Determination of the Frequency of Retroviral Recombination between Two Identical Sequences within a Provirus

TING LI AND JIAYOU ZHANG*
Department of Microbiology and Immunology and Markey Cancer Center,
University of Kentucky, Lexington, Kentucky 40536-0096

Received 18 November 1999/Accepted 22 May 2000

Retroviruses use RNA as their genetic material within viral particles and DNA (provirus) as their genetic material within cells. The rate of recombination during reverse transcription between two identical sequences within the same RNA molecule is very high. In this study, we have developed a sensitive system to study recombination occurring within the proviral sequence. This system includes a murine Moloney leukemia virus vector which contains a neomycin resistance gene (neo) and two mutated green fluorescent protein genes (gfp) in tandem positions. The 3’ end of the first gfp and the 5’ end of the second gfp gene are both mutated, so that neither of these two gfp genes is functional. However, if recombination occurs between the two gfp genes it will create a functional gfp protein. Cells containing such a functional recombinant gfp appear green under fluorescence microscopy. The rate of recombination between the two gfp sequences during a single round of replication is as high as 51%. Green cells appear during proliferation of a clonal clear-cell population and allow a small portion of these recombinations between sequences of proviral DNA to be detected. The frequency of recombination at the proviral DNA level is about 10^{-2} events/cell division, which is very low compared with the frequency of recombination (51%) caused by reverse transcriptase and/or RNA polymerase II.

Retroviruses recombine at a high rate due to the presence of two identical genomic RNA molecules in their virions (5). Reverse-transcribed double-stranded viral DNA which has been integrated into the host chromosomal DNA is called a provirus. Once integrated, the provirus is replicated along with host cell DNA and is genetically transmitted as an integral element of the host genome. During this process, the rate of retroviral recombination within the provirus has been presumed to be very low. As yet, however, this rate has not been accurately measured.

Intrachromosomal recombination between two retroviral long terminal repeats (LTRs) has been described for the somatic and germ line reverse mutation of the retrovirus-induced dilute (d’) coat color mutation of DBA mice (3). The d’ mutation was generated by the spontaneous integration of a Moloney leukemia virus (MLV) into noncoding sequences of the dilute locus. Reversion of the mutation occurred by recombination between the LTRs, leaving a single LTR in each revertant chromosome. The intensely colored wild-type phenotype represented germ line reversion, while the mottled phenotype represented somatic reversion. More than one million mice were screened in 5 years to determine the reverse mutation rates as described above. The germ line reversion rate was 4.5 × 10^{-6} per gamete, and the frequency for detecting somatic reversion was 9 × 10^{-7} events per animal analyzed (17).

MLV provirus excision has also been reported in an in vitro assay (19). When MLV was studied as an insertional mutagen in a Rous sarcoma virus-transformed cell line, several reversion mutations were found. These reverse mutations resulted from various deletions of the MLV provirus, presumably by homologous recombination between the MLV LTRs.

Gene deletions cause human genetic diseases (10). Sequence-directed mutageneses may be mediated by direct repeats. The size of a deletion can vary from a single base to megabases of DNA. Studies of bacterial DNA show that the frequency of deletion is proportional to the sizes of the identical sequences and inversely proportional to the distance between the repeats (15, 18, 20). In mammalian cells, the frequencies of two intrachromosomal H2 gene conversion events range from 10^{-5} to 10^{-6} per DNA molecule analyzed by a PCR assay (8). Spontaneous intrachromosomal recombination rates in somatic cells have been reported as 10^{-5} to 10^{-3} per cell generation with a plasmid encoding two tandem mutated lacZ genes (7). In transgenic animals, the range of frequency of such recombination was determined to be 0.001 to 2% (6, 14). There are several models proposed to explain this deletion process (11). Among these models, some are replication dependent (replication slippage) (9) and some are not (strand exchange and single-strand annealing). It is known that multiple proteins are involved in eucaryotic recombination events.

To study recombination between two identical sequences within the same provirus, a bicistronic MLV-based vector (pJZ481) was constructed. From the 5’ to 3’ direction, pJZ481 carried two mutated copies of the color reporter gene (gfp) (4), an internal ribosome entry segment (IRES) sequence, and a drug resistance gene (neo) (Fig. 1A). The IRES sequence from encephalomyocarditis virus allows the ribosome to bind to an internal AUG and thereby to initiate the translation of the neo gene independently of the upstream sequences (1, 2). The 3’ and 5’ gfp sequences contain individual frameshift mutations which encode nonfunctional proteins. The frequency of the backward frameshift mutation of the gfp gene during a single round of retroviral replication was less than 10^{-5} (22).

Transfection alone caused a high frequency of deletion (or mutation) between the two identical sequences within the same plasmid DNA (21). To avoid deletion during transfection, JZ481 was introduced into the helper cell line PG13 (containing MLV Gag-Pol and Env of gibbon ape leukemia virus) (13) by infection as depicted in Fig. 2. The virus released from each individual PG13 clone, which contained JZ481 virus,
was used to infect D17 cells. The infected D17 cells were selected for the Neo’ phenotype, and visible colonies appeared 10 to 12 days after selection. The individual Neo’ colonies were examined under a fluorescence microscope. Neo’-selected cells that contained the parental provirus (i.e., that contained two mutated gfp sequences) were clear (Fig. 1A and 3A and B).

Recombination between two gfp genes led to the deletion of the mutations, resulting in a functional gfp gene (Fig. 1B). Cells containing the functional gfp gene were green under the microscope (16, 21) (Fig. 3C and D). Recombination rates are the ratio of the number of green colonies to the total number of colonies examined (green colonies and clear colonies). Results (Table 1) indicated that the rate of recombination between the two identical sequences within the same RNA molecule was very high (51% ± 2% per replication cycle). Since the rate of backward frameshift mutations of the gfp gene is lower than 10⁻⁸, most green cells are the result of a deletion. Genomic DNAs from clear (parental type) and green (recombinant) D17 cells were digested with NcoI to determine the nature of the recombinants. There are two NcoI sites within the JZ481 vector (Fig. 1). Parental proviruses should generate a 2.7-kb fragment, while recombinant proviruses with the deletion would create a 1.9-kb fragment. Southern analysis (Fig. 4A) showed that the green cells produced a distinct 1.9-kb fragment, indicating that most of the green cells contained recombinant proviruses, with the deletion occurring between the two gfp gene sequences.

Cells in a single Neo’ D17 colony were offspring of a single infected cell, so the integrated proviruses should be the same as those in the original infected cells. However, by using a fluorescence microscope, the phenotypes of D17 colonies can be divided into three groups (Fig. 3). The first group, which made up 51% of the colonies analyzed (Table 1), consisted completely of green cells that resulted from recombinations before proviral integration (Fig. 3C and D and 5). The second group represented colonies with only clear cells, which carried the parental-type proviruses (Fig. 3A and B). The third group was made up by several colonies that contained mostly clear cells but had a few green cells among them (Fig. 3E and F). The emergence of green cells in a clear colony implied that an event had occurred after the integration of the viral DNA into the host chromosomal DNA. Two possibilities contributed to the appearance of green cells among offspring from their clear-cell ancestor. First, a recombination may have occurred within the provirus during cell proliferation (Fig. 5). A second possibility is that viruses might be released from some infected D17 cells.

### TABLE 1. Microscopic analysis of D17 cells infected with JZ481

<table>
<thead>
<tr>
<th>Clone</th>
<th>No. of colonies</th>
<th>Rate of recombination (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Green</td>
<td>Clear</td>
</tr>
<tr>
<td>1-1</td>
<td>145</td>
<td>144</td>
</tr>
<tr>
<td>1-3</td>
<td>206</td>
<td>171</td>
</tr>
<tr>
<td>3-1</td>
<td>246</td>
<td>264</td>
</tr>
<tr>
<td>3-2</td>
<td>154</td>
<td>157</td>
</tr>
<tr>
<td>3-3</td>
<td>78</td>
<td>72</td>
</tr>
<tr>
<td>Total</td>
<td>829</td>
<td>808</td>
</tr>
</tbody>
</table>

* Each clone was an individual clear clone of PG13 cells containing the parental-type JZ481 provirus. The viruses from each clone were used to infect D17 cells. Infected cells were selected for the Neo’ phenotype. The Neo’ colonies were examined under a fluorescence microscope. The rate of recombination was determined by the ratio of the number of green colonies to the number of total colonies (green colonies and clear colonies). The mean recombination rate was 51% ± 2%.

### TABLE 2. Microscopic analysis of D17 cells infected with JZ481

<table>
<thead>
<tr>
<th>Clone</th>
<th>No. of cells</th>
<th>Frequency of recombination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 0</td>
<td>Day 2</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Green</td>
</tr>
<tr>
<td>1-1</td>
<td>$1.3 \times 10^5$</td>
<td>0</td>
</tr>
<tr>
<td>1-3</td>
<td>$1.4 \times 10^5$</td>
<td>0</td>
</tr>
<tr>
<td>3-1</td>
<td>$1.6 \times 10^5$</td>
<td>0</td>
</tr>
<tr>
<td>3-2</td>
<td>$1.6 \times 10^5$</td>
<td>0</td>
</tr>
</tbody>
</table>

* Neo’ D17 cells were sorted by flow cytometry for the isolation of clear cells. After sorting, the cells were counted and analyzed by fluorescence microscopy. Two days after sorting, the cells were analyzed again using a fluorescence microscope, and the green cells that appeared in these cell populations during these two days were counted. The frequency of proviral recombination is presented as the number of green cells per cell division. The number of cell divisions was estimated by using the difference between total number of cells on day 2 and total number of cells on day 0. The mean frequency was $0.85 \times 10^{-5}$. It is possible that the green cells appearing on day 2 are siblings. In other words, two green cells might have descended from one recombination event. Therefore, the actual frequency should be less than $0.85 \times 10^{-5}$ but larger than $0.29 \times 10^{-5}$.
cells and reinfect D17 cells, with a recombination occurring during reverse transcription (and/or RNA transcription by host RNA polymerase II). To rule out this possibility, a hyg gene expression vector was transfected into D17 cells, and infected cells were selected for Hyg<sup>r</sup>. Hyg<sup>r</sup> colonies were isolated after selection. Ten Hyg<sup>r</sup> D17 cells were mixed with 10 clear (or green) D17 cells infected with JZ481, which were Hyg<sup>s</sup> and Neo<sup>r</sup> (Fig. 6). Hyg<sup>r</sup> D17 cells without JZ481 provirus were cocultured with JZ481 provirus-containing Hyg<sup>s</sup> D17 cells for 15 days. Subsequently, the cells were divided into two portions and selected for Hyg<sup>r</sup> and Neo<sup>r</sup>, respectively. Sixty green cells were found in 8.2 × 10<sup>6</sup> Neo<sup>r</sup> cells, while no green cells were found in the Hyg<sup>r</sup> population 14 days after selection. If there had been virus produced by D17 cells infected with JZ481, it would have had the same opportunity to infect Hyg<sup>r</sup> cells, so that green Hyg<sup>r</sup> cells should have emerged. Furthermore, if

FIG. 3. Microscopic analysis of HCT 116 cells infected with viral vector JZ481 containing the gfp gene. (A) Visible-light micrograph of a neomycin-resistant colony containing parental JZ481 provirus. (B) Fluorescent micrograph of the same colony analyzed in panel A. (C) Visible-light micrograph of a neomycin-resistant colony which contains recombinant JZ481 provirus. (D) Fluorescent micrograph of the same colony analyzed in panel C. (E) Visible-light micrograph of a neomycin-resistant colony containing parental-type provirus. (F) Fluorescent micrograph of the same colony analyzed in panel E. Green cells in this colony are the result of a recombination within the proviral DNA.
green cells in a clear colony (Fig. 3E and F) were the result of reinfection, two proviruses should have integrated into the chromosomal DNA of green cells. One provirus should be the parental type introduced into the cell during the original round of infection, with the second provirus resulting from reinfection followed by a recombination between the two identical gfp sequences. Since retroviral integration is essentially a random process, the parental-type provirus and the recombinant provirus should integrate into different sites in the chromosomal DNA. Cells from a single clear colony with a few green cells (Fig. 3E and F) were serially diluted and plated on a petri dish to separate the green cells from the clear cells. Colonies formed 10 to 12 days after plating. Well-separated green and clear colonies were cloned, and their genomic DNAs were isolated and digested with EcoRI and hybridized with a neo probe. EcoRI cut at the 5′ end of the neo gene within the JZ481 provirus (Fig. 1) and at a site within the host cellular flanking sequence. Southern analysis indicated that only one provirus existed in each cell clone (Fig. 4B). Furthermore, different Neo⁰ colonies resulting from infection and selection contained proviruses integrated into different sites. In contrast, green cells and clear cells from a single colony had proviruses integrated into the same site of host chromosomal DNA, so that the green cells were subclones derived from the clear cells. They were the outcome of the same infection-integration event. To further determine the nature of the green cells within a clear colony, genomic DNAs of green cells and clear cells from the same D17 clear colony were also digested with NcoI. DNAs from clear cells formed a 2.7-kb fragment (Fig. 4A), while DNAs of the green cells from the same colony formed a distinct 1.9-kb band (Fig. 4A). Therefore, green cells in a Neo⁺ D17 clear colony resulted from a homologous recombination event between the two identical gfp sequences at the proviral DNA level (Fig. 5).

To accurately determine the frequency of recombination within the proviral DNA, JZ481-infected colonies were pooled after selection for Neo⁺. The clear cells were isolated by flow

**FIG. 4.** Southern analysis of genomic DNA isolated from D17 cells infected with JZ481. (A) Cellular DNA was cut with NcoI and hybridized with a neo gene probe. Green 1, green 2, clear 1, and clear 2 are individual green or clear colonies formed after infection. Clear 1-g1 and clear 1-g2 are green cells subcloned from the clear 1 colony. Clear 2-g1 and clear 2-g2 are green cells subcloned from the clear 2 colony. Molecular sizes are shown on the left. (B) Cellular DNA was cut with EcoRI and hybridized with a neo gene probe. Green 1, green 2, clear 1, clear 3, and clear 4 are individual green or clear colonies formed after infection. Clear 1-g1 and clear 1-g2 are green cells subcloned from the clear 1 colony. Clear 3-g1 and clear3-g2 are green cells subcloned from the clear 3 colony. Clear 4-g1 and clear4-g2 are green cells subcloned from the clear 4 colony.

**FIG. 5.** Schematic illustration of recombination that occurred at the proviral DNA level. JZ481 virus released from each individual PG13 clone was used to infect D17 cells. Colonies which consisted completely of green cells (gray) represent recombinants that resulted from events before proviral integration. Colonies which contained only clear cells (white) represent cells containing only the parental-type proviruses. Colonies which contained mostly clear cells but had a few green cells among them represent recombination that occurred within the provirus during cell proliferation.
cytometry based on the fluorescent of the green fluorescent protein. Two days after sorting, cells were examined by fluorescence microscopy, and the number of green cells in the population was determined. The number of cells in each petri dish was determined by using a hemacytometer. Out of $1.5 \times 10^6$ cells screened in this assay, only 12 green cells were detected. Therefore, we estimate that the frequency of recombination at the proviral level was about $0.85 \times 10^{-5}$ events/cell division (Table 2).

Since our system does not involve a replication-competent virus, a secondary mutational effect was limited (19). In addition, we used a low multiplicity of infection ($<1:1,000$) that allowed each infected cell to get only one virus (in contrast to an average of three in another study) (19), so the possibility of interchromosomal recombination was also ruled out.

Observations in this study support the notion that retroviral proviruses behave as any normal integrated part of the host chromosome. This retroviral system provides a means to introduce two sequences in tandem positions into the host cell's chromosomal DNA. The transduction of this DNA was stable and allowed for unambiguous determination of the structure and quantification of the integration. This system can provide a simple and fast assay for the study of mammalian homologous recombination.

We thank W. Bargmann, A. Kaplan, M. Purdom, and A. Simmons for helpful comments on the manuscript. We thank J. Strange for flow cytometry analysis.

This research was supported by Public Health Service research grant CA7040.

REFERENCES