10. A. Benninghoven, D. Jaspers, W. Sichterman, *Appl. Phys.* **11**, 35 (1976).
11. J. Pierce, K. L. Busch, R. A. Walton, R. G. Cooks, *J. Am. Chem. Soc.* **103**, 2583 (1981).
12. T. M. Barlak, J. R. Wyatt, R. J. Colton, J. J. Decorpo, J. E. Campana, *ibid.*, in press.
13. M. Barber, R. S. Bordoli, R. D. Sedgewick, A. N. Tyler, *J. Chem. Soc. Chem. Commun.* (1981), p. 325.
14. S. P. Holland, B. J. Garrison, N. Winograd, *Phys. Rev. Lett.* **43**, 220 (1979).
15. P. Eisenberger and L. C. Feldman, *Science* **214**, 300 (1981).
16. See, for example, D. G. Truhlar and J. T. Muckerman, in *Atom-Molecule Collision Theory: A Guide for the Experimentalist*, R. B. Bernstein, Ed. (Plenum, New York, 1979), pp. 505–566; P. Lykos, Ed., *ACS Symp. Ser. 86* (1978); D. J. Rossky and M. Karplus, *J. Am. Chem. Soc.* **101**, 1913 (1979).
17. J. B. Gibson, A. N. Goland, M. Milgram, G. H. Vineyard, *Phys. Rev.* **120**, 1229 (1960).
18. W. L. Gay and D. E. Harrison, Jr., *Phys. Rev. B* **135**, A1780 (1964).
19. D. E. Harrison, Jr., P. W. Kelly, B. J. Garrison, N. Winograd, *Surf. Sci.* **76**, 311 (1978).
20. B. J. Garrison, N. Winograd, D. E. Harrison, Jr., *Phys. Rev. B* **18**, 6000 (1978).
21. ———, *J. Chem. Phys.* **69**, 1440 (1978).
22. D. E. Harrison, Jr., W. L. Moore, Jr., H. T. Holcombe, *Radiat. Eff.* **17**, 167 (1973).
23. N. Winograd, B. J. Garrison, D. E. Harrison, Jr., *Phys. Rev. Lett.* **41**, 1120 (1978).
24. B. J. Garrison, in *Potential Energy Surfaces and Dynamics Calculations for Chemical Reactions and Molecular Energy Transfer*, D. G. Truhlar, Ed. (Plenum, New York, 1981), pp. 843–856.
25. C. A. Andersen and J. R. Hinthorne, *Science* **175**, 853 (1972).
26. M. L. Yu, *Phys. Rev. Lett.* **40**, 574 (1978).
27. A. Blandin, A. Nourtier, D. Hone, *J. Phys. (Paris)* **37**, 369 (1976).
28. J. K. Norskov and B. I. Lundqvist, *Phys. Rev. B* **19**, 5661 (1979).
29. Z. Sroubek, K. Zdansky, J. Zavadil, *Phys. Rev. Lett.* **45**, 580 (1980).
30. M. L. Yu, *ibid.* **47**, 1325 (1981).
31. R. A. Gibbs, S. P. Holland, K. E. Foley, B. J. Garrison, N. Winograd, *Phys. Rev. B*, **24**, 6178 (1981).
32. G. Staudenmaier, *Radiat. Eff.* **13**, 87 (1972).
33. N. Winograd and B. J. Garrison, *Acc. Chem. Res.* **13**, 406 (1980).
34. N. Winograd, D. E. Harrison, Jr., B. J. Garrison, *Surf. Sci.* **78**, 767 (1978).
35. B. J. Garrison, N. Winograd, D. E. Harrison, Jr., *J. Chem. Phys.* **73**, 3473 (1980).
36. B. J. Garrison, *J. Am. Chem. Soc.* **102**, 6553 (1980).
37. N. Winograd, B. J. Garrison, T. Fleisch, W. N. Delgass, D. E. Harrison, Jr., *J. Vac. Sci. Technol.* **16**, 629 (1979).
38. S. P. Holland, B. J. Garrison, N. Winograd, *Phys. Rev. Lett.* **44**, 756 (1980).
39. W. Erley, H. Ibach, S. Lehwald, H. Wagner, *Surf. Sci.* **83**, 585 (1979).
40. K. E. Foley and N. Winograd, to be published.
41. R. D. Macfarlane and D. F. Torgerson, *Science* **191**, 920 (1976).
42. C. J. McNeal and R. D. Macfarlane, *J. Am. Chem. Soc.* **103**, 1609 (1981).
43. G. M. Lancaster, F. Honda, Y. Fukuda, J. W. Rabalais, *ibid.* **101**, 1951 (1979).
44. H. T. Jonkman and J. Michl, *ibid.* **103**, 733 (1981).
45. H. Grade, N. Winograd, R. G. Cooks, *ibid.* **99**, 7725 (1977).
46. J. E. Campana, T. M. Barlak, R. J. Colton, J. J. DeCorpo, J. R. Wyatt, B. I. Dunlap, *Phys. Rev. Lett.* **47**, 1046 (1981).
47. Courtesy of D. J. Surman, Kratos Scientific Instruments.
48. D. H. Williams, C. Bradley, G. Bojesen, S. Santikarn, L. C. E. Taylor, *J. Am. Chem. Soc.* **103**, 5700 (1981); D. H. Williams, G. Bojesen, A. Auffret, L. C. E. Taylor, *FEBS Lett.* **128**, 37 (1981).
49. The financial support of the National Science Foundation, the Office of Naval Research, the Air Force Office of Scientific Research, the Petroleum Research Foundation, the A. P. Sloan Foundation, and the Dreyfus Foundation is gratefully acknowledged.

# Form and Function of Retroviral Proviruses

Harold E. Varmus

*Summary.* Retroviruses have proved to be useful reagents for studying genetic and epigenetic (such as regulatory) changes in eukaryotic cells, for assessing functional and structural relationships between transposable genetic elements, for inducing insertional mutations, including some important in oncogenesis, and for transporting genes into eukaryotic cells, either after natural transduction of putative cellular oncogenes or after experimental construction of recombinant viruses. Many of these properties of retroviruses depend on their capacity to establish a DNA (proviral) form of their RNA genomes as a stable component of host chromosomes, in either somatic or germinal cells.

RNA tumor viruses (now called by the more general term, retroviruses) have long been fascinating to experimentalists from various disciplines because of their capacity to induce neoplasms, their unusual strategy for replication (by way of a DNA intermediate), and their broad distribution in nature as agents transmitted both horizontally and genetically. The goal of much work in retrovirology has been to understand these phenomena in molecular terms (*1*). Two of the themes that pervade such efforts are pertinent to those aspects of retroviral behavior, described in this article: (i) the intimate relation between the genomes of retroviruses and their hosts and (ii) the assignment of many functional attributes of retroviruses to a domain of several hundred base pairs (bp) present at both ends of viral DNA and called long terminal repeats (LTR's).

During infection by retroviruses, viral DNA with its LTR's is generated from an RNA template and covalently joined to host chromosomes by a highly ordered process. Once established at a chromosomal site in a proviral form, viral genes are governed in part by their own LTR's, and subject to changes, including mutation and regulation of expression, that also affect cellular genes. Retroviruses occasionally infect germinal as well as somatic cells, becoming part of the host's genetic endowment in the form of endogenous proviruses. Proviruses can be inserted at many sites in host genomes, carrying potent regulatory signals in their LTR's; these insertions may physically interrupt and inactivate cellular genes, or the regulatory features of viral DNA may alter expression of neighboring genes. Finally, retroviruses not only contribute their genes to the host, they also appropriate cellular genes; several such transduced genes have been implicated in viral oncogenesis. These natural recombinants presage the use of LTR's in genetic vectors to be manipulated in vitro.

## Strategy of Retroviral Replication

The feature that unites retroviruses and distinguishes them from all other animal viruses is the transcription of their single-stranded RNA genomes into double-stranded DNA, later covalently linked to the host genome (*2*) (Fig. 1). The RNA genomes of all replication-competent members of this virus class, regardless of species of origin or pathogenic potential, contain three coding domains that participate in the replicative process, namely, *gag*, for synthesis of

The author is a professor in the Department of Microbiology and Immunology, University of California, San Francisco 94143.


0036-8075/82/0521-0812$01.00/0 

core proteins; *pol*, for synthesis of virion-associated, RNA-directed, DNA polymerase (reverse transcriptase); and *env*, for synthesis of envelope glycoproteins (*3*). Most retroviruses do not kill their host cells; on the contrary, they often confer a growth advantage on them. Hence the integrated viral DNA (the provirus) becomes a stable component of host chromosomes, replicated and inherited in concert with flanking cellular DNA and expressed as RNA and protein mainly through normal cellular mechanisms (Fig. 1).

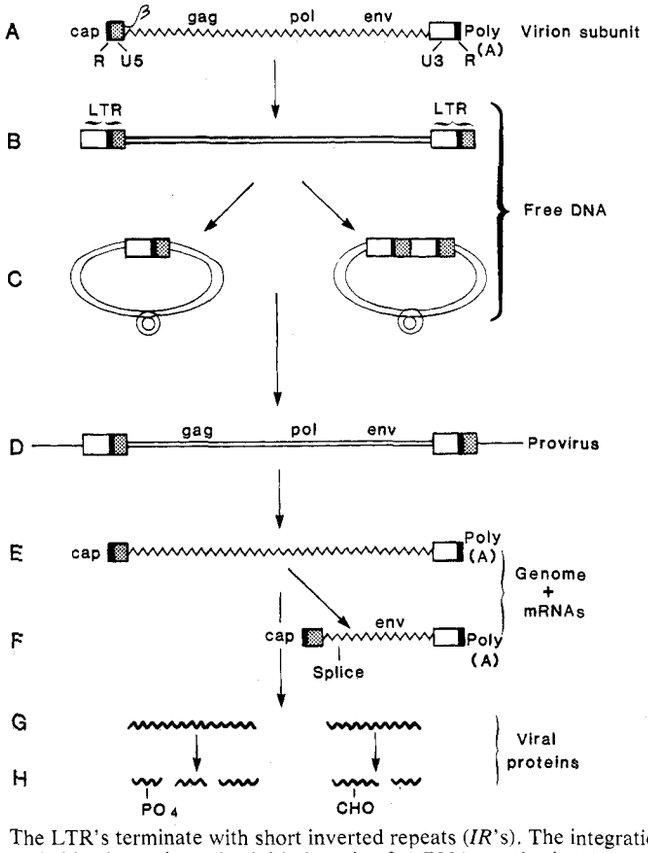
## Reverse Transcription: Problems and Solutions

Nature has devised remarkable solutions to the problems inherent in the conversion of single-stranded viral RNA into double-stranded DNA suitable for integration and full expression of viral genes. Redundancies figure prominently at several stages. Retroviruses are, first of all, genetically redundant in that their genomes are composed of two, usually identical, subunits of about 5 to 9 kilobases (kb) (*3*). Among animal viruses, diploidy is a property only of retroviruses, although its advantages are not well understood. Each subunit is itself terminally redundant, carrying at each end a short sequence (R) that can be up to 80 bases in length (*4*) (see Fig. 1 and Table 1). Many infected cells are haploid for viral genes, because they have acquired only a single provirus (*5, 6*), but a portion of viral RNA is duplicated to form the LTR's found at the 5′ and 3′ end of each provirus (*6, 7*). The LTR unit represents a fusion of sequences characteristic of the 3′ end of viral RNA (U3), the R

Table 1. Comparison of the LTR's of several retroviruses. The sequences and sizes of functionally significant domains of the LTR's of Rous sarcoma virus (RSV) (*11*); the endogenous chicken provirus at the *ev1* locus (*19*); spleen necrosis virus (SNV), an avian reticuloendotheliosis virus (*10*); Moloney strains of murine leukemia and murine sarcoma viruses (MLV, MSV) (*8, 9*); and mouse mammary tumor virus (MMTV) (*24*). The nature of these domains is explained in the text and the legend to Fig. 2. Numbers in parentheses indicate the number of base pairs between two sets of sequences or between sequences and the 5′ boundary of R.

| | | U3 | | | | R | U5 | | |
|---|---|---|---|---|---|---|---|---|---|
| VIRUS | (+) Strand primer sequence | Inverted repeat sequence | Length (bp) | "TATAA" Box | Poly(A) signal | Length (bp) | Length (bp) | Inverted repeat sequence | Binding site for tRNA primer |
| RSV | ...AGGGAGGGGGA | AATGTAGTCTTATGC... | 230 | ...TATTTAG..(16)... | AATAAA(1)... | 21 | 80 | ...GCAGAAGGCTTCATT | TGGTGACCCCGACGTGAT... |
| ev1 | ...AGGGAGGGGGA | AATGTAGTC... | 172 | ...TATATAA..(16)... | AATAAA(1)... | ~21 | 80 | ...GGCTTCATT | TGGTGACCCCGACGTGAT... |
| SNV | ...AGTGGGG | AATGT... | 369 | ...TATAAG...(21) | (53)AATAAA..(21)... | 80 | 100 | ...ACATT | TGGGGGCTCGTCCGGGAT... |
| MoMLV,MSV | ...AGAAAAAGGGGGG | AATGAAAGACCCC... | 371-442 | ...AATAAAAG.(21) | (47)AATAAA..(16)... | ~70 | 75 | ...GGGGTCTTTCATT | TGGGGGCTCGTCCGGGAT... |
| MMTV | ...AAAAAGAAAAAAGGGGGA | AATGCCGC... | 1192 | ...TATAAAAG.(15)... | AGTAAA(3)... | ~13 | 122 | ...GCGGCAGC | TGGCGCCCGAACAGGGAC... |

Initiation site for (+) DNA — Integration site — cap site — Poly(A) site — Integration site — Initiation site for (-) DNA



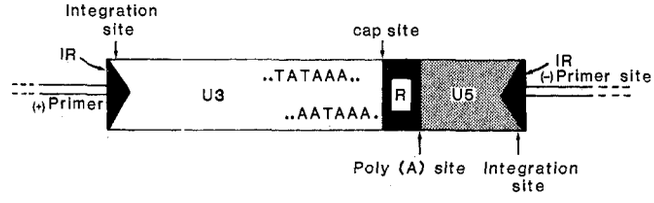




Fig. 1 (left). Replication and expression of retroviral genomes. (A) One of the two identical subunits of a viral RNA genome with its major structural and genetic features: the 5′ cap nucleotide (added to the first encoded nucleotide); the short sequence repeated at both termini (*R*, filled boxes); the sequence unique to the 5′ terminus and repeated in viral DNA (*U5*, shaded box); host tRNA hydrogen-bonded to the genome at the U5 boundary; the coding domains for virion structural proteins (*gag*, *pol*, *env*); the sequence unique to the 3′ terminus and repeated in viral DNA (*U3*, open box), and the tract of polyadenylic acid [*poly(A)*]. (B) The major primary product of reverse transcription, linear duplex DNA, with its long terminal repeats (*LTR*'s) composed of U3, R, and U5. (C) Closed circular DNA, with one or two copies of the LTR. (D) Proviral DNA. (E and F) Genomic and messenger RNA's, derived from the primary transcript by capping, polyadenylation, and splicing; the site at which the 5′ and 3′ domains of subunit RNA are joined to form *env* mRNA is indicated in (F). (G and H) The polyproteins synthesized from viral mRNA's and their mature products after cleavage and, in some cases, glycosylation (*CHO*) or phosphorylation ($PO_4$). Fig. 2 (above). Molecular anatomy of an LTR. The important features of the U3, R, U5, and flanking viral sequences of LTR's are shown in graphic form here and presented in greater detail in Table 1. Features that pertain to a 5′ LTR are shown at the top, those relevant to a 3′ LTR are shown at the bottom. The viral sequence at the 5′ boundary of U3 is the sequence of the putative primer for (+) DNA, and the sequence at the 3′ boundary of U5 is the binding site for the tRNA primer for (−) DNA. The LTR's terminate with short inverted repeats (*IR*'s). The integration sites are 2 bp from each boundary. The sequence resembling TATAAA probably determines the initiation site for RNA synthesis (the cap site), and the nonoverlapping sequence, usually AATAAA, probably determines the polyadenylation site.

sequence, and sequences that are characteristic of the 5′ end of viral RNA (U5), in the order 5′-U3-R-U5-3′ (Figs. 1 and 2); and it is terminated by short sequences forming inverted and often imperfect repeats (*8–12*) (Table 1 and Fig. 2).

Comparison of the structures of viral RNA and viral DNA (Fig. 1) suggests that DNA synthesis requires molecular acrobatics, the transfer of a nascent DNA strand twice between templates during reverse transcription. Although some details are still unsettled, a reasonable picture has been developed of these two "jumps" between templates (*2, 13*); the favored scheme is illustrated in Fig. 3 and summarized below.

Synthesis of the first [or (−)] strand is always initiated with a host transfer RNA (tRNA) primer, where the 3′ end is hydrogen-bonded to an 18-base sequence near the 5′ end of a viral RNA subunit (*14*) (Fig. 3A and Table 1). As will become evident, the point at which synthesis is begun defines the 3′ boundary of U5. When the template is exhausted 100 to 180 bases from the initiation site, the first jump occurs, with the use of DNA complementary to R to form a bridge by base pairing at the 3′ end of one of the RNA subunits (Fig. 3B). Once safely repositioned, the nascent (−) strand is extended, apparently continuously, over the major expanse of the genome. Meanwhile, synthesis of the second, or (+), strand begins at a position that defines the 5′ boundary of U3 (Fig. 3C). Although the primer for this event has not been identified, the priming site is always spanned by a purine-rich tract ending in AATG (A, adenine; T, thymine; G, guanine) (see Table 1), with these four bases probably forming the 5′ end of the (+) strand (*15*). Extension of the (+) strand is also limited by the amount of available template; after the U3, R, and U5 regions of (−) strand DNA are copied, the linked tRNA is probably transcribed up to the position of the first modified base (*13, 15a*), setting the stage for the next major event.

To execute the second "jump," a duplex can be formed between the 3′ end of the growing (−) strand, once it has copied the tRNA binding site, and the 3′ end of the arrested (+) strand (Fig. 3, D and E). Extension of both strands will then produce a blunt-ended, linear duplex, the major product of retroviral DNA synthesis in the cytoplasm of infected cells (Fig. 3F). The linear form is indistinguishable by restriction mapping from proviral DNA, bearing a full complement of genes flanked by LTR's (*6, 7*).

## Retroviral Integration: Mechanism Requirements

Two major unintegrated species have been identified in addition to linear DNA: covalently closed, circular DNA's with one or with two copies of the LTR sequence (*7, 16*) (Fig. 1C). However, it is not yet known which of the unintegrated species of viral DNA is the substrate for integrative recombination. In order to propose realistic mechanisms for inserting various species of viral DNA into host genomes, it is necessary to inspect relevant nucleotide sequences in putative DNA intermediates, host integration sites, and proviruses. Sequencing studies have revealed three features likely to be of major mechanistic significance (*2*).

1) Integration sites in the host chromosome appear to be unrelated to each other or to the termini of proviral DNA (*8–10, 12, 17*), implying that the integrative mechanism joins nonhomologous sequences. This result conforms to the conclusion drawn from physical mapping studies with restriction endonucleases, namely, that proviruses are inserted in many different regions of host genomes (*5, 6, 18*).

2) A short host sequence at the integration site is duplicated during integration, so that the duplicated sequence appears as a direct repeat flanking the provirus (*8, 10, 12, 17, 19*). This sort of duplication could be generated by a staggered cleavage of host DNA at the point of insertion; repair of the resulting single-stranded regions would complete the duplication (*20*). The size of the duplicated host sequence (thus far, 4-, 5-, and 6-bp repeats have been observed) appears to be characteristic of each retrovirus rather than its host cells, suggesting that the enzyme responsible for the proposed staggered cleavage of chromosomal DNA might be virus-coded.

3) The viral nucleotides joined to cellular DNA are rigidly determined. In all published cases, the site of union with host DNA is precisely 2 bp from the predicted ends of linear DNA (*8–10, 12, 17*) (Fig. 2 and Table 1).

In other words, 2 bp appear to be missing from the U3 sequence at the 5′ end of proviral DNA and from the U5 sequence at the 3′ end. The presence of specific viral nucleotides at host-viral junctions implies either that the integrative mechanism acts accurately 2 bp's from the ends of linear DNA to join viral and cellular DNA or that an endonuclease cleaves closed circular viral DNA in a site-specific manner.

Although we do not yet understand the enzymatic mechanisms involved, retroviral integration appears to be a highly ordered event, unlike other processes by which foreign DNA segments are integrated into host chromosomes. For example, integration of papova-, adeno-, and presumably herpesvirus DNA's occurs after infection of certain hosts, with little or no regard for specific sites in host or viral genomes (*21*). Moreover, integrated genomes of these viruses are often rearranged and flanked by cellular DNA, which has itself undergone rearrangements. Likewise, any DNA segment may integrate into host chromosomes after introduction by microinjection or transfection (*22*); however, these events lack precision, even when retroviral DNA's are presented to cells in these artificial ways (*23*). That retrovirus-infected cells can perform integration with conservation, specificity, and symmetry suggests that strong evolutionary forces have been at work on cells, viruses, or both, to shape the integrative mechanism.

## Functions of the LTR

It was appreciated at the time of discovery of LTR's (*7*) that they solved some of the tactical problems that seemed at first glance to be major obstacles to retrovirus replication. (i) Since the provirus contains duplications of sequences present only once in RNA, it is possible to use duplicated sequences, such as U3 in the 5′ LTR, as nontranscribed regulators of expression. (ii) If base pairs must be sacrificed during integration, the losses can be confined to duplicated and therefore expendable regions. (iii) By generating a second copy of the LTR, a copy of R is placed near each end of viral DNA. Thus the copy of R which appeared to be sacrificed during the first jump (Fig. 3B) is restored, and the provirus includes a complete, colinear representation of viral RNA (Fig. 1).

Nucleotide sequencing of LTR's cloned from several strains of retroviruses (*8–11, 19, 24*) suggests that LTR's provide functions fundamental to the expression of most eukaryotic genes, namely, promotion, initiation, and polyadenylation of transcripts (Table 1). Approximately 25 to 30 bp upstream from the 5′ end of R, LTR's invariably exhibit a sequence closely related to the so-called "TATAA box," strongly implicated in the determination of initiation sites for transcription of many eukaryotic genes (*25*). The prediction that the LTR

promotes initiation of transcription by host RNA polymerase II at the 5′ end of the R sequence is supported by several kinds of evidence, including chemical analysis of viral transcripts made in vitro (*26*) or in isolated nuclei (*27*); use of specific inhibitors of RNA polymerase (*28*); and tests of recombinant molecules containing LTR's linked to foreign genes (*29, 30*). Sequencing studies have also revealed the common eukaryotic signal for polyadenylation (AATAAA) (*31*) about 20 bp upstream from the site of polyadenylation in most virus strains (Table 1).

There are unresolved problems posed with regulatory signals in identical termini. It is not known, for example, what dictates the preferential use of the 5′ LTR rather than the 3′ LTR to promote transcription. In most cases, the anatomy of the LTR explains why the 5′ LTR can serve as a promoter without also directing immediate polyadenylation of a minuscule transcript: the polyadenylation signal resides on the 5′ side of the initiation site (Table 1). But in two types of LTR, in which the R region is approximately 70 to 80 bp, the polyadenylation signal is about 50 bp downstream from the transcriptional start (Table 1). A model involving alterations in the secondary structure of viral RNA has been proposed to explain why, in these cases, the polyadenylation signal is respected only in the 3′ LTR (*27*).

Sequences within or bounding LTR's thus influence functions peculiar to retroviruses (reverse transcription and ordered integration) and functions required of eukaryotic genes (synthesis and polyadenylation of RNA transcripts). LTR's may also confer on viral DNA the ability to enhance DNA transformation or modulate transcriptional activity of flanking cellular DNA, and they may serve as sites for homologous recombination. These other functions are discussed below.

## Proviruses Are Subject to Regulatory Influences

Although all proviruses provide their own transcriptional signals, individual proviruses manifest considerable variation in transcriptional activity. Major differences in proviral expression, apparently controlled during transcription, have been observed between related proviruses in the same cell (*32*), between similar proviruses in different cells (*18, 33*), between similar proviruses experimentally introduced into the mouse germ line at different sites (*34*), and after hormonal treatment (*35*) or prolonged growth of cloned cells containing a single provirus (*36*).

What determines such differences? Different levels of expression of some avian viruses probably depend on divergent nucleotide sequences in the U3 regions of otherwise closely related viruses (*37*). In other cases, however, the differences probably reside in nongenetic factors of the sort that have been thought to control cellular genes, such as chemical modification of DNA (for example, methylation) (*38*), alterations in chromatin structure (*39*), or interactions between specific regulatory proteins and DNA (*40*). Three examples of such cases are discussed briefly.

1) *Control of mouse mammary tumor virus (MMTV) RNA synthesis by glucocorticoid hormones.* Addition of glucocorticoid hormones to cells infected with MMTV produces an immediate and dramatic effect on the synthesis of MMTV RNA (*35*). It is likely that this response is mediated by a complex of hormone and host-coded receptor (*40*) that interacts specifically with the MMTV LTR. Strong support for this hypothesis comes from experiments in which various viral or cellular genes have exhibited steroidal control in cultured mammalian cells after linkage to cloned MMTV LTR's (*30*). In vitro mutagenesis of the LTR and binding studies with purified receptor protein should soon define the viral sequence active in the hormonal response.

2) *Modulation of expression of Rous sarcoma virus (RSV) proviruses in mammalian cells.* Cells phenotypically transformed by retroviruses often revert to a normal phenotype, by mechanisms apparently affecting transcriptional activity, with a frequency that would seem to be incompatible with a mutational mechanism (*2*). A high proportion of subclones derived from one RSV-transformed hamster cell exhibit a 10- to a 100-fold reduction in viral RNA and protein without demonstrable change in the structure or the genetic competence of

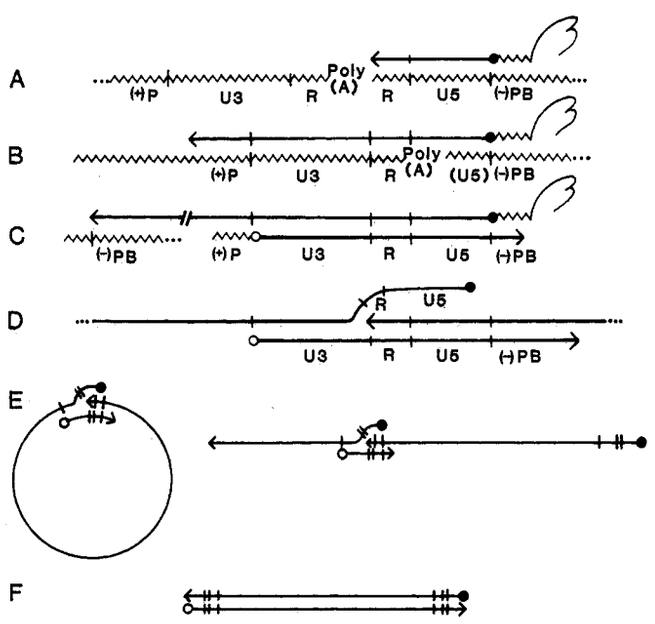



Fig. 3. Critical steps in the synthesis of retroviral DNA. The priming events for (−) and (+) strands of viral DNA (A and C) and the two transfers of nascent strands between templates (B and D), pictured on an expanded scale, lead to production of a linear duplex with two copies of the LTR unit (E and F; reduced scale). RNA is shown as wavy lines, DNA as straight lines with arrows denoting direction of synthesis; the 5′ end of (−) strand DNA is indicated by a filled circle, the 5′ end of (+) strand DNA by an open circle. (−)PB is the transfer RNA primer binding site, (+)P is the putative (+) strand primer sequence. The short vertical lines designate the boundaries of U3, R, and U5 in all panels. (A) A nascent (−) strand copy of R-U5 at the extreme 5′ end of its template. (B) After removal of the 5′ end of the primary template, the nascent (−) strand has base-paired with the R sequence at the 3′ terminus of the same or a companion RNA subunit and is extended along its secondary template of viral RNA. (C) Synthesis of (+) strand DNA commences at a priming site at the boundary of U3, and (+) strand is extended through a portion of the transfer RNA sequence originally bound to viral RNA; the (−) strand concurrently elongates toward the 5′ end of viral RNA, into or beyond the (−)PB site. (D) The nascent (−) strand in (C) has base-paired with the (+) strand, with complementary sequences from the (−)PB region. The (+) strand is then extended on its second template, (−) strand DNA, and the (−) strand is extended by displacement synthesis along its third template, (+) strand DNA. (E) The events in (D) are shown at a reduced scale to indicate the use of either one RNA subunit (left) or two (right) during DNA synthesis. (F) Complete extension of both (+) and (−) strands has produced a linear duplex terminated with LTR's. The molecule is the same as that shown in Fig. 1B.

the single RSV provirus (*36*). There appears to be no compelling explanation for this presumably epigenetic phenomenon. There is, however, some reason to believe that the chromosomal location of the provirus may be a significant determinant; in one RSV-transformed mammalian cell bearing a single provirus, viral gene expression is extremely stable (*41*), but introduction of virus recovered from this cell into fresh cells, with integration at new sites, produces an unstable phenotype (*42*).

3) *Variation in expression of endogenous chicken proviruses*. Two proviruses endogenous to chickens, at the loci called *ev*1 and *ev*2, are usually inefficiently expressed and highly methylated at CpG (C, cytosine) dinucleotides (*43, 44*). In at least three situations, heightened expression of these genetic elements has been associated with hypomethylation: (i) after treatment of cultured cells with 5-azacytidine (*45*); (ii) in an embryo found to be expressing the *ev*1 provirus spontaneously at high levels (*45*); and (iii) in cells horizontally infected by the virus (RAV-O) encoded by *ev*2 (*44*). The activated proviruses also appear to be more sensitive to deoxyribonuclease I digestion of chromatin, particularly in the LTR domains (*43*). However, there is no proof that either methylation or the altered chromatin configuration is directly responsible for the changes in expression.

## Mutations in Proviral Genes

As components of host chromosomes, proviruses are susceptible to the mutational risks as well as the regulatory influences of their hosts. Since viral genes can be readily cloned and sequenced, they provide convenient access to the biochemical basis of mutation in animal cells. There are at least three major difficulties inherent in this approach: (i) retroviral genomes appear to be genetically more labile than host chromosomes, with a high frequency of interviral recombination, deletion formation, and base substitution during passage of virus stocks (*4*); (ii) only the transforming genes of retroviruses produce easily selectable cellular phenotypes; and (iii) alteration in levels of proviral expression may occur frequently and be confused with mutation.

These problems can be largely circumvented with the use of clonal cell lines that can be transformed to a stable neoplastic phenotype by a single provirus, but that are unable to produce virus (and manifest those genetic changes dependent on viral replication), either because the viral genome is defective or because the host cell is nonpermissive for replication. Hence, from one line of RSV-transformed rat cells, called B31, it has been possible to isolate many mutants in which the cell phenotype has reverted to normal and secondary mutants in which the transformed phenotype has been restored (*46, 47*). Examination of these mutants by various biochemical techniques, including molecular cloning and sequencing in some cases, has uncovered a wide range of mutations affecting the expression or genetic composition of the RSV transforming gene (*src*). These include deletions that remove the 5′ LTR and extinguish transcription of the provirus, probable point mutations that inactivate the protein kinase activity of the *src* gene product, a "plus-one" frameshift mutation that is suppressed by a secondary duplication of coding sequence, a few probable nonsense mutations, and insertion mutations discussed below.

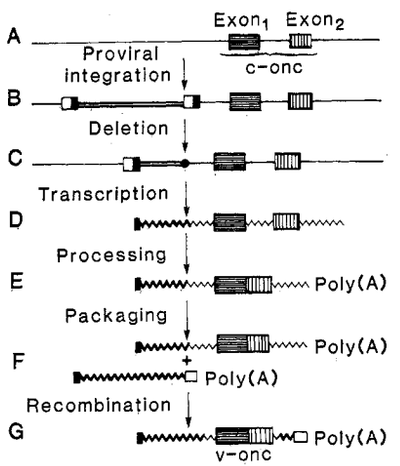



Fig. 4. A possible mechanism for capture of a cellular oncogene by a nontransforming retrovirus (*64*). The horizontally and vertically striped boxes in (A) represent exons in a cellular gene (c-*onc*) destined to be precursor to a viral transforming gene [v-*onc*; (G)]. (B) Proviral DNA of a retrovirus lacking *onc* integrates on the 5′ side of c-*onc* in the same transcriptional orientation. (Viral DNA is indicated by double lines, U3 as open boxes, U5 as closed boxes, host DNA as a single line.) (C) A deletion removes a 3′ portion of proviral DNA and adjacent cellular DNA (the point of rejoining, indicated by a closed circle is arbitrary and could reside in a c-*onc* exon). (D) Transcription from the remaining retroviral LTR produces a primary transcript that links sequences from the 5′ end of the viral genome (heavy wavy line) to c-*onc* sequences. (E) The primary transcript is polyadenylated and spliced. (F) The processed RNA is packaged in retroviral particles, forming heterozygotic dimers with a subunit of wild-type viral RNA. (G) Recombination during reverse transcription occurs on infection of neighboring cells, yielding a genome with viral sequences on both sides of v-*onc*.

## Host Cell DNA Related to Retroviral Genomes

Abundant early evidence for retrovirus-related genes in genomes of normal cells encouraged the view that the histories of retroviruses and their hosts were elaborately entwined (*48*). It is now apparent that two functionally and structurally distinct classes of genetic elements in normal cells are closely related to components of retroviral genomes. First, many animals transmit through their germ lines viral replicative genes organized in the form of proviruses (*49*). Second, all vertebrates harbor a set of well-conserved cellular genes (called c-*onc*'s) that serve as progenitors of retroviral transforming genes (v-*onc*'s) (*50*). These two types of elements are discussed below.

*Endogenous proviruses*. It seems likely that most endogenous proviruses result from infrequent infection of germinal cells and are hence the inadvertent consequence of the unusual form of parasitism that retroviruses exemplify. Close scrutiny of several classes of endogenous proviruses, particularly those in chickens and mice, has unveiled a number of rules governing their origins and functions (*49*). (i) Endogenous proviruses have generally entered the germ line subsequent to speciation (*51–53*) and, occasionally, within recent periods of experimental observation (*54*). (ii) The proviruses appear to be structurally and genetically similar to proviruses acquired by horizontal infection, terminating with LTR's and containing replication genes closely related to those of exogenous viruses (*19, 55*). Like proviruses acquired by horizontal or in vitro infection, endogenous proviruses can be found in many different sites in host genomes (*51–53, 55*) and on multiple chromosomes (*56*). (iii) Endogenous proviruses appear to behave like stable genetic markers, segregating in crosses and being uniformly distributed in all members of an inbred strain (*52, 57*). (iv) Each endogenous provirus has a characteristic level of expression, determined by both host and viral factors acting in cis, and each confers a "phenotype" that reflects its genetic competence (*58*). (v) The competence of endogenous proviruses is sometimes affected by large deletions removing regulatory or coding elements (*55, 58*); it is not known whether such deletions occur at a greater rate in proviral than in cellular DNA. (vi) Proviruses of at least some types do not provide essential functions to their hosts, since occasional animals devoid of such proviruses have been encountered

(*52, 53, 57*). (vi) There may be selection against certain types of proviruses because most endogenous proviruses appear to be fundamentally innocuous: unable to replicate or nontumorigenic. Whether any confer a selective advantage upon their hosts is a matter of dispute.

*The progenitors of many viral transforming genes are cellular genes.* Analysis of certain retroviral genomes has unveiled more than a dozen putative oncogenes (v-*onc*'s) which are unrelated to each other or to viral replicative genes and which appear to mediate rapid oncogenesis in animals and morphological transformation of cultured cells (*50, 59*) (Table 2). Each v-*onc* is closely related to a single cellular sequence (c-*onc*)—or a small family of sequences—from which it is presumably derived (*60*). The c-*onc*'s usually exhibit the properties expected of many cellular genes: the archetypal structure, with intervening sequences between coding domains, and no adjacent LTR's; a high degree of evolutionary conservation, with homologous sequences detected throughout vertebrates; frequent expression as polyadenylated RNA and (where testable) as protein; and little restriction site polymorphism observed during analysis of DNA's from several members of a species (*50*). The number of potential c-*onc*'s is unknown, but repeated isolation of certain *onc*'s (Table 2) and the coincidence of related *onc*'s (*fps* and *fes* or *bas* and *ras*) in viruses isolated from different animal species (*61*) suggest that the number capable of appearing and functioning in viral genomes is probably less than 30. The c-*onc*'s seem to be unlinked to endogenous proviruses (*62*), but it is not known whether any are linked to each other. They may, in any case, constitute a group of genes with related functions, since the products of several v-*onc*'s exhibit an associated protein kinase activity specific for tyrosine residues (*63*). The normal functions of c-*onc*'s and the oncogenic properties of v-*onc*'s are subjects of intense current interest, but beyond the scope of this article.

Table 2. Cellular sequences identified as probable oncogenic sequences in the genomes of retroviruses. Fifteen distinguishable sets of sequences (v-*onc*'s) have been found in the genomes of transforming retroviruses and shown to be homologous to host sequences (c-*onc*'s) with properties of cellular genes. Arbitrary names have been assigned to the *onc* sequences (*59*). The number of probably independent virus isolates containing each element, with an example of each, is listed with the host from which the *onc* sequences have been apparently transduced.

| *onc* gene sequence | Virus isolates (No.) | Virus (example) | Animal origin |
|---|---|---|---|
| *src* | >3 | Rous sarcoma, Prague strain | Chicken, quail |
| *fps* | >3 | Fujinami sarcoma | Chicken |
| *yes* | 2 | Y73 sarcoma | Chicken |
| *ros* | 1 | UR-2 | Chicken |
| *myc* | 4 | Avian myelocytomatosis-29 | Chicken |
| *erb* | 1 | Avian erythroblastosis | Chicken |
| *myb* | 2 | Avian myeloblastosis | Chicken |
| *rel* | 1 | Reticuloendotheliosis, strain T | Turkey |
| *mos* | 2 | Moloney murine sarcoma | Mouse |
| *abl* | 1 | Abelson murine leukemia | Mouse |
| *bas* | 1 | BALB murine sarcoma | Mouse |
| *ras* | >3 | Harvey murine sarcoma | Rat, mouse |
| *fes* | 2 | Snyder-Theilin feline sarcoma | Cat |
| *fms* | 1 | McDonough feline sarcoma | Cat |
| *sis* | 1 | Simian sarcoma | Woolly monkey |

## Retroviruses as Transducing Agents

How do retroviruses acquire *onc*'s? Introduction of an *onc* gene into a retroviral genome, an uncommon occurrence, generally follows passage of weakly or nononcogenic viruses in animals. The detection of new viruses, whose formation may require the collusion of multiple, rare genetic events, is probably facilitated by the induction of tumors. Within the recombinant viral genomes, retroviral sequences are invariably found on both sides of *onc* sequences (a necessary consequence of the mode of replication described earlier); hence there must be at least two recombinations involved. In addition, the intervening sequences found in many c-*onc*'s do not appear in their v-*onc* homologs (*50*).

One recently proposed mechanism for the acquisition of *onc*'s has the appeal of precedents for each of its steps and makes some testable predictions (*64*) (Fig. 4). First, a rare event introduces a nontransforming provirus on the 5′ side of a c-*onc*, in the same transcriptional orientation. Then, a deletion that includes the 3′ LTR fuses the proviral and c-*onc* transcriptional units so that a single, hybrid RNA, reading "R-U5-viral gene(s)-*onc*," can be made and processed to remove c-*onc* introns. These steps would place c-*onc* under control of a viral promoter, alter the cellular phenotype, and propel tumorous expansion of a clone with properties favorable for production of a transforming virus. (Deployment of similar mechanisms in the induction of avian leukosis is discussed below.) Packaging signals proposed to lie near the 5′ end of viral RNA (*65*) would expedite the formation of heterozygous particles containing both the hybrid RNA and replication-competent RNA transcribed from another, normal provirus in the same cell. A high-frequency retroviral recombination system, proposed to depend on the agility of reverse transcriptase in negotiating multiple templates (*4*), would then join the sequences in the hybrid RNA on 3′ side of *onc* to a U3-R sequence from the normal subunit; recombination of precisely this sort is observed experimentally when a helper virus infects cells transformed by a *onc*-containing viral genome that lacks U3-R (*66*). The resulting RNA is likely to be deficient in replicative genes, like the genomes of almost all transforming viruses (*3*), but would be capable of replication if complemented by a helper virus. This scheme is compatible with the introduction of any cellular gene into a retroviral genome but favors those genes that promote cell growth.

## Retroviruses Are Insertion Mutagens

Proviruses have two properties that suggest that they might function as insertion mutagens. (i) They can enter many sites in host genomes, implying a capacity to inactivate genes by physical disruption, and (ii) they possess regulatory signals which could alter transcriptional control of flanking host sequences.

Precedent for the first prediction has been provided (*47*) by the finding of rare insertion mutants among morphological revertants of the RSV-transformed rat line, B31 (see above), after superinfection with a nontransforming murine leukemia virus (MLV). The two isolated mutants (Fig. 5) bear MLV proviruses not in *src* itself, but in a region of the RSV genome normally eliminated by splicing during the production of *src* messenger RNA (mRNA), and the inserts interfere with production of competent mRNA.

Insertion mutants revert if the insert is excised in a manner that restores gene function. In one of the two insertion

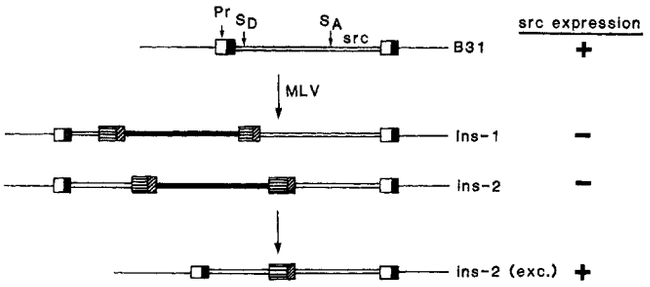



Fig. 5. Insertion mutagenesis by a retrovirus and subsequent excision of proviral DNA. The single Rous sarcoma virus (RSV) provirus in the *B31* line of RSV-transformed rat-1 cells is illustrated with the important landmarks for expression of the transforming gene (*src*): the promoter (*Pr*) in the 5′ LTR, the donor ($S_D$) and acceptor ($S_A$) splice points for processing of *src* mRNA, the coding domain (*src*), and the 3′ LTR. After infection by murine leukemia virus (*MLV*)—a nontransforming retrovirus which, unlike RSV, replicates in rat cells—the illustrated insertions of MLV proviral DNA produced the two mutant lines, ins-1 and ins-2. The bottom line shows the arrangement of proviral DNA in cells [ins-2 (excision)] in which most of the inserted MLV DNA has been lost by recombination between its LTR's. Expression of *src* was determined by measurement of *src* mRNA and the *src* gene product, as well as by observation of the cellular phenotype (*41*, *47*). Symbols for RSV DNA and cellular DNA as in Fig. 4; MLV DNA, solid double line; MLV U3, horizontally hatched boxes; MLV U5, diagonally hatched boxes.

mutants of B31 cells, the transformed phenotype is reinstated at low frequency following removal of most of the MLV provirus, apparently by homologous recombination between the 5′ and 3′ LTR's (*47*) (Fig. 5). Surprisingly, the single remaining MLV LTR has no major effect on synthesis or processing of *src* mRNA.

Do mutating insertions occur in the germ line and can endogenous proviruses be excised? The dilute coat color mutation in DBA/2J mice cosegregates with a single endogenous MLV provirus, and reversion of the mutation is accompanied by loss of at least part of the provirus (*67*). However, direct biochemical verification of insertion mutagenesis in this case is not yet available. Endogenous proviral elements containing sequences homologous only to LTR's have been found occasionally in the germ lines of chickens (*55*) and could be the residue of recombination between LTR's. Excision by this means might lighten the burden of housing an endogenous provirus but would also leave an interrupting and potentially regulating sequence in the germ line.

### Activation of Cellular Genes by Proviruses

The idea that proviruses could influence transcription of flanking host DNA was fostered by their terminally redundant structure; if the 5′ LTR initiates synthesis of viral RNA, what is to keep the 3′ LTR from promoting transcription of downstream host sequences? The predicted transcripts—detectable with probes for only U5 among viral sequences—were first sighted in lines of RSV-transformed mammalian cells (*33*) in which the activated cellular sequences have yet to be defined.

The most provocative evidence for gene activation by proviral DNA has emerged from efforts to understand tumor induction by viruses, such as avian leukosis virus (ALV), that lack transforming genes. ALV-induced lymphomas are composed of clonal cell populations that inevitably carry at least a portion of an ALV provirus; yet in many cases the provirus has been sufficiently deranged by deletion to prevent viral gene expression (*68*).

Thus viral proteins are apparently not required to maintain tumor growth, but even severely truncated proviruses appear to have a positive effect on expression of flanking host DNA (*68*). This is particularly remarkable because the vast majority of ALV-induced lymphomas bear proviruses integrated in the same region of the host genome (*68*), within or adjacent to c-*myc* (*69*), the cellular homolog of the transforming gene (v-*myc*) of some avian transforming viruses (Table 2). Moreover, the insertions are accompanied by an elevated concentration of c-*myc* RNA (*69*). The strong association between the induction of lymphomas and the activation of c-*myc* argues for a functional role for the c-*myc* protein in the oncogenic process. However, the c-*myc* gene product and its function have not been identified.

DNA from ALV-induced lymphomas can transform cultured NIH 3T3 mouse fibroblasts to a neoplastic phenotype (*70*), with the use of an assay developed to identify putative cellular oncogenes in tumors and cell lines from diverse sources, including human neoplasms (*71*). But, surprisingly, viral and c-*myc* sequences from the avian lymphoma DNA do not participate in the transformation of NIH 3T3 cells (*70*). It seems likely then that tumor induction by ALV requires at least two steps—insertional mutagenesis of c-*myc* and activating rearrangements of other cellular oncogenes.

Further surprises have been revealed by efforts to determine how the more efficient expression of c-*myc* is achieved. In the first series of tumors shown to have an activated c-*myc* locus, the instigating proviruses were uniformly positioned on the 5′ side (upstream) of c-*myc*, in the same transcriptional orientation as the c-*myc* gene, and c-*myc* transcripts appeared to anneal with probe for the viral U5 region (*69*). These findings were compatible with a model in which a proviral LTR, presumably the LTR at the right end of viral DNA, provides a more efficient promoter for c-*myc* than its natural promoter (panel I in Fig. 6). The picture has been complicated, however, by the discovery of additional tumors in which ALV proviruses in the vicinity of an activated c-*myc* locus are either positioned downstream (on the 3′ side) of c-*myc* or arranged in the transcriptional orientation opposite to that of c-*myc* on the 5′ side (*72*) (panels II and III in Fig. 6). With a provirus inserted downstream from c-*myc*, the abundant c-*myc* transcript seems to end with a viral U3 sequence (panel II in Fig. 6), probably reflecting use of the polyadenylation signal present in the LTR. When the ALV provirus and c-*myc* are in opposing transcriptional orientations, there is no linkage between viral and c-*myc* RNA (*72*). In these configurations the LTR's cannot simply provide an efficient promoter; instead they may indirectly augment the transcriptional activity of the surrounding domain or confer greater stability upon its transcriptional products. Alternatively, the insertions might disrupt cis-acting control elements.

Other sorts of experiment also suggest that LTR's may enhance gene expression by mechanisms other than provision of a strong promoter. (i) Murine sarcoma virus (MSV) LTR's have a similar dramatic effect on the efficiency of transformation by the cloned viral oncogene, v-*mos*, whether placed on the 3′ or 5′ side of v-*mos* (*29*). The configurations of LTR and v-*mos* in the transformed cells resemble the arrangements of ALV DNA and c-*myc* in Fig. 6, panels I and II; and the resultant transcripts are also analogous (*73*). (ii) The presence of an RSV LTR on either side of a complete herpesvirus thymidine kinase (*tk*) gene potentiates the efficiency of *tk* transformation 20-fold after microinjection of the DNA into cultured $tk^-$ mouse cells (*74*).

Fragments of SV40 DNA (Simian virus 40) encompassing the origin of replication and the 5′ side of the promoter for early genes also have an enhancing effect on the frequency of DNA transformation (*75*). A small region on the 5′ side of the start site for early RNA is required in either orientation for efficient expression of viral or linked cellular genes (*76*). Replacement of this region of SV40 DNA with a portion of an MSV LTR at least partially stimulates early gene expression (*77*). It is not yet known whether the similar functional attributes of these regions of papova- and retroviral DNA's are mediated by the same mechanisms.

### Retroviruses as Genetic Vectors

The capacity of LTR's to direct the efficient initiation and polyadenylation of RNA transcripts and to influence the efficiency of DNA transformation recommends their use in the delivery of genes to cultured cells. However, the probability of transformation is still relatively low for individual cells in most experiments, the transformations have been confined to cell culture, and the naked DNA does not enter the chromosome with the efficiency and specificity evident from the study of proviruses acquired during infection.

One attractive solution to these problems is to use viruses, not simply LTR's, to ferry genes into cells. Transforming viruses, recombinants between c-*onc*'s and retroviral genomes, exemplify the use of viral vectors in nature. By joining cloned fragments of viral DNA, the herpesvirus *tk* gene has recently been placed within retroviral genomes, including an MSV genome which still has its own transforming gene (*78, 79*). The MSV vector is defective for replication, but it can be effectively complemented by an accompanying helper virus and can spread to all susceptible cells in a culture or, presumably, in an animal, transforming both the morphological and biochemical phenotypes. The experimental potential for retroviral vectors has yet to be fully realized; but their pathological capabilities may preclude use in gene therapy in man.

### Retroviruses and Transposable Elements

Many of the properties of retroviruses recounted here have prompted widespread speculation about the extraordinary similarity of proviruses and transposable elements (*2, 6–12, 17–19, 80, 81*). Proviruses are terminally redundant, bear inverted repeats, can be inserted at many sites in host genomes, are flanked by a short host sequence duplicated during integration, act as insertion mutagens, excise by recombination between terminal repeats, undergo deletion mutation, and acquire genes from host genomes. Most or all of these features are exhibited by the transposable elements that inhabit the genomes of bacteria, yeast, and *Drosophila* (*82*). These resemblances raise interesting questions about the origins and functions of retroviruses and about the similarities of mechanisms responsible for these phenomena.

At present there is little rationale for supposing that proviruses undergo transposition in the manner favored for bacterial elements (*20*). Although the net effect of the virus replication cycle is the duplication of a provirus (one copy remaining in its original site, the second introduced at a new site, generally in another cell), there is no evidence that these results can be achieved by any mechanism other than synthesis of viral RNA and its reverse transcription into DNA. Even in chronically infected cells, which are resistant to superinfection, any further synthesis of viral DNA also occurs from an RNA template (*83*), perhaps in nascent virus particles. A more interesting possibility is that nomadic DNA of insects (for example, the *copia* element of *Drosphila*) may behave like proviruses; unintegrated, closed circular *copia* DNA, bearing one or two copies of its terminally repeated sequence, has been found in small amounts in the nuclei of cultured cells (*84*). Despite the striking structural similarity between retroviral and *copia* circular DNA's, nothing is yet known about the mechanisms by which free *copia* DNA is generated.

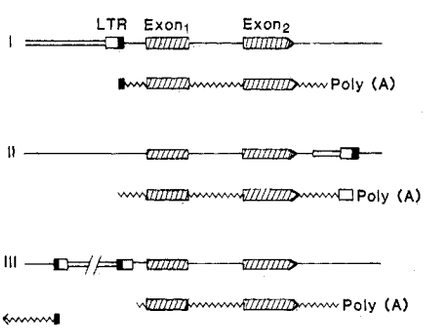



Fig. 6. Arrangements of proviral DNA and a cellular oncogene associated with enhanced expression of the oncogene. Panels I, II, and III depict the three configurations in which avian leukosis virus (ALV) DNA has been found in DNA from chicken bursal lymphomas in the region of c-*myc*, the cellular homolog of a viral transforming gene (v-*myc*) (*69, 72*). The two known exons of c-*myc* are denoted by striped boxes, with the shape of the second exon indicating the direction of synthesis of normal mRNA. Proviral DNA is shown as a double line, with open and closed boxes representing U3 and U5 regions, respectively; cellular DNA (other than the c-*myc* exons) is shown as a single straight line; and RNA is shown as a wavy line. The direction of synthesis of each RNA species is indicated either by the position of the polyadenylate tract [*poly(A)*] at the 3′ terminus or by an arrow (in III). The predicted primary transcripts containing c-*myc* sequences are illustrated for convenience, but the observed abundant c-*myc* RNA's in bursal tumors are processed species in which the region between $exon_1$ and $exon_2$ has probably been removed by splicing. The positions of integration, initiation, and polyadenylation sites are approximate; the drawing is not to scale.

Similarities between proviruses and transposable elements have also been found at the sequence level. For example, all proviruses examined to date and several kinds of transposable elements begin with GT and end with CA, and several snatches of similar sequence can be found within LTR's and terminally redundant sequences in transposable elements (*81*). Are these the conserved fossils of common ancestors or are they the stigmata of functionally significant regions of DNA, independently evolved but engaged in similar activities? Stimulated by such homologies, investigators have proposed that retroviruses emerged from vertebrate DNA when cellular transposable elements, predecessors to LTR's, surrounded and appropriated a host polymerase gene, the forebear of *pol* (*80*). Experimental support for this hypothesis is negligible at the moment. Given the functional and structural similarities among the many contemporary strains of retroviruses, it does seem likely that they descended from a single viral ancestor, but whether retroviruses are more likely than any other type of virus to have arisen from cellular genes is a matter for conjecture.

### Summary

Retroviruses first attracted widespread attention as oncogenic agents that amplify their RNA genomes through DNA intermediates. During the past decade study of these viruses has shown that they have other fascinating roles: as agents with varied pathological potential, dispersed through many species and transmitted by vertical as well as horizontal routes; as parasites well-adapted to host functions, thereby facilitating orderly integration and expression of viral genomes; as intermediates themselves in the relocation of DNA proviruses, which are structural homologs of the transpos-

able elements of other organisms; as mutagens equipped to interrupt or activate cellular genes; and as vectors able to transduce cellular genes and potentially act as agents of evolutionary change. No other class of animal viruses exhibits such profound intimacy with the genomes of their hosts; a detailed account of this relationship seems likely to influence our notions about diverse problems in contemporary biology.

### References and Notes


1. R. A. Weiss, N. Teich, H. E. Varmus, J. M. Coffin, Eds., *The Molecular Biology of Tumor Viruses*. part III. *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., in press). This volume represents a comprehensive account of contemporary retrovirology, including organization and expression of viral genes, nature and function of structural and transforming gene products, viral genetics, pathogenesis, and host range, subjects that are scarcely mentioned in this short article.
2. H. E. Varmus and R. Swanstrom, in (*1*), chap. 5.
3. J. M. Coffin, in (*1*), chap. 4.
4. ———, *J. Gen. Virol.* **42**, 1 (1979).
5. D. Steffen and R. A. Weinberg, *Cell* **15**, 1003 (1978).
6. S. H. Hughes *et al.*, *ibid.*, p. 1397.
7. P. R. Shank *et al.*, *ibid.*, p. 1383; T. W. Hsu, J. L. Sabran, G. E. Mark, R. V. Guntaka, J. W. Taylor, *J. Virol.* **28**, 810 (1978).
8. R. Dhar, W. L. McClements, L. W. Enquist, G. F. Vande Woude, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3937 (1980).
9. C. Van Beveren, J. G. Goddard, A. Berns, I. M. Verma, *ibid.*, p. 3302.
10. K. Shimotohno, S. Mizutani, H. M. Temin, *Nature (London)* **285**, 550 (1980).
11. R. Swanstrom, W. DeLorbe, J. M. Bishop, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 124 (1981).
12. J. E. Majors and H. E. Varmus, *Nature (London)* **289**, 253 (1981).
13. E. Gilboa, S. W. Mitra, S. Goff, D. Baltimore, *Cell* **18**, 93 (1979).
14. J. M. Taylor, *Biochim. Biophys. Acta* **473**, 57 (1977).
15. S. Mitra, M. Chow, J. Champoux, D. Baltimore, *J. Biol. Chem.*, in press.

15a. J. M. Taylor and T. W. Hsu, *J. Virol.* **33**, 531 (1981); R. Swanstrom, J. M. Bishop, H. E. Varmus, *ibid.* **42**, 337 (1982).

16. P. R. Shank, J. C. Cohen, H. E. Varmus, K. R. Yamamoto, G. R. Ringold, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2112 (1978); F. K. Yoshimura and R. A. Weinberg, *Cell* **16**, 323 (1979).
17. S. H. Hughes, A. Mutschler, J. M. Bishop, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4299 (1981); K. Shimotohno and H. M. Temin, *ibid.* **77**, 7357 (1980).
18. G. M. Ringold, P. R. Shank, H. E. Varmus, J. Ring, K. R. Yamamoto, *ibid.* **76**, 665 (1979).
19. F. Hishinuma, P. J. BeBona, S. Astrin, A. M. Skalka, *Cell* **23**, 155 (1980).
20. J. Shapiro, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1933 (1979).
21. R. A. Weinberg, *Annu. Rev. Biochem.* **49**, 197 (1980).
22. M. Wigler *et al.*, *Cell* **16**, 777 (1979).
23. P. Luciw, unpublished data.
24. L. A. Donehower, A. L. Huang, G. L. Hager, *J. Virol.* **34**, 226 (1980); H. Diggelmann, personal communication; J. Majors, unpublished data.
25. R. Breathnach and P. Chambon, *Annu. Rev. Biochem.* **50**, 349 (1981).
26. T. Yamamoto, B. deCrombrugghe, I. Pastan, *Cell* **22**, 788 (1980).
27. E. W. Benz, Jr., R. M. Wydro, B. Nadal-Ginard, D. Dino, *Nature (London)* **288**, 665 (1980).
28. M. Jaquet, Y. Groner, G. Monroy, J. Hurwitz, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3045 (1974); L. Rymo, J. T. Parsons, J. M. Coffin, C. Weissmann, *ibid.*, p. 2782.
29. M. Oskarsson, W. L. McClements, D. G. Blair, J. V. Maizel, G. F. Vande Woude, *Science* **207**, 1222 (1980).
30. F. Lee, R. Mulligan, P. Berg, G. Ringold, *Nature (London)* **294**, 228 (1981); A. L. Huang, M. C. Ostrowski, D. Berand, G. L. Hager, *Cell*, in press; K. Yamamoto, personal communication; J. Majors, unpublished data.
31. N. J. Proudfoot and G. G. Brownlee, *Nature (London)* **252**, 359 (1974).
32. W. S. Hayward and H. Hanafusa, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2259 (1976).
33. N. Quintrell, S. H. Hughes, H. E. Varmus, J. M. Bishop, *J. Mol. Biol.* **42**, 363 (1980).
34. R. Jaenisch *et al.*, *Cell* **24**, 519 (1981).
35. H. E. Varmus, G. Ringold, K. R. Yamamoto, in *Glucocorticoid Hormone Action*, J. Baxter and G. Rousseau, Eds. (Springer-Verlag, Berlin, 1979), pp. 253–278; G. M. Ringold, K. R. Yamamoto, J. M. Bishop, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2879 (1979); H. A. Young, T. Y. Shih, E. M. Scolnick, W. P. Parks, *J. Virol.* **21**, 139 (1977).
36. C. T. Deng, D. Boettiger, I. Macpherson, H. E. Varmus, *Virology* **62**, 512 (1974); D. Boettiger, *ibid.*, p. 522.
37. P. N. Tsichlis and J. M. Coffin, *J. Virol.* **33**, 328 (1980).
38. A. Razin and A. D. Riggs, *Science* **210**, 604 (1980).
39. H. Weintraub and M. Groudine, *ibid.* **193**, 848 (1976).
40. K. R. Yamamoto and B. Alberts, *Annu. Rev. Biochem.* **45**, 721 (1976).
41. H. E. Varmus, N. Quintrell, J. Wyke, *Virology* **108**, 28 (1981).
42. J. Wyke, personal communication; J. Jackson, unpublished data.
43. M. Groudine, R. Eisenman, H. Weintraub, *Nature (London)* **292**, 311 (1981).
44. E. Humphries, C. Glover, R. Weiss, J. Arrand, *Cell* **18**, 803 (1980).
45. K. Conklin, R. Eisenman, J. Coffin, personal communication.
46. H. Oppermann, A. Levinson, H. E. Varmus, *Virology* **108**, 47 (1981); G. Mardon, unpublished data.
47. H. E. Varmus, N. Quintrell, S. Ortiz, *Cell* **25**, 23 (1981).
48. G. J. Todaro and R. J. Huebner, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1009 (1972).
49. J. M. Coffin, in (*1*), chap. 10.
50. J. M. Bishop and H. E. Varmus in (*1*), chap. 9.
51. S. M. Astrin, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 5941 (1978); D. B. Frisby, R. A. Weiss, M. Roussel, D. Stehelin, *Cell* **17**, 623 (1979).
52. J. C. Cohen and H. E. Varmus, *Nature (London)* **278**, 418 (1979).
53. S. H. Hughes *et al.*, *Cell* **18**, 347 (1979).
54. W. P. Rowe and C. A. Kozak, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4871 (1980).
55. S. H. Hughes, Y. Toyoshima, J. M. Bishop, H. E. Varmus, *Virology* **108**, 189 (1981).
56. V. L. Morris *et al.*, *ibid.* **92**, 46 (1979); A. Tereba, L. B. Crittenden, S. M. Astrin, *J. Virol.* **39**, 282 (1981).
57. S. M. Astrin *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **44**, 1105 (1980).
58. W. S. Hayward, S. B. Braverman, S. M. Astrin, *ibid.*, p. 1111; B. Baker, H. Robinson, H. E. Varmus, J. M. Bishop, *Virology* **114**, 8 (1981).
59. J. M. Coffin *et al.*, *J. Virol.* **40**, 953 (1981).
60. D. Stehelin, H. E. Varmus, J. M. Bishop, P. K. Vogt, *Nature (London)* **260**, 170 (1976); J. M. Bishop, *Cell* **23**, 5 (1981).
61. M. Shibuya, T. Hanafusa, H. Hanafusa, J. Stephenson, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6536 (1980); P. R. Anderson *et al.*, *Cell* **26**, 129 (1981).
62. T. G. Padgett, E. Stubblefield, H. E. Varmus, *Cell* **10**, 649 (1977).
63. T. Hunter, B. M. Sefton, K. Beemon, *ICN-UCLA Symp. Mol. Cell. Biol.* **18**, 499 (1980).
64. R. Swanstrom, unpublished hypothesis.
65. P. R. Shank and M. Linial, *J. Virol.* **36**, 450 (1980).
66. M. P. Goldfarb and R. A. Weinberg, *ibid.* **38**, 136 (1981).
67. N. A. Jenkins, N. G. Copeland, B. A. Taylor, B. K. Lee, *Nature (London)* **293**, 370 (1981).
68. G. S. Payne, S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, H. E. Varmus, *Cell* **23**, 311 (1981); B. G. Neel, W. S. Hayward, H. L. Robinson, J. Fang, S. M. Astrin, *ibid.*, p. 323.
69. W. S. Hayward, B. G. Neel, S. M. Astrin, *Nature (London)* **290**, 475 (1980).
70. G. Cooper and P. E. Neiman, *ibid.* **287**, 656 (1980); *ibid.* **292**, 857 (1981).
71. C. Shih, B.-Z. Shilo, M. P. Goldfarb, A. Dannenberg, R. A. Weinberg, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5714 (1979); G. M. Cooper, S. Orenquist, L. Silverman, *Nature (London)* **284**, 418 (1980); C. Shih, L. C. Padhy, M. Murray, R. A. Weinberg, *ibid.* **290**, 761 (1981); J. G. Krontiris and G. M. Cooper, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 1181 (1981).
72. G. S. Payne, J. M. Bishop, H. E. Varmus, *Nature (London)* **295**, 209 (1982).
73. T. Wood and G. F. Vande Woude, personal communication.
74. P. Luciw and M. Capecchi, unpublished data.
75. M. R. Capecchi, *Cell* **22**, 479 (1980).
76. C. Benoist and P. Chambon, *Nature (London)* **290**, 304 (1981); D. Gruss, R. Dhar, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 943 (1981); J. Banerji, S. Rusconi, W. Schaffner, *Cell* **27**, 299 (1981); P. Moreau, R. Her, B. Wosylyk, R. Everett, M. P. Gaub, P. Chambon, *Nucleic Acids Res.* **9**, 6047 (1981); M. Fromm and P. Berg, personal communication.
77. B. Levinson, G. Khoury, G. Vande Woude, P. Gruss, *Nature (London)* **295**, 568 (1982); M. Botchan, personal communication.
78. K. Shimotohno and H. M. Temin, *Cell* **26**, 67 (1981); C. J. Tabin, J. W. Hoffman, S. P. Goff, R. A. Weinberg, *Mol. Cell. Biol.* **2**, 426 (1982).
79. C. M. Wei, M. Gibson, P. G. Spear, E. M. Scolnick, *J. Virol.* **39**, 935 (1981).
80. H. Temin, *Cell* **21**, 599 (1980).
81. G. Ju and A. M. Skalka, *ibid.* **22**, 376 (1980); B. M. Will, A. A. Bayer, D. J. Finnegan, *J. Mol. Biol.* **153**, 897 (1982).
82. M. Calos and J. Miller, *Cell* **20**, 579 (1980).
83. H. E. Varmus and P. R. Shank, *J. Virol.* **18**, 567 (1976); G. R. Ringold, P. R. Shank, K. R. Yamamoto, *ibid.* **26**, 93 (1978).
84. A. J. Flavell and D. Ish-Horowicz, *Nature (London)* **292**, 591 (1981).



85. I thank the many colleagues, both inside and outside retrovirology, who have offered suggestions about this manuscript or provided unpublished work for discussion, and I apologize to those whose work could not be cited within space limitations. R. Weinberg, J. Shapiro, Z. Hall, G. Martin, D. Ganem, R. Swanstrom, J. Coffin, G. Payne, and B. Cook provided special help with the text. This work was supported by grants from the National Institutes of Health and the American Cancer Society.