Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 1230–1234, February 1982
Genetics

Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, an avian retrovirus

(lethal mutations/base pair changes/clustered mutations/*gag* protein/nontandem duplication)

J. J. O'REAR* AND H. M. TEMIN

McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Contributed by Howard M. Temin, November 9, 1981

ABSTRACT **We determined the nucleotide sequence of about 1 kilobase of DNA 3′ to the 5′ long terminal repeat of three noninfectious and one infectious proviral DNA clones of spleen necrosis virus, an avian retrovirus, to determine if the types of nucleic acid changes involved in retrovirus mutation shed light on special features of retrovirus replication. An open reading frame was found starting 411 base pairs from the end of the long terminal repeat. It contained sequences coding for the 36 amino acids at the amino terminus of the p30 of a related reticuloendotheliosis virus [Oroszlan, S., Barbacid, M., Copeland, T., Aaronson, S. A. & Gilden, R. V. (1981) *J. Virol.* 39, 845–854]. Therefore, the open reading frame represents the 5′ end of the *gag* gene. A mutation in one noninfectious provirus changed the initiation codon for the *gag* polypeptide; a mutation in another noninfectious provirus caused premature termination of *gag* polypeptide synthesis; and a nontandem duplication into *gag* resulting from a mistake in initial (+) strand DNA synthesis changed amino acids and the reading frame in a third noninfectious provirus. These mutations appear to be responsible for the lack of infectivity of these provirus clones and indicate a higher relative frequency of mutation in this region of the genome. In addition, all four clones have multiple other mutations. These mutations are mostly base pair substitutions and many are clustered for any one clone, reflecting certain special features of retrovirus replication.**

Retrovirus replication is unique in that it involves RNA-to-DNA information transfer as well as DNA-to-DNA and DNA-to-RNA information transfers. There have been numerous suggestions that retrovirus replication is especially prone to errors (for example, refs. 1, 2, and 3). However, little is known about the types of nucleic acid changes involved in retrovirus mutation. Comparisons of related oligonucleotides obtained after RNase digestion of viral RNAs from different strains of virus have been published (for example, ref. 4).

We previously isolated 10 randomly selected proviruses of spleen necrosis virus (SNV), an avian retrovirus, by molecular cloning from productively infected chicken cells (5). Four of the 10 proviruses are not infectious in DNA transfection assays and 2 proviruses have polymorphisms in restriction enzyme cleavage sites. We showed by genetic mapping that the alterations leading to lack of infectivity in three of the proviruses are in the same 1-kilobase-pair (kbp) fragment of DNA but are not allelic (6).

Therefore, we decided to determine the sequence of this DNA fragment from several noninfectious proviruses to locate the alterations and determine their nature. We also analyzed some additional regions in SNV DNA to determine the nature of the alterations leading to loss of restriction enzyme cleavage sites. A number of mutations were found, mostly base pair substitutions. In addition, one nontandem duplication was found resulting from a mistake in initial (+) strand DNA synthesis. The alterations leading to lack of infectivity are mutation of the sequence for the AUG initiation codon for *gag*, a frameshift introducing premature termination of *gag* polypeptide synthesis, and amino acid changes and a frameshift in the *gag* gene caused by the duplication. In addition, all four clones have multiple other mutations. Many of the mutations are clustered in any one clone.

METHODS AND MATERIALS

Provirus clones 3-73, 13, 63, and 32 from chicken cells infected with an uncloned stock of SNV were described (5). All but 3-73 are from chronically infected cells. All but clone 32 are not infectious.

Subclones in pBR322 of the *Bam*HI fragment from 0.56 to 1.92 kbp from the 5′ end of the proviruses were made from all clones but 13 as described (6). Clone 13 had a new *Bam*HI cleavage site at 1.14 kbp and thus the same region of DNA was prepared as two subclones.

Further restriction enzyme cleavage sites were mapped in the cloned DNA and fragments were subjected to sequence determination, as indicated in Fig. 1, by the Maxam–Gilbert procedure (7).

RESULTS

Maps of SNV DNA are presented in Fig. 2. The region analyzed starts from a *Bam*HI cleavage site in U5 and ends at a *Bam*HI cleavage site at 1.9 kbp from the 5′ end of viral DNA. Most of the sequence was the same in all four clones. In the places where one or two clones were different, it always was possible to specify the wild-type sequence. It is presented in Fig. 3 from the end of the U3 region of the 5′ long terminal repeat (LTR). The first ATG at position 982 in the sequence starts an open reading frame and is assumed to code for the *gag* polypeptide (see *Discussion*). The proposed sequence for the polypeptide is presented underneath the DNA sequence in Fig. 3.

The mutations in the DNA sequences and the consequences for the amino acid sequence of the *gag* polypeptide are summarized in Table 1 and indicated in Fig. 3. Note that in each of the three noninfectious clones the base changes lead to amino acid substitutions (clone 63), a frameshift (clone 3-73), or amino acid substitutions and a frameshift (clone 13) in the *gag* polypeptide sequence. By contrast, the base changes in the infectious clones do not alter the amino acid sequence.

DISCUSSION

As its goal, the present work had understanding the molecular nature and possible origin of some genetic changes in SNV, an avian retrovirus. The cloned proviruses studied were noninfec-

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.

Abbreviations: SNV, spleen necrosis virus; kbp, kilobase pair(s); LTR, long terminal repeat.

* Present address: Kline Biology Tower, Yale University, New Haven, CT 06520.

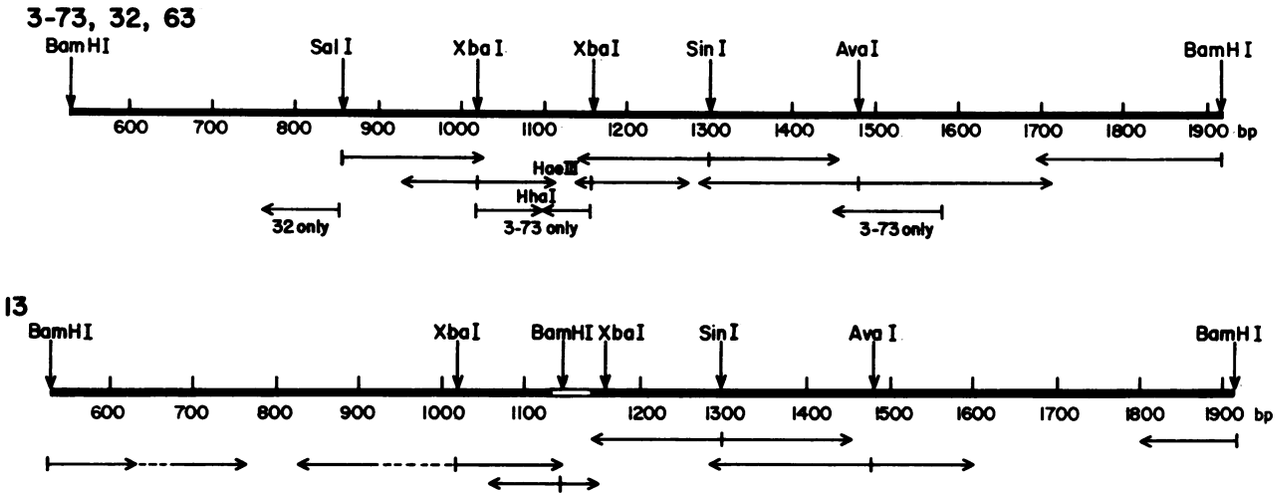

FIG. 1. Sequence determination strategies for *Bam*HI fragments from noninfectious and infectious SNV proviruses. Heavy bars indicate *Bam*HI subclones. Coordinates are from the previously published sequence of clone 14-44. Fragments were isolated and end labeled at the indicated restriction enzyme cleavage sites and, after further digestion, were analyzed by the Maxam–Gilbert technique (7). All chemical reactions were performed at least twice with each fragment except the two at the 5′ end of clone 13. Arrows indicate the direction and approximate extent of sequence determination. All restriction enzyme cleavage sites used for labeling, except *Sal* I, were subjected to sequence determination in a different reaction. (*Upper*) All reactions were done with clones 3-73, 32, and 63 except where indicated. (*Lower*) Reactions with clone 13 are indicated. Open bar indicates the duplication of nucleotides 525–578.

tious (clones 63, 3-73, and 13) or had lost restriction enzyme cleavage sites (clones 32 and 13). All these changes previously had been shown to be the result of genetic alterations in a 1-kbp region immediately 3′ to the 5′ LTR (see Fig. 2 for genetic organization of this region). We chose to investigate the mutations further by determining the sequence in this region from all four provirus clones.

***gag* Polypeptide.** The gene at the 5′ end of all retroviruses is called *gag* and codes for proteins in the interior of the virion. The sequence determination established an open reading frame of 939 nucleotides from the first ATG to the end of our sequence. It seems likely that this open reading frame defines the *gag* protein. Beginning 197 amino acids from the proposed initiator Met, we found a sequence of 37 amino acids almost the same as the sequence of 36 amino acids reported by Oroszlan *et al.* (10) as the NH_2 terminus of reticuloendotheliosis virus strain A p30. (p30 is the major protein specified by the *gag* gene.) The sequences differ only at position 11, Thr versus Gly; position 32, Gln versus X; and positions 35–37, Ser-Ser-Phe versus X-Phe. Both the sequences reported by us and by Oroszlan *et al.* (10) are close to the sequence published by Hunter *et al.* (11) but have several differences. We and Oroszlan *et al.* (10) used different members of the reticuloendotheliosis virus species; Hunter *et al.* (11) also determined the sequence of reticuloendotheliosis virus strain A p30.

The similarity of the sequences determined by us and by Oroszlan *et al.* (10) with different strains of reticuloendotheliosis virus supports the hypothesis that the p30 sequence is relatively conserved (for example, see ref. 10; also see below).

Lethal Mutations. On the basis of the identification of the open reading frame for the *gag* polypeptide, it is now possible to understand why three provirus clones are not infectious. In clone 63, a G-to-A transition changes the initiator AUG to AUA, thus preventing initiation of the *gag* polypeptide. (The next in-phase ATG is not until nucleotide 1523, although out-of-phase ATGs occur at nucleotides 1089, 1107, 1409, and 1458.) In clone 3-73, the insertion of a C leads to a frameshift and termination of synthesis of the *gag* polypeptide. In clone 13, the insertion of nucleotides 525–578 between nucleotides 1129 and 1139 leads to multiple amino acid changes and a frameshift in the *gag* polypeptide. Thus, no *gag* polypeptide can be synthesized.

Three of the four noninfectious proviruses previously cloned (5) are mutant within 600 bp of the NH_2 terminus-encoding region of the *gag* gene and prevent the initiation or cause premature termination of synthesis of the *gag* polypeptide. Thus, either mutations are more frequent in this region or mutations in this portion of the *gag* gene are lethal more often than mutations in other coding sequences. Because two of the lethal

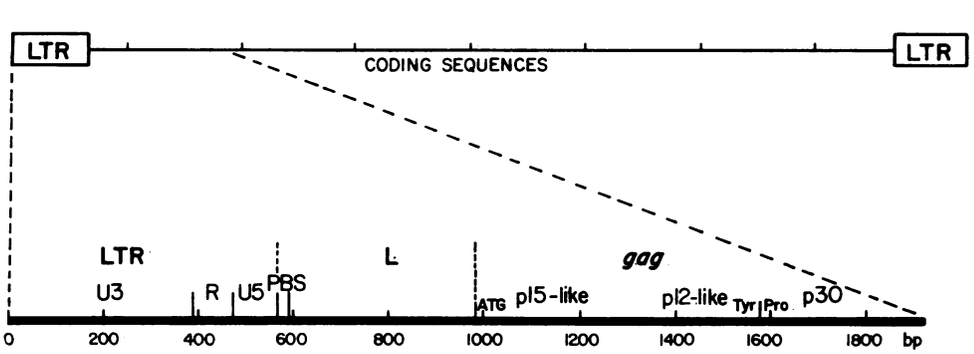

FIG. 2. SNV DNA. (*Upper*) Map of SNV DNA. The long terminal repeat (LTR) is represented by a box. Vertical lines mark 1-kbp intervals. (*Lower*) Detailed map of the 5′ region discussed in this paper. U3, U5, and R are regions that make up the LTR. The sequences of R are found at both ends of viral RNA, U3 is only at the 3′ end, and U5 is only at the 5′ end of viral RNA. PBS, primer binding site; L, region between PBS and the start of the *gag* gene; p15-like, p12-like, and p30, proteins encoded in the *gag* gene. The initiator ATG and the tyrosine/proline boundary between p12-like and p30 are shown.

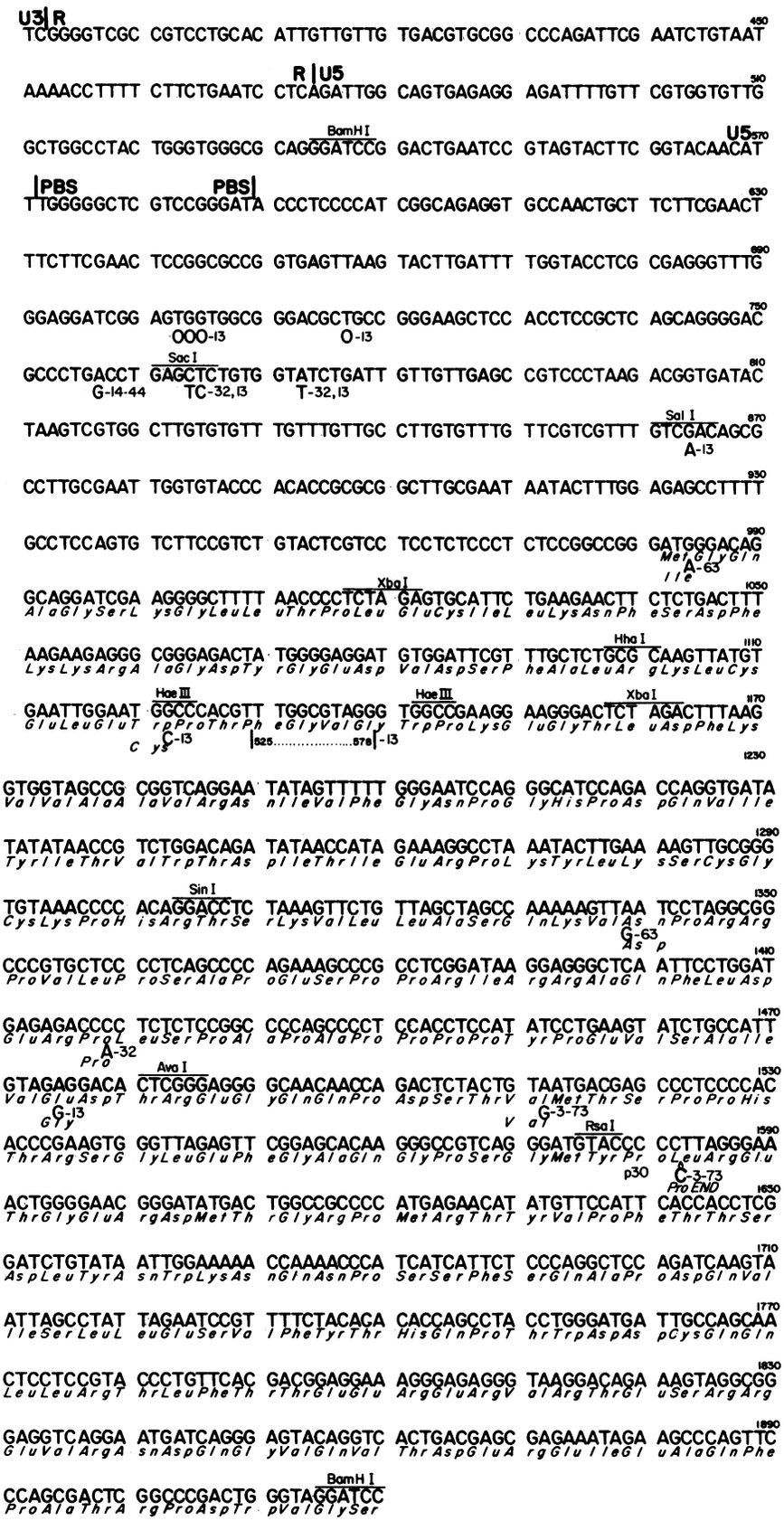

FIG. 3. Wild-type nucleotide sequence of the 5′ end of SNV from the end of the U3 region to 1.92 kbp. Base pairs 391–860 (with the exception of nucleotide 757) are from clone 14-44 (8); the rest and nucleotide 757 are from this paper. Restriction enzyme cleavage sites used or discussed in this paper are indicated. U3, R, U5, and PBS are regions in or next to the 5′ LTR. Mutant nucleotide sequences are listed below the wild-type sequence. O, deletion; ^, 1-bp insertion; number next to the mutant sequence, clone. The insertion in clone 13 is indicated by brackets. A theoretical translation of a *gag* polypeptide is presented under the wild-type nucleotide sequence starting at the first ATG (9). The p30 start, from amino acid data (10, 11), is shown. Proposed mutant amino acids are listed under the mutant nucleotide sequence.

Table 1. Mutations at 5′ end of SNV proviruses

Clone	Location, nucleotide	Mutation in nucleotides	Mutation in amino acids
63	984	G→A	Met→Ile
	1339	A→G	Asn→Asp
3–73	1512	A→G	None, both Val
	1582	Add C	Leu→Pro
			Arg→END
13	703	Delete T	Noncoding region
	704	Delete G	Noncoding region
	705	Delete G	Noncoding region
	717	Delete T	Noncoding region
	764	C→T	Noncoding region
	765	T→C	Noncoding region
	773	A→T	Noncoding region
	864	G→A	Noncoding region
	1122	G→C	Trp→Cys
	1129–1139	Substitution of 525–578	Many, and frameshift
	1475	A→G	Glu→Gly
32	764	C→T	Noncoding region
	765	T→C	Noncoding region
	773	A→T	Noncoding region
	1419	C→A	None, both Pro
14–44	757	G→A	Noncoding region

mutations are frameshifts, the first hypothesis appears more likely.

We do not think that these mutations are a cloning artifact because noninfectious proviruses were found in infected cells (12, 13), the nontandem duplication in clone 13 must have arisen during reverse transcription (see below), and other mutations in clone 13 and 32 are identical (Table 1).

Loss of Restriction Enzyme Cleavage Sites. A second goal of this sequence analysis was to determine the reason for the loss of the *Sal* I cleavage site in clone 13 and the *Sac* I cleavage sites in clones 13 and 32. In both cases, we found that transitions were responsible for the loss of the cleavage sites.

Furthermore, in the sequences of clones 32 and 13, an A was found at nucleotide 757, whereas a G was previously found in clone 14-44 (8). Clone 14-44 has a *Hae* III cleavage site at this position, but clones 60 and 70 do not (14). The G-to-A transition in clones 32 and 13 would destroy the *Hae* III cleavage site. Because A is present in four of the five proviruses, it appears to be the wild-type nucleotide, and clone 14-44 has a mutation.

The similarity of the nucleotide sequences around the missing *Sac* I cleavage site in clones 13 and 32 indicates they probably did not arise independently (see below).

Because single base pair substitutions led to the loss of these restriction enzyme cleavage sites, these results indicate that restriction enzyme cleavage site polymorphism is a reasonable measure of base-pair changes. This assumption had previously been supported by comparison of the results of restriction enzyme digestions and the results of electron microscopic heteroduplex analysis of DNA from sheep and goat mitochondrial DNAs (15).

Other Nonlethal Mutations. In all of the proviruses analyzed, there were other mutations detected in additon to the ones that led to these clones being selected for sequence determination (lack of infectivity or loss of restriction enzyme cleavage sites): one in clone 63; one in clone 3-73; eight in clone 13; and four in clone 32. (Because three mutations are common to clones 13 and 32, they are only counted once.) In addition, there is one mutation in clone 14-44 (see above). These mutations include three transversions, seven transitions, and two small deletions. Seven of the 12 mutations are in noncoding regions, and 2 of the 5 mutations in coding sequences are silent. Thus, there are 3 mutations leading to amino acid changes in 3500 nucleotides identified and 12 mutations leading to changes in RNA structure in 4620 nucleotides identified (this work and ref. 8). These data indicate that nonlethal changes in RNA structure occur approximately 4 times as frequently as nonlethal changes in this region of the *gag* polypeptide. Thus, the RNA primary structure of SNV has fewer constraints than the protein sequence.

Consistent with this hypothesis is the high frequency of variation found in the sequences of LTRs of various retroviruses (summarized in ref. 16).

Mechanisms of Mutation. One striking feature of these mutations is the similarity of the mutations at positions 764–773 in clones 13 and 32. This coincidence is not likely to be the result of independent events and thus must reflect a common ancestry followed by mutation or recombination. A similar conclusion was drawn from the presence of identical changes in the U3 region of LTRs of several SNV clones (14)

All of the alterations in these proviruses leading to a lack of infectivity are located between the *Sal* I cleavage site at 0.86 kbp and the *Bam*HI cleavage site at 1.92 kbp (6). Infectious virus is recovered after cotransfection with a fragment of DNA from a wild-type virus bounded by these sites and the noninfectious clones, indicating that recombination occurs with high frequency. In particular, there must be a high frequency of recombination between the fragment that starts at 0.86 kbp and the mutation that is at 0.984 kbp in clone 63.

Furthermore, infectious virus was recovered from cotransfections of all clones in pairwise combinations (6). The short distance between the lethal mutations in clone 63 at 0.984 kbp and in clone 13 at 1.129–1.139 kbp again indicates a high frequency of recombination after cotransfection.

Some kind of homologous pairing was probably responsible for the insertion, in clone 13, of base pairs 525–578 between base pairs 1129 and 1139. The synthesis of retrovirus DNA is complicated and apparently involves two jumps of short fragments of viral DNA from one end of the genome to the other (see reviews in refs. 16 and 17). In particular, in the synthesis of the (+) strand of viral DNA, a jump of a fragment of DNA from the 3′ end of viral RNA to the 5′ end of viral DNA has been postulated. The correct positioning of the fragment that jumps depends upon homology between the 18 nucleotides in the primer binding site (PBS in Fig. 2) and the complementary nucleotides. Fortuitously, nucleotides 1128–1140 (homologous nucleotides are italic) in the SNV *gag* gene, G-T-*T*-T-*G*-*G*-*C*-*G*-T-*A*-*G*-*G*-*G*, have high homology with nucleotides 523–535 in the U5 region of SNV, *G*-*G*-*T*-G-*G*-*G*-*C*-*G*-*C*-*A*-*G*-*G*-*G*, and both are in G+C-rich regions.

Thus, a likely sequence of events leading to the nontandem duplication in clone 13 is that synthesis of the fragment of (+) strand DNA prematurely terminated and then "jumped" to the wrong position in the *gag* gene rather than to the right position in the primer binding site. A correct jump of another DNA fragment then occurred, and DNA was synthesized from this DNA to the paired region of the first DNA fragment. Action of nuclease and ligase resulted in joining of these two DNAs, and the 3′ end of the first (+) strand DNA fragment acted as a primer for further synthesis of DNA, resulting in the 10-bp deletion found in clone 13. (The duplication of portions of two DNA fragments indicates that reverse transcription took place on both molecules of virion RNA in a single particle.)

Another striking feature of many of the mutations described

here is that they are clustered. For example, in clone 3-73 there are only 71 bp between the two mutations; in clone 13, there are two deletions between positions 703 and 717; and in clones 13 and 32, there are only 10 bp between the three single base pair changes.

These clusters do not have characteristics of the mutational hot spots studied by others (for example, ref. 18). The clusters in SNV DNA are at different positions in different clones and the same base pair is not mutant in any one clone, with the exceptions in clones 13 and 32 discussed above. In addition, the mutations in SNV are not limited to deletions or insertions as is the case of mutational hot spots in bacteria. Thus, some explanation other than the sequence of the DNA is probably responsible for the clusters.

There are two features of the synthesis of retrovirus DNA that might be especially error prone so as to give clustered mutations: removal of internal RNA primers, and recombination. Internal ribonucleotides postulated to be the remnants of primers have been found in unintegrated SNV DNA but not in SNV proviruses (19). Recombination is also frequent in retroviruses (see ref. 20). Both processes might introduce multiple mutations in a restricted region of the viral genome. Alternatively, there could be selection for further mutations in a region after the first one occurred. For example, a closely linked suppressor could restore wild-type function. This hypothesis is less likely than the others because the L region may be analogous to an intervening sequence and not have functions requiring a particular nucleotide sequence along its entire length.

The use of restriction enzyme mapping, marker rescue, and DNA sequence determination has enabled us to define the molecular nature of lethal and nonlethal mutations in SNV proviruses and to explain the occurrence of a nontandem duplication. The type and pattern of mutations reflect the unique life cycle of retroviruses.

We thank E.-C. Tan, Z. Burton, C. Gross, K. Watanabe, and S. Hellenbrand for help, I. Chen, R. Fitts, J. Mertz, and B. Sugden for comments on the manuscript, and C. Gross for major help in revising the manuscript. This investigation was supported by U.S. Public Health Service research grants from the National Cancer Institute. H.M.T. is an American Cancer Society Research Professor.

1. Temin, H. M. (1961) *Virology* **13**, 158–163.
2. Temin, H. M. (1974) *Adv. Cancer Res.* **19**, 47–104.
3. Goff, S., Traktman, P. & Baltimore, D. (1981) *J. Virol.* **38**, 239–248.
4. Lee, W.-H., Nunn, M. & Duesberg, P. H. (1981) *J. Virol.* **39**, 758–776.
5. O'Rear, J. J., Mizutani, S., Hoffman, G., Fiandt, M. & Temin, H. M. (1980) *Cell* **20**, 423–430.
6. O'Rear, J. J. & Temin, H. M. (1981) *J. Virol.* **39**, 138–149.
7. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–580.
8. Shimotohno, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285**, 550–554.
9. Queen, C. L. & Korn, L. J. (1980) *Methods Enzymol.* **65**, 595–609.
10. Oroszlan, S., Barbacid, M., Copeland, T., Aaronson, S. A. & Gilden, R. V. (1981) *J. Virol.* **39**, 845–854.
11. Hunter, E., Bhown, A. S. & Bennett, J. C. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2708–2712.
12. Keshet, E. & Temin, H. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3372–3376.
13. Keshet, E., O'Rear, J. J. & Temin, H. M. (1979) *Cell* **16**, 51–61.
14. Shimotohno, K. & Temin, H. M. (1982) *J. Virol.* **41**, 163–171.
15. Upholt, V. B. & Dawid, I. B. (1977) *Cell* **11**, 571–583.
16. Temin, H. M. (1981) *Cell* **27**, 1–3.
17. Varmus, H. & Swanstrom, R. (1982) in *Molecular Biology of Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Part 3, in press.
18. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. (1978) *J. Mol. Biol.* **126**, 847–863.
19. Chen, I. S. Y. & Temin, H. M. (1980) *J. Virol.* **33**, 1058–1073.
20. Coffin, J. M. (1979) *J. Gen. Virol.* **42**, 1–23.