Direct Proof of the 5' to 3' Transcriptional Jump During Reverse Transcription of the Avian Retrovirus Genome by DNA Sequencing

CHARLES A. OMER,¹ J. THOMAS PARSONS,² AND ANTHONY J. FARAS^{1*}

Department of Microbiology, University of Minnesota Medical School, Minneapolis, Minnesota 55455,¹ and University of Virginia Medical School, Charlottesville, Virginia 22901²

Received 14 October 1980/Accepted 8 January 1981

DNA sequence analysis of avian sarcoma virus in vitro-synthesized complementary DNA and in vivo-synthesized, molecularly cloned viral DNA provides unequivocal proof that only one copy of the terminally redundant viral RNA sequence is copied during reverse transcription.

RNA tumor viruses synthesize double-stranded covalently closed circular DNA from their genomic RNA during reverse transcription (1). The precise mechanism by which this occurs has been only partially elucidated (1, 12). To date, information on RNA tumor virus DNA (vDNA) synthesis has come from two predominant sources: (i) a structural analysis of vDNA intermediates found in virus-infected cells; and (ii) an analysis of the DNA products synthesized by enzymatic reactions in vitro by either detergent-disrupted virus preparations or purified viral RNA and reverse transcriptase. The in vitro system has recently been shown to synthesize vDNA structures (2, 3, 11, 12, 18) similar to those previously identified in vivo, thereby validating their usefulness as a means of studying the precise details of the reverse transcription reactions.

In an attempt to elucidate the mechanism of vDNA synthesis, we have employed both DNA sequencing and molecular cloning techniques to detail the specific reactions required to convert the retrovirus single-stranded RNA genome into a covalently closed circular DNA molecule. Since vDNA synthesis initiates near the 5' end of one of the two RNA subunits of 70S and immediately proceeds to the 3' end of the same or a second subunit (6, 7, 14, 22), one of the crucial steps in any model of vDNA synthesis is the transcriptional jump between the 5' and 3' ends of the viral genome. In this report we demonstrate, by DNA sequencing of both in vitro- and in vivo-synthesized DNA, that only one copy of the terminally redundant RNA sequence is copied during reverse transcription. The data indicate involvement of the terminal redundancy by providing a complementary sequence to which the 3' end of tRNA^{sup}-initiated DNA transcripts can bind to continue elongation.

By employing detergent-disrupted preparations of avian sarcoma virus (ASV), conditions of enzymatic DNA synthesis in vitro can be manipulated so that the bulk of the vDNA synthesized contains a 101-nucleotide transcript representing nucleotide sequences from the 3'

FIG. 1. ³²P-5'-end-labeled in vitro-synthesized complementary DNA. Total complementary DNA was synthesized in vitro by using detergent-disrupted preparations of a high-yield variant of ASV B-77. The reaction contained 1 mg of virion protein per ml, 50 mM Tris-hydrochloride (pH 8.1), 10 mM MgCl₂, 0.025% Nonidet P-40, 1.2% β-mercaptoethanol, 100 μg of actinomycin D (a gift from Merck Sharpe & Dohme) per ml, and 100 μM each dATP, dCTP, dGTP, TTP, and [³H]TTP (0.15 Ci/mmol). Actinomycin D was added to minimize double-stranded DNA formation. Total nucleic acid from the reaction was treated with 0.3 N NaOH for 16 h at 37°C to hydrolyze RNA. The DNA product was then passaged over a G-75 fine column and collected. Total DNA was 5'-end-labeled with [γ-³²P]ATP (~6,000 Ci/mmol) and T4 polynucleotide kinase (17, 27). The labeled DNA was electrophoresed on an 8% polyacrylamide gel as shown. (16). Specific species of DNA were then eluted from the gel for DNA sequence analysis (17).

OH of the tRNA^{trp} primer molecule to the 5' terminus of the viral RNA up to, but not including, the m⁷G capping structure (Fig. 1) (4, 7, 14, 22, 24). These DNA transcripts are the result of reverse transcriptase reaching the 5' end of the viral genome before elongation at the 3' end and are commonly referred to as strong-stop DNA (14, 22, 24). DNA transcripts somewhat larger

(e.g., 150 to 500 nucleotides) than strong-stop DNA can also be identified under these reaction conditions and contain nucleotide sequences representing both 5' and 3' genomic RNA sequences (Fig. 1) (7, 13, 19). In an attempt to determine whether one (5') or both (5' and 3') copies of the RNA genomic terminal redundancy are reverse transcribed during elongation from

FIG. 2. DNA sequencing of 5'-3' junction regions of *in vitro* complementary DNA and molecularly cloned *in vivo* vDNA. DNA sequencing was performed by using the technique of Maxam and Gilbert (17). All numbering of the nucleotide sequences is from the initiation point for DNA synthesis from the tRNA^{trp} primer. (A) Strong-stop DNA (right) and band 3 (left) from Fig. 1 were run in parallel and show that only one copy of the terminally redundant viral RNA sequence is copied at the 5'-3' end junction *in vitro*. Bands 2 and 4 have also been sequenced and represent different size classes of tRNA^{trp}-initiated DNA that are elongated at the 3' end of the viral RNA containing only one copy of the terminally redundant viral RNA sequence (data not shown). Band 1 is neither tRNA^{trp}-initiated nor complementary to the viral genome but rather of the same polarity as the viral genome, and thus represents plus-strand DNA. A complete copy of the Prague ASV (subgroup A) genome has been molecularly cloned in the lambda vector λWES-λB (15). Digestion of cloned DNA (λRPA 101) with EcoRI yields a 2.3-kilobase viral DNA fragment containing the 5'-3' junction region of the viral genome. This 2.3-kilobase fragment was subcloned in pBR322, and the junction region was sequenced as shown in Fig. 3. The sequence of the molecularly cloned Prague A DNA (B) shows that only one copy of the terminally redundant viral RNA sequence is copied at the 5'-3' end junction *in vivo*.

the 5' to the 3' end of viral RNA, we have sequenced these transcripts by the method of Maxam and Gilbert (17). From the sequence data presented in Fig. 2A and 4A, it is clear that only one copy of the terminal redundant RNA sequence is reverse transcribed during the 5' to 3' transcriptional jump in vitro. That these DNA transcripts truly represent the product of elongation from the 5' to the 3' end of the viral genome was also indicated by a direct comparison of their nucleotide sequences with those of the 5' (14, 22) and 3' (20) (D. Schwartz, personal communication) ends of avian retrovirus genomic RNA.

To determine whether a similar transcriptional event was occurring during viral DNA

FIG. 3. Schematic diagram of the 5'-3' junction region of the viral genome, as described in the legend for Fig. 2.

synthesis in vivo, the 5'-3' junction region of molecularly cloned restriction endonuclease fragments of covalently closed circular DNA from ASV-infected cells was also sequenced as described in the legend to Fig. 2B. As was the case for the 5' to 3' elongated DNA transcripts synthesized in vitro, ASV vDNA synthesized in vivo also contains only one DNA copy of the viral genomic terminal redundancy. This finding is consistent with those obtained for two retrovirus groups unrelated to ASV, spleen necrosis virus and Moloney murine leukemia-sarcoma virus (10, 21, 23, 25, 26). Cloned DNA was also sequenced through the tRNA^{trp} primer binding site. The DNA sequence which represents the tRNA^{trp} binding site was identical to that of the 3'-terminal 17 bases of the tRNA^{trp} primer (8, 9). These data confirm the nature of the tRNA^{trp} binding site of the ASV RNA and extend these observations to define the nucleotide complementary to the 3'-adenylic acid residue of tRNA^{trp}, as has been shown for the tRNA primer in spleen necrosis virus and murine retroviruses (10, 21). Direct nucleotide sequencing of this region of molecularly cloned ASV DNA suggests that a uridylic acid residue exists on the viral RNA genome in this position, substantiating

FIG. 4. Detailed sequence of the 5' end of genomic RNA and the terminal redundancy as compared to the 5'-3'-end junction DNA in vitro and in vivo. The 5' ends of genomic RNA were determined from strong-stop complementary DNA sequences. For the variant B-77 ASV this sequence was deduced from Fig. 2A, and for Prague ASV subgroup C (Pr C) the sequence was obtained from previously published data (14, 22). The 5'-3'-end junction sequences for the B-77 variant and Prague ASV subgroup A (PrA) are from Fig. 2A and B, respectively. Prague ASV subgroups A and C differ primarily at the envelope portion of the genome by RNA oligonucleotide mapping and by only 1 base in the 5'-terminal 101 nucleotides. As in Fig. 2, all numbering is from the initiation point for DNA synthesis on the tRNA^{trp} primer. (A) The sequence of the B-77 variant 5'-end RNA and 5'-3'-end junction region complementary DNA as described above. (B) The sequence of Prague ASV subgroup C 5'-end RNA and molecularly cloned in vivo Prague ASV subgroup A 5'-3'-end junction DNA (both strands) as described above. The 3'-terminal 17 bases of tRNA^{trp} are from previously published data (9).

conclusions inferred by less direct analysis (5). Thus, although the 3'-adenylic acid residue on the tRNA^{pp} primer is thought not to base pair with the genomic RNA, it nevertheless exhibits the potential to base pair at this position unless the uridylic acid is modified as suggested previously (12).

The presence of only one copy of the terminal redundancy in the 5'-3' junction of vDNA synthesized *in vitro* or *in vivo* provides direct support for the contention that the 5' to 3' elongation reaction occurs by transcription of the 5' genomic terminal redundant sequence followed by attachment of the complement of this sequence at the 3' end of the strong-stop DNA to the redundant sequence at the 3' terminus of the viral RNA genome. The second part of this reaction, removal of the 5' end of the viral RNA genome complementary to the 3' end of strong-stop DNA by the reverse transcriptase-associated RNase H activity, has been documented previously and involves nucleolytic cleavage of at least 12 to 15 nucleotides of the capped 5' terminus (6). Thus, the initial details of the transcriptional jump from the 5' to the 3' end of the viral genomic RNA are now established. The identification of identical nucleotide sequences in the 5'-3' junction regions of vDNA synthesized both *in vitro* and *in vivo* substantiate the validity of the *in vitro* reverse transcription reactions in delineating the events that occur during vDNA synthesis in virus-infected cells. This is an important consideration in light of the difficulties of studying the precise details of vDNA synthesis in retrovirus-infected cell systems (1, 12).

We thank B. Bulbulian and M. Niels for excellent technical assistance and C. Kosman for typing this manuscript.

This work was supported by Public Health Service grants from the National Institutes of Health (CA-18303 and CA-20011, to A.J.F.) and by a grant from the American Cancer Society (MV29B to J.T.P.). C.A.O. was supported by a fellowship from the Bush Foundation.

LITERATURE CITED

1. Bishop, J. M. 1978. Retroviruses. *Annu. Rev. Biochem.* 47:35-88.
2. Clayman, C. H., E. Mosharrafa, D. Anderson, and A. J. Faras. 1979. Circular forms of DNA synthesized by Rous sarcoma virus *in vitro*. *Science* 206:582-584.
3. Clayman, C. H., E. Mosharrafa, and A. J. Faras. 1979. *In vitro* synthesis of infectious transforming DNA by the avian sarcoma virus reverse transcriptase. *J. Virol.* 29:242-249.
4. Coffin, J., and W. Haseltine. 1977. Terminal redundancy and the origin of replication of Rous sarcoma virus RNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:1908-1912.
5. Coffin, J., and W. Haseltine. 1977. Nucleotide sequence of Rous sarcoma virus RNA at the initiation site of DNA synthesis. The 102nd nucleotide is U. *J. Mol. Biol.* 117:805-814.
6. Collett, M., P. Dierks, J. Parsons, and A. J. Faras. 1978. RNase H hydrolysis of 5'-terminus of the avian sarcoma virus genome during reverse transcription. *Nature (London)* 272:181-184.
7. Collett, M., and A. J. Faras. 1978. Avian retrovirus RNA-directed DNA synthesis: transcription at the 5'-terminus of the viral genome and the functional role for the viral terminal redundancy. *Virology* 86:297-311.
8. Cordell, B., E. Stavnezer, R. Friedrich, J. Bishop, and H. Goodman. 1976. Nucleotide sequence that binds primer for DNA synthesis to the avian sarcoma virus genome. *J. Virol.* 19:548-558.
9. Dahlberg, J., F. Harada, and R. Sawyer. 1974. Structure and properties of an RNA primer for initiation of Rous sarcoma virus DNA synthesis *in vitro*. *Cold Spring Harbor Symp. Quant. Biol.* 39:925-932.
10. Dahr, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude. 1980. Nucleotide sequence of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proc. Natl. Acad. Sci. U.S.A.* 77:3937-3941.
11. Gilboa, E., S. Goff, A. Shields, F. Yoshimura, S. Mitra, and D. Baltimore. 1979. *In vitro* synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus. *Cell* 16:863-874.
12. Gilboa, E., S. Mitra, S. Goff, and D. Baltimore. 1979. A detailed model of reverse transcription and tests of crucial aspects. *Cell* 18:93-100.
13. Haseltine, W., J. Coffin, and J. Hageman. 1979. Products of the Moloney murine leukemia virus endogenous DNA polymerase reaction. *J. Virol.* 30:375-383.
14. Haseltine, W., A. Maxam, and W. Gilbert. 1977. Rous sarcoma virus is terminally redundant: the 5'-sequence. *Proc. Natl. Acad. Sci. U.S.A.* 74:989-993.
15. Highfield, P. E., L. F. Raffeld, T. M. Gilmer, and J. T. Parsons. 1980. Molecular cloning of avian sarcoma virus closed circular DNA: structure and biological characterization of three recombinant clones. *J. Virol.* 36:271-279.
16. Maniatis, T., A. Jeffrey, and H. van deSande. 1975. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry* 14:3787-3794.
17. Maxam, A., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:560-564.
18. Mitra, S., S. Goff, E. Gilboa, and D. Baltimore. 1979. Synthesis of a 600-nucleotide-long plus-strand DNA by virions of Moloney murine leukemia virus. *Proc. Natl. Acad. Sci. U.S.A.* 76:4355-4359.
19. Novak, U., R. Friedrich, and K. Moelling. 1979. Elongation of DNA complementary to the 5'-end of the avian sarcoma virus genome by the virion-associated RNA-dependent DNA polymerase. *J. Virol.* 30:438-452.
20. Schwartz, D., P. Zamecnik, and H. Weith. 1977. Rous sarcoma virus genome is terminally redundant: the 3'-sequence. *Proc. Natl. Acad. Sci. U.S.A.* 74:994-998.
21. Shimotohno, K., S. Mizutani, and H. M. Temin. 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. *Nature (London)* 285:550-554.
22. Shine, J., A. Czernilofsky, R. Friedrich, J. Bishop, and H. Goodman. 1977. Nucleotide sequence at the 5'-terminus of the avian sarcoma virus. *Proc. Natl. Acad. Sci. U.S.A.* 74:1473-1477.
23. Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore. 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc. Natl. Acad. Sci. U.S.A.* 77:3932-3936.
24. Stoll, E., M. Billeter, A. Palmenberg, and C. Weissman. 1977. Avian myeloblastosis virus RNA is termi-

- nally redundant: implications for the mechanism of oncornavirus replication. *Cell* **12**:57-72.
25. **Sutcliffe, J. G., T. M. Shinnick, I. M. Verma, and R. A. Lerner.** 1980. Nucleotide sequence of Moloney leukemia virus: 3'-end reveals details of replication, analogy to bacterial transposons, and an unexpected gene. *Proc. Natl. Acad. Sci. U.S.A.* **77**:3302-3306.
26. **van Beveren, C., J. G. Goddard, A. Berns, and I. M. Verma.** 1980. Structure of Moloney murine leukemia viral DNA: nucleotide sequence of the 5' long terminal repeat and adjacent cellular sequences. *Proc. Natl. Acad. Sci. U.S.A.* **77**:3307-3311.
27. **Walseth, T., and R. Johnson.** 1979. The enzymatic preparation of [α - 32 P]nucleotide triphosphates, cyclic [32 P]AMP, and cyclic [32 P]GMP. *Biochim. Biophys. Acta* **562**:11-31.