THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 256, No. 3, Issue of February 10, pp. 1115–1121, 1981
*Printed in U.S.A.*

# The Terminal Redundancy of the Retrovirus Genome Facilitates Chain Elongation by Reverse Transcriptase*

(Received for publication, May 27, 1980, and in revised form, September 19, 1980)

**Ronald Swanstrom, Harold E. Varmus, and J. Michael Bishop**

*From the Department of Microbiology and Immunology, University of California, San Francisco, California 94143*

**Transcription of DNA from the RNA genome of retroviruses by reverse transcriptase involves an unusual translocation of the growing chain from the 5′ end to the 3′ end of the RNA template. In order to elucidate the mechanism by which this translocation occurs, we have used chain termination to analyze nascent viral DNA synthesized *in vitro* by avian sarcoma virus, and we have determined the nucleotide sequence of appropriate regions of viral DNA isolated from infected cells and cloned into prokaryotic vectors. Our results provide direct experimental evidence for a previously proposed model in which a short terminal redundancy in viral RNA, and a DNA copy of the redundant sequence, are used to allow the growing DNA chain to move from the 5′ to the 3′ end of the template.**

**Transcription of avian sarcoma virus RNA with purified reverse transcriptase also generates an anomalous product, a hairpin DNA that arises when the initial DNA transcript folds back on itself to continue synthesis. The foldback is mediated by an inverted repeat of 5 nucleotides in the sequence of nascent DNA. Anomalous hairpin DNA is not produced by detergent-activated virions. Thus, constituents of the virions or the configuration of encapsidated viral RNA must facilitate correct transcription.**

The replication of retroviruses is mediated by a virus-specific DNA intermediate (1), synthesized during the early hours of infection and subsequently integrated into the chromosomal DNA of the host cell. Viral DNA is transcribed from the single-stranded RNA genome of retroviruses by "reverse transcriptase" (2), an RNA-directed DNA polymerase encoded in a viral gene and encapsidated in virus particles. The products of viral DNA synthesis in cells infected with retroviruses include both linear and circular duplex molecules (3–6). The linear form of viral DNA is bounded by a direct terminal redundancy composed of nucleotide sequences representing domains of both the 3′ and 5′ ends of the viral RNA genome (7, 8); some of the circular molecules contain both copies of the redundant domains, others contain only one copy (7–10).

Our knowledge of how these forms of retrovirus DNA are generated from a linear single-stranded template remains incomplete. In particular, we need to account for the replication of the ends of the linear RNA template by primer-dependent polymerization (11), the genesis of a terminal redundancy in the linear DNA product, and the circularization of a portion of the products. It has been argued previously that mechanisms for the first and second of these events are suggested, at least in part, by events that occur during the course of viral DNA synthesis *in vitro* or *in vivo* (for a review, see Ref. 12). The present communication confirms and extends this view by providing both a detailed analysis of early events during the transcription of DNA from the genome of avian sarcoma virus *in vitro* and a correlative analysis of DNA produced in infected cells.

Transcription of the ASV[1] genome inititates on a $tRNA^{Trp}$ primer located about 100 nucleotides from the 5′ end of the RNA (13), proceeds to the end of the template, and then moves to the vicinity of the 3′-terminus of the template (14–16). As a result, both ends of the RNA template are copied in tandem, shortly after initiation of DNA synthesis. Fig. 1 illustrates a model of how these events might occur (14, 17–21). The model derives from the finding that the RNA genome of ASV (and of closely related viruses) possesses a direct terminal redundancy, each copy composed of 16 to 21 nucleotides ($DTR_{21}$) (17–19, 22) and included in the terminal redundancy in viral DNA described above (7, 8). In the model, RNA at the 5′-terminus of the viral genome is removed from the complementary DNA transcript (perhaps by the action of RNase H activity associated with reverse transcriptase (2)), and the DNA base-pairs with the $DTR_{21}$ sequence at the 3′ end of the template. As a consequence, the nascent DNA is in position for continued transcription from the full length of the template. In addition, the joining of sequences from the 5′ and 3′ ends of the RNA template in the DNA creates the sequence organization as it appears at one end of the final DNA product.

On the basis of this model, we can make two predictions concerning the structure of the initial DNA transcript that joins the 5′ and 3′ domains of the template. First, the DNA will contain only one copy of $DTR_{21}$, whereas the template had two copies at the outset. Second, transcription from the 3′ domain of the template will begin with the nucleotide immediately adjacent to the 5′ boundary of $DTR_{21}$. Data consistent with the first prediction were obtained previously by studying the pyrimidine tracts of DNA synthesized *in vitro* by murine leukemia virus (23), and more recently by nucleotide sequence analysis of cloned murine leukemia virus cDNA (24). However, in these studies the precise site of transcription from the 3′ region of the RNA template was not determined. We now demonstrate that the nucleotide sequence of ASV DNA synthesized either *in vitro* or *in vivo* fulfills both predictions. In addition, we show that transcription of the ASV genome by purified reverse transcriptase (reconstructed reaction) frequently copies nucleotide sequences at the 5′ end of the viral genome into a hairpin structure, and we describe the structural basis of this phenomenon. We have not observed

* This work was supported by United States Public Health Service Grants CA 12705, CA 19287, and Training Grant 1T32 CA 09043, and American Cancer Society Grant VC-70. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] The abbreviations used are: ASV, avian sarcoma virus; $DTR_{21}$, the 16- to 21-nucleotide-long terminal redundancy in the ASV genome.

this hairpin species during synthesis by the holoenzyme in detergent-disrupted virions (endogenous reaction). Our data conform to previous indications that the synthesis of retrovirus DNA proceeds by identical mechanisms *in vitro* and *in vivo*, although the use of purified reverse transcriptase may introduce artifacts not encountered in either DNA synthesized by detergent-disrupted virions or viral DNA synthesis in infected cells.

## RESULTS[2]

*Early Events in the Synthesis of ASV DNA*—We used the dideoxy chain terminator technique for DNA sequencing (25) to examine early events during the synthesis of ASV DNA. This sequencing technique was adapted to allow the use of reverse transcriptase, rather than *Escherichia coli* polymease I, so that nascent DNA transcripts from viral RNA could be examined. We determined the structure of viral DNA transcripts under five different circumstances (listed below), including the endogenous reaction, reconstructed reaction, and DNA synthesized *in vivo*, and compared the sequences obtained to the known sequences at the 5′ (18, 22) and 3′[3] (17) ends of viral RNA. In each instance the data provided support for the model described above.

(i) "Endogenous reactions" with detergent-disrupted virions. Examination of the sequence of nascent DNA transcripts synthesized in the endogenous reaction shows that only one copy of the $DTR_{21}$ sequence is present (Fig. 2). The sequence of the first two nucleotides transcribed beyond the 5′ end of the template could not be determined due to experimental artifact (see miniprint, Fig. 3*A*). The third nucleotide beyond the 5′ end of the template represents transcription from the third position upstream of the $DTR_{21}$ at the 3′ end of the template. We assume that the two obscured positions represent synthesis from the two nucleotides immediately adjacent to the 3′-$DTR_{21}$ sequence, an assumption confirmed by sequencing DNA synthesized *in vivo* (see below).

(ii) The sequence of DNA synthesized in the endogenous reaction in the presence of actinomycin D, an inhibitor of DNA-dependent DNA synthesis (26), was identical with the sequence obtained in the absence of the drug, indicating that RNA is serving as the template for the transcripts extended beyond the 5′ end of the template (Fig. 2). We observed a substantial decrease in the frequency of transcripts extended beyond the 5′ end of the template in the presence of actinomycin D (see miniprint, Figs. 3*B* and 5). This is probably due to the drug binding to the DNA complement of the $DTR_{21}$ sequence and inhibiting base-pairing with the 3′-$DTR_{21}$ sequence (27).

(iii) DNA transcripts synthesized in the reconstructed reaction have a sequence similar to the sequence determined in the endogenous reaction (Fig. 2), although the situation is complicated by the presence of a second type of transcript. Beyond the 5′ end of the template, transcription occurs from two templates, giving rise to a double sequence (see miniprint, Fig. 4). The double sequence can be divided into extended transcription from the 3′ end of the template (Fig. 2), as with the endogenous reaction, and a foldback transcript giving rise to a hairpin product (see below).

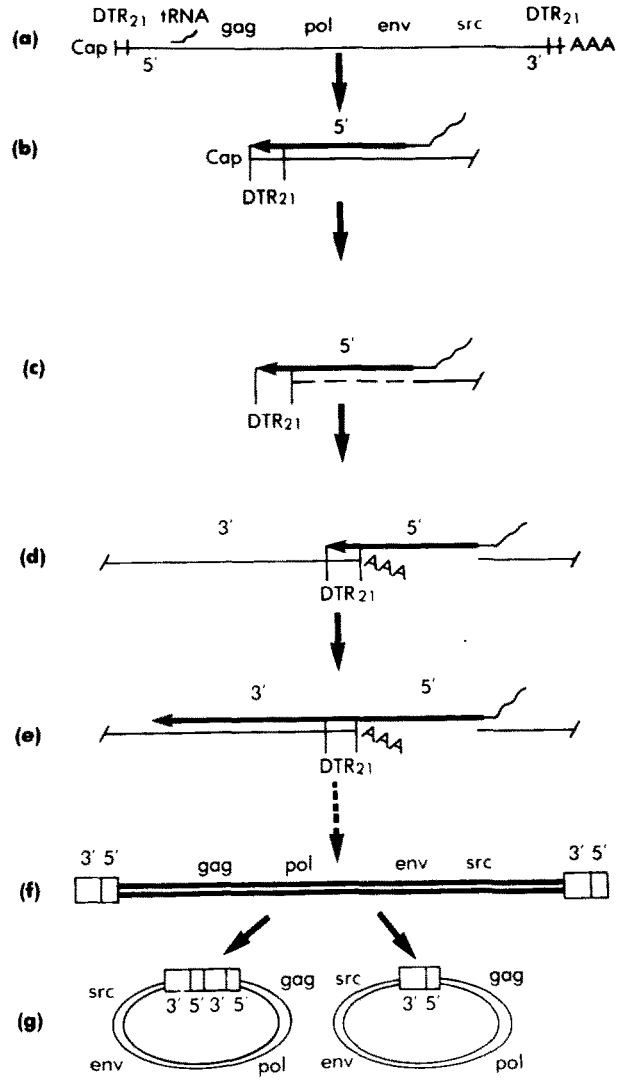


FIG. 1. **Model for elongation of the initial DNA transcript beyond the 5′ end of the template.** The model is explained in the text and is essentially as previously described (14, 17–21). *a*, ASV genome RNA. *b*, initiation of DNA synthesis and transcription to the 5′ end of the template. *c*, removal of the template RNA by RNase H activity. *d*, pairing of the $DTR_{21}$ sequence at the 3′ end of the template with the complement of the $DTR_{21}$ sequence present in the nascent DNA. *e*, elongation of the nascent DNA using the 3′ end of the RNA as template. *f*, duplex linear DNA containing a terminal redundancy. *g*, two forms of circular DNA. The two forms differ by the absence in one of the forms of one of the copies of the redundant domains present at the ends of the linear DNA. The site at which the redundant domains in the linear DNA are joined in the circular DNA is referred to as the circle junction. *$DTR_{21}$*, the 16- to 21-nucleotide-long terminal redundancy present in the ASV genome (17–19, 22). *gag-pol*, *env*, and *src* refer to the known viral genes (29).

(iv) Our interpretation of the double sequence is supported by experiments with viral RNA lacking the 3′ end; this template supports the synthesis of only the hairpin DNA (see miniprint, Figs. 5 and 6). This last result also supports the model; removal of the 3′ end of the template blocks the synthesis of the extended transcripts that are observed in the endogenous reaction and predicted by the model. The viral polymerase and viral 70 S RNA appear to be the only virion components required to accurately carry out the initial steps of DNA synthesis.

(v) We examined the sequence of viral DNA which had been synthesized *in vivo* and amplified by molecular cloning

[2] Portions of this paper (including "Experimental Procedures," Figs. 3 to 7 (with text) of "Results," and additional references) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, Md. 20014. Request Document 80 M-1058, cite author(s), and include a check or money order for $7.60 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

[3] D. Schwartz and W. Gilbert, personal communication.

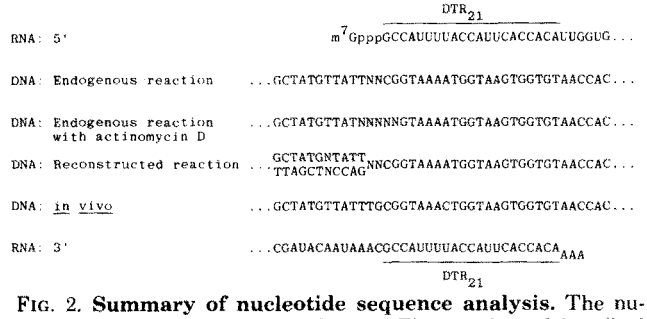


FIG. 2. **Summary of nucleotide sequence analysis.** The nucleotide sequence determinations shown in Figs. 3 and 4 and described in Fig. 7 are compared to the known nucleotide sequences at the 5′ end (18, 22) and the 3′ end[4] (17) of the template RNA. *DNA: endogenous reaction,* DNA sequence determined from the endogenous reaction (Fig. 3*A*). *DNA: endogenous reaction with actinomycin D,* DNA sequence determined from the endogenous reaction containing actinomycin D at 100 μg/ml (Fig. 3*B*). *DNA: reconstructed reaction,* DNA sequence determined from the reconstructed reaction using purified reverse transcriptase and 70 S viral RNA as template (Fig. 4). *DNA: in vivo,* DNA sequence determined from the region of cloned ASV DNA as described in the legend to Fig. 7. The nucleotide sequence of DNA synthesized *in vitro* is extended beyond the sequence shown in Figs. 3 and 4 to facilitate comparison of the entire $DTR_{21}$ sequence.

(28). The structure of the cloned viral DNA also supports the model; the region of this DNA analogous to the region studied *in vitro* contained one copy of the $DTR_{21}$ sequence, and the upstream sequence represented the nucleotides immediately upstream from the $DTR_{21}$ sequence at the 3′ end of the RNA template (Fig. 2). Thus, in each of the settings studied all of the sequences present at the 5′ and 3′ ends of the template are present in the nascent DNA transcript, fused by one copy of the $DTR_{21}$ sequence.

## DISCUSSION

*Early Events in the Synthesis of ASV DNA*—Our findings substantiate and extend previous accounts of the early events in the synthesis of retrovirus DNA (29). In particular, we conclude that DNA transcribed from the 5′ end of the viral genome subsequently base-pairs with a complementary nucleotide sequence at the 3′ end of the genome. In this way primer-dependent replication of the template ends is accomplished, and nascent DNA is in position for continued transcription from the full length of the template. These conclusions provide a function for the terminal redundancy in the viral genome and, therefore, account for conservation and cosegregation of the redundant nucleotide sequences during genetic recombination among retroviruses (30, 31). A similar conclusion has been reached after analyzing pyrimidine tracts (23), and more recently a cDNA clone (24), of DNA transcripts produced in the endogenous reaction of murine leukemia virus.

As can be seen in Fig. 1, the early intermediate we have studied has the sequence organization found at the right end of the linear viral DNA (7, 8). The generation of these sequences is accomplished with the apparent sacrifice of one copy of the $DTR_{21}$ sequence, since two copies of the $DTR_{21}$ sequence in the template are used to generate one copy in the early DNA intermediate. The loss of information is rectified during the generation of the terminal redundancy in the DNA (7, 8, 32). Our own sequencing studies support this view, showing that each copy of the redundant domains in the DNA contains one copy of the $DTR_{21}$ sequence.[4]

Since the genome of retroviruses is diploid (12), two templates for viral DNA synthesis are available in each virion. Does each of these templates give rise to a haploid unit of viral DNA, or do they collaborate in the production of a single molecule of viral DNA? The mechanism illustrated in Fig. 1 permits either possibility and, therefore, leaves the puzzle unsolved. It has been suggested previously that both genomes of a heterozygous retrovirus particle are represented in the progeny of a single infectious event (33, 34). This suggestion can be correct only if each haploid subunit of a retrovirus genome is completely transcribed into biologically active DNA.

*Viral DNA Synthesis in Vitro and in Vivo*—The synthesis of retrovirus DNA has been studied in three settings: the infected cell, "endogenous reactions" with detergent-disrupted virions, and "reconstructed reactions" with purified reverse transcriptase and viral RNA. Viral DNA synthesis in the first and second of these settings may differ by very little; previous reports indicate that endogenous reactions with murine leukemia/sarcoma virus can produce linear duplexes that are identical with those synthesized in infected cells (35, 36), and our present findings indicate that the early events during viral DNA synthesis *in vivo* are probably identical with those in the ASV endogenous reactions.

Reconstructed reactions, while capable of carrying out the initial steps of DNA synthesis (see miniprint, Fig. 4), usually fail to produce mature forms of viral DNA (29). We have shown here that when purified 70 S RNA is used as template, perhaps one-half of the viral DNA products extended beyond the 5′ end of the template are anomalous, consisting of hairpin copies of the sequences near the 5′ end of the RNA. It has previously been shown that when ASV 70 S RNA is used as template, a significant amount of hairpin DNA is generated (37); under certain conditions with 38 S subunit RNA as template, hairpin DNA is the only identifiable DNA product longer than 101 nucleotides (38). It appears possible that constituents of the virion or the configuration of encapsidated viral RNA facilitate correct transcription.

*Anomalous Viral DNA Synthesis in Vitro*—We attribute the synthesis of hairpin ASV DNA to a 5-nucleotide-long inverted redundancy in the nucleotide sequence of nascent DNA that allows the chain to fold back upon itself (see miniprint, Fig. 6*C*). Chain progagation can then continue, using DNA as template. This scheme is supported by the fact that actinomycin D inhibits the genesis of hairpin DNA and by the sequence of the DNA itself (38, and Figs. 4, 5, and 6*A* in miniprint).

There are two arguments that can be made for RNase H playing an obligatory role in the genesis of the hairpin DNA we observed. First, the pairing of the 5-nucleotide-long inverted repeat would appear to require that the initial DNA transcript be single-stranded from position 101 to position 56. Second, since the polymerase is not known to carry out strand displacement synthesis *in vitro* (39), the nascent DNA is probably single-stranded through position 1, allowing transcription to proceed to the primer binding site (Fig. 6). We attribute the complete removal of the RNA template from the nascent DNA transcript to the RNase H activity associated with the polymerase (2). Our initial results with an inhibitor of RNase H activity, sodium fluoride (40), support this view. The synthesis of the foldback species is much more sensitive to inhibition by this drug than is synthesis of the doublet termination product (strong stop DNA) at positions 102 and 103.

The hairpin DNA characterized here is probably identical with that described previously by Collett and Faras (38). These authors suggested that foldback in the nascent DNA might be the means by which synthesis of the second strand

[4] R. Swanstrom, W. DeLorbe, J. M. Bishop, and H. E. Varmus, manuscript in preparation.

of ASV DNA is initiated. We doubt that this suggestion is correct: hairpin DNA has not been found in either infected cells or endogenous reactions (1); the foldback described above shortens duplex viral DNA so that it no longer represents the full extent of viral RNA (see Fig. 6*C*); and formation of the hairpin DNA would preclude genesis of the terminal redundancy in the linear viral DNA.

Foldback synthesis has been observed frequently in the reverse transcription of cellular mRNAs and has been exploited to generate double-stranded DNA for subsequent cloning into prokaryotic vectors (41). We presume that foldback synthesis in these instances is a fortuitous event, akin to the process described here.

Transcription from retrovirus RNA *in vitro* frequently pauses or terminates at the 5′ end of the viral genome. The resulting DNA has been known variously as "short stop DNA," "strong stop DNA," and $cDNA_{5'}$ (42, 43); its length provides a measure of the distance from the point of initiation to the 5′-terminus of the viral genome (13, 44). We and others have previously determined the nucleotide sequence of ASV "strong stop DNA" and have reported its length as 101 nucleotides (18, 22). In the present study, however, we found two apparent species of "strong stop" DNA, both longer than expected (102 and 103 nucleotides; Figs. 3 and 4 in miniprint). What is the origin of this discrepancy? We can envision several possible explanations, none of which appear satisfactory. (i) Heterogeneity in the virus stock. We discount this explanation; we used the same virus as in the previous study, and beyond position 104, we obtained a single nucleotide sequence in phase with the alleged end of the "strong stop" DNAs. (ii) An error in the previous analyses. The agreement among results of several laboratories makes this explanation unlikely, although the previous work was performed in a manner that makes identification of the penultimate and ultimate 3′ residues in "strong stop DNA" less than certain. (iii) Transcription from the "cap" nucleotide in the RNA template. This explanation would not account for the existence of two species of "strong stop DNA," both longer than the uncapped template. In any event, we have not yet identified the nucleotides at positions 102 and 103 and, hence, cannot anticipate the composition of their template(s).

*Aknowledgments*—We thank J. Majors and P. Czernilofsky for helpful discussions, L. Levintow, R. Parker, and J. Majors for reading the manuscript, and B. Cook for excellent stenographic assistance. We also thank W. DeLorbe and R. Parker for making cloned ASV DNA available, and D. Schwartz for communication of results prior to publication.

REFERENCES

1. Weinberg, R. A. (1977) *Biochim. Biophys. Acta* **473,** 39–55
2. Verma, I. M. (1977) *Biochim. Biophys. Acta* **473,** 1–38
3. Gianni, A. M., Smotkin, D., and Weinberg, R. A. (1975) *Proc. Natl. Acad. Sci. U. S. A.* **72,** 447–451
4. Smotkin, D., Yoshimura, F. K., and Weinberg, R. A. (1976) *J. Virol.* **20,** 621–626
5. Guntaka, R. V., Mahy, B. W. J., Bishop, J. M., and Varmus, H. E. (1975) *Nature (Lond.)* **253,** 507–511
6. Guntaka, R. V., Richards, O. C., Shank, P. R., Kung, H.-J., Davidson, N., Fritsch, E., Bishop, J. M., and Varmus, H. E. (1976) *J. Mol. Biol.* **106,** 337–357
7. Shank, P. R., Hughes, S. H., Kung, H.-J., Majors, J. E., Quintrell, N., Guntaka, R. V., Bishop, J. M., and Varmus, H. E. (1978) *Cell* **15,** 1383–1396
8. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V., and Taylor, J. M. (1978) *J. Virol.* **28,** 810–818
9. Shank, P. R., Cohen, J. C., Varmus, H. E., Yamamoto, K. R., and Ringold, G. M. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 2112–2116
10. Yoshimura, F. K., and Weinberg, R. A. (1978) *Cell* **16,** 323–332
11. Watson, J. D. (1972) *Nat. New Biol.* **239,** 197–200
12. Coffin, J. M. (1979) *J. Gen. Virol.* **42,** 1–26
13. Taylor, J. M. (1977) *Biochim. Biophys. Acta* **473,** 57–71
14. Junghans, R. P., Hu, S., Knight, C. A., and Davidson, N. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 477–481
15. Varmus, H. E., Heasley, S., Kung, H.-J., Oppermann, H., Smith, V. C., Bishop, J. M., and Shank, P. R. (1978) *J. Mol. Biol.* **120,** 55–82
16. Taylor, J. M., Illmensee, R., Trusal, L. R., and Summers, J. (1976) in *Animal Virology: ICN-UCLA Symposia on Molecular and Cellular Biology* (Baltimore, D., Huang, A., and Fox, C. F., eds) Vol. 4, pp. 161–174, Academic Press, New York
17. Schwartz, D. E., Zamecnik, P. C., and Weith, H. L. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 994–998
18. Haseltine, W. A., Maxam, A., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 989–993
19. Stoll, E., Billeter, M. A., Palmenberg, A., and Weissman, C. (1977) *Cell* **12,** 57–72
20. Coffin, J. M., and Haseltine, W. A. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 1908–1912
21. Collett, M. S., Dierks, P., Cahill, J. F., Faras, A. J., and Parsons, J. T. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 2389–2393
22. Shine, J., Czernilofsky, A. P., Freidrich, R., Goodman, H. M., and Bishop, J. M. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 1473–1477
23. Haseltine, W. A., Coffin, J. M., and Hageman, T. C. (1979) *J. Virol.* **30,** 375–383
24. Sutcliffe, J. G., Shinnick, T. M., Verma, I. M., and Lerner, R. A. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 3302–3306
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
26. McDonnell, J. P., Garapin, A., Levinson, W. E., Quintrell, N., Fanshier, L., and Bishop, J. M. (1970) *Nature (Lond.)* **228,** 433–435
27. Novak, U., Friedrich, R., and Moelling, K. (1979) *J. Virol.* **30,** 438–452
28. DeLorbe, W. J., Luciw, P. A., Varmus, H. E., Bishop, J. M., and Goodman, H. M. (1980) *J. Virol.* **36,** 50–61
29. Bishop, J. M. (1978) *Annu. Rev. Biochem.* **47,** 35–88
30. Wang, L.-H., Duesberg, P. H., Robins, T., Yokota, H., and Vogt, P. K. (1977) *Virology* **82,** 472–492
31. Joho, R. H., Billeter, M. A., and Weissmann, C. (1978) *Virology* **85,** 364–377
32. Gilboa, E., Mitra, S., Goff, S., and Baltimore, D. (1979) *Cell* **18,** 93–100
33. Weiss, R. A., Mason, W. S., and Vogt, P. K. (1973) *Virology* **52,** 535–552
34. McCarter, J. A. (1977) *J. Virol.* **22,** 9–15
35. Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mitra, S., and Baltimore, D. (1979) *Cell* **16,** 863–874
36. Dina, D., and Benz, E. W. (1980) *J. Virol.* **33,** 377–389
37. Taylor, J. M., Faras, A. J., Varmus, H. E., Levinson, W. E., and Bishop, J. M. (1972) *Biochemistry* **11,** 2343–2351
38. Collett, M. S., and Faras, A. J. (1978) *Virology* **86,** 297–311
39. Hurwitz, J., and Leis, J. P. (1972) *J. Virol.* **9,** 116–129
40. Brewer, L. C., and Wells, R. D. (1974) *J. Virol.* **14,** 1494–1502
41. Efstratiadis, A., Kafatos, F. C., Maxam, A. M., and Maniatis, T. (1976) *Cell* **7,** 279–288
42. Haseltine, W. A., Kleid, D. G., Panet, A., Rothenberg, E., and Baltimore, D. (1976) *J. Mol. Biol.* **106,** 109–131
43. Friedrich, R., Kung, H.-J., Baker, B., Varmus, H. E., Goodman, H. M., and Bishop, J. M. (1977) *Virology* **79,** 198–215
44. Haseltine, W. A., and Kleid, D. G. (1978) *Nature (Lond.)* **273,** 358–364

Additional references are found on p. 1121.

SUPPORTING DATA TO:

## THE TERMINAL REDUNDANCY OF THE RETROVIRUS GENOME FACILITATES CHAIN ELONGATION BY REVERSE TRANSCRIPTASE

Ronald Swanstrom, Harold E. Varmus, and J. Michael Bishop

### EXPERIMENTAL PROCEDURES

Materials. [$^{32}$P]TTP was purchased from New England Nuclear; dideoxynucleoside triphosphates (ddNTP), oligo(dT) cellulose (T3) and actinomycin D were from Collaborative Research; acrylamide and methylene bis-acrylamide were electrophoresis grade from Bio-Rad Laboratories; urea (ultra pure) was from Schwarz-Mann. Prague strain of ASV (subgroup C) was obtained as concentrated suspensions from University Laboratories, Highland Park, New Jersey, through the auspices of the Office of Program Resources and Logistics, National Cancer Institute. AMV reverse transcriptase (deoxynucleoside triphosphate: DNA deoxynucleotidyltransfrase, EC 2.7.7.7) was the kind gift of Dr. J. Beard, Life Sciences, Inc.

Isolation of viral RNA. RNA was extracted from purified virus as described previously and separated into 70S RNA and low molecular weight fractions by rate-zonal centrifugation (1). Viral RNA deficient in sequences from the 3' end of the genome was prepared as follows. The 70S RNA was heated to 60°C for 10 min in a solution containing 0.02M NaCl, 0.001M EDTA, and 0.01M Tris-HCl, pH 7.4. Under these conditions the subunits of the 70S complex are dissociated, but the tRNA primer remains bound to the genome (2,3). Since most of the RNA in these preparations contains covalent chain breaks (2 to 3 per $10^4$ bases), subgenomic fragments of RNA are generated after heating. The heated RNA was then passed over an oligo(dT)-cellulose column in 0.4M NaCl, 0.5% sodium dodecyl sulfate, 0.001M EDTA, 0.01M Tris-HCl, pH 7.4. The RNA that did not bind to the column was passed over another oligo(dT)-cellulose column. The RNA that failed to bind to both columns was precipitated twice with ethanol, resuspended in 0.1 mM EDTA, 0.01M Tris-HCl, pH 7.4, and used as template for DNA synthesis. RNA prepared in this fashion is deficient in sequences from the 3' end of the genome, as shown by its reduced ability to support oligo(dT)-primed DNA synthesis (data not shown).

Sequencing viral DNA transcripts. Viral DNA was synthesized in 20 µl reactions containing 0.06M NaCl, 0.007M $MgCl_2$, 0.005M dithiothreitol, 0.05M Tris-HCl, pH 8.1; each unlabeled dNTP precursor was present at 100 µM, 20 µCi of [α-$^{32}$P]TTP (500 Ci/mmol) was also present. When reactions were carried out in the presence of the ddNTPs, one reaction was run for each ddNTP with the homologous dNTP present at 25 µM (except for TTP, which was never changed). A ddNTP concentration which gave a 70% to 80% inhibition of TMP incorporation was chosen for synthesizing ddNTP terminated transcripts for sequencing. ddATP, ddCTP, and ddGTP were used at between 3 and 5 µM; ddTTP was used at a much lower concentration since the [α-$^{32}$P]TTP concentration was reduced relative to the unlabeled dNTPs. The endogenous reaction contained virus at a protein concentration of about 1 mg/ml and 0.05% NP-40 (Shell). The reconstructed reaction contained 1 µg of viral RNA and about 10 units of AMV DNA polymerase (4). Some reactions contained actinomycin D at a concentration of 100 µg/ml. Reactions were for 30 min at 37°C at which time unlabeled TTP was added to 100 µM in all reactions except the ddTTP reaction. After another 30 min the reactions were terminated either by addition of 10 µl of 1N NaOH (reconstructed reaction) or phenol extraction followed by ethanol precipitation of nucleic acids and treatment of the precipitates with 0.3N NaOH. Alkaline hydrolysis of RNA was done at room temperature for 12 hr. The solution was then neutralized with acetic acid and the DNA precipitated with ethanol. The DNA was finally resuspended in an aqueous solution containing 50% formamide, heated to 100°C for 3 min, then analyzed in an 8% acrylamide gel containing 8 M urea (5). The gel was fixed in 10% acetic acid and the DNA products visualized by autoradiography using Kodak X-OMAT R film.

Molecular cloning and sequencing of ASV DNA. Covalently-closed, circular viral DNA, prepared from cells infected with the Schmidt-Ruppin strain of ASV, has recently been cloned in a prokaryotic vector; details of the cloning and amplification will be presented elsewhere (6). In the present work, we used a 1600 base-pair Pvu II fragment that encompasses the region of circular viral DNA where the redundant termini of linear molecules are fused as a consequence of circularization. A subclone of this fragment (pPvDG) in pBR322 has been used as the source of DNA for sequencing. The amplification of the recombinant plasmid was done in a P2 containment facility in accordance with the National Institute of Health Guidelines for Recombinant DNA Research. The cloned DNA was sequenced by the chemical cleavage method of Maxam and Gilbert (7).

### RESULTS

The use of chain-termination with reverse transcriptase.

Our work exploited the chain-terminator technique developed by Sanger et al., for sequencing DNA (8). In this procedure, DNA synthesis is carried out in the presence of one or another ddNTP, which, when incorporated into DNA, terminates chain propagation. At suitable ratios of ddNTP:dNTP, propagation terminates with roughly equal frequency within the first several hundred nucleotides wherever a nucleotide complementary to the ddNTP occurs in the template for DNA synthesis. We modified the original procedure by using reverse transcriptase, rather than E. coli polymerase I, in order to examine directly the DNA transcribed from viral RNA. Our adaptations of the procedure were similar to those reported recently by other investigators for the purpose of indirect RNA sequencing (9,10).

We found that reverse transcriptase incorporates ddNTPs more efficiently than does E. coli polymerase I. It was therefore necessary to redetermine the ratio of ddNTP:dNTP that would give a satisfactory frequency of chain termination; the effective ratio with reverse transcriptase proved to be ca. 1:5 (ddNTP:dNTP), although the value was slightly different for each of the four nucleotides. In each instance, the optimal ratio inhibited total DNA synthesis by ca. 70%.

Three problems were encountered in interpreting the pattern of ddNTP specific termination products. First, chain propagation by reverse transcriptase terminates at reproducible positions along the template, even in the absence of ddNTP (e.g., see Figure 4, positions 90 and 108); we do not known why these terminations occur, but they pose a problem to sequence analysis because they can obscure the identity of the template nucleotide at the site of termination. This problem can be partly circumvented by using higher concentations of ddNTP so that specific terminations caused by the analogue are even more frequent than the spontaneous terminations. Under these circumstances, however, the inhibition of synthesis of the longer transcripts was in some cases so profound that we were not always able to apply this approach in the present study; we were therefore forced to accept occasional ambiguities in the determination of nucleotide sequence. The second problem encountered was the presence of a series of weak ddNTP specific termination products giving rise to a faint second sequence (see in particular the ddTTP lane of Figure 3A). This faint series of bands was most noticeable below position 100 with the long exposure required to see the sequence longer than 103 nucleotides. The second sequence can be accounted for as a series of transcripts one nucleotide longer than the dominant series of ddNTP termination products. We attributed the second sequence to $tRNA^{Trp}$ molecules lacking the 3' terminal adenosine priming the synthesis of DNA products one nucleotide longer than each pronounced ddNTP termination product. In some instances both light and dark exposures were used in determining the sequence. The third problem was a small amount of sequence heterogeniety in the virus stock used (see the faint extra band in the ddC lane, position 94, Figure 3).

Structure of initial DNA transcripts: endogenous reaction.

Under the proper conditions synthesis of viral DNA by detergent-disrupted virions (the "endogenous reaction") appears to approximate events that occur in the infected cell. We therefore began our analysis of nascent DNA with the endogenous reactions using conditions that were designed to produce an appreciable amount of DNA propagated beyond the 5' end of the viral genome. Four reactions, each with one of the ddNTP's present, were carried out using detergent-disrupted virions as the source of template and polymerase. In each reaction, radioactive TTP was used as a precursor. After the reactions, the tRNA primer was removed from the DNA transcripts by hydrolysis with alkali and the DNA products examined by polyacrylamide gel electrophoresis followed by autoradiography.

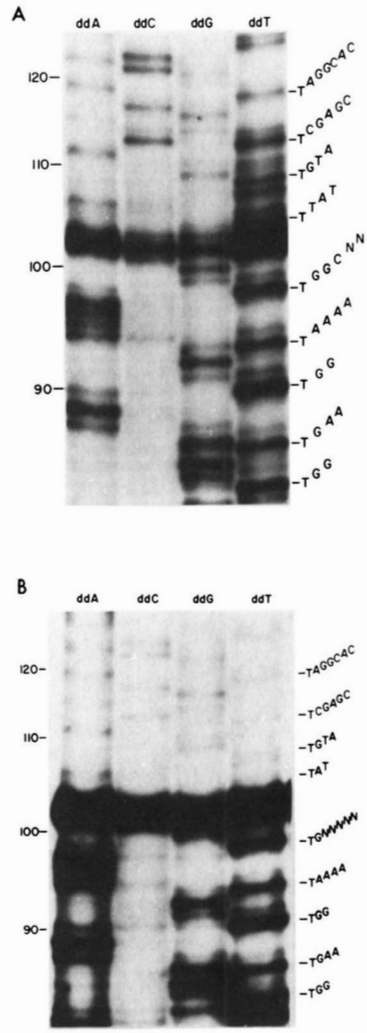


Figure 3. Nucleotide sequence analysis of nascent ASV DNA: endogenous reaction. Purified virions were disrupted with a nonionic detergent and DNA synthesized in the presence of ddNTPs and radioactive TTP as described in Experimental Procedures. After removal of protein and RNA, the DNA products were analyzed by polyacrylamide gel electrophoresis in an 8% gel containing 8M urea. The DNA products were visualized by autoradiography. ddA, ddC, ddG, and ddT refer to the ddNTP present in each reaction. The interpretation of the nucleotide sequence is shown on the right; N refers to a position where specific terminations are obscured by the presence of a non-specific termination product. The length assignments correspond to previous determinations (11,12) and represent the number of nucleotides from the site of initiation. The film exposure time was chosen to allow detection of transcripts longer than 103 nucleotides. a) Endogenous reactions carried out in the absence of actinomycin D. b) Endogenous reactions carried out in the presence of actinomycin D (100 µg/ml).

The autoradiogram and interpretation of the sequence are shown in Figure 3A. The length assignments are based on previous sequence data (11,12) and represent the number of nucleotides from the site of initiation on $tRNA^{Trp}$. The sequence is easily read through position 101, but the identity of the nucleotides at the next two positions is obscured by non-specific terminations. At least one of these terminations must correspond to the previously described "strong stop" that marks the last nucleotide transcribed from the 5' end of the viral RNA (11,12). According to all previous data, this "stop" should occur at position 101 - the alleged distance from the site of initiation to the 5' end of the viral genome (11,12).

We cannot account for either the additional nucleotides in the strong stop DNA from our experiments (positions 102 and 103) or the apparent heterogeneity of this DNA (see Discussion). Starting with position 104, the identity of nucleotides at individual positions is again clear. The DNA sequence shown in Figure 3A is compared in Figure 2 to the known nucleotide sequences at the 5' and 3' ends of the RNA..

Actinomycin D had no qualitative effect on the synthesis of DNA in the endogenous reaction (Figs. 3B and 2). However, chain propagation beyond the "strong stops" (i.e., transcription from the 3' end of the viral RNA) was reduced by actinomycin D, and the relative amount of "strong stop" DNA was increased.

Structure of initial DNA transcripts: reconstructed reaction.

We next examined the structure of nascent DNA transcripts synthesized in the "reconstructed reaction" utilizing purified reverse transcriptase and 70S viral RNA. Four reactions, each with a specific ddNTP, were carried out using the same molar ratio of ddNTP to dNTP as described above. Analysis of the termination products from these reactions is shown in Figure 4. The pattern of stops through position 103 is similar to that shown in Figure 3A. Starting with position 104, there are two bands in each position, indicating that by position 104 there are two different templates being used. The pairs of ddNTP stops are of approximately equal intensity, suggesting that both templates are used equally. There is a ddNTP-independent termination product at position 108 in all four reactions which obscures the specific termination at this sites; otherwise, specific terminations could be read through position 114.

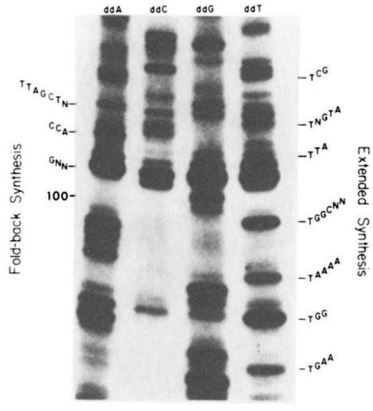


Figure 4. Sequence analysis of nascent ASV DNA transcripts: reconstructed reaction. DNA was synthesized in the presence of ddNTPs using purified reverse transcriptase and ASV 70S RNA as the source of polymerase and template. Reactions were carried out and the DNA isolated and analyzed by electrophoresis in an 8% polyacrylamide gel containing 8M urea as described in Experimental Procedures. ddA, ddC, ddG, and ddT refer to the specific ddNTP present in each reaction. The interpretation of the nucleotide sequence is based on the pattern of ddNTP termination products and is shown on the side. The interpretation of the double sequence beyond position 104 is described in the text. N refers to a position where no ddNTP termination product can be discerned due to the presence of a non-specific termination product. The length assignments are based on earlier sequence analysis (11,12) and represent the number of nucleotides from the primer.

The terminations could be divided into two sets. One set comprised a sequence identical to the sequences obtained using the endogenous reaction (Fig. 4, "Extended Sequence"; Fig. 2). The other set of terminations describe the sequence illustrated in Figure 4 as "Fold-back". We were able to define a template for this sequence when we recognized that nascent DNA 101 nucleotides in length could fold-back to form a hairpin structure by virtue of a five nucleotide long inverted repeat at positions 57 to 61 and 97 to 101. The hairpin has an unpaired loop of 35 nucleotides and aligns the 3' end of the DNA so that chain propagation can continue with the 5' domain of nascent DNA as template. We substantiated this proposal by means of two experimental strategies.

(i) Extension of nascent DNA beyond position 101 by fold-back synthesis is dependent upon a DNA template and should therefore be sensitive to inhibition by actinomycin D. We found that the presence of the antibiotic in reconstructed reactions greatly diminished the yield of DNA longer than 101 nucleotides (Figure 5, lanes 1 and 2); the effect was striking on a DNA of 160 nucleotides, which represented a major product of the unperturbed reconstructed reactions (Figure 5, lane 1) but which was not conspicuous among the products of the endogenous reaction (Figure 5, lane 4). We conclude that fold-back synthesis is directed by a DNA template, and that the 160 nucleotide DNA is probably a major product of that synthesis (see below).

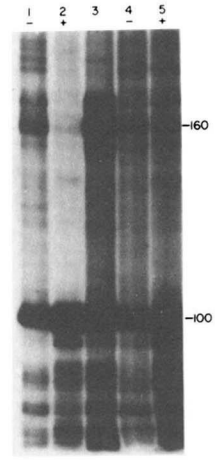


Figure 5. The effect of actinomycin D on ASV DNA synthesis in vitro. Viral DNA was synthesized as described in Experimental Procedures but in the absence of ddNTPs. At the end of the reactions, protein was removed by phenol extraction, RNA removed by hydrolysis with alkali, and the DNA products analyzed by polyacrylamide gel electrophoresis in an 8% acrylamide gel containing 8M urea. Length assignments were made based on comparisons of electrophoretic mobilities with DNAs of known molecular weight (not shown). Reconstructed reaction with 70S viral RNA as template in the absence, Lane 1, and presence, Lane 2, of actinomycin D (100 μg/ml). Lane 3, reconstructed reaction with viral RNA deficient in sequences near the 3' end of the genome as template. Endogenous reaction in the absence, Lane 4 and presence, Lane 5, of actinomycin D (100 μg/ml).

(ii) We prepared an RNA template which should restrict DNA synthesis to the putative fold-back product. The strategy for preparing this template was to remove the 3' end of the genome (linked to the poly A) to prevent the DNA chain elongation mediated by the 3' $DTR_{21}$ sequence. Transcription of this specialized template in a reconstructed reaction gave rise to relatively large amounts of DNA longer than 101 nucleotides (Figure 5, lane 3); in particular, DNA of 160 nucleotides was a prominent product, along with DNA's slightly smaller and slightly larger than 160 nucleotides (Figure 5, lane 3). Thus, the specialized template directs the synthesis of a large amount of a transcript that comigrates in a polyacrylamide gel with the actinomycin D sensitive (and presumed fold-back) transcript synthesized using 70S viral RNA (Figure 5, compare lanes 1 and 3).

Analysis of nucleotide sequence by chain termination confirmed the identity of the 160 nucleotide long DNA species. As seen in Figure 6A, a single sequence is generated from the specialized template, and this sequence is identical to sequences in the 5' domain of viral RNA (Figure 6B). We conclude that when the 3' terminal redundancy of viral RNA is not available to facilitate chain propagation beyond the 5' end of the viral genome, fold-back synthesis is favored by default. Figure 6B summarizes the nucleotide sequences involved in the fold-back and the sequences shown in Figures 4 and 6A; Figure 6C illustrates the events that probably occur during the generation of the fold-back transcript.

The major products of fold-back synthesis described above (Figure 5, lane 3) were heterogeneous. Figure 6A illustrates the heterogeneity in greater detail: three non-specific terminations are apparent, one at position 157 (denoted "a"), one at ca. position 160 (the most prevalent termination; denoted "b") and one well beyond position 160 (denoted "c"). Termination a lies at a point where further transcription would encounter the RNA duplex formed by the binding of $tRNA^{Trp}$ to the viral genome (see Figure 6B). By contrast the locations of terminations b and c remain enigmatic. The DNAs defined by the three terminations (a, b and c) contain completely overlapping sets of pyrimidine tracts (data not illustrated). We therefore assumed that the longer DNAs represented extensions beyond the point of termination a - extensions that would occur by the transcription of $tRNA^{Trp}$ (see Figure 6C). We were unable to sustain this presumption, however: the sequence analysis illustrated by Figure 6A (and others similar to it) inexplicably failed to identify nucleotides at positions beyond termination a; and the DNA defined by termination c did not contain a pyrimidine tract ($C_4$) predicted from the sequence of $tRNA^{Trp}$, the ostensible template for chain extension beyond termination a (data not illustrated).

Comparison of viral DNA synthesized in vivo and in vitro.

Circular ASV DNA synthesized in vivo has recently been cloned in a prokaryotic vector (6). We determined the nucleotide sequence of the region within the cloned DNA that corresponds to the site at which sequences from the 5' and 3' ends of the RNA template are joined during viral DNA synthesis. Mature viral DNA contains two copies of the nucleotide sequences representing the domains of the 5' and 3' ends of the template. By following the replication scheme shown in Figure 1 it can be seen that the initial DNA trancript becomes the right end of the linear DNA and is proximal to the viral src gene sequences. In the circular DNA the right end of the linear DNA is placed at the left side of the circle junction, and it is this region that is represented in Figure 7. The sequence of the cloned DNA is presented in Figure 2 along with the sequence of DNA synthesized in vitro and the sequences at the 5' and 3' ends of the RNA template. The results show that viral DNA synthesized in vivo contains only one copy of the $DTR_{21}$ sequence at the site joining the 3' and 5' domains of the RNA. In addition, it is apparent that transcription at the 3' end of the template starts with the nucleotide immediately next to the 3' $DTR_{21}$ sequence. The nucleotide sequences of DNA synthesized in vitro and in vivo differ at one position; this difference can be attributed to the different strains of virus used in the two experiments.

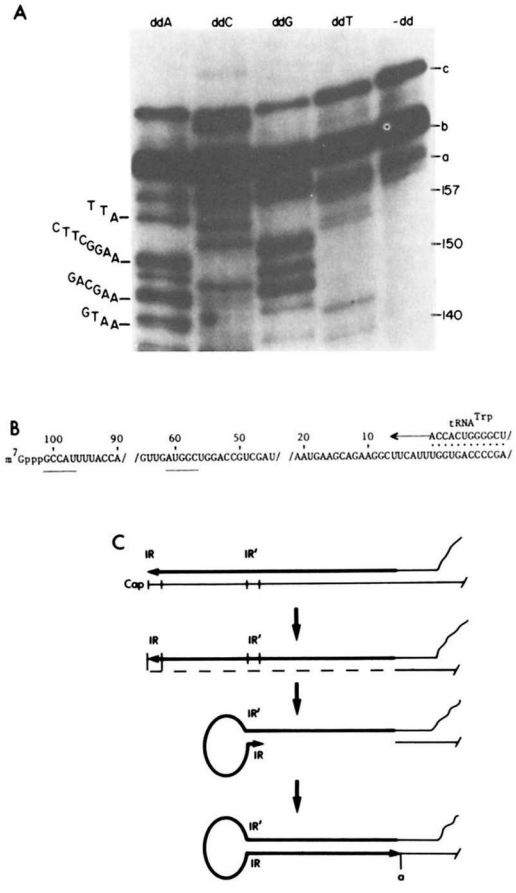


Figure 6. Characterization of DNA synthesized in the fold-back reaction.
a) Nucleotide sequence analysis of DNA synthesized using viral RNA deficient in sequences near the 3' end of the genome as template. DNA was synthesized in reconstructed reactions in the presence of ddNTPs as described in Experimental Procedures. Viral RNA prepared to be deficient in sequences near the 3' end of the genome was used as template. Analysis of DNA products was as described in Figure 3.
b) Portions of the nucleotide sequence at the 5' end of the RNA (11,12) that are relevant to the synthesis of the hairpin DNA. The numbers refer to the number of nucleotides from the tRNA primer. The underlined sequences show the five nucleotide long inverted repeat. The arrow indicates the site of initiation of viral DNA synthesis. The nucleotide sequence involved in base-pairing between the genome RNA and the $tRNA^{Trp}$ primer has been previously described (13,14).
(c) Schematic representation of events leading to the genesis of the fold-back DNA product. The steps are i) synthesis of DNA to the 5' end of the template, ii) removal of the RNA template by RNase H, iii) pairing of the sequences in the inverted repeat, and iv) continued synthesis of DNA using the 5' domain of the initial DNA product as template.

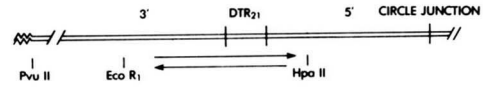


Figure 7. Disposition of the $DTR_{21}$ sequence in viral DNA synthesized in vivo. The figure represents a portion of the physical map of cloned DNA (6). Sequencing studies were performed with a subclone of the DNA, cloned into the Pvu II site of pBR322. The subclone contains both copies of the redundant domains present in viral DNA; the site at which the redundant domains are fused during formation of the circular DNA is referred to as the circle junction. The restriction enzyme sites Eco RI and Hpa II bound the $DTR_{21}$ sequence and, after cleavage with these enzymes, these sites were used for end labeling followed by sequencing using the chemical cleavage method exactly as described by Maxam and Gilbert (7). The sequence of this region of the DNA is presented in Figure 2.

REFERENCES

1. Bishop, H.M., Levinson, W.E., Quintrell, N., Sullivan, D., Fanshier, L., and Jackson, J. (1970) Virology 42, 182-195.
2. Canaani, E., and Duesberg, P. (1972) J. Virol. 10, 23-31.
3. Dahlberg, J.E., Sawyer, R.C., Taylor, J.M., Faras, A.J., Levinson, W.E., Goodman, H.M. and Bishop, J.M. (1974) J. Virol. 13, 1126-1133.
4. Tsiapalis, C.M., Houts, G.E., and Beard, J.W. (1976) Nucleic Acids Res. 3, 2267-2276.
5. Sanger, F., and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
6. DeLorbe, W.J., Luciw, P.A., Varmus, H.E., Bishop, J.M., and Goodman, H.M. (1980) J. Virol. 36, 50-61.
7. Maxam, A.M., and Gilbert, W. (1980) Methods in Enzymol. 65, 499-560.
8. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
9. Zimmern, D., and Kaesberg, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4257-4261.
10. Hamlyn, P.H., Brownlee, G.G., Cheng, C.C., Gait, M.J., and Milstein, C. (1978) Cell 15, 1067-1075.
11. Shine, J., Czernilofsky, A.P., Freidrich, R., Goodman, H.M., and Bishop, J.M. (1977) Proc. Natl. Acad. Sci. USA 74, 1473-1477.
12. Haseltine, W.A., Maxam, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 989-993.
13. Eiden, J.J., Quade, K., and Nichols, J.L. (1976) Nature 259, 245-247.
14. Cordell, B., Stavnezer, E., Friedrich, R., Bishop, J.M., and Goodman, H.M. (1976) J. Virol. 19, 548-558.