

Review

Retrovirus variation and reverse transcription: Abnormal strand transfers result in retrovirus genetic variation

Howard M. Temin

McArdle Laboratory, University of Wisconsin, 1400 University Avenue, Madison, WI 53706

Contributed by Howard M. Temin, May 28, 1993

ABSTRACT Human immunodeficiency virus variation is extensive and is based on numerous mistakes in reverse transcription. All retrovirus replication requires two strand transfers (growing point jumps) to synthesize the complete provirus. I propose that the numerous mistakes in reverse transcription are the result of this requirement for the two strand transfers needed to form the provirus.

Retroviruses vary at a notoriously high rate. For example, antibody- and drug-resistant human immunodeficiency virus type 1 (HIV-1) strains rapidly appear in infected and treated persons, and it is estimated that the HIV-1 sequences (*env* gene) in an infected person change $\approx 1\%$ per year (1, 2). Retroviruses recombine frequently, and simpler retroviruses often contain captured cellular protooncogenes (3, 4). [Simpler retroviruses contain only genes for virion proteins—*gag*, *pol*, and *env*. More complex retroviruses, like HIV-1, encode additional genes involved in regulation (5, 6).]

I propose that this high rate of retrovirus variation is a direct consequence of the requirement for transfer of the nascent strand at the reverse transcriptase growing point during retrovirus DNA synthesis. [A similar suggestion was made by Bebenek *et al.* (7) on the basis of studies with purified HIV-1 reverse transcriptase.] Of course, selection and other processes will finally determine the effects of this variation (8, 9). However, the high rate of genetic change in each replication cycle ensures that there is a wide field for selection and other processes. Reverse transcriptase is coded for by the retroviral *pol* gene and has associated RNase H activity, which may be required for one of the primer transfers (10–12). Because of its multiple roles—RNA-directed DNA synthesis, DNA-directed DNA synthesis, digesting RNA-DNA hybrid molecules, and strand transfers—reverse transcriptase must be quite flex-

ible in structure and action (H. Buc, personal communication). I shall also consider in this paper the hypotheses (i) that misincorporation promotes strand transfers (13) and (ii) that misincorporation accompanies strand transfers (14).

Retroviruses are a family of animal viruses that alternate their genetic material between RNA in the virion and DNA in the infected cell (15, 16). In addition, all retrovirus virions contain two identical molecules of virion RNA—the dimer RNA. The DNA form of a retrovirus, the provirus, is larger than the viral RNA form (Fig. 1). During reverse transcription, promoter/enhancer sequences found at the 3' end of viral RNA within the unique 3' RNA sequences (u3) are duplicated at the 5' end of viral DNA to form the U3 DNA sequences (capital U indicates DNA rather than RNA), and downstream polyadenylation sequences at the 5' end of viral RNA within the unique 5' RNA sequences (u5) are duplicated at the 3' end of viral DNA to form the U5 DNA sequences. These duplications result in the formation of long terminal repeats (LTRs) at both ends of the proviral DNA and provide autonomy in the cis-acting sequences needed for transcription and replication, which are the same for retroviruses. This autonomy results from the virus U3 sequences containing promoter/enhancer elements that are recognizable by cellular transcription factors and 3' LTR sequences that are recognizable by cellular polyadenylation factors.

The duplications in the LTRs are a result of two jumps, switches, or transfers of the reverse transcriptase growing point from one end of each template to the other end during replication (17) (Fig. 2). (In this article, I use the term strand transfers for these processes.)

Retrovirus genetic variation consists of base-pair substitutions, frameshifts, deletions, deletions with insertions, homologous recombination, and nonhomologous recombination. I shall discuss, in relation to the strand-transfer hypothesis, minus-strand and plus-strand DNA primer transfers and each of these types of genetic variation. All of these processes with the exception of some deletions with insertions and the two types of recombination involve only one molecule of the retrovi-

rus dimer RNA (J. S. Jones, R. W. Allan, and H.M.T., unpublished data).

Another way to state the strand-transfer hypothesis is that, instead of steady processive polymerization, the reverse transcriptase growing point frequently pauses and enters a metastable state, leaving this metastable state to continue polymerization either at the next base or at another base at a different location. This transfer can be a result of the growing point moving or of another portion of the template displacing the template at the growing point. Polymerization at locations other than the next base gives rise to all of these types of variation except some base-pair substitutions that require misincorporation before continuing polymerization. Other base-pair substitutions involve dislocation (18).

Primer Transfers

To synthesize the LTR and then have a primer for copying the bulk of the viral genome, retroviruses start minus-strand DNA synthesis near the 5' end of viral RNA using a base-paired cellular tRNA as a primer. This primer is annealed to the primer binding site (pbs) in viral RNA. After copying of the u5 and repeat (r) regions, the nascent minus-strand DNA transfers to the r sequences at the 3' end of the same molecule of viral RNA, next to the poly(A) sequence. RNase H activity, associated with the reverse transcriptase molecule, may be involved in this transfer, removing the RNA r and u5 sequences (10–12). [Because a retrovirus virion contains two molecules of viral RNA the minus-strand primer DNA could theoretically transfer to the same molecule or to the other one (19). Recent work has clearly established that, in the absence of breaks, the minus-strand primer DNA always transfers from the 5' to the 3' end of the same RNA molecule, designated intramolecular minus-strand

Abbreviations: HIV-1, human immunodeficiency virus type 1; LTR, long terminal repeat; u3 (U3), unique 3' RNA (DNA); u5 (U5), unique 5' RNA (DNA); pbs (PBS), primer binding site in RNA (DNA); ppt (PPT), polypurine tract in RNA (DNA); r (R), repeat region in RNA (DNA).

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.

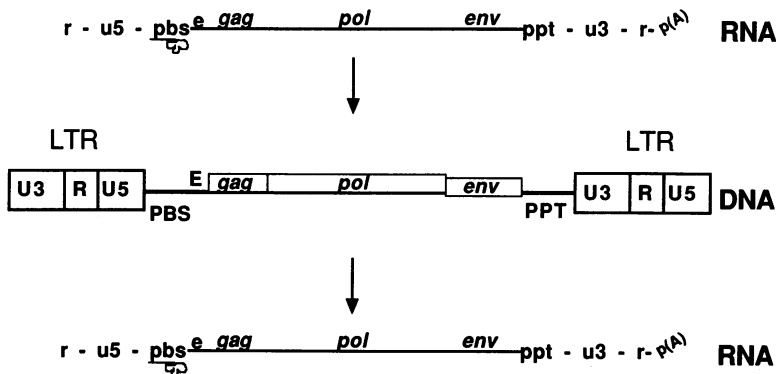


FIG. 1. RNA and DNA genomes of a simpler retrovirus. Infecting viral RNA is reverse transcribed to viral DNA, which is transcribed to form progeny viral RNA. *gag*, *pol*, and *env* are genes for virion proteins. The other named genes act in cis and are described in the text. p(A) is a polyadenylate tail, which is not reverse transcribed. The tRNA primer is shown annealed to pbs. In the virion, there are two copies of the genomic RNA.

DNA primer transfer (J. S. Jones, R. W. Allan, and H.M.T., unpublished data).] Other recent work indicates that the minus-strand and plus-strand DNA primer transfers take place during the elongation or synthesis phase of DNA synthesis rather than at the end of the template (refs. 13, 17, 20, and 44; G. Pulsinelli and H.M.T., unpublished data; J. Zhang and H.M.T., unpublished data). Therefore, minus- and plus-strand strong stop DNAs, as they are traditionally termed, are not the usual intermediates for the reverse transcriptase growing point primer transfers. The transferred minus-strand DNA can then be used as a primer to copy the viral RNA up to the end of the remaining

5' RNA sequences, thus generating most of the minus-strand DNA. After the minus-strand DNA is elongated through the R U3 regions, a reverse transcriptase RNase H activity cleaves the RNA template near its 3' end just after a polypurine tract (ppt). The 3' end of the viral RNA ppt forms a primer for plus-strand DNA synthesis. Elongation from this point occurs. At some point during the copying of U5 and the tRNA primer, the reverse transcriptase growing point transfers to the 3' end of the minus-strand DNA molecule, annealing to the complementary PBS sequences. [In ≈20% of cases the strand transfer happens upon reaching the end of the pbs sequences in the minus-strand tRNA primer (G.

Pulsinelli and H.M.T., unpublished data). Much more rarely, the transfer is to DNA copied from the other molecule in the virion, designated intermolecular plus-strand primer transfer (J. S. Jones, R. W. Allan, and H.M.T., unpublished data). The low rate of intermolecular transfers may reflect the low rate of synthesis of complete minus-strand DNA from both RNA molecules in one virion.]

Base-Pair Substitutions

Base-pair substitution mutations involve dislocation mutagenesis or misincorporation by reverse transcriptase at the growing point, followed by polymerization beyond the misincorporation. There are definite hot spots for substitution mutations by reverse transcriptase as there are with other DNA polymerases (7, 18, 21–24). The retroviral reverse transcriptase does not have any error-correcting function (25, 26), perhaps because it lacks necessary accessory proteins and nuclease activities (27).

I propose that after misincorporation, the surrounding sequence determines whether or not there is polymerization at the base adjacent to the mismatch, thereby maintaining the reading frame, or transfer to another position on the template, forming a deletion, insertion, or recombinant. Reverse transcriptases appear to differ from other DNA polymerases more by the frequency of extension from a misincorporation than from the frequency of misincorporation itself (18, 22–24). This observation indicates that the reverse transcriptase can add some base-paired nucleotides relatively efficiently to a nucleotide that is not base-paired.

Frameshifts

Frameshifts, the additions or subtractions of 1 base, commonly occur during retrovirus replication, as in all other replication, within runs of a single nucleotide, and their frequency increases as the runs become longer (21, 28). Thus, with spleen necrosis virus, a simpler avian retrovirus, runs of 9 or 10 thymines or of 9 or 10 adenines result in frameshifts in 20–40% of replications (21, D. P. W. Burns and H.M.T., unpublished data). Frameshifts usually add or delete 1 base from the run itself. Dislocation mutagenesis (7, 18), where the frameshift involves a base-pair substitution incorporating the nucleotide next to the run, is a good illustration of the process.

Deletions

Deletions in retrovirus replication, as in many other systems, usually involve removal of nucleotides between small direct repeats (Fig. 3) (29–31). In addition, misincorporation can lead to deletions

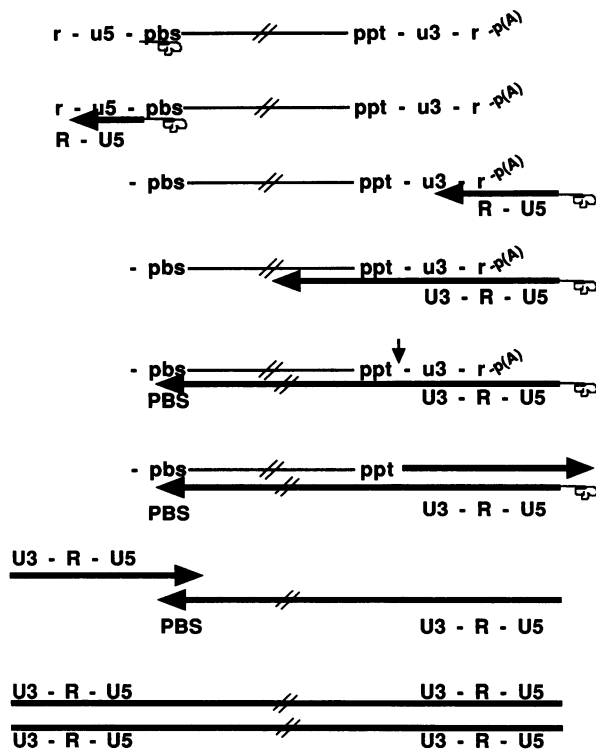


FIG. 2. Synthesis of retrovirus DNA. Thin lines, RNA; thick lines, DNA. Complete description is given in text.

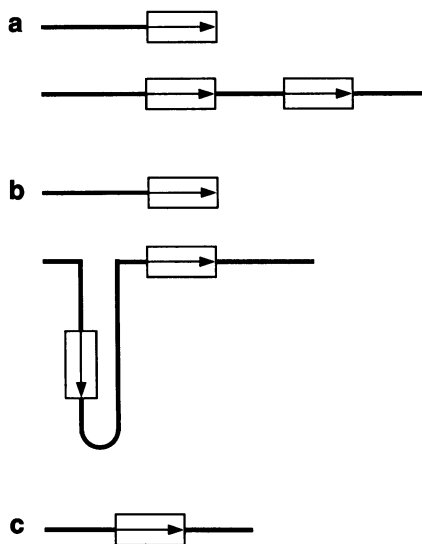


FIG. 3. Simple deletion. Misalignment to an identical sequence (box with arrow) on the template can result in deletion as illustrated.

when the reverse transcriptase growing point scans downstream for an identical sequence rather than polymerizing through the misincorporation (13). However, the high rate, almost 100%, of deletions of long tandem repeats makes it unlikely that misincorporation is required for all deletions (32).

Deletions with Insertions

In retrovirus replication, it is not uncommon to find extra nucleotides inserted in the deletion, substituting for the deleted bases (31). Analysis of these inserted sequences reveals that they result from the reverse transcriptase growing point transferring to a small region of sequence identity on another template molecule in the virion (refs. 13, 31, 37, and 38; L. M. Mansky and H.M.T., unpublished data). This new template is the RNase H-digested plus-strand virion RNA or another RNA that is encapsidated in the virion. A second reverse transcriptase growing point transfer is then required to return to the original template. Often several abnormal strand transfers are required before returning to the original template.

Dimer RNA

All of the processes discussed above—primer transfers, base-pair substitutions, frameshifts, deletions, and deletions with insertions—involve only one RNA template molecule. The second viral RNA molecule does not seem to be required for normal reverse transcription (J. S. Jones, R. W. Allan, and H.M.T., unpublished data). However, recombination between the two strands of RNA provides a strong positive selective advantage for retroviruses, allowing them to repair breaks in the RNA and to exchange nucleic acid

sequences. Retroviruses do not have a pool of replicative intermediates or other molecules that can recombine (15). Thus, they have evolved dimer virion RNA to provide substrates for recombination (35).

Homologous Recombination

Homologous recombination during retrovirus replication almost always occurs during the original minus-strand DNA synthesis (36, 37). Homologous recombination results from the reverse transcriptase growing point transferring to an identical sequence on the other RNA molecule of the dimer RNA. Homologous recombination can be the result of usual reverse transcriptase growing point transfer, called copy-choice, or the result of an RNA break that forces the reverse transcriptase growing point to transfer, called forced copy-choice (38, 39). It has also been proposed that misincorporation is necessary for recombination (14). This hypothesis is based on experiments with purified HIV-1 reverse transcriptase, which showed that when the reverse transcription growing point transfers from RNA to RNA at a blunt-ended RNA-DNA hybrid molecule there is addition of an untemplated nucleotide. Since such reverse transcription growing point transfers do not usually occur at a blunt end, except possibly during forced copy-choice recombination, the hypothesis is unlikely to apply generally. In fact, when a modification of the system described by Zhang and Temin (34) was used, direct sequencing of recombinants in a region of sequence identity in the midst of nonidentical sequences showed no base-pair substitutions in 22 of 22 recombinants (J. Zhang and H.M.T., unpublished data).

Thus, an earlier misincorporation is not necessarily involved in homologous recombination. This is not surprising, since the rate of recombination is so high that the rate of base-pair substitution would be too high for viability if misincorporation were a necessary precursor for homologous recombination (refs. 40 and 41; J. S. Jones, R. W. Allan, and H.M.T., unpublished data).

Nonhomologous Recombination

When the retrovirus virion contains nonviral RNA sequences, the reverse transcriptase growing point can transfer to this RNA. When the nonviral RNA sequences are in a chimeric RNA molecule, a single reverse transcriptase growing point transfer will result in formation of a virus capable of replication with helper virus or helper cell proteins. The chimeric RNA usually results from read-through of transcription past the normal retrovirus polyadenylation sequences. The transfer is usually to a short region of

sequence identity in the otherwise nonidentical sequence (33, 34). This process has given rise to naturally occurring, highly oncogenic retroviruses, which contain an insertion of cellular protooncogene sequences (34, 42).

Increasing the size of the region of sequence identity in the midst of an otherwise nonidentical sequence increases the rate of such nonhomologous recombination (J. Zhang and H.M.T., unpublished data). At its maximum, however, the rate of such nonhomologous recombination is 1000 times less than that of homologous recombination. This result, together with other evidence that the relative location of the regions of sequence identity in the midst of otherwise nonidentical sequences affects the recombination rate (J. Zhang and H.M.T., unpublished data), indicates a higher order of virion organization that is not yet described and that can influence the reverse transcriptase growing point transfers.

Is One Property of Reverse Transcriptase Responsible for All of These Processes?

In this article, I have suggested that the necessity for the reverse transcriptase growing point to transfer from one place on the template to another place on the template, in order to form the primer molecules for much of the DNA synthesis and LTRs, underlies all of these processes of genetic variation. Some evidence in favor of one underlying process comes from a comparison of rates of mutations and types of mutations in two different viruses. As mentioned earlier, spleen necrosis virus is a simpler avian retrovirus, similar to murine leukemia viruses. Bovine leukemia virus is a more complex retrovirus, similar to human T-cell leukemia viruses. The overall rate of forward mutations in bovine leukemia virus replication is significantly less than the rate for spleen necrosis virus (L. M. Mansky and H.M.T., unpublished data). However, the distribution of different types of mutations is the same for both bovine leukemia and spleen necrosis viruses (L. M. Mansky and H.M.T., unpublished data). Thus, the bovine leukemia virus reverse transcriptase growing point seems to have a lower propensity to transfer during normal viral replication than the spleen necrosis virus reverse transcriptase growing point, but the results of the transfers are similar.

Attempts to Measure Kinetic Parameters

Numerous attempts have been made to model these processes in cell-free systems with purified reverse transcriptase and defined templates (for a recent review, see ref. 18). The results are similar to those found in experiments that ana-

lyze a single cycle of retrovirus replication (but see ref. 43).

All experiments with defined templates run into the inescapable problem of local sequence effects, which I have already indicated are an important feature of the reverse transcriptase growing point transfers. Thus, any experimental rates are the average for a particular template. [It should be noted that the comparisons of spleen necrosis and bovine leukemia viruses discussed above was done with the exact same template but in the opposite orientation (L. M. Mansky and H.M.T., unpublished data).]

Given this problem, we have measured the rates of each of these steps in a single cycle of replication of a simpler avian retrovirus. The rates are expressed as mutations per base pair per replication cycle and are as follows: base-pair substitutions, 1×10^{-5} ; frameshifts, 1×10^{-6} ; deletions, 2×10^{-6} ; deletions with insertions, 1×10^{-6} ; homologous recombination, 2×10^{-4} ; non-homologous recombination, 5×10^{-8} ; recombination of a limited sequence identity in the midst of otherwise nonidentical sequence, 6×10^{-6} (refs. 21, 31, and 41; J. S. Jones, R. W. Allan, and H.M.T., unpublished data; J. Zhang and H.M.T., unpublished data).

In terms of the strand-transfer hypothesis, the most informative rates are perhaps the rates of frameshifts. In the formation of a frameshift within a run of 10 thymines or 10 adenines, the sum of the rates of formation of the metastable state and the probability of continuing misincorporation is $\approx 20\%$ (ref. 21; D. P. W. Burns and H.M.T., unpublished data). A simple interpretation of this result would be that there is a 40% probability of the reverse transcriptase growing point entering the metastable state for each 10 thymines or adenines incorporated and a 50% probability of slippage within the run. (I assume that the probability of a mistaken polymerization is $< 50\%$.) The lower rates of the genetic processes other than frameshifts discussed in this article would reflect the lower probability that the reverse transcriptase growing point would make an inappropriate transfer to resolve the metastable state in the absence of a nearby run of the same nucleotide.

In contrast, the rate of base-pair substitution would first include misincorporation, which would induce the metastable state of the reverse transcriptase growing point, and then resolution of the metastable state by readthrough or transfer controlled by the local and nearby sequences.

Summary

Retroviruses developed reverse transcriptase growing point transfers to form a provirus that is autonomous with respect to cis-acting sequences for transcription; that is, the enhancer/promoter

sequences in viral DNA are copied from the viral RNA genome. This strand-transfer process can occur during polymerization of internal sequences as well as during primer synthesis. The rate of transfer and the genetic effects of the transfers depend on the local nucleotide sequence, distant or foreign sequences, and the sequence of the second RNA strand in the dimer. On the average, a simpler retrovirus seems to have for each round of replication at least one additional transfer that can have a genetic effect in addition to the two transfers required to make the LTRs and primers.

Retroviruses have made a virtue of necessity by using the reverse transcriptase growing point transfer mechanism both in their replication and in their high rate of mutation. The AIDS epidemic is just one striking expression of this ability.

I thank K. Boris-Lawrie, S. Broder, H. Buc, D. Burns, J. Jones, T. Kunkel, D. Loeb, L. Mansky, G. Pulsinelli, B. Sugden, H. Varmus, S. Yang, and J. Zhang for useful comments on the manuscript. The research in H.M.T.'s laboratory is supported by U.S. Public Health Service Grants CA-22443 and CA-07175 from the National Cancer Institute. H.M.T. is an American Cancer Society Research Professor.

- Albert, J., Abrahamsson, B., Nagy, K., Aurelius, E., Gaines, H., Nyström, G. & Fenyö, E. M. (1990) *AIDS* **4**, 107–112.
- Myers, G. & Pavlakis, G. N. (1992) in *The Retroviridae*, ed. Levy, J. A. (Plenum, New York), Vol. 1, pp. 51–106.
- Hu, W. S., Pathak, V. K. & Temin, H. M. (1993) in *Reverse Transcriptase*, eds. Skalka, A. M. & Goff, S. P. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 251–274.
- Bishop, J. M. & Varmus, H. (1985) in *RNA Tumor Viruses: Molecular Biology of Tumor Viruses*, eds. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., Vol. 2, pp. 249–356.
- Cullen, B. R. (1991) *J. Virol.* **65**, 1053–1056.
- Temin, H. M. (1992) in *The Retroviridae*, ed. Levy, J. A. (Plenum, New York), Vol. 1, pp. 1–18.
- Bebenek, K., Abbotts, J., Roberts, J. D., Wilson, S. H. & Kunkel, T. A. (1989) *J. Biol. Chem.* **264**, 16948–16956.
- Coffin, J. (1993) in *Reverse Transcriptase*, eds. Skalka, A. M. & Goff, S. P. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 445–479.
- Sánchez-Palomino, S., Rojas, J. M., Martínez, M. A., Fenyö, E. M., Nájera, R., Domingo, E. & López-Galíndez, C. L. (1993) *J. Virol.* **67**, 2938–2943.
- DeStefano, J. J., Mallaber, L. M., Rodriguez-Rodriguez, L., Fay, P. J. & Bambara, R. A. (1992) *J. Virol.* **66**, 6370–6378.
- Tanese, N., Telesnitsky, A. & Goff, S. P. (1991) *J. Virol.* **65**, 4387–4397.
- Luo, G. & Taylor, J. (1990) *J. Virol.* **64**, 4321–4328.
- Pulsinelli, G. & Temin, H. M. (1991) *J. Virol.* **65**, 4786–4797.
- Peliska, J. A. & Benkovic, S. J. (1992) *Science* **258**, 1112–1118.

- Weiss, R., Teich, N., Varmus, H. & Coffin, J. eds. (1985) *RNA Tumor Viruses: Molecular Biology of Tumor Viruses* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., Vol. 2.
- Coffin, J. M. (1990) in *Virology*, eds. Fields, B. N., Knipe, D. M., Chanock, R. M., Hirsch, M. S., Melnick, J. L., Monath, T. P. & Roizman, B. (Raven, New York), Second Ed., Vol. 1, pp. 1437–1500.
- Tekesnitsky, A. & Goff, S. P. (1993) in *Reverse Transcriptase*, eds. Skalka, A. M. & Goff, S. P. (Cold Spring Harbor Lab. Press, Plainview, NY), 1st Ed., Vol. 1, pp. 49–83.
- Bebenek, K. & Kunkel, T. A. (1993) in *Reverse Transcriptase*, eds. Skalka, A. M. & Goff, S. P. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 85–102.
- Panganiban, A. T. & Fiore, D. (1988) *Science* **241**, 1064–1069.
- Lobel, L. I. & Goff, S. P. (1985) *J. Virol.* **53**, 447–455.
- Pathak, V. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6019–6023.
- Preston, B. D., Poiesz, B. J. & Loeb, L. A. (1988) *Science* **242**, 1168–1171.
- Mendelman, L. V., Petruska, J. & Goodman, M. F. (1990) *J. Biol. Chem.* **265**, 2338–2346.
- Ricchetti, M. & Buc, H. (1990) *EMBO J.* **9**, 1583–1593.
- Battula, N. & Loeb, L. (1976) *J. Biol. Chem.* **251**, 982–986.
- Roberts, J. D., Bebenek, K. & Kunkel, T. A. (1988) *Science* **242**, 1171–1173.
- Sancar, A. & Hearst, J. (1993) *Science* **259**, 1415–1420.
- Ripley, L. S. (1990) *Annu. Rev. Genet.* **24**, 189–213.
- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77–84.
- Streisinger, G. & Owen, J. (1985) *Genetics* **109**, 633–659.
- Pathak, V. K. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6024–6028.
- Rhode, B. W., Emerman, M. & Temin, H. M. (1987) *J. Virol.* **61**, 925–927.
- Zhang, J. & Temin, H. M. (1993) *Science* **259**, 234–238.
- Zhang, J. & Temin, H. M. (1993) *J. Virol.* **67**, 1747–1751.
- Temin, H. M. (1991) *Trends Genet.* **7**, 71–74.
- Hu, W.-S. & Temin, H. M. (1992) *J. Virol.* **66**, 4457–4463.
- Jones, J. S., Allan, R. W. & Temin, H. M. (1993) *J. Virol.* **67**, 3151–3158.
- Coffin, J. M. (1979) *J. Gen. Virol.* **42**, 1–26.
- Xu, H. & Boeke, J. D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8553–8557.
- Nowak, M. (1990) *Nature (London)* **347**, 522.
- Hu, W.-S. & Temin, H. M. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1556–1560.
- Swain, A. & Coffin, J. M. (1992) *Science* **255**, 841–845.
- Varela-Echavarria, A., Garvey, N., Preston, B. D. & Dougherty, J. P. (1992) *J. Biol. Chem.* **267**, 24681–24688.
- Ramsey, C. M. & Panganiban, A. T. (1993) *J. Virol.* **67**, 4114–4121.