Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 8553–8557, December 1987
Genetics

High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*

(reverse transcription/retrotransposons/yeast)

HUA XU AND JEF D. BOEKE

Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Communicated by Hamilton O. Smith, August 13, 1987

ABSTRACT **By following the fates of genetically marked Ty elements, we observed a very high frequency (80–90%) of deletion between directly repeated marker sequences during transposition. From blot hybridization analyses of Ty RNA and DNA species found in the Ty virus-like particles, we determined that the deletion events occurred during or immediately after reverse transcription of Ty RNA but before integration of Ty DNA. The results suggest that the Ty reverse-transcription machinery can recognize homologous sequences in the template. This capacity may be utilized in the replication and recombination processes of retrotransposons and retroviruses.**

Ty elements are a family of transposons found in the yeast *Saccharomyces cerevisiae*. They are about 6 kilobases (kb) in length and consist of a central coding region flanked by long terminal direct-repeat (LTR) sequences of 335 base pairs (bp). Ty elements transpose by a mechanism similar to retrovirus reverse transcription and integration (1). Virus-like particles (VLPs) have been found in yeast cells that overproduce Ty transcripts and consequently undergo many transposition events (2, 3).

Retroviruses undergo recombination at an unusually high frequency (4, 5). Genetic evidence suggests that the recombination process is an intermolecular event requiring the formation of heterozygous virions containing two kinds of parental RNA molecules in the same particle. A similar process may occur in the case of Ty. The yeast *SPT3* gene product is required for the transcription of full-length Ty transcripts. In *spt3* mutants only a small amount of shorter Ty transcripts is seen (6). Transposition of genomic Ty elements is abolished in *spt3* mutants, as expected, since Ty RNA is an intermediate in transposition. Ty transcription and therefore transposition can be reestablished by introduction of pGTyH3 plasmids, in which transcription of TyH3 under the control of the *GAL1* promoter (a galactose-inducible promoter), into *spt3* cells followed by galactose induction. However, the transposition events recovered in the *spt3* strains differ from those recovered in the *SPT3* strains, in that transposition in *SPT3* strains results in a high frequency of sequence changes within newly transposed Ty elements. Such sequence changes are rare or absent in the *spt3* mutants (1, 7). These data suggest that transcripts from both the genomic Ty elements and the plasmid-borne Ty are required for such "transpositional recombination" between Ty elements to occur. Several models have been proposed for retrovirus recombination; these include template switching, or breakage–repair, and strand displacement and assimilation during reverse transcription (4, 5).

We originally began this work in order to determine the maximum length of the RNA molecules that can be packaged into Ty VLPs. Various marker genes were inserted into the TyH3 element, whose transcription is under the control of the *GAL1* promoter. After galactose induction, the ability of these marked Ty elements to transpose was checked by blot hybridization analyses of the restriction endonuclease-digested genomic DNA (8) and phenotypic assays of the cells. We found that up to 1.7 kb of non-Ty sequence can be inserted into the Ty element without a significant effect on transposition. However, when the Ty element was marked with two copies of the same gene in tandem orientation, only one copy of the marker was left after transposition. We determined the stage in the transposition process at which the deletion events occurred, and we propose a model for Ty recombination and replication on the basis of these observations. A similar phenomenon, generation of deletions between direct repeats during viral replication, was recently reported in the case of spleen necrosis virus (9).

MATERIALS AND METHODS

Yeast Strains and Media. Yeast strains used in this work are listed in Table 1. Yeast media were as described (10, 11).

Transposition Assay. The yeast strains were grown selectively on SC-ura glucose medium. To induce transposition, the cells were streaked onto SC-ura galactose medium. After 5 days at 22°C, individual colonies were streaked separately onto SC-ura glucose plates to select for cells that retained the plasmid. These were replica-plated to YPD (rich) medium to allow the loss of the plasmid. After single-colony purification, genomic DNAs were made from plasmid-free cells (1). Southern hybridization analysis (8) of the genomic DNAs was then carried out. Following plasmid segregation, the phenotypes of cells in regard to the marker genes were also checked by growing the cells on rich medium containing G418 (for *neo* gene-marked Ty elements) or synthetic medium lacking tryptophan (for *TRP1* marked Ty elements).

Isolation of Ty VLPs. We used a slight modification (D. Eichinger and J.D.B., unpublished) of the procedure of Garfinkel *et al.* (2) to isolate Ty VLPs from galactose-induced cells.

RESULTS

Marker Sequences of Up to 1.7 kb Can Be Transposed Efficiently When Inserted into a Ty Element. It has been shown (1) that some short marker sequences such as the *lac* operator fragment (40 bp) can be inserted into the noncoding region of a Ty element without affecting its transposition. We were interested in determining the maximum length of the marker sequences that remain transposable when inserted into a Ty element—i.e., the "packaging limit" of Ty VLPs. We constructed a series of pGTyH3-derived plasmids by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: VLP, virus-like particle.

Table 1. Yeast strains

Strain	Genotype*	Plasmid†
JB516	1	pJEF1105 (= pGTyH3-neo)
JB567	1	pJEF1111 (= pGTyH3-neo-neo)
YH46	1	pX61 (= pGTyH3-TRP-neo-TRP)
YH23	2	pX3 (= pGTyH3-TRP)
YH24	2	pX4 (= pGTyH3-TRP)
YH40	2	pX56 (= pGTyH3-TRP-neo)
YH41	2	pX58 (= pGTyH3-TRP-neo)
YH42	2	pX59 (= pGTyH3-neo-TRP-neo)
YH43	2	pX60 (= pGTyH3-neo-TRP-neo)

*Genotype 1 is *a, ura3-52, his4-519, leu2-3/112, ade1-100, GAL*$^+$. Genotype 2 is *α, ura3-167, his3Δ200, leu2Δ1, trp1Δ1, GAL*$^+$.
†Detailed structures of the plasmids are given in Fig. 1.

marking the Ty element with (*i*) the bacterial *neo* gene (12), (*ii*) the yeast *TRP1* gene (13), and (*iii*) different combinations of the *TRP1* and *neo* genes (Fig. 1). The results of transposition assays showed that the *TRP1* (0.7-kb) and *neo* (1-kb) segments can be used to mark Ty elements without inhibiting transposition frequency significantly (J.D.B., H.X., and G. R. Fink, unpublished data). Moreover, a doubly marked Ty, containing both the *TRP1* and *neo* segments, also transposes at a near-normal frequency (Fig. 2 and Table 2). As will be shown below, occasionally a Ty carrying as much as 2.7 kb of insert can also transpose.

High-Frequency Excision Between Directly Repeated Genes During Ty Transposition. When two copies of the same gene in the same orientation were inserted into the *GAL1*/Ty construct as the marker sequences, in most cases only one copy remained after transposition had occurred (Fig. 3 and Table 2). For example, when two copies of the *neo* gene in tandem were inserted into TyH3 sequences, the apparent transposition frequency was comparable to that of Ty marked with a single *neo* gene. However, Southern analyses of the structures of the progeny marked Ty elements that had transposed into the genome showed that only one copy of the *neo* gene remained in these Ty elements; the second copy was lost during transposition (data not shown). Moreover, even when the two directly repeated *neo* genes were separated by the *TRP1* gene, this excision still occurred at a frequency as high as 90%, resulting in the loss of both the *TRP1* gene and one copy of the *neo* gene. Apparently this kind of deletion is not specific for the *neo* gene, since the Ty element marked with two copies of the slightly shorter *TRP1* gene interrupted by one *neo* gene behaves similarly. A small fraction of the progeny Ty elements in each case correspond to the parental configuration (i.e., *neo-TRP1-neo* or *TRP1-neo-TRP1*).

The Excision Events Occur Posttranscriptionally. There are

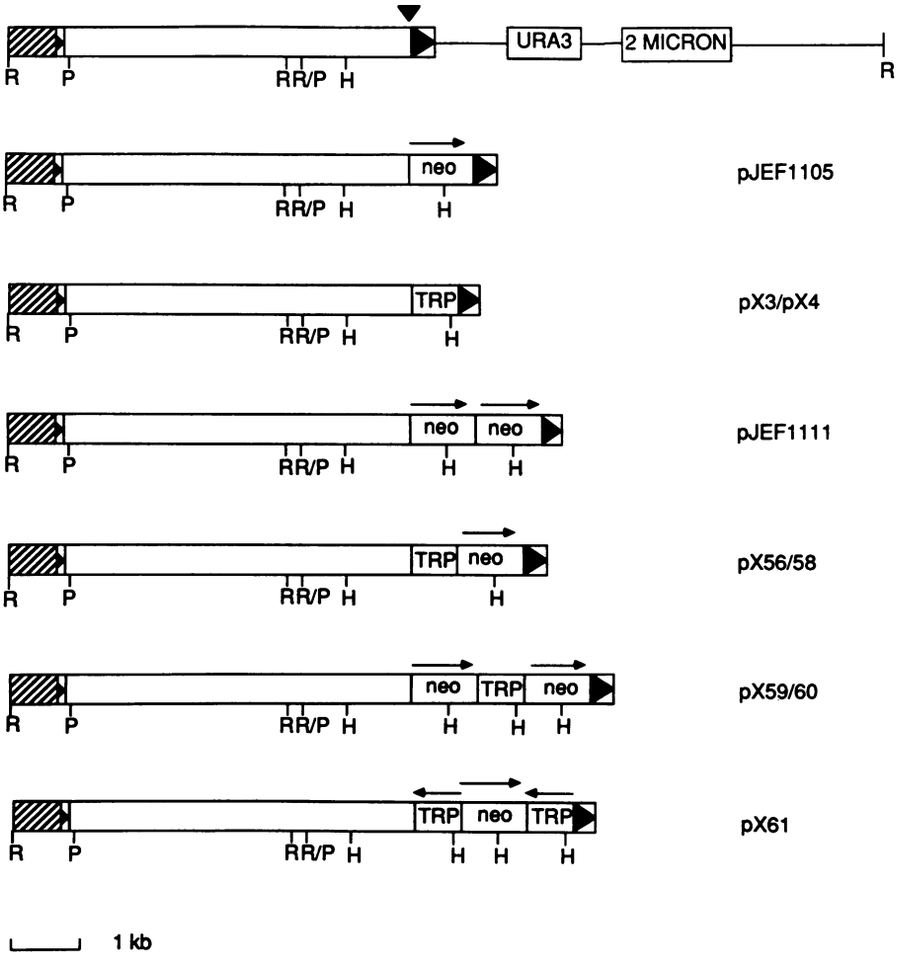

FIG. 1. Structures of pGTyH3 and its derivatives. TyH3 sequences are shown as an open box. Boxed triangle represents the long terminal repeats. Hatched box represents the *GAL1* promoter. *TRP* and *URA3* are the yeast *TRP1* and *URA3* genes, respectively; 2MICRON is a segment of the yeast 2-μm plasmid containing the replication origin; *neo* is the bacterial neomycin-resistance gene; pBR322 sequences are represented by straight lines. Arrows indicate direction of transcription. Plasmid pGTyH3 is shown in its entirety at the top of the figure. The filled triangle indicates the *Bgl* II site where markers are inserted. For the remaining pGTy plasmids, only the *GAL1*/Ty region is shown. Restriction sites are abbreviated as follows: H, *Hin*dIII; P, *Pvu* II; R, *Eco*RI.

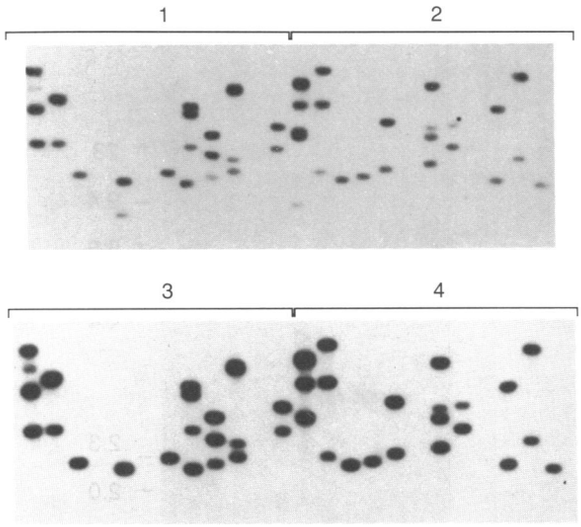

FIG. 2. Ty elements carrying both *TRP1* and *neo* can transpose. Southern blots of genomic DNA of derivatives of YH40 and YH41 that underwent transposition induction were probed with either the *neo*-specific (*Upper*) or the *TRP1*-specific (*Lower*) probe. DNA samples were cut with *Pvu* II, electrophoresed in 0.6% agarose, and blotted. Lane groups: 1 and 3, derivatives of YH40; 2 and 4, derivatives of YH41.

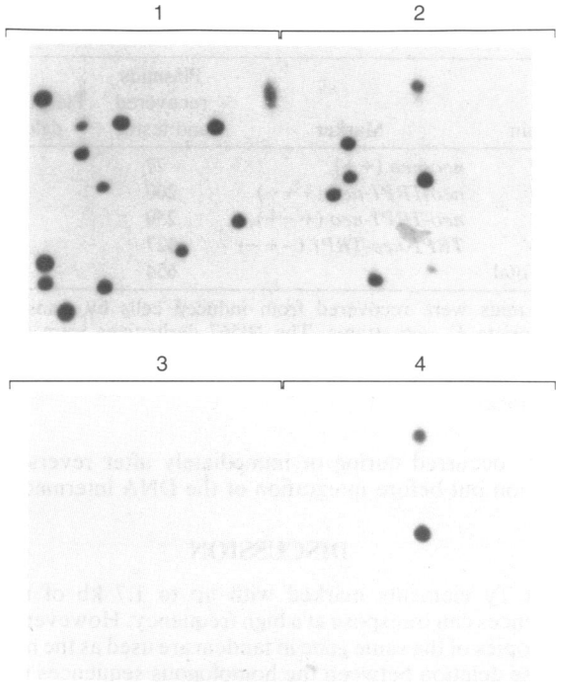

FIG. 3. Specific loss of *TRP1* gene when flanked by repeats of the *neo* gene. Southern blots of genomic DNA of derivatives of strains YH42 and YH43 that have undergone transposition were probed with either the *neo*-specific (*Upper*) or the *TRP1*-specific (*Lower*) probe. DNA preparation and hybridization were as for Fig. 2. Lane groups: 1 and 3, derivatives of YH42; 2 and 4, derivatives of YH43.

several possible stages during transposition where the excision events could occur. First, the marked pGTy (parental) plasmids might not be stable inside the cells, since *S. cerevisiae* is known to be very active at homologous recombination. To rule out this possibility, multiple independent plasmids were recovered from galactose-induced or uninduced yeast cells by transformation of *Escherichia coli* with total yeast DNAs. Analyses of these plasmids revealed that less than 0.25% of them were the deletion products of homologous recombination (Table 3). There are 30–50 copies of the pGTyH3-derived plasmids per cell. Therefore, most cells contain only the original plasmids introduced by transformation.

Other points at which the deletion may occur include transcription and reverse transcription. We isolated Ty VLPs from strains JB516 and JB567, which contain plasmids carrying Ty elements marked with a single copy and two copies of the *neo* gene, respectively, after galactose induction. Nucleic acids, including the putative RNA and DNA intermediates of Ty transposition, were isolated from Ty VLP preparations and subjected to hybridization analyses. The RNA molecules are full-length Ty transcripts. The DNAs found in the VLPs primarily consist of linear, full-length, double-stranded molecules (D. Eichinger and J.D.B., unpublished data). It is clear that RNA transcripts from the strain containing pGTy plasmids marked with tandem copies of the *neo* gene contain both copies; RNA molecules containing a single *neo* gene cannot be detected above background (Fig. 4). In contrast, the majority of the Ty DNA molecules in the same VLP preparation contain only one *neo* gene (Fig. 5). Densitometric quantitation of this autoradiograph revealed that there was 10-fold more Ty DNA marked with only a single copy of *neo* than of full-length Ty DNA (containing both copies of *neo*). These results suggest that the deletion

Table 2. Transposition of marked Ty elements

Strain	Marker in Ty*	No. of colonies analyzed	No. of G418^r colonies†	No. of Trp$^+$ colonies†	No. of bands hybridizing to *neo* probe‡	No. of bands hybridizing to *TRP* probe‡
YH23	*TRP1* (−)	24	NA	24	NA	52
YH24	*TRP1* (+)	10	NA	6	NA	25
JB516	*neo* (+)	18	15	NA	47	NA
JB567	*neo-neo* (++)	20	14	NA	46	NA
YH40	*TRP1-neo* (++)	12	11	0§	21	20
YH41	*TRP1-neo* (−+)	12	12	12	23	22
YH42	*neo-TRP1-neo* (+−+)	12	NT	NT	14	0
YH43	*neo-TRP1-neo* (+++)	12	NT	NT	6	2
YH46	*TRP1-neo-TRP1* (−+−)	24	NT	NT	4	24

NA, not applicable; NT, not tested.

*Orientation of the marker sequence with respect to Ty sequence is indicated in parentheses: +, transcription direction same as that of Ty; −, transcription direction opposite that of Ty.

†Number of colonies remaining G418-resistant or Trp$^+$ following segregation of the pGTyH3 plasmid.

‡Southern blot hybridization was carried out as for Figs. 2 and 3.

§The *TRP1* gene is not expressed well in this construct.

Table 3. Stability of tandemly repeated marker sequences in parental plasmids

Strain	Marker	Plasmids recovered and tested	Plasmids with deletions
JB567	*neo-neo* (++)	77	0
YH42	*neo-TRP1-neo* (+++)	200	0
YH43	*neo-TRP1-neo* (+−+)	250	2
YH46	*TRP1-neo-TRP1* (−+−)	127	0
Total		654	2

Plasmids were recovered from induced cells by transforming appropriate *E. coli* strains. The JB567 derivatives were checked directly by restriction digestion; the other strains were checked phenotypically for loss of kanamycin resistance or tryptophan prototrophy.

events occurred during or immediately after reverse transcription but before integration of the DNA intermediates.

DISCUSSION

Yeast Ty elements marked with up to 1.7 kb of non-Ty sequences can transpose at a high frequency. However, when two copies of the same gene in tandem are used as the marker, precise deletion between the homologous sequences usually occurs during the transposition process. The data of Rhode *et al.* (9), concerning similar results with spleen necrosis virus, agree with our data and suggest that the observed phenomena apply to many if not all retroviral systems.

We determined that these excision events happen during or after reverse transcription of Ty RNA, inside the VLPs. This recombination-like phenomenon is characterized by (*i*) its nonreciprocality, (*ii*) its extremely high frequency, (*iii*) its requirement for repeated sequences in the donor Ty, and (*iv*) its formation of precise deletions. Since the small circular product expected from an ordinary reciprocal recombination event is not found (Fig. 5), we believe that the deletion occurs during reverse transcription and that the deleted sequence is never converted to the double-stranded DNA form. Therefore, the excision seems to reflect the ability of the Ty reverse-transcription machinery to recognize homologous sequences in the RNA templates.

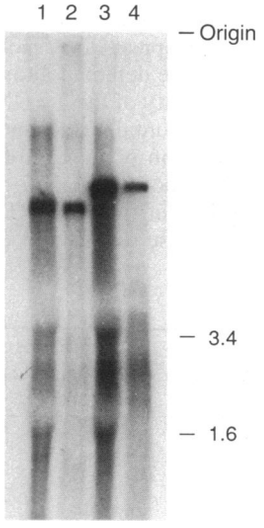

FIG. 4. Deletion of information does not occur in RNA. Strains JB516 and JB567 were grown on SC-ura raffinose medium overnight and then diluted into SC-ura galactose medium and allowed to grow for 24 hr. Nucleic acids were purified from VLPs or from total yeast-cell lysate by repeated phenol/chloroform extraction and ethanol precipitation. The RNA was fractioned in a 6% formaldehyde/1% agarose gel and transferred to nitrocellulose membrane. The blot was probed with the *neo*-specific probe. Lanes 1 and 3: 20 μg of total RNA from JB516 and JB567, respectively. Lanes 2 and 4: 2 μg of VLP RNA from JB516 and JB567, respectively. Molecular size markers run in parallel are indicated in kb.

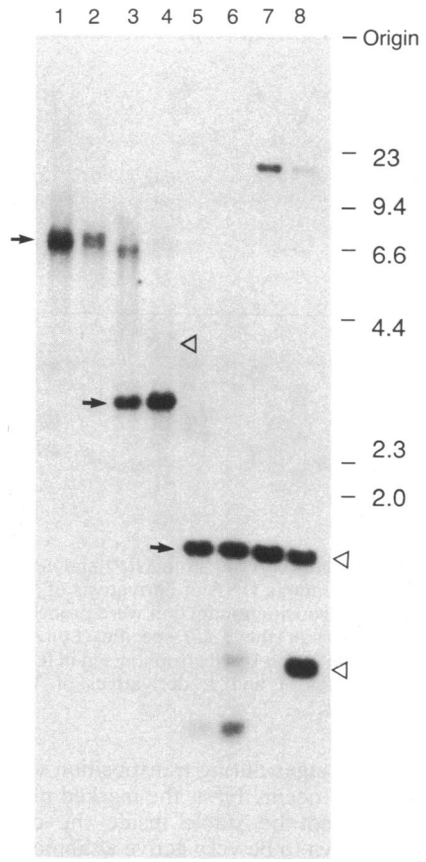

FIG. 5. Deleted form of DNA found within VLPs. Nucleic acids were purified as described in the legend to Fig. 4. The DNA was digested with restriction enzymes and subjected to Southern analysis. The probe used was specific for the *neo* gene. Plasmids contained in JB516 (pJEF1105) and JB567 (pJEF1111) were also blotted for comparison. Lanes 1, 3, and 5: JB516 DNA (uncut, *Eco*RI-digested, and *Hin*dIII-digested). Lanes 2, 4, and 6: JB567 DNA (uncut, *Eco*RI-cut, and *Hin*dIII-cut). Lane 7: pJEF1105 (*Hin*dIII-cut). Lane 8: pJEF1111 (*Hin*dIII-cut). Molecular size markers are indicated in kb. The larger band in lane 3 is not reproducible and probably represents a minor, circular form of Ty DNA. The 12-kb bands in lanes 7 and 8 are plasmid-derived. Arrows point to the bands expected for linear full-length Ty DNA marked with a single *neo* gene; the triangles indicate those expected for Ty marked with two copies of the *neo* gene.

In the accepted model for reverse transcription of retrovirus genomes, the (−)-strand strong stop DNA is thought to be transferred from one end of the template to the other by virtue of a directly repeated region, called R, in the long terminal repeat sequences (14). We propose a model (Fig. 6) in which the deletion events described above occur via a similar mechanism. During (−)-strand synthesis, the reverse-transcription machinery may constantly search for regions in the template(s) complementary to the nascent DNA strand. An opportunity for template switching or "jumping" between segments of template(s) occurs every time such a homologous region is found. The polymerase may traverse along the RNA template discontinuously in a series of "runs" and "pauses." During the runs, the reverse transcriptase polymerizes DNA continuously. During the pauses, the enzyme stops polymerizing and the nascent DNA chain is partially peeled from the template. The freed short DNA

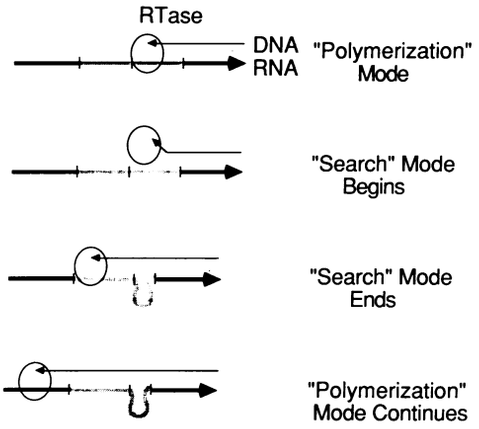

FIG. 6. Model for generation of precise deletions between direct repeats during (−)-strand reverse transcription. Reverse transcriptase (RTase; RNA-directed DNA polymerase) traverses the RNA template in a series of "runs" and "pauses." During the runs, the enzyme is in the "polymerization" mode and polymerizes DNA continuously. During the pauses, the enzyme is in the "search" mode and stops polymerizing. Part of the nascent DNA chain is peeled from the template and becomes available for base-pairing to complementary sequences, shown here as two copies of the same gene. If the nascent DNA hybridizes to the copy of the gene that is ahead of the pause site (i.e., the first copy of the gene), then deletion will occur. The deleted region forms a loop that is no longer available as template. Note that this model also allows intermolecular template switching to occur.

fragment base-pairs to the same region of the template or to another region of the same template or to a second template. In the first case, nothing would happen and the enzyme would continue to reverse transcribe the RNA. In the latter cases, template switching or "jumping" on the template(s) would occur. The net result of such processes might include (*i*) (−)-strand strong stop DNA transfers from the 5′ end of the template to the 3′ end if the homology used is the R region; (*ii*) "recombination" between two templates; and (*iii*) deletion between repeat sequences, an intramolecular reaction that may occur very frequently. This model could also explain why we always observe deletions between homologous marker sequences but very few if any duplications. When the reverse transcriptase "jumps" in the 5′-to-3′ direction (with respect to the nascent chain), the region between the homologous sequences would be looped-out, which could make it unavailable for further template switching. In this way, 3′-to-5′ jumping would be prevented, perhaps sterically, or perhaps because a free RNA end is required for scanning. Intermolecular jumping could give rise to duplication products, but if this happens, it does so at a much lower frequency undetectable in our assays.

Alternatively, size selection could explain the observed deletion events and the lack of duplications. The packaging limit of Ty VLPs may be such that although they can hold Ty transcripts with up to 2.7 kb of marker sequence, they will be too full to allow efficient reverse transcription to take place. Deletions between the repeated marker sequences by inter- or intramolecular recombination would release some of these constraints and thus be selected for. Naturally, duplications would not survive such constraints. Further work will be required to test the possible role of size selection.

Several models have been proposed for replicational recombination of retrovirus genomes. According to the break–repair model, template-switching occurs only when the polymerase encounters breaks in the RNA template (4). Since breaks in the template would presumably be randomly positioned, and template switching would occur between corresponding regions of the two virus genomes, it is hard to explain our results with this model because the deletion frequency is so high, yet the Ty element marked with two copies of the *neo* gene transposes just as well as the Ty marked with only one *neo* gene. In the displacement–assimilation model, recombination is initiated by the generation of a branch in the donor molecule via strand-displacement synthesis, followed by assimilation of the branch into a transient D-loop in a recipient molecule. The resolution of the "H structures" formed is presumed to take place after exposure to specific nucleases and ligase in the nuclei of infected cells. This model does not easily explain our data, since it does not accommodate intramolecular recombination; also, the final recombination products could be formed only after the intermediates get into the nucleus. Yet we found the deletion products in the VLPs, which are located mainly in the cytoplasm (2, 3). In our "search-and-jump" model (Fig. 6), recombination does not require breaks in the templates and can occur anywhere in the genome. It also allows both inter- and intramolecular recombination to take place. It seems reasonable that a common mechanism would be utilized to achieve replication and recombination in retroviral systems.

We thank D. Shortle for helpful discussions. This work was supported by National Institutes of Health Grant GM36481 and by a grant from the Searle Scholars Program/Chicago Community Trust to J.D.B.

1. Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) *Cell* **40**, 491–500.
2. Garfinkel, D. J., Boeke, J. D. & Fink, G. R. (1985) *Cell* **42**, 507–517.
3. Mellor, J., Malim, M. H., Gull, K., Tuite, M. F., McCready, S., Dibbayawan, T., Kingsman, S. M. & Kingsman, A. J. (1985) *Nature (London)* **318**, 583–586.
4. Coffin, J. M. (1979) *J. Gen. Virol.* **42**, 1–26.
5. Skalka, A. M., Boone, L., Junghans, R. & Luk, R. (1982) *J. Cell. Biochem.* **19**, 293–304.
6. Winston, F., Durbin, K. J. & Fink, G. R. (1984) *Cell* **39**, 675–682.
7. Boeke, J. D., Styles, C. A. & Fink, G. R. (1986) *Mol. Cell. Biol.* **6**, 3575–3581.
8. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
9. Rhode, B. W., Emerman, M. & Temin, H. M. (1987) *J. Virol.* **61**, 925–927.
10. Sherman, F., Fink, G. R. & Lawrence, C. W., eds. (1978) *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Rev. Ed.
11. Jimenez, A. & Davies, J. (1980) *Nature (London)* **287**, 869–871.
12. Oka, A., Sugisaki, H. & Takanami, M. (1981) *J. Mol. Biol.* **147**, 217–226.
13. Tschumper, G. & Carbon, J. (1980) *Gene* **10**, 157–166.
14. Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. (1979) *Cell* **18**, 93–100.