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

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+ genome initiates on a (RNA)n 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 (DTRs) (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 (21)), and the DNA base-pairs with the DTRs 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 DTRs, 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 DTRs. 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

1 The abbreviations used are: ASV, avian sarcoma virus; DTRs, 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

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 polymerase 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 vitro, and compared the sequences obtained to the known sequences at the 5' (18, 22) and 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 DTR21 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, 3A). The third nucleotide beyond the 5' end of the template represents transcription from the third position upstream of the DTR21 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'-DTR21 sequence, an assumption confirmed by sequencing DNA synthesized in vitro (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. 3B and 5). This is probably due to the drug binding to the DNA complement of the DTR21 sequence and inhibiting base-pairing with the 3'-DTR21 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).

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.

D. Schwartz and W. Gilbert, personal communication.

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 DTR21 sequence at the 3' end of the template with the complement of the DTR21 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. DTR21, 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).
\[ \text{DNA sequence determined from the endogenous reaction contains...} \]

**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 (17) of the template RNA. DNA: endogenous reaction, DNA sequence determined from the endogenous reaction (Fig. 3A). DNA: endogenous reaction with actinomycin D, DNA sequence determined from the endogenous reaction containing actinomycin D at 100 μg/ml (Fig. 3B). 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 vitro, 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 DTR2 sequence.

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 DTR2 sequence, and the upstream sequence represented the nucleotides immediately upstream from the DTR2 sequence at the 5' 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 DTR2 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 DTR2 sequence, since two copies of the DTR2 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 DTR2 sequence.

Since the genome of retroviruses is diploid (12), two terminal redundancy in the DNA might be the means by which synthesis of the second strand.

\[ \text{DNA: endogenous reaction...} \]

**Retrovirus DNA Synthesis**

plates 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 Vivo and in Vitro**—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 vitro 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. 6C). Chain propagation can then continue, using DNA as template. This scheme is supported by the fact that actinomycin D inhibits the synthesis of hairpin DNA and by the sequence of the DNA itself (38, and Figs. 4, 5, and 6A 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 double-stranded 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...
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. 6C); 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; 42, 43); its length provides a measure of the distance from the point of initiation to the 5’-terminus of the viral genome (12, 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 from the "cap" nucleotide in the RNA template. This explains the "strong stop DNA," both longer than the uncapped template, and, hence, cannot anticipate the composition of their template(s).

Acknowledgments—We thank J. Majors and P. Czerniakofsky 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., Yoshuruma, F. K., and Weinberg, R. A. (1976) J. Virol. 20, 621–626
5. Gunata, R. V., Mahy, B. W. J., Bishop, J. M., and Varums, H. E. (1975) Nature (Lond.) 253, 507–511
6. Gunata, R. V., Richards, O. C., Shank, P. R., Kung, H.-J., Davidson, N., Fritsch, E., Bishop, J. M., and Varums, H. E. (1976) J. Mol. Biol. 106, 327–337
7. Shank, P. R., Hughes, S. H., Kung, H.-J., Majors, J. E., Quintrell, N., Gunata, R. V., Bishop, J. M., and Varums, H. E. (1978) Cell 15, 1383–1396
8. Hsu, T. W., Sabran, J. M., Mark, G. E., Gunata, R. V., and Taylor, J. M. (1978) J. Virol. 28, 810–818
9. Shank, P. R., Cohen, J. C., Varums, 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. Varums, H. E., Heasly, S., Kung, H.-J., Oppermann, H., Smith, V. C., Bishop, J. M., and Shank, P. R. (1978) J. Mol. Biol. 126, 55–82
16. Taylor, J. M., Illmensee, R., Trusal, L. S., and Summers, J. (1976) in Animal Virology: ICM-UCLA Symposium on Molecular and Cellular Biology (Baltimore, D., Huang, A., and Fox, C. F., eds) pp. 41, 61–74
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, 988–993
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, 2380–2383
22. Shine, J., Czerniakofsky, A. P., Freidrich, R., Goodson, 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. 36, 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. (1979) Nature (Lond.) 278, 433–435
27. Novak, U., Friedrich, R., and Moelling, K. (1979) J. Virol. 30, 438–452
28. DeLorbe, W. J., Luciw, P. A., Varums, 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. John, R. H., Billeter, M. A., and Weissmann, C. (1978) Virology 85, 364–377
32. Gubio, E., Mitra, S., Gott, S., and Baltimore, D. (1979) Cell 18, 93–100
33. Weiss, R. A., Mason, W. S., and Vogt, P. K. (1973) Virology 52, 553–552
34. McCarter, J. A. (1977) J. Virol. 22, 9–15
35. Gubio, E., Gott, S., Skelshis, S. Yoshuruma, F., Mitra, S., and Baltimore, D. (1979) Cell 16, 863–874
36. Dina, D., and Benz, D. (1980) J. Virol. 33, 377–389
37. Taylor, J. M., Faras, A. J., Varums, H. E., Levinson, W. K., 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., Ralibergen, E., and Baltimore, D. (1976) J. Mol. Biol. 106, 109–131
43. Friedrich, P., Kung, H.-J., Baker, B., Varums, H. E., Goodman, H. M., and Bishop, J. M. (1979) Virology 79, 198–215
44. Haseltine, W. A., and Kleid, D. G. (1978) Nature (Lond.) 273, 368–369

Additional references are found on p. 1121.
Retrovirus DNA Synthesis

EXPERIMENTAL PROCEDURES

MATERIALS

[List of materials and reagents used in the experiment, not provided in the image]

METHODS

[Detailed methods and procedures used in the experiment, not provided in the image]

RESULTS

[Presentation of results, data, and figures, not provided in the image]

DISCUSSION

[Analysis and interpretation of the results, not provided in the image]

Figure 3. Nucleotide sequence analysis of人在DNA endonuclease reaction. Purified vi-rions were disrupted with a nuclease-deficient and nuclease beta-synthesized in the presence of pancreatic and non-specific nuclease 1 and 2. The method employed a polyethylene glycol electrophoretic and agarose gel electrophoresis to separate bands and determine the positions of nucleotides. The result showed a clear separation between the two bands, indicating the presence of specific and non-specific terminations. The figure is a diagrammatic representation of the experiment and the data obtained.

The autoradiograph 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 of the endonuclease reaction (11,12). The alignment shown in the diagram is a comparison of the autoradiograph and the nucleotide sequence data. The positions of specific and non-specific terminations are marked by arrows and boxes, respectively. The autoradiograph shows a clear separation between the two bands, indicating the presence of specific and non-specific terminations. The figure is a diagrammatic representation of the experiment and the data obtained.

The authors also note that the accuracy of the method used for sequence determination is limited by the resolution of the autoradiograph and the accuracy of the measurement of the nucleotide positions. The method is useful for the study of specific and non-specific terminations, but further refinement may be required to improve the accuracy of the method. The figure is a diagrammatic representation of the experiment and the data obtained.
Retrovirus DNA Synthesis

We cannot account for the additional nucleotides in the strong stop DNA from our experiments (positions 102 and 103) or in the apparent heterogeneity of this DNA (see Discussion). Starting with position 101, 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 DNA.

Anticodon B had no qualitative effect on the synthesis of DNA in the endogenous reaction (Fig. 3B and 3C). However, chain propagation beyond the "strong stop" (i.e., transcription from the 3' end of the viral RNA) was reduced by anticodon B, and the relative amount of "strong stop" DNA was increased. 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 anticodon were carried out using the same molar ratio of DNA to DNA as described above. Analysis of the termination products from these reactions is shown in Figure 4. The pattern of stops through position 101 is similar to that shown in Figure 3A. Starting with position 101, there are now two bands in each position, indicating that by position 101 there are two different templates being used. The pairs of DNA stops are of approximately equal intensity, suggesting that both templates are used equally. There is a 70S RNA-independent termination product at position 103 in all four reactions which obscures the specific termination at this site; otherwise, specific terminations could be read through position 101.

Figure 4. Sequence analysis of nascent DNA template from reconstructed reaction. DNA was synthesized in the presence of 70S RNA using purified reverse transcriptase and 70S viral 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 8% urea as described in Experimental Procedures. A, B, C, D, and E refer to the specific anticodon present in each reaction. The interpretation of the nucleotide sequence is based on the pattern of DNA termination products and is shown on the side. The interpretation of the double sequence beyond position 101 is described in the text. A, B, C, and D refer to a position where no DNA termination products are observed 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 sequence obtained using the endogenous reaction (Fig. 4, "Monoclonal 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 25 nucleotides and an 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 the 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 anticodon B. We found that the presence of the anticodon 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 100 nucleotides, which represented a major product of the unpurified 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 101 nucleotide DNA is probably a major product of that synthesis (see below).

(ii) Figure 5. The effect of anticodon B on DNA synthesis in vitro. Viral DNA was synthesized as described in Experimental Procedures but in the absence of 70S RNA. A, B, C, D, E, F, G, H, and I refer to the specific anticodon present in each reaction. Reactions were analyzed by electrophoresis in an 8% polyacrylamide gel containing 8% urea. Lane 1, no RNA present; Lane 2, of anticodon B (500 μM); Lane 3, reconstructed reaction with viral RNA present; Lane 4, reconstructed reaction in the presence of the 1' and the 3' end of the genome as template. The effect is to diminish the yield of DNA longer than 101 nucleotides (see below). Lane 5, of anticodon C (100 μM/μl).

(iii) 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 DNA chain propagation mediated by the 3' DNA 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 100 nucleotides was a prominent product, along with a DNA slightly shorter and slightly larger than 101 nucleotides (Figure 5, lane 3). Thus, the specialized template directs the synthesis of a large amount of transcript that comigrates with the polyacrylamide gel with the anticodon 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 100 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 sequence, 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.

(iv) The major products of fold-back synthesis described above (Figure 5, lane 3) were heterogenous. Figure 6A illustrates the heterogeneity in greater detail; three non-specific terminations are apparent, one at position 157 (boxed g), one at position 148 (boxed h), and one well beyond position 160 (boxed i). The quaternary structure at a point where further transcription would encounter the RNA duplex formed by the binding of UMAP to the viral genome (see Figure 6B). By contrast the locations of the terminations g, h, and i contain overlapping sets of priming tract sequences (data not illustrated). We therefore reasoned that the longer nascent RNA represented extension beyond the point of termination g and that the DNA formed by termination g did not contain a priming tract. g' predicted from the sequence of UMAP, the canonical template for chain extension beyond termination g (data not illustrated).

Compensation of viral DNA synthesized in vitro and in vivo. Circular AVH DNA synthesized in vitro has recently been cloned in a plasmid vector (18). We determined the nucleotide sequence of the region within the cloned DNA that corresponds to the site at which sequencing from 5' and 3' ends of the DNA template is joined during viral DNA synthesis. Circular viral DNA contains two copies of the nucleotide sequences representing the domains of the 5' and 3' ends of the template. By following the reaction scheme shown in Figure 7 it can be seen that the sequence synthesized in the right end of the linear DNA is in phase with the viral gene sequence. In the circular DNA the right end of the linear DNA is placed at the left side of the circle function, and it is in this region that we are interested. 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 DNA template. The results show that viral DNA synthesized in vitro contains only one copy of the 5' DNA, sequence at the site joining the 5' and 3' domain of the DNA. In addition, it is apparent that transcription at the 3' end of the template starts with the nucleotide immediately next to the 5' DNA, sequence. The nucleotide sequence of DNA synthesized in vitro and in vivo differ at the positions; this difference can be attributed to the different strains of virus used in the two experiments.
Retrovirus DNA Synthesis

Figure 7. Comparison of the 

Figure 6. Characterization of DNA synthesized in the fold-back reaction.

REFERENCES

1. Bishop, R.H., Lewin, R.B., Quinones, R., Sullivan, D., Fambrough, D., and Jackson, J. (1976) Nature 259, 182-190.
2. Cisarri, R., and Duesberg, P. (1972) J. Virol. 32, 13-31.
3. Daeljberg, C., Sawyer, R.C., Taylor, J.M., Pernu, A.J., Levinson, M.E., Goodman, R.M., and Bishop, J.M. (1979) J. Virol. 9, 313-313.
4. Teshima, C.K., Muto, E., and Boyd, J. (1974) J. Virol. 5, 2771-2778.
5. Sanger, F., and Coulam, A. (1978) FEBS Lett. 91, 199-209.
6. Debrice, H., Jolico, P.A., Yara, R., Bishop, J.M., and Goodman, R.M. (1980) J. Virol. 34, 50-61.
7. Mazur, A.M., and Gilbert, W. (1980) Methods in Enzymology 55, 498-508.
8. Sanger, F., Nicklen, S., and Coulam, A.B. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
9. Elorns, S., and Hendrick, F. (1981) Proc. Natl. Acad. Sci. USA 75, 4241-4244.
10. Nishiyama, H., Tsuchiya, C.C., Cheng, C.C., Gait, M.J., and Kellman, C. (1978) Cell 15, 1067-1075.
11. Shih, I., Czajkowsky, A.P., Friedrich, R., Goodman, R.M., and Bishop, J.M. (1977) Proc. Natl. Acad. Sci. USA 74, 1473-1477.
12. Benelkine, M.A., Mazur, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 989-993.
13. Elden, J.D., Oake, R., and Nicola, J.L. (1979) Nature 279, 245-247.
14. Goddell, R., Sternler, R., Fried, R., Bishop, J.M., and Goodman, R.M. (1979) J. Virol. 33, 549-558.