Proc. Natl. Acad. Sci. USA
Vol. 78, No. 7, pp. 4299–4303, July 1981
Biochemistry

A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular DNA sequence present in only one copy prior to integration

(retroviruses/transposable elements/long terminal repeats/DNA sequences/integration sites)

STEPHEN H. HUGHES*, ANN MUTSCHLER*, J. MICHAEL BISHOP†, AND HAROLD E. VARMUS†

*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724; and †Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Contributed by J. Michael Bishop, April 16, 1981

ABSTRACT **The Rous sarcoma virus (RSV)-transformed rat cell line RSV-NRK-2 contains a single complete RSV provirus. We have obtained recombinant λ clones that contain both ends of the RSV provirus and the flanking rat sequences. The provirus is integrated in unique DNA and is present in only one of the two homologous chromosomes. The rat sequences into which the RSV provirus integrated were also cloned from the RSV-NRK-2 cell line. The sequences of the regions involved in the recombination event have been determined and compared. Our data suggest that, compared with the sequence of viral DNA in the large circular form of unintegrated viral DNA, the provirus lacks two base pairs at each end and that the provirus is flanked by a six-base-pair direct repeat of cellular DNA. This six-base-pair repeat was apparently created during the integration event because this sequence was present only once at the integration site before the provirus was inserted. A survey of eight other independent RSV transformed rat cell lines demonstrates that, in agreement with earlier results, the RSV proviruses have entered different segments of rat cell DNA. We have also determined the sequence of a second virus DNA–host cell DNA junction from a second RSV-transformed rat cell line (RSV-NRK-4) and find that there are no obvious similarities between the two integration sites or between the integration sites and the termini of viral DNA.**

Retroviruses insert a DNA copy of their genome into host DNA as an obligate step in their life cycle. This recombination event is known to be relatively specific with respect to the site in viral DNA and relatively nonspecific with respect to the site in the host DNA (1, 2). The retrovirus genome is RNA; before integration, the viral RNA-dependent DNA polymerase must copy the RNA genome into DNA. The first stable intermediate is a double-stranded linear DNA molecule made in the cytoplasm of infected cells. This viral DNA, which has a full-length (−) strand and segmented (+) strands, is slightly larger than the RNA from which it is derived, and sequences from both the 3′ and 5′ ends of the genomic RNA form a direct repeat (called the LTR) at the termini of the linear DNA which varies in size from one species of retrovirus to another. Each repeat has the structure U_3RU_5; U_3 is a unique sequence from the 3′ end of viral RNA and U_5 is a unique sequence from the 5′ end of viral RNA (1, 3, 4) (see Fig. 1). The R sequences are repeated at both the 5′ and 3′ ends of viral RNA. In the case of Rous sarcoma virus (RSV), U_3 is 230 base pairs (bp), R is 21 bp, and U_5 is 80 bp (5–8).

The linear viral DNA migrates from the cytoplasm of an infected cell to the nucleus where it is converted to two major forms of circular molecules. The smaller circle has a single LTR and presumably is formed by homologous recombination between the LTRs. The larger circle has two copies of the LTR and presumably is formed by a ligation event that brings together the ends of the linear viral DNA (3, 4). The nature of this event is unknown, partly because the exact termini of the linear DNA have not been defined. It is not known whether the linear form or one of the circular forms of DNA is the direct antecedent of the integrated provirus.

The structure of the integrated provirus is similar to that of unintegrated linear RSV DNA, with cellular DNA joined to viral DNA at or near the ends of the LTR (1, 2). The structures of viral RNA, the three forms of unintegrated viral DNA, and the provirus are given in Fig. 1.

To learn more about the integration of RSV DNA, we cloned segments of DNA that included the junctions between viral and host DNA and compared these to the unaltered host sequences into which the viral DNA integrated. These clones were derived from the DNA of the transformed rat cell line RSV NRK-2 which contains a single complete RSV provirus (1).

The ends of the integrated provirus are flanked by a 6-bp repeat of cellular DNA. This repeat is apparently created by the integration event because this 6-bp sequence is present only once in the host DNA into which the provirus integrated. The provirus appears to have lost 2 bp from the right and left LTRs, based on comparisons with the sequence of cloned circular viral DNA (8). However, there are bases at the junction between cellular and viral DNA whose origin we cannot determine unambiguously because the same base pairs exist at the corresponding sites in viral and rat DNA. It therefore is possible that our estimate of the number of base pairs contributed to the 6-bp repeat by cellular DNA is wrong by 1. If this were true, there would be a corresponding change of 1 bp in the number of base pairs lost from one LTR. We also determined the sequence of a junction between virus and host cell DNA from a second RSV-transformed rat cell line (RSV-NRK-4) and found that there are no obvious similarities between the two integration sites.

Integration of RSV DNA is similar to the integration of the DNAs of other retroviruses, murine sarcoma virus, spleen necrosis virus, and mouse mammary tumor virus, which lose 2 bp at each end during integration. These three proviruses are flanked by repeats of cellular DNA of 4, 5, or 6 bp, respectively (9–11). The product of integrative recombination of retrovirus DNA resembles the product of integrative recombination of bacterial or yeast transposons (12–16), the bacterial virus Mu, (17–19), and the *Drosophila* movable elements such as *copia*

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

Abbreviations: LTR, long terminal repeat; RSV, Rous sarcoma virus; bp, base pair(s); kb, kilobase(s).

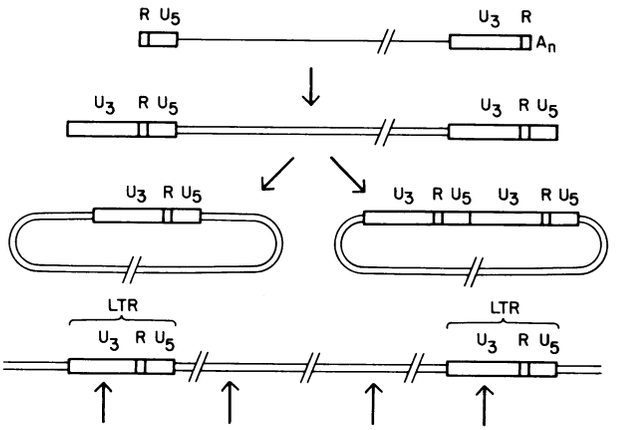

FIG. 1. Intermediates in replication of RSV genome. The RNA genome of RSV (*Top*) is approximately 9.5 kilobases long. The 3′ end of the RNA is polyadenylylated (A_n), and a 21-base sequence (R) adjacent to the poly (A) is also present at the extreme 5′ end of the RNA. A unique stretch of 230 bases adjacent to R at the 3′ end of the molecule is called U_3; a unique stretch of 80 bases adjacent to R at the 5′ end of the genome is called U_5. When the RNA genome is copied into DNA, the U_3 and U_5 sequences are found at both ends of the linear DNA, and the sequence U_3RU_5 is the LTR. The linear molecule is synthesized in the cytoplasm and migrates to the nucleus when it is the precursor to two circular forms, one with one copy and the other with two copies of the LTR. One of these three molecules (the linear form or one of the two circular forms) is probably the direct precursor of the provirus (*Bottom*), in which the ends of the LTRs are now attached to cellular DNA. The arrows under the provirus indicate the *Eco*RI sites in the provirus.

(20, 21). This suggests that the mechanisms of integration of these diverse DNA molecules are likely to be related.

MATERIALS AND METHODS

Recombinant DNA Experiments. High molecular weight DNA was prepared from rat cell lines RSV-NRK-2 and RSV-NRK-4 (transformed with the Schmidt–Ruppin D strain RSV). The DNA was digested to completion with *Eco*RI and the resulting fragments were fractionated by electrophoresis on a 0.8% agarose gel (HGPT grade, Sea-Kem). Aliquots from each fraction were analyzed by electrophoresis on a small 0.8% agarose gel and transferred to nitrocellulose paper. Fractions containing fragments of interest were identified by hybridization with appropriate ^{32}P-labeled DNA probes.

λgtWES was digested with *Eco*RI, the ends were caused to cohere by annealing, and the insert was removed by fractionation on a 10–30% sucrose gradient. The purified λgtWES "arms" (2–10 μg) were ligated to an appropriate aliquot (2–10 μg) from the agarose gel-fractionated DNA. Most of the fragments we wished to clone were small, and we adjusted the ratio of arms to inserts to give multiple fragments of rat cell DNA in most of the recombinant bacteriophage. Ligation was documented by fractionation on an agarose gel. The ligated DNA was packaged *in vitro*, the resulting recombinant bacteriophage (we typically obtained 10^5–10^6 recombinants) were plated on DP 50 SupF at 10,000–30,000 per 150-mm plate, and a portion of the λ DNA was transfered to nitrocellulose filters. The nitrocellulose filters were hybridized (48–72 hr) with 1–4 × 10^6 cpm of DNA (specific activity, 2–20 × 10^8 cpm/μg) per filter. We used the same hybridization and washing conditions for plaque hybridization that we had used previously for the analysis of restriction fragments from RSV-transformed cell DNA (1). Positive plaques were purified by a second round of plating. We obtained multiple independent recombinants containing each of the desired fragments. The individual λ recombinants were digested with *Eco*RI and the insert was identified by hybridization. In all cases the cloned *Eco*RI fragment was exactly the same size as the corresponding fragment in the digest of genomic DNA and contained the predicted restriction sites.

The *Eco*RI fragments were subcloned into pBR322. The *Eco*RI-digested pBR322 was treated with an excess of calf intestinal alkaline phosphatase (Boehringer) to reduce the frequency of self-ligation. All pBR322 derivatives were cloned and propagated in the recA$^-$ host HB101. All procedures involving recombinant DNA were performed in accordance with the National Institutes of Health guidelines.

Sequence Determination. DNA fragments were isolated on agarose or acrylamide gels and purified by chromatography on DEAE-Sephacryl. The purified fragments either were labeled at their 3′ ends with the Klenow fragment of *E.coli* polymerase I (Boehringer) or were treated with calf intestinal alkaline phosphatase (Boehringer) and labeled at their 5′ ends with polynucleotide kinase (PL Biochemicals). The complementary strands of the labeled fragments were separated or the fragments were cut again with a second restriction endonuclease. The asymmetrically labeled fragments were chromatographed on DEAE-Sephacryl after electrophoretic purification. The DNA sequence was determined by the chemical methods of Maxam and Gilbert (22).

Southern Filter Hybridization. The conditions for complete digestion of cellular DNA, electrophoresis of the resulting DNA fragments, transfer to nitrocellulose, and hybridization with ^{32}P labeled DNA probes have been described (1, 3). In these experiments some of the DNA probes were prepared from gel-purified DNA fragments by nick-translation.

RESULTS

Cloning and Characterization of DNA Flanking the Provirus in RSV-NRK-2. The RSV-transformed rat cell RSV-NRK-2 contains a single complete RSV provirus. *Eco*RI fragments containing the ends of the provirus and the flanking cellular DNA were cloned in λgtWES and subcloned in pBR322. The *Eco*RI fragment from the right end of the provirus (RE-2; Fig. 2) was

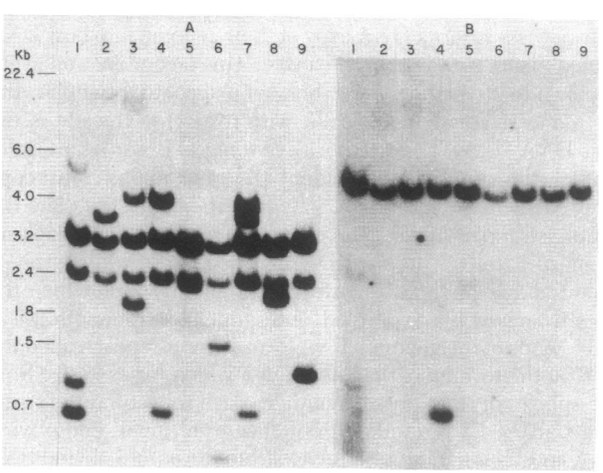

FIG. 2. RSV provirus is not found in both equivalent sites in the homologous chromosomes; proviruses enter different regions in different cell lines. DNA from a series of RSV-transformed NRK cell lines was digested to completion with *Eco*RI, fractionated on a 0.8% agarose gel, and transferred to nitrocellulose. (*A*) Filter was first incubated with $cDNA_3$ specific for the U_3 region of RSV. (*B*) After the radioactivity had decayed, the same filter was reincubated with the RE-2 probe. Lanes: 1, RSV NRK-15; 2, RSV NRK-1; 3, RSV NRK-7; 4, RSV NRK-2; 5, RSV NRK-3; 6, RSV NRK-4; 7, RSV NRK-8; 8, RSV NRK-5; 9, RSV NRK-6. Size markers are shown in kb.

700 bp long and contained 150 bp of viral DNA linked to 550 bp of cellular DNA. A similar protocol was used to obtain a 4.0-kilobase (kb) fragment (LE-2) which contained the left end of the provirus and flanking host cellular DNA.

The 700-bp RE-2 fragment was labeled by nick-translation and hybridized to *Eco*RI-digested DNA from a series of RSV-transformed rat lines, including RSV-NRK-2. The 550-bp segment of cellular DNA in the RE-2 clone was present only once in the genome, in a region that yielded a 4.5-kb *Eco*RI fragment (Fig. 2B). Digestion of RSV NRK-2 DNA with *Eco*RI produced two fragments that hybridized with RE-2: a 4.5-kb fragment and a 700-bp fragment [the one that was cloned as RE-2 (Fig. 2*B*, lane 4)]. This 700-bp fragment also hybridized with a probe specific for the U_3 portion of the LTR of the provirus, but the 4.5-kb fragment did not. The presence of the 4.5-kb fragment in *Eco*RI-digested DNA from the transformed line RSV-NRK-2 demonstrates that the equivalent sequence in the homologous chromosome is not interrupted by a provirus. The RE-2 probe detected the 4.5-kb *Eco*RI fragment in digests of DNA from the other nine RSV-transformed lines, confirming earlier conclusions from restriction mapping (1, 2) that the proviruses in the other lines have not integrated into the region occupied by the RSV-NRK-2 provirus. In addition to the cellular sequences, the RE-2 probe contains the right half of a LTR so it also hybridizes weakly to virus-specific fragments that contain the same region of the LTR (Fig. 2).

We used the RE-2 clone containing the unique 550-bp segment of rat DNA to identify λgtWES clones containing the 4.5-kb *Eco*RI fragment from RSV-NRK-2 cells. The cloned 4.5-kb fragment (US-2) hybridized not only to the 550-bp segment of cellular DNA adjacent to the right end of the provirus but also to a cloned 4.0-kb segment (LE-2) that contained the left end of the provirus in RSV-NRK-2 and the adjacent cellular sequences (data not shown). The 4.5-kb segment of cellular DNA did not hybridize to any viral DNA probe; however, the two fragments from the right and left ends of the provirus, the 0.7-kb RE-2 clone and the 4.0-kb LE-2 clone, contained 150 and 180 bp of viral DNA, respectively.

The DNAs from the US-2, LE-2, and RE-2 clones were mapped with restriction enzymes; the restriction sites in the 4.5-kb fragment precisely conformed to the restriction sites in the cellular DNA portions of the 4.0-kb and 0.7-kb fragments (data not shown). There was a *Hin*dIII site about 700 bp from the *Eco*RI site in viral DNA in the 4.0-kb LE-2 clone. The corresponding *Hin*dIII site was about 1000 bp from the *Eco*RI site in the 4.5-kb fragment; this 1.0-kb *Hin*dIII – *Eco*RI fragment was subloned in pBR322. This fragment, US-2H, hybridized to cellular sequences from both sides of the provirus—i.e., to both the 0.7-kb RE-2 and 4.0-kb LE-2 cloned segments. This demonstrates that the 1.0-kb US-2H fragment cloned into pBR322 contains the cellular DNA sequences into which the RSV provirus integrated.

RSV Provirus Appears To Have Lost 2 bp of Viral DNA and Is Flanked by a 6-bp Direct Repeat of Cellular DNA That Was Present in One Copy Prior to Integration. The sequences at the ends of the provirus and the flanking cellular DNA were determined by the protocols of Maxam and Gilbert (22) from the LE-2 and RE-2 clones (Fig. 3). The sequence of the corresponding region from the large circular form of RSV DNA is shown for comparison. The sequence of the integration site for the RSV provirus was determined from fragments derived from the 1.0-kb *Hin*dIII–*Eco*RI clone US-2H.

The simplest interpretation of our data would be that 2 bp of viral DNA are missing at each end of the provirus; these 4 bp are present at the joint between U_3 and U_5 in the large circular form of unintegrated viral DNA and are presumed to be present at the ends of unintegrated linear DNA (Fig. 3). We also conclude that the integration of the RSV provirus creates the 6-bp direct repeat from a cellular sequence present only once in the unoccupied site. However, there are two possible but

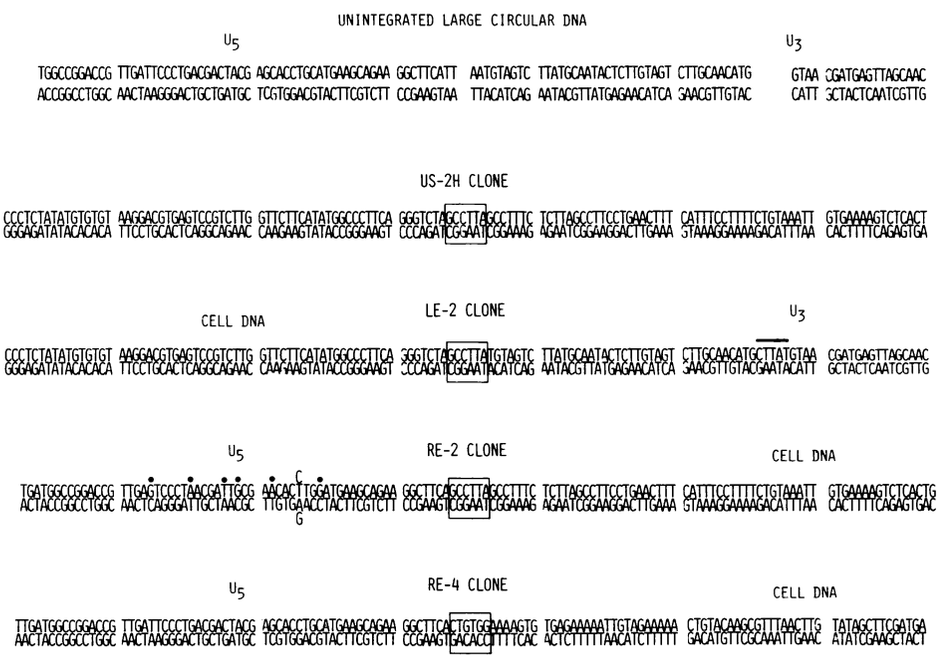

FIG. 3. Sequences of unintegrated RSV DNA, host–virus junctions, and site in host DNA with which the viral DNA recombined. The top sequence includes the junction between U_3 and U_5 in the large circular form of RSV DNA reported by Swanstrom *et al.* (6) for the SR-A strain of RSV. To facilitate comparisons with the sequences at the ends of the proviruses, the sequence of the unintegrated large circular DNA is presented with two gaps, one at the $U_3 - U_5$ junction and one in the U_3 region at a position that corresponds to a 5-bp insertion in the U_3 sequence of the LE-2 clone. This insert in the U_3 region of the RE-2 sequence is indicated by a bar and discussed in the text. In the sequence of the host DNA from the US-2H clone, the 6-bp sequence (enclosed in a box) is present only once. In RE-2, there are six changes in the U_5 sequence compared to the SR-A U_5, but the U_5 of the RE-4 clone is identical to that of the SR-A U_5.

unlikely alternatives. (*i*) The rightmost base within the 6-bp repeat at the left end of the provirus (an adenine) could have a viral and not a cellular origin because there is an adenine at the corresponding positions in both the cellular DNA and the viral DNA sequences. This would mean that the repeat of cellular DNA would be 5 bp, not 6. (*ii*) The first base on the 5′ side of the 6-bp repeat at the right end of the provirus (also an adenine) could be derived from cellular DNA and not from viral DNA because both the cellular and viral DNAs have adenine at the corresponding positions. If the adenine were derived from cellular DNA, the repeat of cellular DNA would be 7 bp, not 6. Either of these alternative hypotheses requires asymmetrical recombination events; thus, we favor a 6-bp repeat of cellular DNA and the loss of 2 bp at each end of the provirus.

The sequence of the U_3 segment in the left LTR has been compared with the published sequences of LTRs from an SR-A (8) and SR-D (23) strains of RSV (Fig. 3) The SR-D provirus from RSV-NRK-2 appears to contain a 5-bp insert, C-T-T-A-T, located 38 bp from the junction between viral and cellular DNA. These 5 inserted bases have considerable homology with the repeat at the ends of the provirus, forming a perfect direct repeat of a 5-base sequence that includes an adenine of ambiguous origin and the first base that was clearly derived from viral sequences. To test whether this 5-bp insert is found in the U_3 segments at both ends of the provirus, we cloned the 3.1-kb *Eco*RI segment that contained most of the U_3 portion of the right LTR and determined the sequence of the U_3 portion of this clone. The same 5-bp insert found in the left LTR also was present in the right LTR (data not shown). Because the 5-bp insert was present in the U_3 segments of the right LTR it is unlikely, although not impossible, that it was produced by or influential during this particular integration event.

There is no clear homology between the viral DNA and the host sequences immediately adjacent to the integration site. There is some limited homology between U_5 and the host sequences to the left of the provirus; 9 of the first 20 bp are identical (Fig. 3). However, chance would give 5 of 20, and homology at this level may be a coincidence. The same type of comparison between U_5 and the cellular sequences to the right of the provirus shows only 6 of the first 20 base pairs as matching.

There Is No Obvious Sequence Homology Between Two Integration Sites. DNA from a second cell line, RSV-NRK-4 (1), which also contains a single complete RSV provirus, was digested with *Eco*RI, and clones containing the right end of the provirus and adjacent cellular DNA were isolated. The cloned fragment, RE-4, was ≈1600 bp long and contained about 1450 bases of cellular DNA. The RE-4 segment was labeled, hybridized to *Eco*RI-digested cellular DNA, and found to contain a repeated DNA sequence that reacted with a large number of fragments in an *Eco*RI-digest of NRK DNA (data not shown). We were able to isolate a subfragment of unique DNA from the 1.6-kb fragment, and we found that the single RSV provirus in RSV-NRK-4 entered only one of the two homologous chromosomes (data not shown). We also determined the sequence of the joint between viral and cellular DNA in the cloned fragment (Fig. 3). Again, 2 bases of U_5 were missing at the end of the provirus. Comparison of the sequences in the cellular DNA flanking the two proviruses reveals no obvious homologies between the two integration sites or between viral and host DNA. Five of the first six bases flanking the proviruses differ. The two cellular DNA sequences flanking the right side of both proviruses contain a preponderance of A·T base pairs—47 of 74 (63%) in RE-2, and 45 of 68 (66%) in RE-4. However, the region to the left of the provirus (LE-2) is not A+T-rich [40 of 74 (54%) base pairs are A·T]. Because the sample is small and the overall A+T composition of mammalian DNA is ≈60%, these observations are probably not significant.

DISCUSSION

Several other retroviral proviruses exhibit the types of symmetries we have observed at host–virus junctions of an RSV provirus. The proviruses of murine sarcoma virus, spleen necrosis virus, and mouse mammary tumor virus are flanked by 4-, 5-, and 6-bp repeats, respectively, of cellular DNA (9–11). In each of these cases, 2 bp are lost from LTRs at each end of the proviral DNA. In the two cases that have been directly tested, with murine sarcoma virus (24) and mouse mammary tumor virus (11), the repeated cellular sequences are present only once in the virgin integration site; thus, the integration events include creation of the duplication. Analysis of aberrant unintegrated murine leukemia virus DNA suggests that its integration follows these same rules (25). The chicken endogenous provirus *ev*-1 is also flanked by a 6-bp direct repeat, and these 6 bp are present only once in the homologous region of the genome of a chicken that lacks *ev*-1 (26). If it is assumed that this line of chickens never acquired *ev*-1 and that the lack of *ev*-1 was not due to an ancestral excision event, the 6 bases flanking *ev*-1 were also duplicated during integration.

We have shown here that 2 bp appear to be lost from each end of RSV DNA during integration and that integration creates a 6 bp direct repeat of cellular DNA. However, there are 2 bases whose origins are ambiguous, implying that the duplicated sequence could be 5 or 7 bp (see *Results*). We favor the interpretation that the duplication is 6 bp and that the provirus loses 2 bp from each LTR, because the alternatives would require asymmetrical recombination at the two ends of the provirus and because this interpretation is most consistent with the structure of other retroviral proviruses.

Structural analyses suggest that integrative recombination may involve similar mechanisms for several apparently diverse DNA molecules. (*i*) The bacterial transposons, the yeast element Ty-1, movable elements in *Drosophila* such as *copia*, the bacteriophage Mu, and the proviruses of retroviruses all recombine with a large set of sites in the host genome but utilize a specific site (or pair of sites) in their own DNA.

(*ii*) The termini of several of these elements [e.g., Ty-1, *copia*, Tn9 (a bacterial transposon), and retroviral proviruses] are composed of substantial direct repeats. This rule is not universal, however, because some bacterial transposons terminate in large inverted repeats and the ends of Mu DNA have sequences that can be arranged as a short inverted repeat only with significant manipulation (17, 18). In three of the cases in which elements terminate with direct repeats (Ty-1, Tn9, retroviruses), sequences that appear to be identical to the terminal repeats are found at various places in the host genome unassociated with the body of the transposable elements or provirus (12, 13, 19, 27, 28). Such "free repeats" could arise by homologous recombination; in at least one case, a single LTR was left behind when the rest of a murine leukemia virus provirus was excised (28). Such events, however, are not universal; the 276-bp sequence that forms the direct repeats of *copia* has been found only as part of a complete *copia* element (20).

(*iii*) The yeast element Ty-1, several of the bacterial transposons, integrated Mu viral DNA, *copia*, and the retrovirus proviruses are all flanked by short repeats (4–11 bp) of cellular DNA. In the cases that have been examined closely, this cellular sequence is repeated as a consequence of the insertion event—i.e., the unoccupied site in the host DNA contains only a single copy (15–18, 20, 21).

(*iv*) The retrovirus proviruses, *copia*, and the bacterial transposon Tn9 terminate with short imperfect inverted repeats; this means that each of the direct terminal repeats in these elements includes short inverted terminal repeats. In the RSV provirus, 10 of the 13 bp in these short inverted repeats are identical. Only one of the three mismatches is strongly antibonding, a purine·purine pair at position 10 (Fig. 3; ref. 8). However, the yeast element Ty-1 does not have this type of terminal inverted repeat (although the two terminal base pairs form a tiny inversion), and Mu does not have a simple inverted repeat.

(*v*) Several of the bacterial transposons appear to have an affinity for A+T-rich regions of host DNA. There is no apparent homology among the cellular sequences into which an individual species of retrovirus integrates (10, 11, 30); however, the sequences to the right of the two RSV proviruses we have examined are both slightly A+T-rich. This is not a large enough sample, and more data will be required to determine the relevance of these observations.

Although it is tempting to speculate that the mechanism of integration of these DNA molecules of diverse origins is similar, several fundamental questions remain unanswered. In the case of the retroviruses, it is not yet known which of three possible candidates is the immediate precursor to the integrated provirus (Fig. 1). We have previously argued that the most likely candidates are the larger circular DNA or the linear DNA (1, 3). However, in bacterial systems, a plasmid carrying a single insertion sequence (structurally equivalent to unintegrated circular DNA with one LTR) yields a structure that physically resembles a provirus after a recombination event. If the analogy applies to retroviruses, the small circular viral DNA would be the precursor (25). However, if the small circular DNA integrated via this mechanism, so should the large circle; if the large circle integrated in this way, one or both ends of resulting proviruses would be flanked by two LTRs in tandem. Proviruses of this type have not been observed, which suggests that models for retrovirus integration that are the direct descendants of the proposals of Shapiro (31) for transposons should be treated cautiously.

The recombination events in the bacterial transposons are catalyzed by enzymes encoded within the transposon, in some cases within the terminal repeats (14, 32). The precision of integration of retrovirus DNA strongly implies the involvement of specific proteins. The integration of different types of retroviral DNAs creates cellular DNA repeats of different length; this suggests that a viral enzyme is involved. This hypothesis has yet to receive definitive support from studies of different types of proviruses in a single species of host cell. Mouse mammary tumor virus and RSV proviruses have been examined in infected rat cells (ref. 11; this report), but both types of proviruses appear to generate a 6-bp repeat. Spleen necrosis virus proviruses and the endogenous provirus *ev*-1 have been studied in chicken cells; the former DNA is flanked by 5-bp repeats (10, 30) and the latter, by a 6 bp repeat (26). However, these two types of proviruses were acquired under different circumstances: the former by infection of cultured fibroblasts and the latter by natural infection of the germ line.

We are grateful to Howard Goodman for advice about cloning, and to Yasha Gluzman, Kim Nasmyth, and John Fiddes for helpful suggestions about sequence determination. This research was supported by grants from the National Institutes of Health and the American Cancer Society.

1. Hughes, S. H., Shank, P. R., Spector, D. H., Kung, H. J., Bishop, J. M., Varmus, H. E., Vogt, P. K., & Breitman, M. L. (1978) *Cell* **15**, 1397–1410.
2. Sabran, J., Hsu, T., Yeater, C., Kaji, A., Mason, W. S. & Taylor, J. (1979) *J. Virol.* **29**, 170–178.
3. Shank, P. R., Hughes, S. H., Kung, H. J., Majors, J. E., Quintrell, N., Guntaka, R. V., Bishop, J. M. & Varmus, H. E. (1978) *Cell* **15**, 1283–1395.
4. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. (1978) *J. Virol.* **28**, 810–819.
5. Shine, J., Czernilofsky, A. P., Friedrich, R., Goodman, H. M. & Bishop, J. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1473–1477.
6. Schwartz, D., Zamecnik, P. & Weith, H. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 994–998.
7. Haseltine, W., Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 989–993.
8. Swanstrom, R., DeLorbe, W., Bishop, J. M. & Varmus, H. E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 124–128.
9. Dhar, R., McClements, W. L., Enquist, L. W. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3937–3941.
10. Shimotohno, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285**, 550–554.
11. Majors, J. & Varmus, H. E. (1981) *Nature (London)* **289**, 253–258.
12. Cameron, J., Loh, E. & Davis, R. (1979) *Cell* **16**, 739–751.
13. MacHattie, L. A. & Jackowski, J. (1977) in *DNA Insertion Elements, Plasmids and Episomes*, eds. Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 219–228.
14. Calas, M. & Miller, J. (1980) *Cell* **20**, 579–595.
15. Gafner, J. & Phillipsen, P. (1980) *Nature (London)* **286**, 414–418.
16. Farabaugh, P. J. & Fink, G. R. (1980) *Nature (London)* **286**, 352–356.
17. Allet, B. (1979) *Cell* **16**, 123–129.
18. Kahmann, R. & Kamp, D. (1979) *Nature (London)* **280**, 247–250.
19. Chow, L. T. & Bukhari, A. I. (1977) in *DNA Insertion Elements, Plasmids and Episomes*, eds. Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (Cold Spring Harbor Laboratory, Spring Harbor, NY), pp. 295–306.
20. Levis, R., Dunsmuir, P. & Rubin, G. M. (1980) *Cell* **21**, 581–588.
21. Dunsmuir, P., Brorein, W., Jr., Simon, M. & Rubin, G. M. (1980) *Cell* **21**, 575–579.
22. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
23. Yamamoto, T., Jay, G. & Pastan, I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 176–180.
24. McClements, W., Dhar, R., Blair, D., Enquist, L., Oskarsson, M. & Vande Woude, G. (1981) *Cold Spring Harbor Symp. Quant. Biol.* **45**, 699–705.
25. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. & Baltimore, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3932–3936.
26. Hishinuma, F., DeBona, P. J., Astrin, S. & Skalka, A. M. (1981) *Cell* **23**, 155–165.
27. Hughes, S. H., Vogt, P. K., Robinson, H. L., Bishop, J. M. & Varmus, H. E. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **44**, 1077–1089.
28. Hughes, S. H., Toyoshima, K., Bishop, J. M. & Varmus, H. E. (1981) *Virology* **108**, 189–207.
29. Varmus, H. E., Quintrell, N. & Ortiz, S. (1981) *Cell*, in press.
30. Shimotohno, K. & Temin, H. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7357–7361.
31. Shapiro, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1933–1937.
32. Heffron, F., McCarthy, B. J., Ohtsubu, H. & Ohtsubu, E. (1979) *Cell* **18**, 1153–1163.