Journal of Cellular Biochemistry 19:293–304 (1982)
Tumor Viruses and Differentiation 75–86

Genetic Recombination in Avian Retroviruses

A.M. Skalka, L. Boone, R. Junghans, and D. Luk

Department of Cell Biology, Roche Institute of Molecular Biology, Nutley, New Jersey 07110

The avian retroviruses—and probably other retroviruses as well—undergo a variety of recombinational events with relatively high efficiency. An understanding of the molecular basis of these events should provide insight into the important biological properties these agents exhibit when they become integrated into somatic or germ-line host cells, when they exchange genetic information among themselves, or when they transduce host cell genes. In this article we review molecular models for homologous recombination, against a background of the other types of recombination events that are typical of these viruses. It seems probable that the retroviruses will provide useful models for analysis of a variety of DNA rearrangements known to occur in eukaryotic cells.

Key words: reverse-transcription, strand-displacement synthesis, heteroduplex DNA, DNA H-structures, proviral integration, homologous recombination, transduction, recombination models, RNA tumor viruses

The recent expansion of our knowledge of eukaryotic gene structure and function has provided new insight into the biological importance of DNA rearrangement. We now know of many instances in which specific gene expression is regulated by apparently programmed recombination reactions. Some examples are the rearrangements leading to formation of antibody genes, transposable elements that govern the development of pigment and kernel size in maize, and yeast mating-type casettes [1]. In addition, it now seems probable that similar sorts of DNA rearrangements, involving sizeable stretches of DNA and perhaps similar molecular mechanisms, are a driving force in evolution [2].

This report will consider various aspects of the molecular biology of avian retroviruses. However, as outlined below, one major interest in this system stems from its possible use as a model for the mechanisms and biochemistry involved in genetic recombination in eukaryotes.

L. Boone is now at the National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.

R. Junghans is now at the University of Miami School of Medicine, R123, PO Box 016960, Miami, FL 33101.

Received June 10, 1982; accepted July 28, 1982.

0730-2312/82/1903-0293$03.50

HIGH FREQUENCY EVENTS

The avian retroviruses undergo two types of recombination with relatively high efficiency: (1) *Homologous recombination* is the type of genetic exchange that occurs between largely homologous retroviral genomes. The frequency of this event has been estimated at approximately 10% per mixed infection [3], the highest known value for any RNA or DNA virus. Genetic evidence suggests that this reaction requires the formation of heterozygous virions, with the two parental molecules encapsidated in the same diploid particle. The actual recombination event takes place after infection with the heterozygote, probably at the time of reverse-transcription. This review will focus on possible mechanisms of homologous recombination and studies from our own laboratory that have allowed us to generate a new model for this event. (2) Retroviral *integration* into host cell DNA involves the formation of new linkages between nonhomologous viral and host DNAs. No estimates on the absolute frequency of this event are available, but it is assumed to be efficient since integration is an obligatory step in the virus's life cycle. Studies of integration have generally involved two approaches. The first includes structural analysis of unintegrated and integrated viral DNAs. Numerous reports of such analyses, including those from our own laboratory [4,5], have revealed striking similarities between the retroviruses and prokaryotic and eukaryotic transposable genetic elements (see [6] for review). The second approach, aimed at clarifying the biochemistry through utilization of molecular clones generated in the structural analyses, is just beginning [7].

LOW FREQUENCY EVENTS

Retroviruses undergo a third type of recombination that, although apparently infrequent, is of substantial biological importance. This event—viral *transduction* of host DNA—results from another type of interaction between host and viral genetic material whereby host genes are eventually incorporated into viral genomes. In this new arrangement, the host genes become subject to viral control elements. Their resulting aberrant expression is apparently the basis for oncogenicity of the acute transforming retroviruses (see [8] for review). Some recent studies provide clues to the possible mechanism of this event [9–11] that may be related to the first type of recombination mentioned above.

REPLICATION OF THE VIRAL GENOME

Figure 1 shows a simplified diagram summarizing the major stages in the replication of retroviral genomes (see [12] for a detailed review). The genetic material in a retrovirus particle is a 70S RNA complex of two 35S genomes encapsidated in a protein core, surrounded by a cell-derived membrane envelope. The two RNA genomes in this diploid particle can serve as messenger RNA; therefore, by convention, they are designated to be of positive strand polarity.

Studies of retroviral replication in infected cells have revealed the following sequence of events: Reverse-transcription of a single-stranded viral RNA genome (A) results in the formation of a double-stranded DNA molecule (B) with one intact and one interrupted strand and with long terminal repeats (LTRs) of sequences unique to the 5′and 3′ends of the viral RNA. The intact strand is complementary to the viral

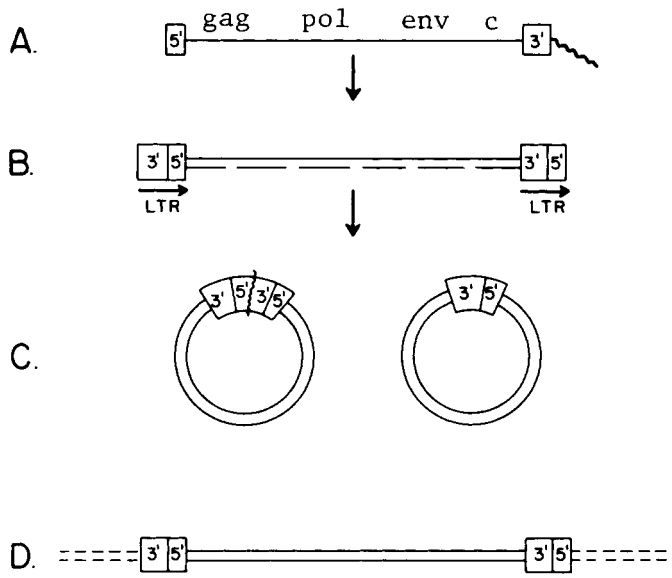

Fig. 1. Replication of retroviral genomes. (A) The viral genome, a single-stranded RNA molecule of positive (mRNA) polarity. Relative order of essential genes, *gag*, *pol*, and *env* are noted above the line. The wavy line is the poly (A) tract at the 3′ terminus. Boxes represent 5′ and 3′ terminal sequences. (B) Linear double-stranded viral DNA synthesized in the cytoplasm of infected cells by reverse-transcription. The gene order is colinear with the viral RNA genome. The DNA termini have direct repeats (LTRs) derived from the boxed sequences unique to 5′and 3′ends of the viral RNA. (C) Circular viral DNA, covalently closed circular molecules found in the nucleus of infected cells, presumably formed by ligation of the termini of linear molecules. Two species that differ by the number of LTRs have been detected. The wavy line represents the junction between the tandem LTRs in molecules that contain two copies. (D) Integrated proviral DNA, covalently linked to cellular DNA (dashed lines). The gene order is colinear with that of the unintegrated linear molecule (B). LTRs are found at the junctions with cellular DNA at the site of integration. Transcription of the integrated provirus by cellular RNA polymerase II, starting in the left and ending in the right LTR, produces viral RNA genomes and viral mRNAs as well. Adapted from [4].

genome and, therefore, of negative polarity. The discontinuous strand is copied from the negative DNA template presumably using oligoribonucleotide primers formed during the degradation of viral RNA by the RNAseH activity of the viral reverse transcriptase [13]. Unintegrated circular DNA molecules (C), found somewhat later in the nucleus of infected cells, contain either one or two copies of the LTRs. In the latter structures, the LTRs appear as tandem repeats, presumably the result of ligation of linear molecules. LTRs also flank the integrated provirus at the junction between cellular and viral DNA. It is not yet known which form is the substrate for integration. Expression of integrated provirus (D) results in production of progeny RNA genomes identical in structure to the parent genome (A), and smaller messenger RNA molecules as well.

Study of the mechanism of formation of the early intermediates shown in Figure 1 has been facilitated by the fact that purified virions possess all of the enzymatic machinery necessary to carry out the first steps in the process. To initiate synthesis it

is merely necessary to gently disrupt the virus envelope and permit deoxynucleotides and small molecular weight cofactors to enter. In our hands the DNA products formed seem to resemble the linear cytoplasmic form [14,15a,15b]. We have proposed, therefore, that cellular components are required for later steps in the replication program and that our in vitro system provides us with intermediates that are the substrates for these subsequent reactions. Recent detailed analysis of these intermediates has revealed a structure that was somewhat unexpected.

Figure 2A provides additional details regarding the structure of a full-length viral DNA molecule with LTRs (as in Fig. 1B). The somewhat complicated series of reactions leading to its formation have been summarized elsewhere [16]. The unexpected feature discovered by analysis of the products of our in vitro system was the single-stranded branches, illustrated in Figure 2B. We have proposed that these are generated by strand-displacement that occurs as the leading edge of one positive strand segment meets the tail end of its neighbor. Biochemical analysis of linear cytoplasmic viral DNA made in infected cells indicates that similar branched structures exist in vivo and, therefore, are not an artifact of our system [15a,15b]. We know that integrated provirus is an intact structure and we have proposed two possible pathways for its formation. In one (Fig. 2C), the terminal positive segment gradually replaces all of the others. In the second (Fig. 2D), cellular nuclease and ligase repair

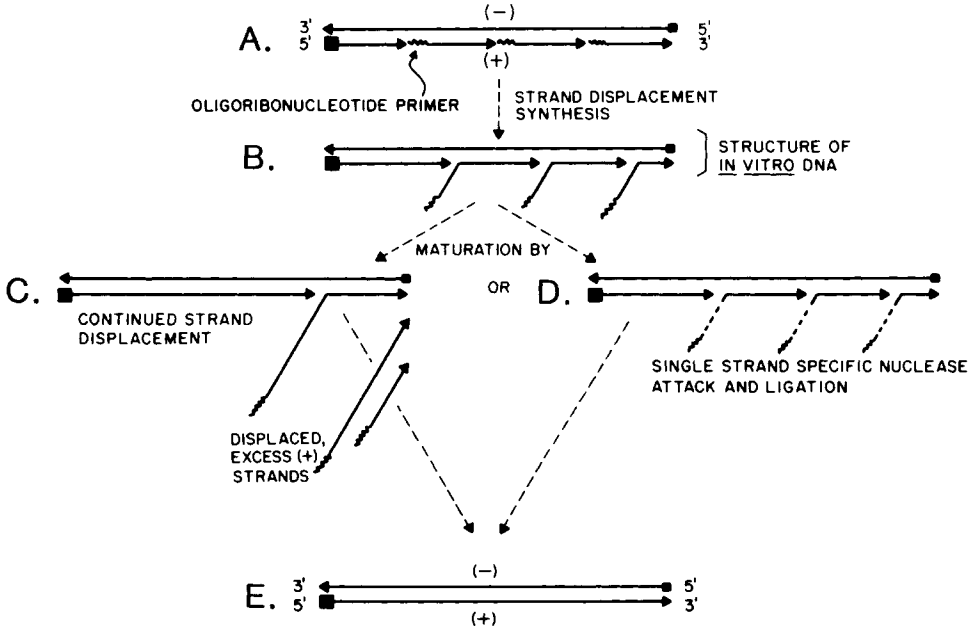

Fig. 2. Schematic representation of strand displacement synthesis and possible mechanism of maturation to completely intact duplex DNA. The numerous steps that lead to the linear structure shown at the top have been omitted here for clarity. Synthesis is in the direction of the arrows, and wavy lines represent presumed oligoribonucleotide primers. The small box at the beginning of the negative strands represents negative strand strong-stop sequences. The larger box at the 5′ end of the terminal positive strand represents positive strand strong-step sequences. Adapted from [15a,15b].

the intermediate. Although a small proportion of full length positive strands have been detected in DNA products of murine viruses incubated in vitro [17], our analysis of the avian product has revealed few, if any, intact positive strands even after long periods of synthesis. Thus, we believe this second pathway to be most likely.

The strand-displacement synthesis, which apparently takes place during reverse transcription, is of biological interest for at least two reasons. First, it helps to explain several unresolved steps in the generation of the full length linear structures—in particular the LTRs—and this has been discussed elsewhere [16]. Second, and of importance in the current context, is the possible relevance to homologous recombination.

Single-Strand Branches and Recombination

There is convincing evidence in prokaryotes and supportive evidence in eukaryotes that DNA recombination is accomplished through formation of a "heteroduplex joint." Figure 3C, which is adapted from a review by Radding [18], shows such an intermediate, and one mechanism proposed for its formation. Strand separation—in this scheme accomplished by displacement synthesis (Fig. 3A)— is a critical first step in the procedure. This step and the subsequent formation and removal of D-loops have all been demonstrated in vitro using proteins purified from prokaryotic sources. The similarity between the structure proposed for the branched linear retroviral DNA molecule and that in Figure 3A is striking. The fact that two retroviral DNA molecules might be synthesized concurrently in the same diploid particle made it seem likely that interactions between the two molecules could take place. Because the frequency of homologous recombination is very high in avian retrovirus we were encouraged to look for such interactions in the electron microscope.

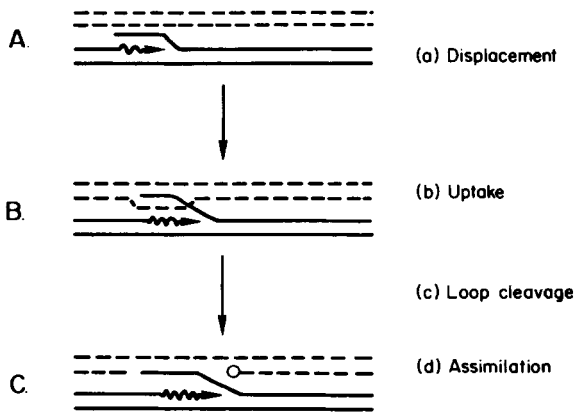

Fig. 3. Formation of a "heteroduplex joint," a key intermediate in general recombination. (A) The event may be initiated by strand-displacement synthesis at a nick in the donor DNA. Other methods of generating single strands can also be envisioned; the important element is that a free single strand is available for the next step. (B) Uptake of the single strand into a recipient molecule. In E coli this step is facilitated by the *rec* A protein. The displaced recipient strand forms a D-loop that may be digested by single strand-specific nucleases. (C) The donor strand is now stably assimilated into the recipient molecule in a "heteroduplex joint." Branch migration amd molecular rotation can then generate the double-strand exchange proposed for a Holliday structure. These and subsequent events are discussed in detail by Radding [18] from which this figure was adapted.

Electron Microscope Analysis of Retroviral DNA

A branched linear structure synthesized in permeabilized avian retroviral particles is shown in Figure 4. Under the conditions employed [14,15a,15b], synthesis of the duplex is quite rapid and completed by about 1 hour at 41°C. Branches, which occur at apparently random locations, are synthesized at a slower rate, but continually thereafter growing longer with increasing time. The molecule in Figure 4 is from a product of 18 hours; some of the branches are as long as 2 kb. The single-stranded nature of the branches was confirmed by spreading the product in the presence of bacteriophage T4 gene 32 protein (Fig. 5 and [19]). These studies showed preferential binding of the gene 32 protein and subsequent thickening of the branches as expected for single-stranded DNA [20].

Among the products of synthesis, we observed linear duplexes linked together by single-strand bridges at points equidistant from their termini; we have called these novel forms H structures. Detailed electron microscopy (EM) and gel analysis of viral DNA products showed that the proportion of larger DNA molecules (defined as greater than half genome-length) contained in H structures was approximately 40% at 1 hr and remained roughly the same thereafter. Molecules that looked like H structures were also seen at the earliest time points, but then nascent DNA chains were often too short to make the identifications unambiguous. The observations seemed most consistent with a mechanism whereby H structures were being formed continuously as the two genomes were reverse-transcribed, with the probability of formation increasing as DNA synthesis proceeded. Long duplexes complexed with shorter DNA molecules were common (Fig. 6). It seems probable that the short molecules originated from the broken RNA genomes that are known to be present in large proportion in retroviral preparations [21]. EM and gel data from the preparations analyzed indicated that only 20–30% of the DNA was reverse-transcribed from full length genomes. Thus, as calculated (from Poisson distribution), there was an average of 1.1–1.5 breaks per RNA genome. It is worthwhile to note that if these H structures are recombinational intermediates, as long as one genome is full length the second genome need not be intact to contribute to the genetic pool.

From calculations based on the value of percentage of full-length DNA and percentage of H structures containing larger than half-lengh duplexes, it can be estimated that approximately 10% of all viral DNA molecules may be involved in H structure formation. Finally, in one survey of 20 H structures, each selected to contain one genome-length duplex, 8 (40%) were found to contain a second genome length duplex. This result suggests that both RNA genomes within a single diploid virus can be full reverse-transcribed. This would rule out certain models of reverse transcription (see for example [22]) where a single viral DNA molecule is produced at the expense of two RNA genomes. It also suggests that with some (probably low) frequency a single virus could give rise to a clone of cells containing two integrated proviruses, a fact that might complicate genetic analyses as will be discussed below.

Displacement/Assimilation Model of Recombination

Two distinctive features of retroviral H structures—their relatively high frequency in the population and the fact that interactions involve homologous regions of viral DNA duplexes—are consistent with intermediates in homologous recombination. We have, therefore, proposed a mechanism for their formation called the displacement/assimilation model to identify the two critical steps in the process. The first step

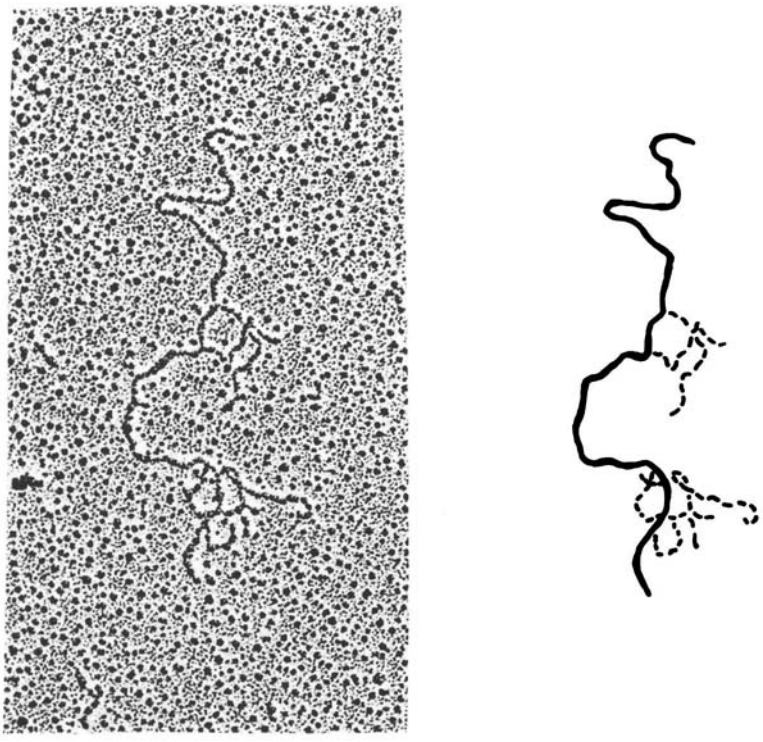

Fig. 4. Electron micrograph of linear retroviral DNA with several branches. Product was obtained from an 18-hr synthesis. Conditions are described in [19].

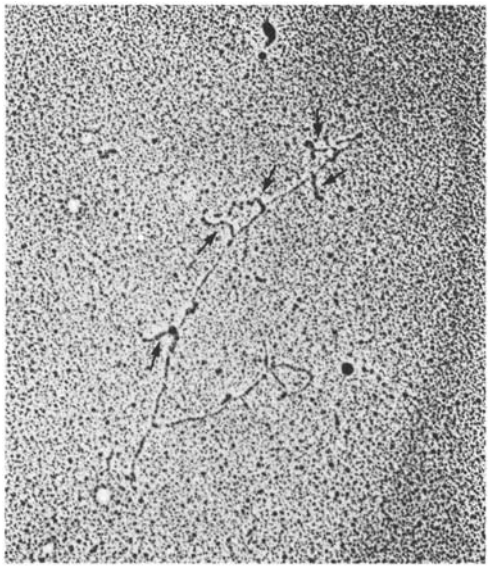

Fig. 5. Electron micrograph of linear retroviral DNA spread in the presence of T4 gene 32 protein. Product was obtained from 4-hr synthesis. Arrows point to thickened branches where the gene 32 protein has bound preferentially. Conditions are described in [19].

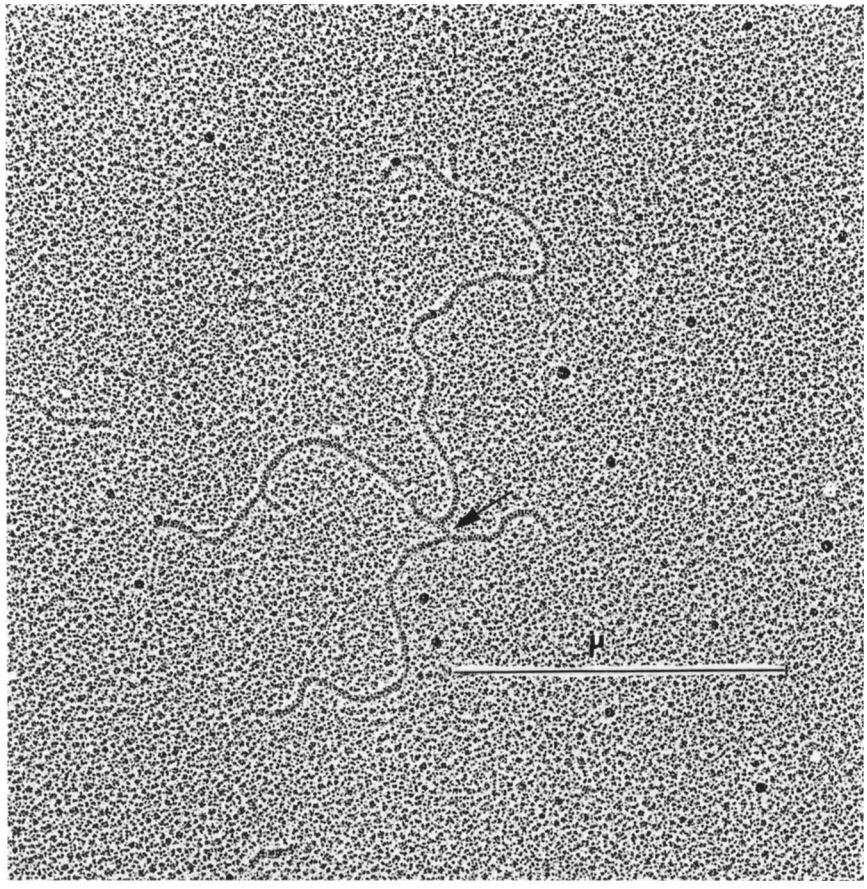

Fig. 6. Electron micrograph of an H structure. Product was obtained from a 4-hr synthesis. In this example, one of the duplexes is full length. The second is shorter and presumably derived from a damaged genome. The bridge (arrow), occurs at a point equidistant (approximately 3.3 kb) from what the simplest interpretation suggests, is the 3′ end of both genomes. Conditions and additional details are described in [19].

(Fig. 7) is formation of single-strand branches by positive strand displacement synthesis. In the second step, DNA from the branch is assimilated into the "recipient" duplex by annealing to a homologous regions of the negative strand at a transient gap in which positive strand synthesis has not yet occured. Failure to observe D-loops or whiskers in the EM at the origin of bridges makes it seem unlikely that strand assimilation occurs by "invading" homologous duplex regions as proposed in Radding's scheme (Fig. 3). Other differences which relate to the length of branches and proposed subsequent reactions have been discussed in detail elsewhere [19]. As with branches on simple linear duplexes, we presume that the bridges of H structures are eventually digested by cellular nucleases and the resulting "free" duplexes repaired by cellular ligases. The resolved products would be a donor parental duplex and a heteroduplex recombinant. The genetic consequences of these events would then depend on the fate of the heteroduplex. If the mismatch were extensive, its repair

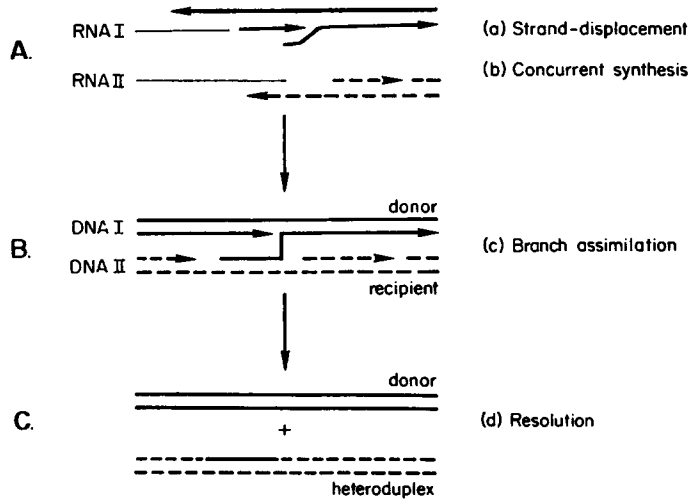

Fig. 7. Displacement/assimilation model of recombination. (A) As in the general model of Figure 3, this exchange is initiated by the generation of a single-strand branch in the donor molecule (I), via strand-displacement synthesis. Concurrent DNA synthesis on a second RNA genome (II) in the virion core produces a transient single-strand gap into which the single strand branch may be assimilated. (B) An H structure that is conceptually analogous to the "heteroduplex joint" in Figure 3C. This structure accumulates and appears to be stable in virus cores in vitro and presumably also in the cytoplasm of infected cells in vivo. (C) Resolution of the H structures is presumed to take place after exposure to single-strand specific nucleases and DNA ligase in the nuclei of infected cells. This shows the products from the simplest reactions. Other, more complicated results involving heteroduplex repair or even branch migration and subsequent formation of Holliday structures cannot be excluded.

might lead to a phenomenon known as gene "conversion," in which one or the other strand of the mismatch is removed and replaced by a copy of the other. If the heteroduplex persists beyond the point at which the viral DNA is integrated and replicated, it would give rise to two proviruses, one the recipient parent and the second a recombinant, both at the same genetic loci.

Comparison With Other Models for Retroviral Recombination

Two other models for retroviral recombination have been described in detail recently and these are illustrated in Figure 8. One of them, the DNA exchange model [3], was proposed at the time when it first seemed clear that circular DNA molecules were intermediates in the formation of proviral DNA. This, and an awareness of the discontinuity of the positive strands, made plausible a type of exchange that had been suggested earlier for the circular DNA of bacteriophage ϕx174 [23]. It can be seen that this model is similar in concept, though different in detail, to the displacement/assimilation model. One important similarity is the proposed production of a heteroduplex recombinant. The second scheme, the RNA breakage–repair model [22], postulates genetic exchange during synthesis of the first negative strand of viral DNA. This model has several attractive features, one of which is its ability to explain how retroviruses can survive such extensive damage in their RNA genomes. The ability to use intact parts of both genomes provides a rationale for the diploid nature of the

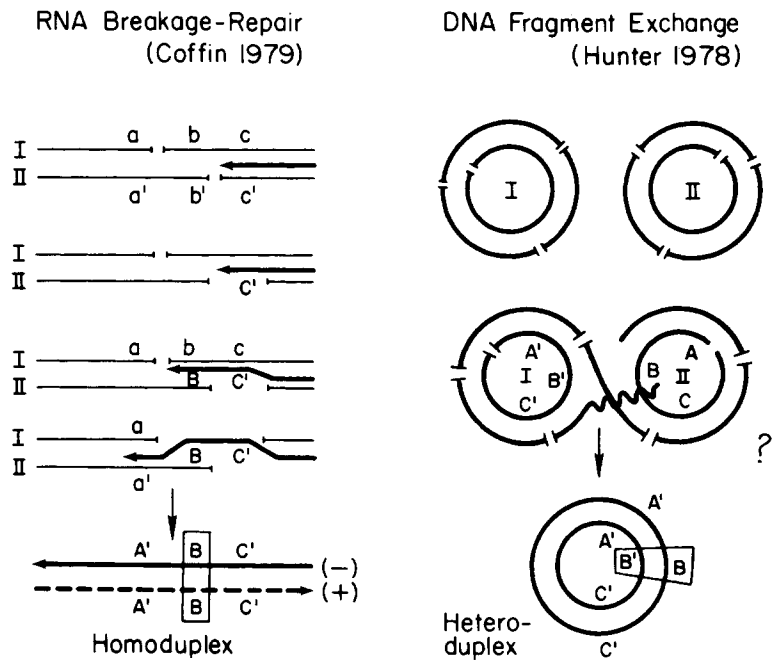

Fig. 8. Two earlier models for homologous recombination in retroviruses. In RNA breakage–repair (left), DNA synthesis (heavy arrow) is initiated on one of the two RNA genomes in the virion core, but must stop when it encounters a break in the template. The blunt-ended DNA/RNA hybrid duplex is then a substrate for the RNAseH activity that is an integral part of the reverse-transcriptase. Removal of RNA from the end of the duplex leaves a free single strand of DNA that can then hybridize to the second RNA genome in the core. Synthesis then proceeds on the second template, joining markers c′ and b, until a break is again encountered and the whole procedure repeated once more joining a′ to b to c′ on the negative strand. Synthesis of the complementary positive strand produces a single homoduplex recombinant DNA. Adapted from Coffin [22]. In DNA fragment exchange, the interacting substrates are two circular DNA duplexes presumed to have been synthesized from the two RNA genomes in a virion. The fragment containing marker B′ on molecule I is exchanged for that containing B from molecule II. The resulting recombinant, presumed to be repaired by cellular ligase, is a heteroduplex containing both B and B′. The fate of molecule II is unspecified. Adapted from Hunter [3].

virus. Furthermore, the biochemistry seems plausible, since transcription is known to jump from one end of the RNA genome to the other utilizing the same sort of mechanism proposed. Finally, formation of only one viral DNA per particle could explain the pseudo-haploid genetic behavior of the virus [24,25]. The model differs from the DNA exchange model in that a homoduplex recombinant is produced.

No physical data are available to support the RNA breakage–repair model. However, we cannot exclude that some genetic exchanges originate through this mechanism in addition to the displacement/assimilation route. A similar "copy choice" scheme has been proposed to explain the low frequency events presumed to give rise to acute transforming viruses [26], but these could also be explained by displacement/assimilation where short regions of homology facilitate assimilation.

TABLE I. Predictions of the Displacement/Assimilation Model Compared With Those of the RNA Breakage-Repair and DNA Fragment Exchange Models.

Displacement/assimilation recombination	RNA breakage-repair	DNA fragment-exchange
1. Recombination requires genotypically mixed virions	Yes	Yes
2. Recombination is not reciprocal	NA*	Yes
3. Recombinant clones may be parental/recombinant heterozygotes	No	Yes
4. Viability of the recombinant requires intactness of at least one of the parental genomes	No	Yes
5. Positive strands are recombinant negative strands are parental	No	Yes

*Not applicable. Since only one provirus/virion is predicted, reciprocity is not possible.

Table I lists some of the predictions of the displacement/assimilation model and compares them with those of the RNA breakage-repair and DNA fragment exchange models. The data that are available so far appear to support or are consistent with most of the predictions. The most careful genetic experiments designed to minimize multiple infection and multiple rounds of replication seem to support predictions 1, 2, and 3 [27,28,29]. However, interpretation of the results relevant to prediction 3, which would appear to rule out the RNA breakage-repair model, is not entirely unambiguous. The probability that subgenomic fragments of defective genomes can contribute to the genetic pool and the possibility that with some (low) frequency two proviruses could originate from one virus particle make it difficult to evaluate these genetic data. The last two predictions have not been tested, but transfection experiments may shed some light on prediction 5.

FINAL COMMENTS

One feature of the displacement/assimilation model seems worth emphasizing because it relates to the general view of recombination in retroviruses with which this review began. In this model, a critical initiation step is the strand-displacement synthesis catalyzed by reverse-transcriptase. In a sense then, the reverse-transcritptase can be considered a "recombinase." This same versatile protein has another interesting activity—an endonuclease that acts preferentially on supercoiled DNA—documented most thoroughly by Grandgenett and coworkers [30]. Discovery of this activity has led to the proposal that reverse-transcriptase may also be involved in integrative recombination. The availability of molecular clones of DNA containing presumed substrates of the integration reaction should make it possible to test this notion directly. At present, however, it seems possible that a single viral protein may relate all three types of recombination in retroviruses.

REFERENCES

1. Cold Spring Harbor Symposium on Movable Genetic Elements, Part 2: New York: Cold Spring Harbor, Vol 45, 1980.
2. Starlinger P: Plasmid 3:241, 1980.
3. Hunter E: Curr Top Microbiol Immunol 79:295, 1978.
4. Ju G, Skalka AM: Cell 22:379, 1980.
5. Hishinuma F, DeBona PJ, Astrin S, Skalka AM: Cell 23:155, 1981.
6. Temin H: Cell 28:3, 1982.
7. Leis J, Duyk G, Longiaru M, Skalka AM: J Virol (submitted).
8. Bishop JM: Cell 23:5, 1981.
9. Takeya T, Hanafusa H, Junghans R, Ju G, Skalka AM: J Cell Molec Biol 1:1024, 1981.
10. Shalloway D, Zelenetz AD, Cooper GM: Cell 24:531, 1981.
11. Parker RC, Varmus HE, Bishop JM: Proc Natl Acad Sci USA 78:5842, 1981.
12. Coffin J: New York: Cold Spring Harbor Laboratory, 1982.
13. Olsen JC, Watson KF: Biochem Biophys Res Comm 97:1376, 1980.
14. Boone LR, Skalka AM: J Virol 37:109, 1981a.

15a. Boone LR, Skalka AM: J Virol 37:117, 1981b.

15b. Hsu TW, Taylor JM: J Virol 44:47, 1982.

16. Junghans RJ, Boone LR, Skalka AM: J Virol (in press).
17. Giboa E, Mitra SW, Goff S, Baltimore D: Cell 18:93, 1979.
18. Radding C: Ann Rev Biochem 47:847, 1978.
19. Junghans RJ, Boone LR, Skalka AM: Cell 30:53, 1982.
20. Delius H, Mantell NJ, Allberts B: J Mol Biol 67:341, 1982.
21. Mangel WG, Delius H, Duesberg PH: Proc Natl Acad Sci USA 71:4541, 1974.
22. Coffin JM: J Gen Virol 42:1, 1979.
23. Benbow RM, Zuccarelli AJ, Sinsheimer RL: Proc Natl Acad Sci USA 72:235, 1975.
24. Rubin H, Temin HM: Virology 7:75, 1959.
25. Kelloff G, Aaronson SA, Gilden RU: Virology 42:1133, 1970.
26. Swanstrom R, cited in Varmus HE: Science 216:812, 1982.
27. Weiss RA, Mason WS, Vogt PK: Virology 52:535, 1973.
28. Wyke JA, Beamand JA: J Gen Virology 43:349, 1979.
29. Carloni G, Kaczorek M, Hill M: Proc Natl Acad Sci USA 77:3014, 1980.
30. Golomb M, Grandgenett DP, Mason W: J Virol 38:548, 1981.