sedimentation on glycerol gradients. Discrepancies in the extent of oligomerization reported by our laboratory (*16*) and by Perisic and co-workers (*6*) are most likely related to the greater tendency of bacterially expressed HSF to aggregate as hexamers and higher oligomers and to the anomalous migration of HSF when analyzed by nondenaturing polyacrylamide gradient gel electrophoresis.
12. S. K. Rabindran, J. Clos, R. I. Haroun, J. Wisniewski, unpublished observations.
13. B. K. Jakobsen and H. R. B. Pelham, *Mol. Cell. Biol.* **8**, 5040 (1988); C. Szent-Györgyi, D. B. Finkelstein, W. T. Garrard, *J. Mol. Biol.* **193**, 71 (1978); D. S. Gross and W. T. Garrard, *Annu. Rev. Biochem.* **57**, 159 (1988).
14. P. K. Sorger, *Cell* **62**, 793 (1990).
15. J. S. Larson, T. J. Schuetz, R. E. Kingston, *Nature* **335**, 372 (1988).
16. J. Clos *et al.*, *Cell* **63**, 1085 (1991).
17. S. K. Rabindran, G. Giorgi, J. Clos, C. Wu, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6906 (1991).
18. In addition to the cloning of two human HSF cDNAs [HSF1 (*17*) and HSF2 (*19*)], several HSF cDNAs have been cloned from tomato (*7*) and mouse [K. D. Sarge, V. Zimarino, K. Holm, C. Wu, R. I. Morimoto, *Genes Dev.* **5**, 1902 (1991)]. The COOH-terminal leucine zipper is conserved in the mouse and to some extent in the tomato HSFs. The necessity for multiple HSFs is unclear because yeast and *Drosophila* HSFs are apparently single-copy genes; it may reflect a cell or tissue restriction or a functional diversity related to the varied nature of the stress signals.
19. T. J. Schuetz, G. J. Gallo, L. Sheldon, P. Tempst, R. E. Kingston, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 6911 (1991).
20. The cDNA for human HSF1 (including the 5′ and 3′ untranslated regions) was subcloned into pCMV5 (a derivative of pCMV3), which has a multiple cloning site downstream of the CMV promoter-enhancer. Point mutations in the leucine zipper motif were introduced by site-directed mutagenesis. For constructing the deletion mutants, we subcloned the entire cDNA into pCMV6 (derived from pCMV5), which has stop codons downstream of the multiple cloning site. COOH-terminal nested deletions were obtained with the double-stranded nested deletion kit (Pharmacia). The end points of all deletion mutants and the sequence of the point mutation were confirmed by sequencing. For transfection experiments, plasmids were prepared by alkaline lysis and purified by polyethylene glycol precipitation and ethidium bromide–CsCl gradient centrifugation.
21. Human embryonic kidney 293 cells transformed with adenovirus [F. L. Graham and J. Smiley, *J. Gen. Virol.* **36**, 59 (1977)] were grown for 18 to 24 hours in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) in T-25 tissue culture flasks coated with poly-lysine (1×10^6 cells per flask). Cells were synchronized by a thymidine block (3.3 mM for 18 to 20 hours). Two hours after release from the block, cells were washed twice with opti-MEM (Gibco/BRL) and incubated with the lipofectin-DNA complex for 5 hours. We prepared the complexes by combining 10 μl of lipofectin (BRL) and 1 μg of DNA in a final volume of 1.2 ml. Transfections were ended with the addition of 1.2 ml of DMEM containing 20% FBS, and we allowed the cells to express transfected DNA for 12 to 14 hours. Before the cells were harvested, flasks were either kept at room temperature or heat-shocked at 44°C for 20 min. Cells were harvested by gentle scraping, and whole-cell extracts were prepared as described [V. Zimarino, C. Tsai, C. Wu, *Mol. Cell. Biol.* **10**, 752 (1990)].
22. In order to achieve optimal regulation of HSF1 in 293 cells, we adjusted the conditions for DNA transfection and expression. In general, the amount of protein expressed in transfected cells was limited by the use of small amounts of DNA (1 μg) and by short transfection and expression times. Longer transfection or expression times resulted in a high constitutive amount of DNA-binding activity in transfected cells that were not subjected to heat stress.
23. This finding is consistent with our observation that the double amino acid substitution in the fourth zipper alone does not lead to a completely constitutive DNA-binding activity.
24. The molecular size of HSF1 monomer and trimers are deceptively high when analyzed by gel-filtration chromatography, which only measures the Stokes radius. The positions of the globular marker proteins indicated in the figure serve as reference points and do not accurately estimate the native size of the HSF protein. The resolution of the microbore Superose 6 column (Pharmacia) was decreased because we applied a 25-μl sample volume in order to detect the expressed HSF1 protein.
25. Expression plasmids for *Drosophila* HSF were constructed by insertion of polymerase chain reaction (PCR) fragments containing the full-length or truncated ORF (the sequence 5′-AATTCAAA-3′ was added upstream of the ATG codon) into pRmHa-3 vector between the *Drosophila* metallothionein promoter-leader and the alcohol dehydrogenase gene 3′ end. We transfected CsCl-purified plasmids into SL2 cells using lipofectin according to the protocol of L. Sondergaard (personal communication). Briefly, about 3×10^6 cells in M3 medium supplemented with 10% heat-inactivated FBS were grown in T-25 flasks at 22°C for 24 hours. The medium was replaced with serum-free M3 containing 50 μl of lipofectin-DNA complex. We prepared complexes by combining 30 μl of lipofectin with 20 μl of diluted DNA (2 μg), and allowed the transfections to proceed for 24 hours. Transfections were stopped by the addition of FBS to a final concentration of 10%, and we allowed cells to recover for 24 hours before adding $CuSO_4$ to a final concentration of 0.7 mM. After 16 hours of induction, the cells were collected either directly or after 15 min of heat shock at 37°C, and whole-cell extracts were prepared as described [V. Zimarino and C. Wu, *Nature* **327**, 727 (1984)].
26. The amount of DNA-binding activity of mutant 1–585 is significantly higher than the activity of the full-length *Drosophila* HSF and is due to overexpression of the mutant protein in transfected cells. However, even when the expression was reduced, this protein was fully active under nonshock conditions.
27. The 12-residue element is not found in the sequence of *Drosophila* HSF. It is possible that another element may serve the same function.
28. The molecular chaperone Hsp70 has been suggested to act as a negative regulator of HSF trimer formation (*2, 3, 16*), and there is evidence for interactions between Hsp70 and HSF in vitro, although the biological significance of these interactions remains to be determined [K. Abravaya, M. P. Myers, S. P. Murphy, R. I. Morimoto, *Genes Dev.* **6**, 1153 (1992); R. Baler, W. J. Welch, R. Voellmy, *J. Cell Biol.* **117**, 1151 (1992); S. K. Rabindran *et al.*, in preparation]. However, the size of the HSF monomer in cell extracts as measured by gel filtration and sedimentation analysis is incompatible with a stable association between HSF and Hsp70 (*11*).
29. The induction temperature of hHSF1 is lowered to 37°C when hHSF1 is expressed in *Drosophila* cells (J. Clos, S. K. Rabindran, J. Wisniewski, C. Wu, in preparation).
30. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
31. We thank B. Howard (National Institutes of Health) for pCMV5, T. Bunch for pRmHa-3, W. Herr and L. Goldstein for gifts of expression vectors, T. J. Schuetz and R. E. Kingston for the HSF2 cDNA clone, L. Sondergaard for the SL2 cell transfection protocol with lipofectin, J. Eldridge for oligonucleotide synthesis, and C. Klee for suggestions. R.I.H. was supported by a Howard Hughes Medical Institute–NIH Research Scholarship.

20 July 1992; accepted 3 November 1992

Rate and Mechanism of Nonhomologous Recombination During a Single Cycle of Retroviral Replication

Jiayou Zhang and Howard M. Temin*

Oncogenes discovered in retroviruses such as Rous sarcoma virus were generated by transduction of cellular proto-oncogenes into the viral genome. Several different kinds of junctions between the viral and proto-oncogene sequences have been found in different viruses. A system of retrovirus vectors and a protocol that mimicked this transduction during a single cycle of retrovirus replication was developed. The transduction involved the formation of a chimeric viral-cellular RNA, strand switching of the reverse transcription growing point from an infectious retrovirus to the chimeric RNA, and often a subsequent deletion during the rest of viral DNA synthesis. A short region of sequence identity was frequently used for the strand switching. The rate of this process was about 0.1 to 1 percent of the rate of homologous retroviral recombination.

Highly oncogenic retroviruses have incorporated cellular proto-oncogene sequences between their long terminal repeats (LTRs). Most hypotheses for the origin of highly oncogenic retroviruses from cellular proto-oncogenes and replication-competent retroviruses propose an initial formation of a chimeric retrovirus–proto-oncogene RNA (*1, 2*). This chimeric RNA results either from transcription of DNA after a deletion that fuses 5′ viral sequences to cellular sequences or from readthrough transcription, which is often followed by abnormal splicing. An additional recombination step is then needed to add 3′ viral sequences and to form a highly oncogenic retrovirus. The 3′ viral–proto-oncogene junctions, when compared to the parental viral and proto-onco-

McArdle Laboratory for Cancer Research, University of Wisconsin–Madison, Madison, WI 53706.

*To whom correspondence should be addressed.

gene sequences, fall into three groups: those having no sequence identity, those having a short sequence identity, and those having an insertion (3). Nonhomologous retroviral recombination has been studied by transfection either of a truncated viral DNA, followed by infection with a replication-competent virus (4, 5), or of a replication-competent viral DNA with a selectable marker on its 3′ end (6). These systems allowed more than one cycle of replication and did not allow unambiguous determination of the structure and amount of the precursors to recombination (7).

To determine the rate and mechanism of nonhomologous recombination during a single round of retroviral replication, we developed a model system. Plasmid pJZ211 (Fig. 1) (8), a vector derived from spleen necrosis virus (SNV) and Moloney murine leukemia virus (MLV) (9, 10), contains a deletion in the U3 region of its 3′ SNV LTR and an Xho I restriction site linker in the deletion site. This vector also contains a truncated MLV vector between the two SNV LTRs, in the opposite transcriptional orientation to the SNV LTRs. In this truncated MLV vector, the hygromycin resistance gene (*hyg*^{R}) is expressed from an MLV LTR, and a herpes simplex virus thymidine kinase (TK) termination sequence replaces the 3′ MLV LTR, which is completely deleted.

DNA of pJZ211 was transfected into the SNV C3A2 helper cell line (containing the SNV *gag-pol* and *env* genes) (11) (Fig. 1). The cells were selected for hygromycin resistance (Hyg^R), and the Hyg^R cells were pooled (12) and designated step 1 cells. We used virus from step 1 cells to infect the MLV (xenotropic) helper cell line PG13 (containing a gibbon ape leukemia virus envelope) (13, 14). Infected cells were selected for Hyg^R, and individual clones were isolated and designated step 2 cells (15). The titers of virus from step 1 cells on PG13 were low [about ten colony-forming units (CFU) per milliliter], as a result of the low susceptibility of mouse cells to infection by SNV (16).

To test whether any virus capable of forming Hyg^R colonies was produced by the step 2 cells, we used the supernatant medium (3 ml) from each step 2 cell clone to infect D17 cells (a dog osteosarcoma cell line) and the infected cells were selected for Hyg^R. No Hyg^R colonies were detected (17). Because of the U3 deletion in SNV, there is no transcription from the SNV 5′ LTR after one round of replication (9).

An MLV vector (pLN) (10) containing a *neo* gene was transfected into the amphotropic MLV PA317 helper cell line (Fig. 1) (18). The cells were selected for neomycin resistance (Neo^R), and the Neo^R cells were pooled (19). We used virus (3.2×10^4 CFU/ml) from the PA317 cells transfected with pLN to superinfect the step 2 cells containing JZ211, and the infected cells were selected for Neo^R. Two clones of Neo^R cells were isolated from each step 2 cell clone; these clones were designated step 3 cells. Each step 3 cell clone contained a single JZ211 provirus and a single LN provirus (20).

The supernatant medium of each step 3 clone was used to infect D17 cells (the target cells) (21) and the infected cells were selected separately for Hyg^R and for Neo^R. The resulting resistant cells were designated step 4 cells. Hyg^R colonies form only when nonhomologous recombination has occurred between the JZ211 and LN genomes so that the *hyg* gene is between two LTRs (Fig. 1, step 4). The target cells do not contain viral *gag-pol* and *env* gene products for retrovirus replication, and no progeny virus was released from them (22). Therefore, the vector virus had undergone only one cycle of replication. The Neo^R titers were about 10^5-fold larger than the Hyg^R titers, but all step 3 clones produced some virus capable of forming Hyg^R colonies (Table 1).

In order to evaluate the ratio of *hyg* RNA to *neo* RNA, we determined the amounts of *hyg* and *neo* RNAs in step 3 cells and virions using dot blot hybridization (23) (Table 2). In most of the pools of virions, the ratio of *hyg* RNA in virions to *hyg* RNA in cells (Table 2) was higher than the corresponding ratio for *neo* RNA (Table 2). Therefore, the packaging of viral RNA does not discriminate against viral RNA without U3 and R (terminal repeat) sequences at the 3′ end.

The relative ratios of *neo* RNA to *hyg* RNA in step 3 viruses were determined (Tables 1 and 2), the ratios of virions containing heterodimeric RNA to the virions resulting in a Neo^R colony were calculated (24), and the rate of nonhomologous recombination (overall rate of formation of Hyg^R colonies) during a single retroviral replication cycle was determined (24) (Table 1). The overall rate of nonhomologous recombination during a single retroviral replication cycle is about 5×10^{-5}.

The sequences of the junctions between JZ211 and LN in recombinant (Hyg^R) proviruses in step 4 cells were determined. DNA from 56 individual step 4 colonies was isolated. The DNA sequences at the *hyg*-MLV junctions were amplified with the polymerase chain reaction (PCR) with two primers: one primer was located upstream from the 3′ end of the *hyg* gene, and the other primer was located in the MLV U5 region (Fig. 1) (25). Fifty-two of the 56 genomic DNA samples were successfully amplified. The sequences of 31 of these amplified fragments were determined.

The junction sequences between JZ211 and LN can be divided into three groups. The largest group, designated the ppt-type (polypurine tract–type) because these recombinants contain sequences from LN up to just before the ppt sequence next to the 3′

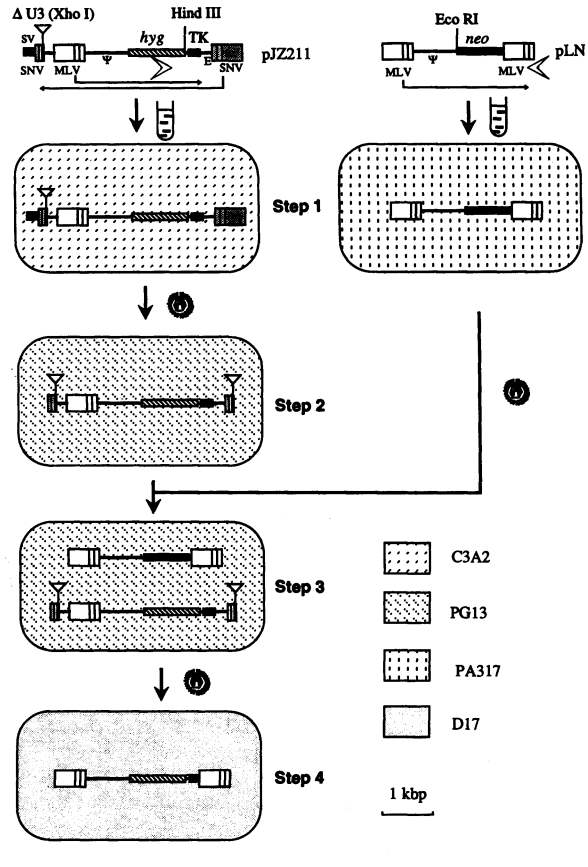


Fig. 1. Outline of an experimental approach for the determination of the rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. Plasmid backbone sequences are not shown. The transcriptional orientations of SNV and MLV are shown by the long, thin arrows. The large, open arrowhead under the *hyg* gene represents the location of the PCR primer Hyg 1658. The large arrowhead under the 3′ LTR of pLN represents the location of the PCR primer MLV U5 709. Transfections are indicated as a test tube, and infections are indicated as a virion. The different backgrounds represent the indicated cell lines. SV, late polyadenylation signal of simian virus 40; ψ and E, encapsidation sequences of MLV and SNV. The lines in the LTR separate U3, R, and U5 regions.

Table 1. Assay of step 3 clones on D17 cells. ND, not determined.

Clone	Hyg^R (CFU)	Neo^R (10^5 CFU)	Hyg^R/Neo^R (10^{-5})	*hyg/neo* (RNA)*	Rate†
2-1	1.0	1.4	0.7	0.3	1.5×10^{-5}
2-2	2.2	2.4	0.9	0.1	5.0×10^{-5}
4-1	48	12	4	0.4	7.0×10^{-5}
4-2	15	3.1	4.8	0.4	8.4×10^{-5}
5-1	9.2	1.9	4.8	0.8	5.3×10^{-5}
5-2	5.8	6.3	0.9	0.5	1.4×10^{-5}
6-1	4.2	1.4	3.0	ND	
6-2	5.4	0.57	9.5	ND	
				Average	4.8×10^{-5}

*Viral *neo* and *hyg* RNAs from step 3 cells (Table 2) were analyzed as described (*23*). †The rate of nonhomologous recombination was determined as described (*24*).

Table 2. Relative amounts of *hyg* and *neo* RNAs in step 3 cells and virions (*23*). V, virus; C, cell.

Clone	*hyg*			*neo*		
	V	C	V/C	V	C	V/C
2-1	5	5	1.0	18	11	1.7
2-2	8	4	2.0	89	63	1.4
4-1	19	9	2.1	45	27	1.6
4-2	7	8	0.9	19	35	0.6
5-1	23	14	1.7	29	20	1.5
5-2	98	20	4.9	208	48	4.3

LTR, comprised 17 of 31 step 4 clones analyzed. The junctions of ppt-type recombinants consist of the following: (i) all 112 bp from JZ211 after the *hyg* stop codon and a variable length (0 to 93) of polyadenylate [poly(A)] sequence, followed by the whole sequence of the LN 3′ LTR (Fig. 2A, ppt/U3); (ii) some part of the *hyg* translated region up to a run of purines, followed by the MLV LTR (purine sequence/U3); or (iii) some part of the *hyg* translated region and an inserted sequence (not present in LN or JZ211) ending with a purine-rich stretch, followed by the MLV LTR (Fig. 2A, insert).

The second group, designated the general-type, comprised 10 of 31 step 4 clones analyzed. The junction of general-type recombinants consists of different lengths of the 3′ end of JZ211 attached to different regions of the LN sequence upstream of the MLV 3′ LTR and all of the 3′ LTR. One of the ten general-type recombinants formed without the use of a short region of sequence identity (Fig. 2B, clone 7′). Six of ten general-type recombinants (Fig. 2B shows, as an example, clones 7 and 41) used a short region of sequence identity, 5 to 8 bp in length, between JZ211 downstream of *hyg* and in the 3′ end of the *neo* sequence in LN (*26*). Two pairs of different clones (7, 41 and 21, 29) used the same short regions of sequence identity for recombination, which indicates that there are hotspots for recombination (*27*). Three of ten general-type recombinants (Fig. 2B shows, as an example, clone 44) contained an insertion of 2, 6, and 30 bp between the JZ211 and LN sequences. The inserted sequences of 6 or 30 bp from the last two clones are not present in LN or JZ211.

The third group, designated the undefined-type, comprised 4 of 31 step 4 clones analyzed. The 5′ end of these recombinants consisted of different lengths of the 3′ end of JZ211 and a deleted LN 3′ LTR. Two of the four undefined-type recombinants contained a short sequence (2 to 5 bp) inserted between the JZ211 and LN sequences (Fig. 2C shows, as an example, clone 8′). One of the four undefined-type recombinants had 5 bp of sequence identity between JZ211 downstream of *hyg* and the LN 3′ LTR (Fig. 2C, clone 20′). Clone 32 (Fig. 2C) contained all 112 bp from JZ211 after the *hyg* stop codon and 221 bp downstream from the TK termination sequence in JZ211, followed by the 3′ LTR of LN with the first 16 bp deleted (Fig. 3B, undefined-type with insertion). This sequence indicated that clone 32 was a recombinant between LN and a readthrough transcript of JZ211. The position at the junction in the Hyg^R provirus sequence that diverges from the LN sequences is called the leaving site. The leaving sites are indicated for each clone in Fig. 2, A through C, and are summarized in Fig. 2D. The LN leaving sites cluster in three short regions in the 3′ half of the vector.

The entire sequence insertion in clone 44 (30 bp) is not in the GenBank–European Molecular Biology Laboratory libraries (*28*). However, the sequence insertion between JZ211 and LN in clones 10 (Fig. 2A) and 44 (Fig. 2B) contained a 14-bp identity (underlined in Fig. 2A). This 14-bp identity was found in the dihydrouridine arm of lysine and phenylalanine tRNA sequences. Clone 10 and 44 step 4 cells were infected with viruses from different step 3 cells (step 3 cell lines, 5-1 and 5-2). However, these two step 3 cell lines were developed from the same step 2 cells (step 2 cell line 5), which suggests that the 14-bp identity resulted from a readthrough transcript.

To form a recombinant between JZ211 and LN, the chimeric (JZ211 or Hyg^R) and viral (LN or Neo^R) RNAs must be copackaged. After infection, the growing point of synthesis of the DNA complementary to viral RNA, minus-strand DNA, jumps from the viral RNA to the chimeric RNA (Fig. 3A). The 5′ end of U3 is a hotspot for this jump, but there are other hotspots as well (Fig. 2D). When the growing point leaves from the 5′ end of U3 just before the ppt and jumps to the poly(A) sequence of the chimeric RNA, a ppt recombinant results. When the growing point leaves from the 5′ end of U3 just before the ppt and jumps to a purine-rich sequence [rather than the poly(A) sequence], a ppt recombinant with a purine sequence results. When the growing point leaves from the 5′ end of U3 just before the ppt and jumps to a readthrough transcript of the chimeric RNA, an insertion results. When the reverse transcription growing point transcribes the whole 3′ LTR, continues to the 3′ *neo* sequence in LN, and then jumps to a sequence downstream of *hyg* in the chimeric RNA, usually at a short stretch of sequence identity, a general-type recombinant results [Fig. 3A, general- (and ppt-) type]. Undefined-type recombinants are formed as the result of a deletion that occurs during plus-strand DNA synthesis (Fig. 3A, undefined-type).

The insertions between JZ211 and LN sequences were probably formed by readthrough transcription and deletion after strand switching (*29*) (Fig. 3B) by the following steps. (i) Readthrough transcription occurs to form a chimeric viral *hyg* RNA containing 3′ sequences downstream from the TK termination sequence. (ii) The reverse transcription growing point transcribes the 5′ end of U3 just before the ppt (ppt-type with insertion) or transcribes the whole 3′ LTR and continues to the 3′ *neo* sequence in LN (general-type with insertion). (iii) The growing point jumps to the readthrough transcript downstream of the *hyg* termination sequence. (iv) Reverse transcription continues, with a deletion in the nonviral sequences. The deletion probably occurs during plus-strand DNA synthesis (*3*). The sequence downstream of the TK termination sequence found in clone 32 and the identity of some of the inserted sequences in clones 10 and 44 demonstrate that the readthrough frequency was high enough to be detected in this system and that the source of the inserted sequence is readthrough followed by deletion, rather than a third template (*6*).

No ppt-type or undefined-type recombinants have been observed outside the laboratory. Formation of highly oncogenic retroviruses requires that the virus express the oncogene and replicate as an infectious virus (*30*). To test if proviruses resulting from the three types of nonhomologous recombination were capable of producing infectious

progeny, we cloned junctions of each of the three types of recombinants into a vector that consisted of two MLV LTRs derived from pLN and a *hyg* gene (pJZ206). The sequences between *hyg* and the 3′ end of pJZ206 were replaced by the junction sequences from the three types of nonhomologous recombination. Only vectors with 3′ sequences from general-type recombination produced infectious progeny efficiently (*31*). However, the recombinants occurring in nature are the product of long-term growth of viruses and may not directly reflect the original recombination events. Selection is operating on the natural events but not on the single-cycle viruses characterized here. Therefore, ppt-type or undefined-type events might occur, with subsequent recombination or deletion steps leading to the viruses observed in nature.

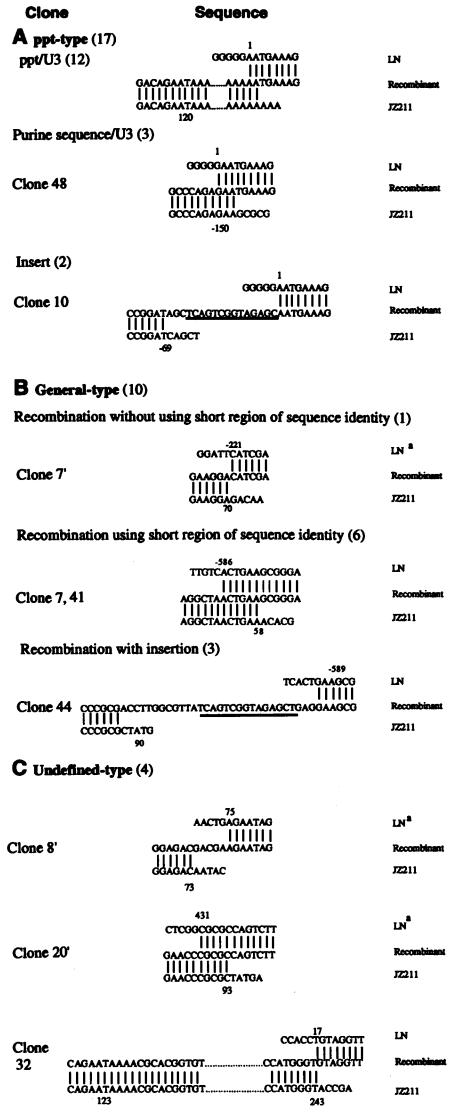
The overall rate of nonhomologous recombination is 5×10^{-5} per replication cycle (Table 1). This rate is similar to that measured in other experimental systems (*4*, *5*). Ten of 31 step 4 proviruses were general-type recombinants that were capable of producing infectious progeny. Thus, the rate of transduction during a single cycle of retrovirus replication to form an infectious *hyg* virus is 2×10^{-5} ($5 \times 10^{-5} \times 10/31$), or about 5×10^{-8} per base pair per cycle (*32*). The rate of homologous recombination is about 4×10^{-5} per base pair per cycle (*33*). Thus, nonhomologous recombination occurs at approximately 0.1 to 1% of the rate of homologous recombination.

The formation of the 3′ viral sequences of highly oncogenic retroviruses has been hypothesized to occur at the RNA level (*6*, *34*) or at the DNA level (*5*, *35*). We found three types of nonhomologous recombination: the ppt-type, the general-type, and the undefined-type. In some of the ppt-type recombinants, the recombination junction contains a poly(A) sequence. Therefore, the recombination probably occurred between RNAs during minus-strand DNA synthesis.

In undefined-type recombination, after recombination of the ppt- or general-type, a deletion occurs during plus-strand DNA synthesis (*29*). The rate of deletion is 2×10^{-6} per base pair per cycle (*29*). We found four out of 31 recombinants of this type, which indicates a higher than expected rate of deletion.

The leaving sites for general-type recombination from LN are clustered in three short regions in the 3′ end of the LN vector (Fig. 2D). The three short regions and the 5′ end of the LTR seem to represent hotspots for the reverse transcription growing point to leave its template. The clustering of recombination in the 3′ half of the LN vector may indicate that the two copies of viral RNA are packaged into a virion in parallel to form a dimer structure. Therefore, only when the reverse transcription growing point leaves the 3′ end of LN and docks on the 3′ end of JZ211 does it form a functional *hyg* gene.

The 3′ proto-oncogene–virus junctions in highly oncogenic retroviruses fall into three groups. The first group has no sequence identity—for example, the Abelson murine leukemia virus (*36*). The second group has a short region of sequence identity—for example, the 3611 murine sarcoma virus (*37*). And the third group has an insertion—for example, the Fujinami sarcoma virus (*38*). The relative frequencies of these three types in highly oncogenic retroviruses were about the same as found here for the three types of general-type recombinants (*39*).



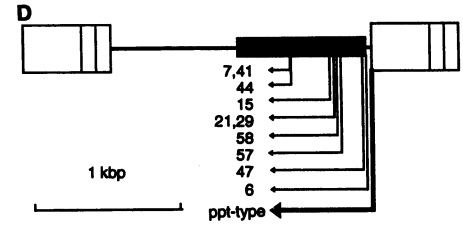


Fig. 2. Nucleotide sequence of ppt-type (**A**), general-type (**B**), and undefined-type (**C**) junctions and the leaving sites of general-type recombinants. The sequences in (A) through (C) were obtained by dideoxy nucleotide sequencing from clones amplified by PCR on step 4 cellular DNA with two primers. One primer was located near the 3′ end of the *hyg* gene, and the other primer was located in the MLV U5 region (Fig. 1). The upper sequences represent the sequence of LN, and the lower sequences represent the sequence of JZ211. The middle sequences represent the junction of the recombinants. The lines between the sequences indicate identity. The numbers on the LN and JZ211 sequences represent the relative locations of the junctions with respect to these two viruses. The first A at the 5′ end of the 3′ LTR is designated as nucleotide 1 in the LN vector, and the first G after the stop codon for the *hyg* gene is designated as nucleotide 1 in the JZ211 vector. The underlined sequences in clone 10 and 44 are identical. The dots in ppt/U3 and in clone 32 represent identical sequences that are not shown. Clones 7′, 8′, and 20′ resulted from recombination between JZ211 and an LN (LN^a) with a slightly different LTR. Numbers in parentheses represent the number of clones analyzed in each group. (**D**) Summary of leaving sites in general-type (and ppt-type) recombinants. The bases of the arrows represent the leaving sites from the LN vector; the numbers represent the clones. The exact site for some clones is indicated in (A) and (B).

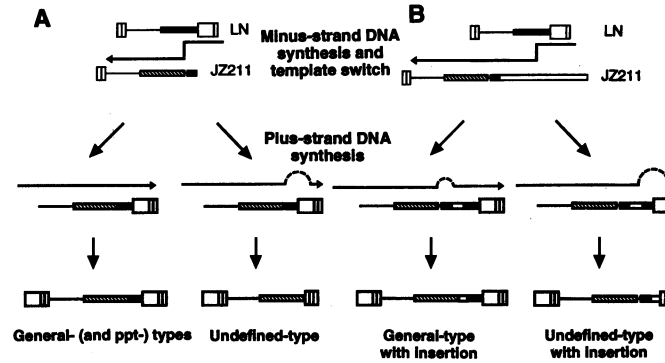


Fig. 3. Models of recombination between JZ211 and LN without (**A**) or with (**B**) insertions. The structures of JZ211 and LN are as described in Fig. 1. Thin lines represent RNA and thick lines represent DNA. The horizontal arrows represent minus- or plus-strand DNA syntheses. The empty box in (B) in JZ211 represents non–MLV sequences resulting from readthrough transcription.

REFERENCES AND NOTES

1. H. M. Temin, *Biochem. Int.* **18**, 1 (1989).
2. W.-S. Hu, V. K. Pathak, H. M. Temin, in *Reverse Transcriptase*, A. M. Skala and S. Goff, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, in press).
3. J. Zhang and H. M. Temin, *J. Virol.*, in press.
4. H. Stuhlmann, M. Kieckmann, P. Berg, *ibid.* **64**, 5783 (1990).
5. D. W. Goodrich and P. H. Duesberg, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 3604 (1990).
6. A. Swain and J. M. Coffin, *Science* **255**, 841 (1992).
7. J. P. Murnane, M. J. Yezzi, B. P. Young, *Nucleic Acids Res.* **18**, 2733 (1990).
8. Plasmids have small p's before their names (for example, pJZ211); viruses made from those plasmids do not (for example, JZ211).
9. R. Dornburg and H. M. Temin, *Mol. Cell. Biol.* **8**, 2328 (1988).
10. A. D. Miller and G. J. Rosman, *BioTechniques* 7, 980 (1989).
11. S. Watanabe and H. M. Temin, *Mol. Cell. Biol.* **3**, 2241 (1983).
12. We used pJZ211 DNA (10 μg) to transfect the C3A2 helper cells (2 × 10^5 cells per 60-mm plate). Transfection was performed by the dimethyl sulfoxide–polybrene procedure [S. Kawai and M. Nishizawa, *ibid.* **4**, 1172 (1984)]. Twenty-four hours after transfection, the medium was replaced with medium that contained hygromycin (0.08 mg/ml). About 30 colonies appeared 15 days after transfection. The cells were pooled and transferred to a 100-mm plate, and the supernatant medium was harvested when the cells became confluent.
13. A. D. Miller *et al.*, *J. Virol.* **65**, 2220 (1991).
14. PG13 cells are derived from NIH 3T3 cells. The virus released from PG13 cells does not infect NIH 3T3 cells (*13*). We determined virus titers as described (*21*) without freezing and thawing the virus solutions. To remove cells, we centrifuged the virus solutions for 10 min at 3000*g*, and the upper half of each solution was carefully removed and used for infections.
15. The structures of the proviruses formed from the U3$^-$ vector in the PG13 cells were monitored by Southern (DNA) analysis. The Xho I linker in pJZ211 is duplicated in the 5′ LTR during formation of the step 2 provirus. Cellular DNA of each step 2 cell clone was digested with Xho I. Six of eight step 2 cell clones analyzed contained the expected 3.1-kbp fragment, which hybridized to a *hyg* probe. These six clones were used for further study.
16. J. Embretson and H. M. Temin, *J. Virol.* **61**, 3454 (1987).
17. In an experiment with similar vector pRD18, the step 2 cells produced Hyg^R virus at a low rate because pRD18 contains two promoters for the *hyg* gene (*9*).
18. A. D. Miller and C. Buttimore, *Mol. Cell. Biol.* **6**, 2895 (1986).
19. We used pLN DNA (10 μg) to transfect the PA317 helper cells as described (*21*). About 50 colonies appeared 15 days after transfection. The cells were pooled and transferred to a 100-mm plate. When the cells became confluent, the supernatant medium was collected and frozen until used for infection.
20. DNA from each step 2 clone DNA was digested with Hind III. Hind III digests the JZ211 proviral DNA 3′ to the *hyg* sequence and digests the cellular flanking sequence 5′ to the JZ211 provirus (Fig. 1). The number of JZ211 provirus copies was determined by hybridization to a *hyg* probe. All of the step 2 cells contained only one copy of the JZ211 provirus, and these proviruses were integrated into different sites in the step 2 cellular genomes (*22*). We used Eco RI to analyze the step 3 cell clones. Eco RI digests the LN provirus 5′ to the *neo* sequence and digests the cellular flanking sequence 3′ to the LN provirus (Fig. 1). The number of LN provirus copies was determined by hybridization to a *neo* probe. All of the step 3 cell clones contained only a single copy of the LN provirus, and these proviruses were integrated into different sites in the step 3 cells (*22*).
21. J. P. Dougherty and H. M. Temin, *Mol. Cell. Biol.* **6**, 4387 (1986).
22. J. Zhang and H. M. Temin, unpublished data.
23. Viral RNA and cellular RNA were isolated, and dot blot analyses were performed on duplicate filters. One filter was hybridized with a *hyg* probe, and the other filter was hybridized with a *neo* probe. To normalize different hybridization of the *neo* and *hyg* probes, we used the same amount of hybridization buffer containing both *hyg* and *neo* probes to hybridize to the same amount of *hyg* and *neo* DNA (0.2 μg). We used a densitometer to measure the relative amounts of JZ211 (*hyg*) and LN (*neo*) RNAs and the standard. The ratio of *hyg* RNA in virions to *hyg* RNA in the cells was calculated for each clone (Table 2).
24. The relation between *hyg* and *neo* RNA frequencies in step 3 virions and the frequencies of *hyg* and *neo* heterodimers in step 3 virions can be described in algebraic terms by means of the Hardy-Weinberg equation as follows. Let h be the frequency of *hyg* RNA in step 3 virions and n be the frequency of *neo* RNA in step 3 virions, so that $h + n = 1$. Assuming random packaging, the frequencies of different *hyg* and *neo* dimers are given by

$$h^2 + 2hn + n^2 = 1$$

where h^2 is the frequency of virions containing two *hyg* RNAs, $2hn$ is the frequency of virions containing one *hyg* RNA and one *neo* RNA (F_{hn}), and n^2 is the frequency of virions containing two *neo* RNAs. Neo^R colonies result from infection by virions containing two *neo* RNAs and one-half of the virions containing one *neo* and one *hyg* RNA. Therefore, the frequency of virions capable of forming Neo^R colonies (F_n) = $n^2 + hn$. The rate of nonhomologous recombination is the ratio of the Hyg^R titer (T_h) to the frequency of virions with heterodimer RNA (F_{hn}), as compared with the ratio of the Neo^R titer (T_n) to the frequency of virions with *neo* RNA (F_n). Therefore,

$$\text{ate} = \frac{T_h}{F_{hn}} \cdot \frac{F_n}{T_n} = \frac{T_h}{T_n} \cdot \frac{n^2 + hn}{2hn} = \frac{T_h}{T_n} \cdot \left(\frac{1}{2} \cdot \frac{n}{h} + \frac{1}{2}\right)$$

where n/h is the reciprocal of *hyg/neo* RNA in step 3 virions (Table 1).
25. PCR was performed as described [J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989)]. Primer Hyg 1685 (GCTTGTATGGAGCAGCAGACG) is located 351 bases upstream from the stop codon of the *hyg* gene, and primer MLV U5 709 (GTCAATCACTCAGAGGAGACC) is located 30 bases downstream from the 3′ end of the R region of LN (Fig. 1). The sizes of amplified fragments ranged from 400 bp (clone 25, undefined type) to 1.6 kb (clone 44, Fig. 2B). The MLV LTR is about 600 bp.
26. About 200 perfect matches of five nucleotides are found between the 112-bp sequence of JZ211 [after the stop codon of the *hyg* gene and before the poly(A) sequence] and 1.8 kb of LN (the sequence between the two LTRs); two perfect matches of eight nucleotides are found between these two sequences.
27. To determine if these different clones with the same sequence were the result of contamination during PCR, we performed Southern analysis with a *hyg* probe. Cellular DNAs were digested with Hind III. Hind III digests the recombinant proviral DNA 3′ to the *hyg* sequence and digests the cellular flanking sequence 5′ to the provirus. The Hind III *hyg* fragments in clones 7 and 41 were different in size, which indicates that these two clones contained proviruses integrated in different sites. In addition, the cellular DNAs of clones 7 and 41 were digested with Eco RI and Eco RV. The junction fragment was about 1.6 kb in both cases. Therefore, clones 7 and 14 were independent (*22*).
28. The FASTA program (Genetics Computer Group) was used. W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444 (1988).
29. V. K. Pathak and H. M. Temin, *ibid.* **87**, 6024 (1990).
30. H. M. Temin and C. K. Miller, *Cancer Surv.* **3** (no. 2), 229 (1984).
31. We used the resulting pJZ206-derived plasmids to transfect the PG13 helper cell line. The cells were selected for Hyg^R, and the Hyg^R cells were pooled. We used the supernatant medium of each cell clone to infect D17 cells, the infected cells were selected for Hyg^R, and the titer of virus from each cell clone was determined. Retrovirus with 3′ sequences from general-type recombination (pJZ206/PCR58) produced infectious progeny at approximately the same titer (1.4 × 10^4 CFU/ml) as JZ206 (3.6 × 10^4 CFU/ml). Vectors with 3′ sequences from ppt-type recombination (pJZ206/PCR23) produced infectious progeny at a titer (1.5 × 10^2 CFU/ml) 1% of that of JZ206, probably because a ppt-type recombinant contains an additional poly(A) addition sequence in its viral DNA [S. Shimotohno and H. M. Temin, *Cell* **26**, 67 (1981)]. Only viral RNAs that read through the internal poly(A) signal would be capable of forming a provirus. In addition, poly(A) may not be as good a substrate for ribonuclease (RNase) H cleavage as a normal ppt [A. J. Rattray and J. J. Champoux, *J. Mol. Biol.* **208**, 445 (1989)]. The Hyg^R viruses with sequences from the undefined group (pJZ206/PCR25) were not capable of producing infectious progeny (0 CFU/ml) because these recombinants did not have the 5′ U3 sequences required for viral replication.
32. The rate of nonhomologous recombination depends on the following: (i) the length of sequence within which the reverse transcription growing point is able to leave its template, (ii) the length of sequence on a second template into which the reverse transcription growing point is able to switch, (iii) the relative positions of the leaving and docking sites, and (iv) the ability of the resulting viral DNA to form a provirus. In this system, the length of the potential docking sequence is the maximum sequence that can be deleted at the 3′ end of the *hyg* gene (after an insertion occurred) that still allowed the Hyg^R phenotype to be selected, plus the sequence between the stop codon of the *hyg* gene and the poly(A) (112 bp). We isolated a *hyg* gene with a 76-bp deletion in its 3′ end in a Hyg^R cell clone (Fig. 2A, clone 48). Therefore, the potential docking sequence is at least 200 bp (76 + 112 = 188). From Fig. 2D, the potential leaving sequence is about 600 bp. We define the sequence available for nonhomologous recombination as the average of the potential leaving and docking sequences. Thus, the rate is the frequency divided by 400 bp [the average of 600 bp (the potential leaving sequence) and 200 bp (the potential docking sequence)]. The rate of general-type nonhomologous recombination per base per cycle is thus 5 × 10^{-8} (2 × 10^{-5}/400).
33. W.-S. Hu and H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1556 (1990).
34. J. M. Coffin, *J. Gen. Virol.* **42**, 1 (1979).
35. D. W. Goodrich and P. H. Duesberg, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3733 (1988); R. P. Junghans, L. R. Boone, A. M. Skalka, *Cell* **30**, 53 (1982); P. Olson, H. M. Temin, R. Dornburg, *J. Virol.* **66**, 1336 (1992).
36. J. W. J. Wang *et al.*, *Cell* **36**, 349 (1984).
37. U. R. Rapp, J. L. Cleveland, T. I. Bonner, S. M. Storm, in *The Oncogene Handbook*, E. P. Reddy, A. M. Skalda, T. Curran, Eds. (Elsevier, New York, 1988), pp. 213–253.
38. C. C. Huang, C. Hammond, J. M. Bishop, *J. Mol. Biol.* **181**, 175 (1985).
39. We have collected sequences of 29 acutely oncogenic retroviruses. Among the 29, 1 is without a short region of sequence identity, 15 used a short region of sequence identity, and 13 have insertions (*3*). In our general-type recombinants, one was without a short region of sequence identity, six used a short region of sequence identity, and three have insertions (Fig. 2B).

40. We thank J. Schoening, A. Amendt, and B. Pietz for technical assistance and K. Boris-Lawrie, J. Jones, L. Mansky, T. Mitchell, A. Panganiban, G. Pulsinelli, B. Sugden, and S. Yang for helpful comments on the manuscript. Supported by U.S. Public Health Service research grants CA-22443 and CA-07175 from the National Cancer Institute. J.Z. is a fellow of the Leukemia Society of America. H.M.T. is an American Cancer Society research professor.

15 June 1992; accepted 15 September 1992