Proc. Natl. Acad. Sci. USA
Vol. 78, No. 1, pp. 124–128, January 1981
Biochemistry

Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: Viral DNA contains direct and inverted repeats similar to those in transposable elements

(retrovirus DNA/circular DNA/terminal repeat/control regions/primers)

RONALD SWANSTROM, WILLIAM J. DELORBE, J. MICHAEL BISHOP, AND HAROLD E. VARMUS

Department of Microbiology and Immunology, University of California, San Francisco, California 94143

Contributed by J. Michael Bishop, September 16, 1980

ABSTRACT **We have determined the nucleotide sequence of portions of two circular avian sarcoma virus (ASV) DNA molecules cloned in a prokaryotic host–vector system. The region whose sequence was determined represents the circle junction site—i.e., the site at which the ends of the unintegrated linear DNA are fused to form circular DNA. The sequence from one cloned molecule, SRA-2, shows that the circle junction site is the center of a 330-base-pair (bp) tandem direct repeat, presumably representing the fusion of the long terminal repeat (LTR) units known to be present at the ends of the linear DNA. The circle junction site is also the center of a 15-bp imperfect inverted repeat, which thus appears at the boundaries of the LTR. The structure of ASV DNA—unique coding region flanked by a direct repeat that is, in turn, terminated with a short inverted repeat—is very similar to the structure of certain transposable elements. Several features of the sequence imply that circularization to form the SRA-2 molecule occurred without loss of information from the linear DNA precursor. Circularization of another cloned viral DNA molecule, SRA-1, probably occurred by a different mechanism. The circle junction site of the SRA-1 molecule has a 63-bp deletion, which may have arisen by a mechanism that is analogous to the integration of viral DNA into the host genome. Flanking one side of the tandem direct repeat is the binding site for $tRNA^{Trp}$, the previously described primer for synthesis of the first strand of viral DNA. The other side of the direct repeat is flanked by a polypurine tract, A-G-G-G-A-G-G-G-G-G-A, which may represent the position of the primer for synthesis of the second strand of viral DNA. An A+T-rich region, upstream from the RNA capping site, and the sequence A-A-T-A-A-A are present within the direct repeat sequence. These sequences may serve as a promoter site and poly(A) addition signal, respectively, as proposed for other eukaryotic transcription units.**

Three major types of virus-specific DNA have been identified in cells after infection with retroviruses: linear duplex (form III) DNA, the initial product of synthesis by RNA-directed DNA polymerase; covalently closed circular (form I) DNA, derived from linear precursors; and DNA covalently integrated into the host genome (proviral DNA) (1, 2).

Physical maps of the unintegrated and integrated forms of avian sarcoma virus (ASV) DNA have helped to clarify the structural relationships among these forms (cf. Fig. 1). Form III and proviral DNA are coextensive with a subunit of the RNA genome (10–13). These forms also contain a long terminal repeat (LTR) of about 300 base pairs (bp) that is not present in the RNA genome (10–13). The repeated domain is composed of sequences unique to each end of the RNA (U_5 and U_3) joined by a short sequence, R, that is present as a terminal repeat in the RNA (7). Two principal classes of circular DNA have been characterized (10, 11). One class bears a single copy of the LTR sequence and presumably arises by recombination between the LTRs at the ends of linear DNA. The second class contains two copies of the LTR sequence and probably results from fusion of the ends of linear DNA. The site at which sequences from opposite ends of linear DNA have been united will be referred to as the circle junction site.

Numerous uncertainties remain concerning the mechanisms of synthesis, circularization, and integration of retroviral DNA. One approach to these issues is a detailed examination of the structure of the various forms of viral DNA. In this report we present data on the nucleotide sequence at the circle junction site from two molecularly cloned ASV DNA molecules. A comparison of the sequences at the circle junction site of the two clones indicates that circularization probably occurred by different mechanisms in these cases. The sequence at one of the circle junction sites is arranged in a pattern of direct and inverted repeats reminiscent of bacterial transposons. Regions important for the synthesis and transcription of viral DNA are also apparent in the LTR and flanking sequences.

MATERIALS AND METHODS

Materials. [α-^{32}P]dNTPs and [γ-^{32}P]ATP (3000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels) were purchased from New England Nuclear and Amersham/Searle; restriction endonucleases and polynucleotide kinase were from New England BioLabs and were used as described by the supplier; acrylamide and methylenebisacrylamide were electrophoresis grade from Bio-Rad; reverse transcriptase (RNA-dependent DNA polymerase) from avian myeloblastosis virus was the kind gift of J. Beard (Life Sciences, St. Petersburg, FL).

Molecular Cloning of ASV DNA. The isolation and characterization of the SRA-1 and SRA-2 clones of ASV DNA in the (λ)gtWes(λ)B vector have been described (14). Subclones of restriction endonuclease fragments of SRA-2, inserted into the plasmid pBR322, were also used in this study. One subclone (pEcoRI-D), was generated by cleaving SRA-2 with *Eco*RI, then isolating the 330-bp fragment spanning the circle junction site. The other subclone (pPvu II-DG) contains a *Pvu* II fragment, also spanning the circle junction site (R. Parker, personal communication). The amplification of all recombinant DNA molecules was done in accordance with the *National Institutes of Health Guidelines for Recombinant DNA Research.*

Determining Sequences of Viral DNA. Sequence determinations were done by using the chemical cleavage method of Maxam and Gilbert (15). DNA from both the phage and plasmid vectors was used.

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: bp, base pair(s); ASV, avian sarcoma virus; LTR, long terminal repeat; U_5, the sequence in the LTR in viral DNA that is unique to the 5′ end of the viral RNA; U_3, the sequence in the LTR that is unique to the 3′ end of viral RNA; R, the sequence in the LTR present at both ends of viral RNA.

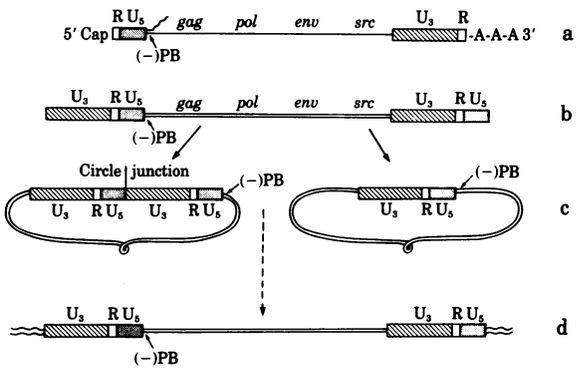

FIG. 1. Outline of ASV DNA synthesis and integration. (*a*) Genetic map of the RNA genome with the regions relevant to viral DNA synthesis. *gag*, *pol*, and *env* are the viral genes required for replication and *src* is the viral transforming gene (3). A 16- to 21-nucleotide terminal repeat is present in the RNA and this sequence is denoted R (4–6). The stippled box represents the U_5 region, defined as the region between the $tRNA^{Trp}$ binding site and the R sequence at the 5′ end of the RNA (7). The hatched box represents the U_3 region, defined as the sequence at the 3′ end of the RNA (excluding R) that is duplicated during the generation of the terminal repeat in the DNA (7). The $tRNA^{Trp}$ binding site is labeled (−)PB, the binding site for the primer of the first (minus) strand of viral DNA (8, 9). In this diagram the U_5, R, and U_3 regions are shown on an expanded scale. (*b*) Structure of unintegrated duplex linear DNA. The linear DNA is coextensive with the RNA but contains a LTR composed of the U_3, R, and U_5 regions (10, 11). (*c*) Structures of the two forms of circular DNA. The two forms differ by the presence of one or two copies of the LTR sequence (10, 11). The site where the ends of the linear DNA are fused is referred to as the circle junction. (*d*) The integrated (proviral) DNA is also coextensive with the RNA and contains the LTR sequence at each boundary with host sequences (12, 13).

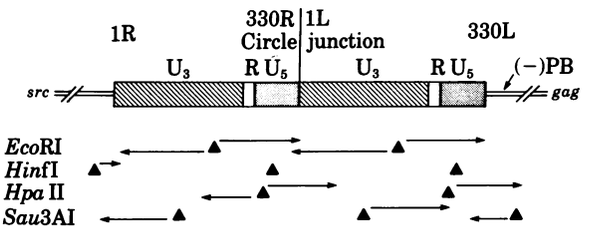

FIG. 2. Strategy for determining the nucleotide sequence of the circle junction region of SRA-2. A physical map of the circle junction region with the restriction endonuclease sites of four enzymes used in DNA sequencing is shown relative to the regions illustrated in Fig. 1. The nucleotide sequences in fragments of DNA generated by restriction endonuclease cleavage were determined by using the chemical cleavage method described by Maxam and Gilbert (15). The arrows indicate the direction of sequencing from the various sites. The numbering system is described in the text and in the legend to Fig. 3.

RESULTS

Strategy for Determining the Sequence of the Circle Junction Region. We have recently cloned circular ASV DNA isolated from cells infected with the Schmidt–Ruppin A strain of ASV (14). Two clones, SRA-1 and SRA-2, were chosen for study because they appeared from restriction endonuclease mapping to contain two complete, or nearly complete, copies of the LTR sequence. These two clones were amplified and subcloned for further mapping and nucleotide sequence analysis.

A sequence determination strategy utilizing restriction endonuclease sites around the circle junction site of SRA-2 is shown in Fig. 2. This figure is drawn so that the region from the right end of linear DNA appears to the left of the circle junction; the region from the left end is to the right of the junction. The sequence determined for 350 nucleotides on either side of this site is presented in Fig. 3. A summary of restriction endonuclease sites within this sequence is available upon request.

The Circle Junction is the Center of a 330-bp Direct Repeat. The most obvious feature of the sequence shown in Fig. 3 is the presence of a 330-bp perfect direct repeat. We have oriented the repeat with respect to the ends of the linear DNA by restriction endonuclease mapping (14). One copy of the repeated domain, numbered 1R to 330R, is the copy proximal to

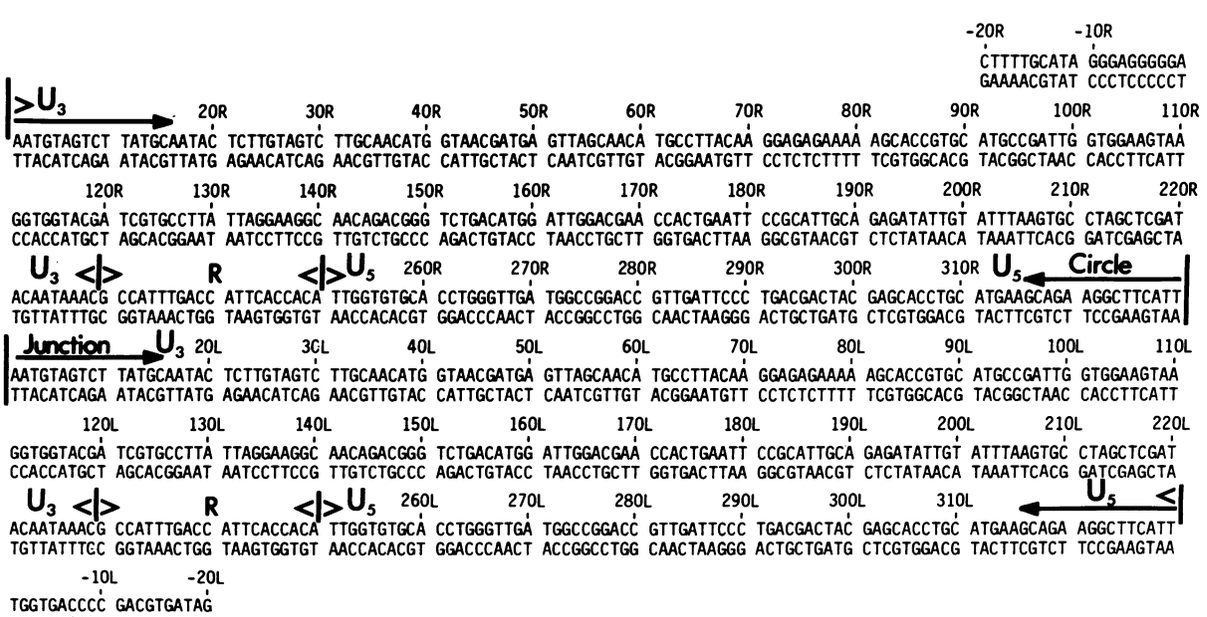

FIG. 3. Sequence surrounding the circle junction site. The sequence determined for 350 nucleotides on either side of the circle junction is illustrated. The repeated domain is numbered 1 to 330, starting with the boundary in the U_3 region. The two copies of the repeated domain are distinguished by an L (left) or R (right) following the number. Left and right refer to the ends of the linear DNA as presented in Fig. 1. Note that the region from the left end of linear DNA is shown to the right of the circle junction site and that the region from the right end of linear DNA is to the left of the circle junction site. Negative numbers refer to sequences flanking the repeated domain, numbering from the boundaries.

src and presumably represents the LTR sequence from the right end of the linear DNA. The other copy of the repeated domain, numbered 1L to 330L, is the copy proximal to *gag* and presumably represents the LTR sequence from the left end of the linear DNA. The flanking unique sequences are denoted with negative numbers, starting at each boundary of the direct repeat.

Portions of the sequence within the repeated domain can be recognized by comparison to previous sequence determination studies. The U_5 region (see Fig. 1) is present at positions 251 to 330. There are a total of seven nucleotide substitutions in this region compared to the sequence of the Prague C strain of virus used in previous cDNA sequence determination studies (4, 5). The R sequence, positions 230 to 250, represents the short redundancy at the ends of the genome RNA (4–6). The U_3 region covers positions 1 to 229. The sequences of portions of the U_3 region have been determined previously by using oligo(dT)-primed cDNA transcripts from the 3′ end of the RNA genome (D. E. Schwartz and W. Gilbert, personal communication) and by using cloned cDNAs synthesized with ASV mRNA as template (16). The portion of the sequence from 170R to −20R has also been determined by the dideoxy method, using a subclone of the SRA-2 clone as template (17).

We have identified the circle junction site by assuming that the first nucleotide of the U_5 sequence should represent the extreme right end of linear viral DNA (see *Discussion*). On the basis of this assumption the circle junction can be recognized as the center of the direct repeat because the entire U_5 sequence is present (251R to 330R), terminating at the center of the direct repeat (Fig. 3). The relevant portions of the U_5 sequences from the Prague C and B77 strains of ASV (4, 5) are compared to the sequence from the cloned DNA, and all give an equivalent sequence (Fig. 4*a*). It would appear that circularization occurred without loss of information from at least the right end of the linear DNA. The sequence starting with position 1L must therefore represent information from the left end, perhaps the exact left end, of the linear DNA (see *Discussion*).

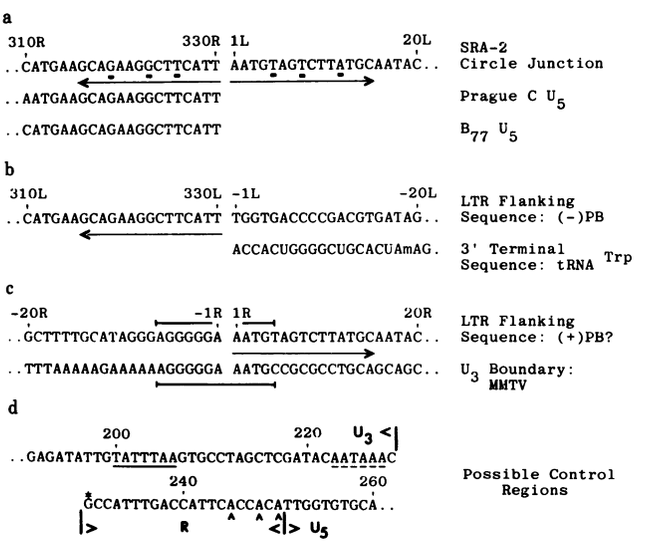

FIG. 4. Summary of the ASV DNA sequence in the circle junction region. The numbering is as described in the legend to Fig. 3. In each case the plus strand sequence is shown. (*a*) The sequence at the circle junction site is compared to a portion of the published U_5 sequences of two closely related viruses (4, 5). The arrow denotes the inverted repeat present at the circle junction. The underlined nucleotides are mismatched in the repeat. (*b*) The sequence at one of the boundaries of the direct repeat (proximal to the *gag* gene) and the flanking sequence are compared to the 3′ terminal sequence of chicken $tRNA^{Trp}$ (8). The homology with $tRNA^{Trp}$ identifies this region of viral DNA as coding for the primer binding site, (−)PB. (*c*) The sequence at the other boundary of the direct repeat (proximal to the *src* gene) and the flanking sequence containing the polypurine tract are compared to the sequence at the equivalent site in murine mammary tumor virus (MMTV) DNA (18, 19). The boundary may represent the site of priming of the second strand of viral DNA, (+)PB. (*d*) Sequences within the LTR sequence that may be involved in control of transcription are underlined and described in the text. The asterisk denotes the nucleotide adjacent to the m^7G cap in the genome RNA (4, 5, 20–22). The vertical arrowheads show the positions of the polyadenylylation sites in genome RNA (6).

An examination of the circle junction site reveals that it is not only the center of the 330-bp direct repeat but also the center of a 15-bp imperfect inverted repeat, or palindrome, with 12 out of the 15 nucleotides forming the repeat (Fig. 4*a*; 316R to 330R and 1L to 15L). Because the circle junction site is the center of the direct repeat, the sequences representing the inverted repeat are also present at the boundaries of the direct repeat (1R to 15R and 316L to 330L) (Fig. 3).

Sequences Flanking the Direct Repeat. The primer for synthesis of the first strand of viral DNA is a cellular $tRNA^{Trp}$ which is bound to the genome 101 nucleotides from the 5′ end of the viral RNA (4, 5, 8, 9). Flanking the right boundary of the direct repeat, proximal to the *gag* gene, is an 18-nucleotide stretch having perfect homology to the 3′ end of chicken $tRNA^{Trp}$ (8) (−1L to −18L; Fig. 4*b*). The region of homology terminates at the position of a m^1A in the $tRNA^{Trp}$ sequence.

The left boundary of the direct repeat, proximal to the *src* gene, may represent the site of initiation of synthesis of the second strand of viral DNA ("plus strong stop DNA," see *Discussion*). The sequence flanking this boundary should then represent the position of the primer for plus strong stop DNA. The left boundary of the direct repeat is flanked by a polypurine tract, A-G-G-G-A-G-G-G-G-G-A, in the plus strand (−1R to −11R; Fig. 4*c*). It is likely that this polypurine tract represents a portion of the sequence of the as yet unidentified primer for the second strand of viral DNA.

Transcriptional Control Regions Within the LTR Sequence. The LTR sequence may serve regulatory functions in the synthesis and processing of viral RNA (7). An examination of the LTR sequence suggests that control regions are present. (*i*) An A+T-rich region 25–30 nucleotides upstream from the RNA capping site may serve a promoter-like function in the synthesis of eukaryotic mRNAs (23). The sequence on the 5′ side of the ASV RNA capping site contains an A+T-rich region (T-A-T-T-T-A-A) starting 24 nucleotides upstream from the capping site (Fig. 4*d*). (*ii*) The sequence A-A-T-A-A-A appears in the LTR domain (Fig. 4*d*), and this sequence has been implicated as a poly(A) addition signal in eukaryotic mRNAs (24). Seventeen to 22 nucleotides downstream from this sequence are three poly(A) addition sites in viral RNA (6).

A Circle Junction Sequence Lacking Portions of U_3 and U_5. We examined the nucleotide sequence at the circle junction site of another cloned molecule, SRA-1. Previous mapping experiments had shown that SRA-1 has a short deletion within the LTR sequence (14). Further mapping experiments placed the deletion at or near the circle junction site (data not shown). The nucleotide sequence of the circle junction site of SRA-1 is shown in Fig. 5. By comparison to the SRA-2 sequence (Fig. 3) it can be seen that there is a deletion at the circle junction site, with 2 nucleotides missing from the U_5 sequence and 61 nucleotides missing from the U_3 sequence. The mapping data indicate that the remainder of each copy of the LTR sequence is intact. Because we have not observed similar deletions during propagation of the SRA-2 clone in either the phage or plasmid vector, we assume that the deletion present in SRA-1 occurred *in vivo* and is not an artifact of molecular cloning.

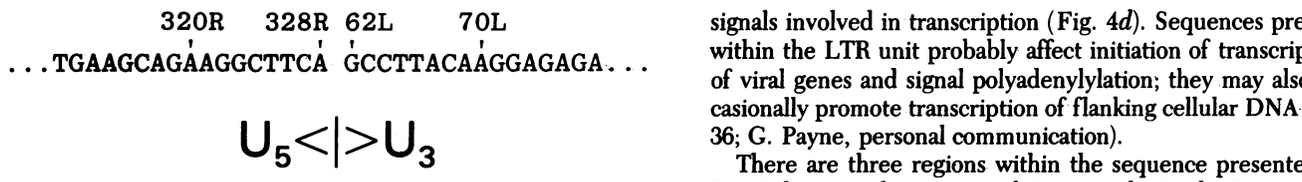

FIG. 5. Sequence at the circle junction site of SRA-1. The numbering is the same as for the sequence determined by using the SRA-2 clone (Fig. 3).

DISCUSSION

The circle junction site of the cloned ASV DNA molecule (SRA-2) is the center of a 330-bp direct repeat. We believe the direct repeat in the circular DNA arose by fusion of the ends of linear viral DNA and that the direct repeat represents the complete copies of the repeated domains composing the terminal repeat (LTR) at the ends of the linear DNA. The circle junction is also the center of a 15-nucleotide-long inverted repeat that concludes the LTR. The overall sequence organization of ASV DNA—a unique sequence (here the ASV coding region), flanked by a direct repeat (the LTR), with the repeated domains concluded by a short inverted repeat—is analogous to the structure of certain prokaryotic and eukaryotic transposable elements (25–28). It is tempting to speculate that homologous structures have been selected as the consequence of an event common to the replication of retroviruses and the mechanism of transposition (for example, the insertion of DNA into new sites in the host genome). However, at present there are no data to suggest any common mechanism utilizing these homologous structures.

The interpretation that the direct repeat centered at the circle junction site represents the entire sequence within the LTR is based upon the presence of the entire U_5 sequence at the circle junction site. Studies *in vitro* and *in vivo* suggest that the right end of the linear DNA is defined by the region immediately adjacent to the $tRNA^{Trp}$ binding site, the U_5 sequence (refs. 3, 10, 29, and 30; Fig. 1). The entire U_5 sequence is present at the circle junction site, suggesting that circularization occurred without loss of information from at least the right end, and, by inference, from either end of the linear DNA.

Three other observations argue that the entire sequence from the left end of linear DNA is present at the circle junction in SRA-2 DNA. First, restriction endonuclease mapping of linear DNA (10) places the left end 150 to 200 bp to the left of the *Eco*RI site at position 178 in the U_3 region. Second, the presence of inverted repeats at the ends of the DNA gives striking symmetry to the structure of viral DNA. Inverted repeats of various lengths between 3 and 21 nucleotides have been observed at the circle junction of a cloned MuLV DNA molecule (31) and at the host–virus junction of several cloned proviral DNAs (18, 19, 32–34). The third relevant observation also concerns the structure of proviral DNA. Nucleotide sequence analyses of the junctions of host and proviral DNA have shown that the provirus is usually missing two base pairs from the U_5 region (16, 18, 19, 32, 33, 35); these analyses have also shown that the last two nucleotides of viral sequence at the left end of proviral DNA are T-G (18, 19, 32–34). Assuming that two base pairs are lost from each end of viral DNA during integration, then the left end of linear DNA is two base pairs from this dinucleotide. ASV may follow this pattern, since a T-G dinucleotide appears in the U_3 sequence, two base pairs from the circle junction site. Thus we infer that the sequences at the exact ends of linear ASV DNA are represented by the sequence composing the circle junction site in the SRA-2 molecule.

The LTR, like several transposable elements (25), may carry signals involved in transcription (Fig. 4*d*). Sequences present within the LTR unit probably affect initiation of transcription of viral genes and signal polyadenylylation; they may also occasionally promote transcription of flanking cellular DNA (ref. 36; G. Payne, personal communication).

There are three regions within the sequence presented in Fig. 3 that are relevant to viral DNA synthesis: the R sequence and the regions flanking either side of the direct repeat.

(*i*) We have noted that the sequence representing the short terminal repeat in genomic RNA is present only once in the LTR sequence in DNA synthesized *in vitro* or *in vivo* (37). This observation conforms to the model for DNA synthesis in which a DNA copy of the R sequence, from the 5′ end of the RNA genome, base-pairs with the R sequence at the 3′ end of the RNA to allow DNA synthesis to continue beyond the 5′ end of the template (7).

(*ii*) The regions flanking either side of the direct repeat probably represent sequences related to the primers for the first and second strands of viral DNA. Immediately flanking the right boundary of the direct repeat, proximal to *gag*, is an 18-nucleotide stretch homologous to the 3′ terminus of $tRNA^{Trp}$, the primer for synthesis of the first strand of viral DNA (8). The length of homology between the viral sequence and $tRNA^{Trp}$ that we have detected extends slightly the length of homology previously throught to be present, as determined by RNA sequencing (38, 39). The region of homology terminates with the appearance of the first modified nucleotide, m^1A, in the tRNA.

(*iii*) The major species of plus strand DNA identified in infected cells is about 300 nucleotides long and is synthesized using the right end of the minus strand (the U_3-R-U_5 region) as template (29). This species has been referred to as plus strong stop DNA (40). A model has been proposed in which the minus strand is completed by copying plus strong stop DNA (10, 40). In this way the U_3-R-U_5 region that is present at the right end of the minus strand is duplicated at the left end, forming the second copy of the LTR sequence. In this model for DNA synthesis the site at which plus strong stop DNA is primed defines the boundary of the U_3 region in the LTR sequence. Using the circle junction site to define the boundary of the U_3 region, we find a polypurine tract, A-G-G-G-A-G-G-G-G-G-A, in the plus strand at the site expected for the primer for plus strong stop DNA. A polypurine tract 21 nucleotides long is present in an equivalent position in murine mammary tumor virus DNA, with 11 nucleotides at the U_3 boundary being identical between ASV and murine mammary tumor virus (Fig. 4*c*; refs. 18 and 19). It is likely that this polypurine tract represents a portion of the sequence of the primer for plus strong stop DNA, although we do not yet know the nature of that primer.

Our data indicate that circularization can occur by more than one mechanism. In addition to circular DNA molecules that appear to contain either one or two copies of the LTR sequence (10, 11, 14), we have identified a molecule that contains two copies of the LTR sequence, each of which has suffered a deletion. The loss of two base pairs from the U_5 region is analogous to the absence of the two equivalent base pairs of other retroviruses in proviral DNA (16, 18, 19, 32, 33, 35). For this reason we suspect that the DNA molecule carrying the deletion at the circle junction site arose by recombination involving a mechanism similar to integration. Shoemaker *et al.* (31) have also encountered unusual circular molecules of murine leukemia virus DNA containing rearrangements that they interpret as arising by integration of one region of the DNA into another region of the DNA. Those authors emphasize models in which the rearrangement of viral DNA follows circularization. In the case we have examined it is simpler to imagine that one end of linear DNA was integrated (with the loss of 2 base pairs) 61 base pairs

from the other end. Ju *et al.* (41) and Highfield *et al.* (42) have observed another class of circular DNA molecules; most of the molecules they found are not similar to either SRA-1 or SRA-2 but contain variable deletions at the circle junction site within the U_3 and U_5 regions. It seems likely that circular viral DNA molecules found *in vivo* are a heterogeneous population of structures arising by several mechanisms.

We thank J. Majors for many helpful discussions, J. Majors, R. Parker, D. Schwartz, and G. Payne for communication of results prior to publication, and H. Martinez and P. Czernilofsky for assistance with computer analysis of the sequence. We also thank J. Migneault for excellent stenographic assistance. This work was supported by U.S. Public Health Service Grants CA 12705 and CA 19287, Training Grant IT32 CA 09043, and American Cancer Society Grant VC-70. W.J.D. holds a fellowship from the Leukemia Society of America.

1. Weinberg, R. A. (1977) *Biochim. Biophys. Acta* **473**, 39–56.
2. Shank, P. R. & Varmus, H. E. (1978) *J. Virol.* **25**, 104–114.
3. Bishop, J. M. (1978) *Annu. Rev. Biochem.* **47**, 35–88.
4. Shine, J., Czernilofsky, A. P., Friedrich, R., Goodman, H. M. & Bishop, J. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1473–1477.
5. Haseltine, W. A., Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 989–993.
6. Schwartz, D. E., Zamecnik, P. C. & Weith, H. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 994–998.
7. Coffin, J. M. (1979) *J. Gen. Virol.* **42**, 1–26.
8. Harada, F., Sawyer, R. C. & Dahlberg, J. D. (1975) *J. Biol. Chem.* **250**, 3487–3497.
9. Taylor, J. M. & Illmensee, R. (1975) *J. Virol.* **16**, 553–558.
10. 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**, 1383–1395.
11. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. (1978) *J. Virol.* **28**, 810–818.
12. 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.
13. Sabran, J. L., Hsu, T. W., Yeater, C., Kaji, A., Mason, W. S. & Taylor, J. M. (1979) *J. Virol.* **29**, 170–178.
14. DeLorbe, W. J., Luciw, P. A., Goodman, H. M., Varmus, H. E. & Bishop, J. M. (1980) *J. Virol.* **36**, 50–61.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
16. Yamamoto, T., Tyagi, J. S., Fagan, J. B., Jay, G., deCrombrugghe, B. & Pastan, I. (1980) *J. Virol.* **35**, 436–443.
17. Czernilofsky, A. P., DeLorbe, W., Swanstrom, R., Varmus, H. E., Bishop, J. M. Tischer, E. & Goodman, H. M. (1980) *Nucleic Acids Res.* **8**, 2967–2984.
18. Majors, J. E. & Varmus, H. E. (1980) in *Animal Virus Genetics*, ICN-UCLA Symposia on Cellular and Molecular Biology, eds. Fields, B., Jaenisch, R. & Fox, C. F. (Academic, New York), in press.
19. Majors, J. E. & Varmus, H. E. (1980) *Nature (London)*, in press.
20. Beemon, K. L. & Keith, J. M. (1976) in *Animal Virology*, ICN-UCLA Symposia on Cellular and Molecular Biology, eds. Baltimore, D., Huang, A. & Fox, C. F. (Academic, New York), Vol. 4, pp. 97–105.
21. Furuichi, Y., Shatkin, A. J., Stavnezer, E. & Bishop, J. M. (1975) *Nature (London)* **257**, 613–620.
22. Keith, J. & Fraenkel-Conrat, H. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3347–3350.
23. Ziff, E. B. & Evans, R. M. (1978) *Cell* **15**, 1463–1475.
24. Proudfoot, N. J. & Brownlee, G. G. (1974) *Nature (London)* **252**, 359–362.
25. Calos, M. P. & Miller, J. H. (1980) *Cell* **20**, 579–595.
26. Farabaugh, P. J. & Fink, G. R. (1980) *Nature (London)* **286**, 352–356.
27. Gafner, J. & Philippsen, P. (1980) *Nature (London)* **286**, 414–418.
28. Finnegan, D. J., Rubin, G. M., Young, M. W. & Hogness, D. S. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1053–1063.
29. Varmus, H. E., Heasley, S., Kung, H.-J., Oppermann, H., Smith, V. C., Bishop, J. M. & Shank, P. R. (1978) *J. Mol. Biol.* **120**, 55–82.
30. Taylor, J. M., Illmensee, R., Trusal, L. R. & Summers, J. (1976) in *Animal Virology*, ICN-UCLA Symposia on Cellular and Molecular Biology, eds. Baltimore, D., Huang, A. & Fox, C. F. (Academic, New York), Vol. 4, pp. 161–174.
31. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. & Baltimore, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3932–3936.
32. Dhar, R., McClement, W. L., Enquist, L. W. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3937–3941.
33. Shimotohno, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285**, 550–554.
34. Van Beveren, C., Goddard, J. G., Berns, A. & Verma, I. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3307–3311.
35. Hager, G. L. & Donehower, L. A. (1980) in *Animal Virus Genetics*, ICN-UCLA Symposia on Cellular and Molecular Biology, eds. Fields, B., Jaenisch, R. & Fox, C. F. (Academic, New York), in press.
36. Quintrell, N., Hughes, S. H., Varmus, H. E. & Bishop, J. M. (1980) *J. Mol. Biol.*, in press.
37. Swanstrom, R., Varmus, H. E. & Bishop, J. M. (1980) *J. Biol. Chem.*, in press.
38. Cordell, B., Stavnezer, E., Friedrich, R., Bishop, J. M. & Goodman, H. M. (1976) *J. Virol.* **19**, 548–558.
39. Coffin, J. M. & Haseltine, W. A. (1977) *J. Mol. Biol.* **117**, 805–814.
40. Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. (1979) *Cell* **18**, 93–100.
41. Ju, G., Boone, L. & Skalka, A. M. (1980) *J. Virol.* **33**, 1026–1033.
42. Highfield, P. E., Rafield, L. F., Gilmer, T. M. & Parsons, J. T. (1980) *J. Virol.* **36**, 271–279.