$ for 10 min and layered over a 3-ml 20% sucrose cushion. Virus was pelleted by centrifugation at 25,000 rpm for 2 h in an SW27 rotor (Beckman Instruments, Inc., Fullerton, Calif.). The virus pellet was suspended in 4 ml of a solution of 100 mM NaCl, 10 mM Tris hydrochloride, pH 7.4, and 1 mM EDTA, and RNA was extracted according to the method of Palese and Schulman (25).

Focus-forming unit (FFU) assay. A 2.5-cm dish of confluent CEFs was incubated with serial dilutions of the supernatant for 40 min at 37°C. The cells were then overlaid with medium

containing 0.75% agar, incubated at 37°C, and monitored for focus formation. After 7 days of incubation, no foci were seen at any dilution. To allow further time for foci to appear and to prevent cell death, the agar overlay was removed and replaced with liquid medium. On day 23, foci were seen in dishes of various dilutions. Because of the opportunity for viral spread across the monolayers during this procedure, the titer of FFU per milliliter thus obtained is only an estimate.

Colony formation in soft agar. A culture of confluent CEFs growing in a 10-cm tissue culture dish (7×10^6 cells per dish) was prepared and infected with 2 ml of virus (approximately 400 FFU) in the presence of 0.3 ml of DEAE-dextran (1:40) for 4 h at 37°C. Six plates, each with a bottom agar layer of 80% F-10 medium, 10% tryptose phosphate broth, 6% heat-inactivated calf serum, 2% heat-inactivated chick serum, 88 μ g of folic acid per ml, 0.04% dimethyl sulfoxide, 0.18% sodium bicarbonate, 100 U of penicillin per ml, 100 μ g of streptomycin per ml, 0.25 μ g of fungizone per ml, and 0.56% agar, were prepared during this time. Virus-infected cells were trypsinized and suspended in 6.5 ml of liquid F-10 medium. To each agar-containing plate, 1 ml of infected cell suspension was added. A top agar layer consisting of 50% F-10, 6% tryptose phosphate broth, 0.09% sodium bicarbonate, 4% heat-inactivated calf serum, 2% heat-inactivated chick serum, 0.25 μ g of vitamins per ml (GIBCO Laboratories, Grand Island, N.Y.), 40% cultured supernatant from uninfected CEFs, 0.5% dimethyl sulfoxide, 100 U of penicillin per ml, 100 μ g of streptomycin per ml, 0.13 μ g of fungizone per ml, and 0.44% agar was added to these six plates and shaken immediately to evenly distribute cells. The entire culture was allowed to solidify for 15 min at room temperature and then incubated at 37°C. After 14 days, well-isolated colonies containing 100 to 500 cells were seen. With a 1-ml syringe containing less than 0.1 ml of 0.05% trypsin and 0.2 mM EDTA, colonies were picked and then added to a 2.5-cm dish of confluent susceptible CEFs containing 1 ml of liquid medium. Transformation of these dishes was monitored daily, and expansion of these dishes was performed as required.

RESULTS

Experimental strategy for determination of the RSV mutation rate. To determine the RSV mutation rate, variation within the progeny of a parental clone was examined after a single replication cycle. The replication cycle began with a single colony of transformed cells, the parental clone (Fig. 1), and virus produced from that single colony was used to establish multiple transformed cell clones. In this single replication cycle, proviral DNA of the original parental colony was transcribed by RNA polymerase II to produce viral RNA for progeny virus. After infection of cells with this progeny virus, reverse transcription allowed synthesis and integration of proviral DNA. Mutations occurring during this single cycle of replication (from a single colony of transformed cells to transformed progeny cells) may be the result of errors made by RNA polymerase II or reverse transcriptase or both. Since mutations rarely happen during provirus replication (18), it is unlikely that mutations caused by host cell DNA polymerase during growth of the original colony would be detected. By using denaturing-gradient gel analysis, mutations in the progeny were screened and a mutation rate for the retrovirus for a single replication cycle was calculated accordingly.

Melting map of the dsDNA copy of the RSV Pr-C genome. The first step in analysis of the RSV mutation rate by

FIG. 1. Experimental design for measuring the mutation rate of RSV during a single replication cycle. The single cycle of replication involves transcription first of the parental provirus by RNA polymerase II to produce the RNA of the viral genome and then of the viral RNA by reverse transcriptase to make the provirus that is integrated in progeny colonies. Analysis by denaturing-gradient gel electrophoresis of RNA from the amplified virus detects the consensus sequence, which reflects that of the provirus in the progeny colonies. Individual mutations arising during the amplification would thus not be detected by this method.

denaturing-gradient electrophoresis was the generation of a full-length melting map of the virus genome. Such a map is dependent on the availability of the nucleotide sequence. For this study, the RSV Pr-C strain, which had been completely sequenced by Schwartz et al. (29), was used. An expected melting progression along the dsDNA form of RSV Pr-C was calculated as described by Lerman et al. (21). The abscissa in Fig. 2 represents nucleotide position within the RSV Pr-C genome, while the ordinate describes the temperature at which each base pair has an equal probability of

helix or random chain configuration. Cooperativity among adjacent base pairs is strong enough to permit rather long domains of contiguous helix to undergo the transition from the helical to the random coil state within relatively narrow temperature intervals, giving a stepwise appearance to the melting map. Experimental results obtained with various dsRNA and DNA-RNA duplexes have shown that relative stabilities of different domains within the duplexes correlate well with those predicted by the theoretical melting behavior for the dsDNA duplexes (32). The sensitivity of denaturing-gradient gel electrophoresis for the detection of single-point mutations has been previously well characterized for both DNA (21) and RNA molecules (32). Using a range of sequenced mutants, these workers showed that heteroduplexes between wild-type and mutant molecules are less thermodynamically stable than perfectly matched duplexes and undergo melting of mismatched domains at lower concentrations of denaturant. In both studies, all single-point mutations screened by heteroduplex analysis on denaturing gradient gels were accurately and easily detected.

Initial characterization of viral regions chosen for analysis. With the aid of the melting map, seven regions of the RSV Pr-C genome were chosen for further analysis (Fig. 2). Regions containing a relatively low-melting domain abutting a high-melting domain were selected. It is only in the low-melting domains of such heteroduplex regions that mutations can be detected reliably (21). Another factor influencing our choice of regions for analysis was that we wanted to study at least one portion of each gene. Three regions of the *env* gene were chosen for analysis, corresponding to regions that were previously reported to be conserved (*env4*) or hypervariable (*env1*) among different strains of avian sarcoma and leukemia viruses (2, 5). These regions, therefore, were subcloned from pATV-8 into RNA polymerase transcription vectors pGEM-4. Control wild-type duplexes were made by hybridizing ³²P-labeled transcripts of opposite polarity with one another (for specifics on the construction of wild-type duplexes, see Materials and Methods).

Since the melting map only predicts the location of the various domains, perpendicular denaturing-gradient gels were used to determine experimentally the percent denaturant at which melting of the domains of the RNA-RNA duplexes occurred (Fig. 3). An S1-treated wild-type control heteroduplex for each region examined was loaded in a single well spanning the width of the cathode edge of the gel.

FIG. 2. Melting map of the dsDNA form of the RSV Pr-C genome. The map indicates the temperature at which each base pair will be at equilibrium between helix or random chain configuration for the dsDNA molecule in an aqueous environment containing about 0.02 M Na⁺. Also indicated are the boundaries of the four retrovirus genes and the seven regions selected for mutational analysis. Numbering is according to that of Schwartz et al. (29).

FIG. 3. The experimental melting pattern of dsRNA hybrids. The mobility of the radioactively labeled dsRNA hybrid was examined at 65°C on a 6.5% polyacrylamide gel containing a linearly increasing gradient of denaturant from 0 to 100%. Electrophoretic migration of the dsRNA hybrid is from top to bottom (perpendicular to the denaturing gradient). By using the first inflection point of this curve, it is possible to estimate the percent denaturant at which melting of the lowest melting domain, predicted from the theoretical melting map, occurs.

Migration was from top to bottom, perpendicular to a denaturing gradient that increases from left to right. For example, Fig. 3A shows an analysis of the *env1* region. RNA molecules on the left side of the curve (below 50% denaturant) were double stranded, while molecules on the right side of the curve (above 55% denaturant) were completely melted. Molecules between these two extremes had a completely melted low-melting domain and an intact high-melting domain and migrated the most slowly. At the first inflection point of the experimental curve, melting of the low-melting domain, predicted from the theoretical melting map for the *env1* region, occurred. The perpendicular denaturing gels from the remaining regions showed patterns of various sigmoidal shapes (for example, the curves for the various *env* regions are shown in Fig. 3). The theoretical melting map was used to identify the location and extent of the lowest-melting domain for each region.

Isolation of the parental clone. The 9.3-kilobase fragment, representing the full-length RSV Pr-C genome, was isolated from a partial *Hind*III digestion of pATV-8. A large dish of confluent CEFs was transfected with this DNA by electroporation. In 14 days, complete transformation of the entire dish was observed. Transformed cells were then cloned at a limiting dilution (0.2 transformed cells per 2.5-cm dish) on cells genetically resistant to infection by RSV Pr-C. Out of 98 2.5-cm dishes, 13 were found to contain a single focus. These clones represented parental candidates for the following experiments. Virus liberated into the supernatants of these dishes was collected and aliquoted. One aliquot of each candidate was amplified on susceptible CEFs. Within 3 days, complete transformation of these cells occurred. Virus was collected from the supernatants for analysis. Heteroduplexes were formed by hybridizing the amplified viral RNA of a candidate to 32 P-labeled transcripts derived from each of the subcloned regions. Nuclease S1 digestion followed by nondenaturing polyacrylamide gel electrophoresis did not detect any deletion-insertion mutations. Samples were then loaded onto a gel containing a gradient of denaturant parallel to the direction of electrophoresis. The presence of a muta-

tion would cause premature melting of the low-melting domain, leading to retardation of the sample in the gel. Our definition of a wild-type clone, therefore, was a candidate for which all seven regions being probed would exhibit patterns of migration identical to that of the control hybrid (Fig. 4). Additionally, although a candidate may have harbored multiple integrated proviruses (of which some may have contained mutations), we estimate that denaturing-gradient gel electrophoresis is capable of detecting mutant bands at up to one-tenth the intensity of the major band. Since the number of proviruses integrated per cell as a result of avian retrovirus infection is usually lower than 10 (20), it is unlikely that the chosen parental clone contained an undetected mutant provirus. Furthermore, if the mutations detected in the progeny clones were directly attributable to the presence of a mutated provirus in the parental clone, the mutations should all be identical. Because most mutations detected could be shown to be clearly different, due either to their locations in different regions of the genome or to the different

FIG. 4. Screening parental candidates in the *env4* region with parallel denaturing-gradient gel electrophoresis. RNA-RNA hybrids were constructed by using radioactively labeled minus-sense probes and viral RNA from parental candidates as described in Materials and Methods. All hybrids were initially examined for deletion-insertion mutations by using electrophoresis with 6.5% polyacrylamide gels. Hybrids for this region were then electrophoresed at 65°C on a gel containing a denaturant gradient (20 to 80%) parallel to the electric field. Patterns of migration for *env4* were identical to those of the control hybrid (lane C).

TABLE 1. Mutational analysis of selected regions of the RSV genome

Probe	Size of low-melting domain ^a	Nucleotides screened ^b	Mutations detected
<i>gag1</i>	104	6,032	1
<i>gag3</i>	156	9,048	4
<i>pol1</i>	249	14,442	0
<i>env1</i>	125	7,250	2
<i>env4</i>	159	9,222	0
<i>env3</i>	87	5,046	0
<i>src1</i>	245	14,210	2

^a In nucleotides; as predicted by the theoretical melting map.

^b For each region of the genome analyzed, the same 58 progeny RNAs were screened.

mobilities of their heteroduplexes on denaturing-gradient gels (see below), it appears that the chosen parental clone did in fact contain only wild-type provirus. It should also be noted that of the five clones which emerged as possible parental candidates, the one with the highest virus yield (approximately 2.5×10^3 FFU/ml) was chosen as the parent for this study.

Progeny generation and mutation rate calculation. A confluent dish of susceptible CEFs was infected at a multiplicity of infection of approximately 10^{-5} FFU per cell with an original aliquot of the parental virus. After 4 h, cells were split into six 10-cm dishes and grown in soft agar. After 14 days, approximately 60 colonies were seen in each dish. Well-isolated colonies were picked and amplified. Virus was harvested from the supernatant of each of 58 progeny colonies, and viral RNA was examined for mutation. Nuclease S1 analysis of RNA duplexes for the seven subcloned regions did not detect any insertion-deletion mutations. Further analysis was then performed by using denaturing-gradient gel electrophoresis (see Fig. 5). As previously discussed, only the low-melting domains of each region can be analyzed by this method, and from the theoretical melting maps, the target size for each of these domains is known. The data are summarized in Table 1. In all, 65,250 nucleotides of the RSV genome of 58 progeny RNA were analyzed, with the detection of nine mutations, located within selective regions of the *gag*, *env*, and *src* genes. No mutations were detected in the 3' region of the *pol* gene and two regions of the *env* gene. Sequence analysis of a chosen mutant (that showed retarded migration of the *env1* probe) confirmed that we were in fact detecting single-base mutations (data not shown).

The overall mutation rate for the virus was calculated by using the formula: mutation rate = 9 mutations per 65,250 nucleotides per replication cycle = 1.4×10^{-4} mutations per nucleotide per replication cycle.

DISCUSSION

We report here a method that has allowed the rapid and direct measurement of the mutation rate of a retrovirus. Progeny descended from a single virion after one replication cycle were collected, and seven different regions of the genome were analyzed for mutations. CEFs were infected at a low multiplicity of infection, thus avoiding the possible amplification of defective particles which could drive virus evolution (33) and lead to an erroneous calculation of the RSV mutation rate. In all, 65,250 nucleotides were screened, with the detection of nine mutations. The RSV mutation rate was calculated as 1.4×10^{-4} mutations per nucleotide per

replication cycle. Since provirus replication does not appear to be error prone (18), we attribute this rate to mutations incurred during the replication of the RSV genome by RNA polymerase II or reverse transcriptase or both. Our experimental design, however, did not allow us to dissect the proportion of mutations contributed by each enzyme alone.

The use of denaturing-gradient gel analysis allowed several sites throughout the genome to be screened for mutations, overcoming any bias which may be due to the existence of regions of differential variability within viral genomes. An extreme example of such variability is the bacteriophage T4 rII locus, in which 10^4 -fold differences in mutation frequencies were observed for single nucleotides at different sites in the gene (28). Additionally, there are many reports demonstrating that different regions of viral genomes can exhibit different levels of variability in nature (for reviews, see references 31 and 35). Thus, studies analyzing only a small portion of a viral genome may give skewed estimates of the mutation rate of that virus. In this study, mutations were detected only within certain regions of the *gag*, *src*, and *env* genes. The region of the *env* gene in which we observed mutations corresponds to a region of the gp85 protein reported to be extremely variable among different strains of avian sarcoma and leukosis viruses (2, 5). In the *pol* gene, we analyzed a region corresponding to the putative zinc-binding site of the retroviral endonucleases (17). Interestingly, although we analyzed the greatest number of nucleotides for this region, no mutations were detected. However, the numbers of mutations detected in each region are too small to conclude whether their distribution is nonrandom. We did not detect insertion-deletion mutation by nuclease S1 analysis in any region. This may be due to the deleterious effect of such mutations on proteins necessary for competent replication or transformation or both.

The design of these experiments enables us to calculate a minimum mutation rate for RSV under conditions in which the frequency of mutation due to recombination with endogenous viruses is very low. The $v^{-}gs^{-}chf^{-}$ chicken cells used express only very low levels of endogenous proviral message (approximately 0.5 to 1 copy of mRNA per cell; 1, 14, 36). In contrast, the amount of viral RNA in a cell infected by an exogenous retrovirus is several orders of magnitude higher (3,000 to 20,000 copies per cell; 13). Recombination between retroviruses is dependent on the expression of mRNA in a cell and most likely occurs during reverse transcription of RNA of heterozygous virions (that is, virions containing one endogenous and one exogenous viral RNA strand; for a review, see reference 22). Because of the vast excess of exogenous viral RNA in infected $v^{-}gs^{-}chf^{-}$ cells, the occurrence of copackaging heterozygous RNAs would be very low (most optimistically, 1 in 1,500 virions, assuming 1 copy of endogenous RNA per cell and 3,000 copies of exogenous RNA per cell). Even if one assumes that recombination would occur in 100% of such heterozygotes, the frequency of recombinants would be much lower than the number of mutants detected in this study (9 out of the 58 progeny clones examined). However, it is not possible to exclude the possibility that the parental clone chosen for these experiments was aberrant in its expression of endogenous proviral message, and produced amounts of endogenous proviral message-specific RNA orders of magnitude higher than did previously examined cell lines. Nevertheless, because two of the nine mutations were located in the *src* gene, which is not present in endogenous viruses, we believe that we are in fact looking at polymerase error-induced mutations and not at those due to recombination.

FIG. 5. Mutational analysis using parallel denaturing-gradient gel electrophoresis for progeny RNA hybrids. Mutations (lanes M) detected in the *env1* (A), *src1* (B), and *gag3* (C) regions are shown here. Mutant samples have a retarded migration through the gradient with respect to the control hybrid (lanes C). The denaturing gradients used for the regions shown here are as follows: (A), 30 to 70%; (B), 30 to 70%; (C), 50 to 100%.

Studies in which recombinants were selected for by phenotypic changes have shown that although the frequency of recombinants in *chf*⁻ cells was very low (less than 1 virion in 10⁵), up to 5% of RSV progeny have been shown to be recombinant for an endogenous viral envelope marker after passage through *chf*⁺ cells (38). Therefore, although our calculation of the mutation rate for a retrovirus is already very high, it may in fact be much higher in cells expressing large amounts of endogenous viral RNA. Also, because the experiments described in this report selected for the production of viable transforming virus, mutations that would produce nonviable progeny would not be detected. Therefore, the mutation rate presented here should be considered a minimum mutation rate.

It should also be noted that during the generation of progeny colonies, there can be a cycle or two of reinfection of the initially infected clone of cells, leading to more than one provirus in the clones so derived. However, if following insertion of a wild-type provirus, a mutation occurred in a second or subsequent round of infection and gave rise to a mutant provirus, a mixture of both wild-type and mutant virus would be produced. Therefore, a strong band on the denaturing-gradient gels representing the wild-type virus would be present as well as secondary retarded bands due to the mutant(s). However, as can be seen in Fig. 5, progeny characterized as mutants were those that gave a single dominant band of retarded mobility on a denaturing-gradient gel in the absence of a wild-type band. Therefore, the mutants detected in this study are most likely those that arose during a single cycle of replication.

Previous studies on variation in retroviruses have been undertaken. An *in vitro* error rate for reverse transcription of a DNA template from a mutant bacteriophage has been measured to be approximately 10⁻³ mutations per nucleotide polymerized (11). This high error rate, however, may not be able to be extrapolated to *in vivo* conditions. In another study, Coffin et al. (4) repeatedly passaged the Prague B strain of RSV undiluted from one cell culture to another and used oligonucleotide fingerprint analysis to determine the variability of the virus. A frequency on the order of 3 × 10⁻⁴ bases per passage of virus was found, but because of the complexity of this system, this number cannot be used to develop a precise mutation rate (3). Most recently, Dougherty and Temin (6) employed a spleen necrosis virus-based vector that contained less than 1 kilobase of retrovirus sequences and two dominant selectable genes to estimate a mutation rate. By monitoring for changes in expression of these selectable genes, they calculated that 0.5% mutant

virions arose after a single replication cycle. This study, however, did not permit a precise calculation of the mutation rate in terms of mutations per nucleotide per replication cycle, because the target size for mutation was not known. Nevertheless, the results from these studies do predict high variability in cloned retroviral populations, which is consistent with our results.

RNA viruses generally exhibit much higher variation in nature than do organisms that contain a DNA genome, and it has been postulated that this observation may be explained by the lack of fidelity of RNA polymerases (16). However, few direct determinations of mutation rates for RNA viruses have been done to test this hypothesis. Steinhauer and Holland (34) used oligonucleotide mapping with T1 ribonuclease to calculate an error rate for the vesicular stomatitis virus RNA polymerase of 10^{-3.15} substitutions per base incorporated. Parvin et al. (26) measured the mutation rates for influenza A virus and poliovirus type 1 as 1.5 × 10⁻⁵ and <2 × 10⁻⁶ mutations per nucleotide per cycle, respectively, by repeated dideoxy sequencing of a single gene for each virus from numerous virus clones, each derived from the same plaque. Sedivy et al. (30) also calculated a mutation rate for poliovirus (at 7.5 × 10⁻⁶ mutants per nucleotide per plaque generation) by using an inducible mammalian amber suppressor system to monitor the reversion frequency of a mutant of poliovirus. This poliovirus mutation rate is in close agreement with that of Parvin et al. (26), as the latter authors estimated that poliovirus undergoes five replication cycles per plaque generation. However, the mutation rates calculated by Parvin et al. (26) and Sedivy et al. (30) are both much lower than the error rates observed by Ward et al. (37) for the *in vitro* transcription of homopolymeric RNA templates by purified poliovirus RNA polymerase (which ranged from 7 × 10⁻⁴ to 5.4 × 10⁻³ noncomplementary nucleotides incorporated per total nucleotides incorporated, depending on the reaction conditions). This discrepancy probably reflects the difference between *in vivo* and *in vitro* reaction conditions. Durbin and Stollar (7) assessed virus variability by estimating the frequency of reversion of a particular Sindbis virus host-restricted mutant. Direct sequencing analysis of the E2 gene of revertants allowed them to estimate a mutation rate of <10⁻⁶ errors per base incorporated, but they could not exclude the formal possibility that a second mutation outside the E2 gene might be involved in reversion to the ts⁺ phenotype and consequently affect their reversion frequency estimate. Thus, the above results indicate that mutation rates are indeed high for RNA viruses (16), but that marked differences (by at least 2 orders of magnitude) exist

between viruses. RSV, with a mutation rate of 1.4×10^{-4} mutations per nucleotide per cycle, is one of the fastest-mutating viruses yet studied.

Previous results have suggested that mutation rates may correlate with the variability exhibited by viruses in nature and, therefore, with the effectiveness of vaccination against these viruses. There is epidemiological evidence indicating that influenza A virus evolves more rapidly than poliovirus. Specifically, vaccine strains against influenza A viruses have to be changed frequently to protect against an evolving virus population. In contrast, vaccines against poliovirus are based on strains which have been used for the last several decades, ostensibly without loss of efficacy. Parvin et al. (26) suggested that the higher mutation rate found for influenza A virus may be responsible for generating the diversity required for the rare event in which a mutant that could escape neutralization in a previously immunized host is produced. Our determination of the RSV mutation rate (which is approximately 9 and 65 times higher than those estimated for influenza A virus and poliovirus, respectively) indicates that RSV is an extremely mutable virus. This finding, in association with the reports of rapid rates of evolution for human immunodeficiency virus (12), Moloney murine sarcoma virus (9), and RSV (10), predicts that the high mutation rate of the retroviral genomes may be the source of their accelerated evolutionary rates. If mutation rates can indeed be correlated inversely with the effectiveness of vaccination against a given virus, these results suggest that prevention of retroviral-mediated disease via vaccination may be difficult.

ACKNOWLEDGMENTS

We gratefully acknowledge the expert technical assistance of T. Latham. We thank L.-H. Wang for many helpful discussions, suggestions, and his gift of cell cultures needed during the course of this project, as well as for critical reading of the manuscript. We also thank J. Parvin for many stimulating discussions.

J.M.L. is supported by Public Health Service Medical Scientist Training grant GM07280 from the National Institutes of General Medical Sciences. This work was supported by Public Health Service grants AI-11823 and AI-18998 (P.P.) and DK38381 (F.I.S.) from the National Institutes of Health.

LITERATURE CITED

- Baker, B., H. L. Robinson, H. E. Varmus, and J. M. Bishop. 1981. Analysis of endogenous avian retrovirus DNA and RNA: viral and cellular determinants of retrovirus gene expression. *Virology* **144**:8–22.
- Bova, C. A., J. C. Olsen, and R. Swanstrom. 1988. The avian retrovirus *env* gene family: molecular analysis of host range and antigenic variants. *J. Virol.* **62**:75–83.
- Coffin, J. M. 1986. Genetic variation in AIDS viruses. *Cell* **40**:1–4.
- Coffin, J. M., P. N. Tsichlis, C. S. Barker, and S. Voynow. 1980. Variation in avian retrovirus genomes. *Ann. N.Y. Acad. Sci.* **354**:410–425.
- Dorner, A. J., and J. M. Coffin. 1986. Determinants for receptor interaction and cell killing on the avian retrovirus glycoprotein gp85. *Cell* **45**:365–374.
- Dougherty, J. P., and H. M. Temin. 1986. High mutation rate of a spleen necrosis virus-based retrovirus vector. *Mol. Cell. Biol.* **6**:4387–4395.
- Durbin, R. K., and V. Stollar. 1986. Sequence analysis of the E2 gene of a hyperglycosylated, host restricted mutant of Sindbis virus and estimation of mutation rate from frequency of revertants. *Virology* **154**:135–143.
- Fischer, S. G., and L. S. Lerman. 1979. Two-dimensional electrophoretic separation of restriction fragments of DNA. *Methods Enzymol.* **68**:183–191.
- Gojobori, T., and S. Yokoyama. 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and its cellular homologues. *Proc. Natl. Acad. Sci. USA* **82**:4198–4201.
- Gojobori, T., and S. Yokoyama. 1987. Molecular evolutionary rates of oncogenes. *J. Mol. Evol.* **26**:148–156.
- Gopinathan, K. P., L. A. Weymouth, T. A. Kunkel, and L. A. Loeb. 1979. Mutagenesis *in vitro* by DNA polymerase from an RNA tumor virus. *Nature (London)* **278**:857–859.
- Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks. 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**:1548–1553.
- Hayward, W. S. 1977. Size and genetic content of viral RNAs in avian oncovirus-infected cells. *J. Virol.* **24**:47–63.
- Hayward, W. S., S. B. Braverman, and S. M. Astrin. 1980. Transcriptional products and DNA structure of endogenous avian proviruses. *Cold Spring Harbor Symp. Quant. Biol.* **44**:1111–1121.
- Hanafusa, H. 1969. Rapid transformation of cells by Rous sarcoma virus. *Proc. Natl. Acad. Sci. USA* **63**:318–325.
- Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol. 1982. Rapid evolution of RNA genomes. *Science* **215**:1577–1585.
- Johnson, M. S., M. A. McClure, C.-F. Feng, J. Gray, and R. F. Doolittle. 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologues with nonviral enzymes. *Proc. Natl. Acad. Sci. USA* **83**:7648–7652.
- Jolly, D. J., R. C. Willis, and T. Friedmann. 1986. Variable stability of a selectable provirus after retroviral vector gene transfer into human cells. *Mol. Cell. Biol.* **6**:1141–1147.
- Katz, R. A., C. A. Omer, J. H. Weis, S. A. Mitsialis, A. J. Faras, and R. V. Guntaka. 1982. Restriction endonuclease and nucleotide and sequence analyses of molecularly cloned unintegrated avian tumor virus DNA: structure of large terminal repeats in circle junctions. *J. Virol.* **42**:346–351.
- Khoury, A. T., and H. Hanafusa. 1976. Synthesis and integration of viral DNA in chicken cells at different times after infection with various multiplicities of avian oncornavirus. *J. Virol.* **18**:383–400.
- Lerman, L. S., S. G. Fischer, I. Hurley, K. Silverstein, and N. Lumelsky. 1984. Sequence-determined DNA separation. *Annu. Rev. Biophys. Bioeng.* **13**:399–423.
- Linial, M., and D. Blair. 1982. Genetics of retroviruses, p. 649–783. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), *RNA tumor viruses*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7035–7056.
- Palese, P., and J. L. Schulman. 1976. Mapping of the influenza virus genome: identification of the hemagglutinin and the neuraminidase genes. *Proc. Natl. Acad. Sci. USA* **73**:2142–2146.
- Parvin, J. D., A. Moscona, W. T. Pan, J. M. Leider, and P. Palese. 1986. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *J. Virol.* **59**:377–383.
- Potter, H., L. Wier, and P. Leder. 1984. Enhancer-dependent expression of human κ immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Natl. Acad. Sci. USA* **81**:7161–7165.
- Salts, Y., and A. Rosen. 1971. Neighbor effects in the mutation of ochre triplets in the T4 rII gene. *Mutat. Res.* **13**:109–113.
- Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* **32**:853–869.
- Sedivy, J. M., J. P. Capone, U. L. RajBhandary, and P. A. Sharp. 1987. An inducible mammalian amber suppressor: propagation of a poliovirus mutant. *Cell* **50**:379–389.

31. **Smith, D. B., and S. C. Inglis.** 1987. The mutation rate and variability of eukaryotic viruses: an analytical review. *J. Gen. Virol.* **68**:2729–2740.
32. **Smith, F. I., J. D. Parvin, and P. Palese.** 1986. Detection of single base substitutions in influenza virus RNA molecules by denaturing gradient gel electrophoresis of RNA-RNA or DNA-RNA heteroduplexes. *Virology* **150**:55–64.
33. **Spindler, K. R., F. M. Horodyski, and J. J. Holland.** 1982. High multiplicities of infection favor rapid and random evolution of vesicular stomatitis virus. *Virology* **119**:96–108.
34. **Steinhauer, D. A., and J. J. Holland.** 1986. Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J. Virol.* **57**:219–228.
35. **Steinhauer, D. A., and J. J. Holland.** 1987. Rapid evolution of RNA viruses. *Annu. Rev. Microbiol.* **41**:409–433.
36. **Wang, S. Y., W. S. Hayward, and H. Hanafusa.** 1977. Genetic variation in the RNA transcripts of endogenous virus genes in uninfected chicken cells. *J. Virol.* **24**:64–73.
37. **Ward, C. D., M. A. M. Stokes, and J. B. Flanagan.** 1988. Direct measurement of the poliovirus RNA polymerase error frequency in vitro. *J. Virol.* **62**:558–562.
38. **Weiss, R. A., W. S. Mason, and P. K. Vogt.** 1973. Genetic recombinants and heterozygotes derived from endogenous and exogenous avian RNA tumor viruses. *Virology* **52**:535–552.