JOURNAL OF VIROLOGY, Feb. 1986, p. 674–677
0022-538X/86/020674-04$02.00/0
Copyright © 1986, American Society for Microbiology
Vol. 57, No. 2

NOTES

Structure of a Cloned Circular Retroviral DNA Containing a tRNA Sequence between the Terminal Repeats

JOHN COLICELLI AND STEPHEN P. GOFF*

Department of Biochemistry and Molecular Biophysics, Columbia University, College of Physicians and Surgeons, New York, New York 10032

Received 15 July 1985/Accepted 11 October 1985

In the course of analyzing a series of cloned circular retroviral DNAs, we recovered an unusual clone. The molecule consisted of a complete viral genome containing two copies of the long terminal repeat with extra sequences between the repeats. These extra bases proved to be a nearly complete DNA copy of a glycine tRNA, including bases that corresponded to modified and nonpairing bases of the mature tRNA. A model is proposed to account for the formation of the aberrant clone.

In the early phase of retrovirus infection, the single-stranded genomic RNA is reverse transcribed in a complex reaction to form a full-length double-stranded DNA flanked by long terminal repeat (LTR) sequences (4). Synthesis of the first strand of this DNA is initiated by a tRNA primer bound near the 5′ end of the template RNA and results in the formation of a short DNA molecule termed minus-strand strong-stop DNA (7). This DNA is translocated to the 3′ end of the template and elongated; later in the reaction, probably after the second DNA strand has been initiated, the tRNA is removed from the DNA (15). The completed duplex linear DNA is cyclized to generate two circular forms in the nucleus of the infected cell (20, 25). One of these forms contains one LTR sequence, and the other contains two tandem LTR sequences. The latter molecule is apparently formed by the blunt-end ligation of the duplex linear DNA (21) and probably serves as the immediate precursor for the formation of the integrated proviral DNA (16). Many of the details of these steps are poorly understood; for example, it is not known precisely how strong-stop DNA transfers occur nor how the linear duplex DNA is cyclized to form the two circular species.

We are studying the replication of Moloney murine leukemia virus and have generated a series of mutants with alterations at the edge of the LTR near the tRNA primer-binding site (2; Colicelli and Goff, manuscript in preparation). Many of these alterations do not dramatically affect any stage of the life cycle, including reverse transcription, cyclization of the linear DNA, or formation of the integrated provirus. One such silent mutation, designated *in*594-2, consists of a substitution of the two 3′-terminal bases (TT) of the viral LTR by a four-base sequence (TATA), yielding a net insertion of two bases. To characterize the effects of the mutation, we introduced the altered DNA into cells by cotransformation with selectable marker DNA (24), recovered infectious virus, and applied the virus to NIH 3T3 cells. The virus was replication competent and was able to direct the formation of all three preintegrative DNA forms at normal levels. To determine the structure of the circular forms in greater detail, we extracted low-molecular-weight DNA (8), cleaved the DNAs with *Hin*dIII, and prepared clones in the lambda phage vector Charon 30A (18). Clones containing viral DNA were identified by plaque hybridization (1), and 75 of them were examined by gel electrophoresis and blot hybridization (22) to determine the size of the retroviral insert. The majority of the inserts were derived from viral DNAs that had suffered large deletions or from full-length DNAs containing one LTR. Five clones contained inserts approximately 8.8 kilobases in length, the size of the larger viral circles, and were studied further. These inserts were excised from the phage vector and transferred to plasmid pBR322. Restriction enzyme mapping showed that the DNA inserts were all full-length copies of the viral genome containing two copies of the LTR.

A more detailed analysis of the full-length clones revealed that one of the clones contained an insertion of extra sequences at the junction between the two LTRs. The DNA sequence in this region was determined by the procedure of Maxam and Gilbert (12) (Fig. 1). Both LTRs were intact and the silent substitution mutation was still present, but between the two LTRs was a new sequence of 61 base pairs (Fig. 1). Comparison with known mammalian tRNAs (3) revealed that the sequence was a nearly complete copy of mammalian glycine tRNA (5). The 3′-proximal 52 base pairs of the tRNA matched perfectly, the next 9 bases of the tRNA sequence were missing, 7 of the next 9 bases matched, and the last 4 bases were missing. No other tRNA was a possible candidate, because the anticodon was present in the sequence. Analysis of the sequence at the normal primer-binding site of the clone revealed that the sequence complementary to the wild-type proline tRNA primer was still present (data not shown).

How did this new sequence appear in the clone? The fact that the sequence was complementary to a tRNA and appeared next to the edge of the LTR where DNA synthesis is normally initiated by proline tRNA suggests that it arose by reverse transcription of a misutilized glycine tRNA. We propose a mechanism by which the observed structure could have been generated (Fig. 2). First, we propose that a glycine tRNA (rather than a proline tRNA [6, 17]) paired

* Corresponding author.

```
in594-2(8·8) : ··· G G G G T C T T T C A T A T A | T G G T G C A T
tRNAGly :      5'                                  3'OH A C C A C G U A

T G G C C G G G A A A C G A A C C C G G G C C
A C C G G C C C U U m1A G C Ψ T G G G m5C m5C m5C G G

T C C C G C G T G G C A G G C G A G A A T G G
A G G G C G m5C A C C G U C C G C U C U U A C U
                                         A G A D G G U G A

A C C T C C A | A A T G A A A G A C C C C ··· 3'
U G G U G G U Um A C G OP 5'
```

FIG. 1. DNA sequence of the aberrant clone in the region between the two LTRs and comparison with the sequence of glycine tRNA. The top bases show the sequence of the aberrant 8.8-kilobase clone of mutant *in*594-2, reading in the 5'→3' direction. The boxed sequences are within the LTRs. The top LTR contains the original mutation, terminating with the sequence TATA rather than the usual TT; the bottom LTR is the wild-type sequence. The LTRs of the wild-type virus are normally joined to each other with no inserted bases (2, 21). The second row of bases is the sequence of mammalian glycine tRNA (5) aligned for maximal pairing with the inserted sequence. The unusual bases are: $^{1}_{m}$A, 1-methyladenine; Ψ, pseudouridine; $^{5}_{m}$C, 5-methylcytidine; D, dihydrouridine; U_m, 2'-O-methyluridine.

with the genomic RNA at the normal primer-binding site, making only 12 of the usual 18 base pairs. This step would have required the assistance of the reverse transcriptase to make the reaction thermodynamically possible; perhaps sequences flanking the primer-binding site are recognized by the enzyme. According to our model, this tRNA was then extended to form minus-strand strong-stop DNA, and the DNA was translocated to the 3' end of the viral RNA and elongated as usual.

Next, plus-strand strong-stop DNA was made. The 3' site of termination of this molecule is unclear but might be altered by the presence of the aberrant tRNA; normally, a part of the tRNA is copied during the synthesis of this molecule (4, 13, 23). The next step is usually the pairing of the 3' end of the plus-strand strong-stop DNA with the 3' end of the minus strand to allow elongation. Normally, the removal of the primer tRNA exposes a sequence complementary to the 3' end of the minus strand (15). As a key difference from the normal situation, we propose that in this aberrant event, the pairing and transfer occurred without removal of the glycine tRNA. The result is that at the end of synthesis of the full-length plus strand, the retained tRNA was copied into DNA.

The final step to complete the molecule is the cyclization of the linear DNA; normally, this is achieved by the blunt-end ligation of the two termini. In the aberrant reaction, the left end of the linear DNA would be completed as usual by the extension of the minus strand to the end of its template, forming a normal duplex blunt end. The right end of the linear DNA, however, would contain an RNA-DNA duplex; if the tRNA were removed at this time, it would contain a long single-stranded tail. We propose that the DNA strand of the RNA-DNA duplex or the single-stranded DNA tail was ligated to the other, blunt, end. It may be significant that bases of the tRNA sequence were missing at this terminus, and it may be that these four bases were lost in the course of this ligation reaction. After this ligation, completion of the structure was simple; the 3' end of the minus strand was extended and joined to the 5' end to form a covalently closed circle. If the tRNA were not removed earlier, it could be readily removed by displacement or RNase H action at this time.

This aberrant clone has several implications for reverse transcription of the retroviral genome. First, the structure of the clone suggests that unusual tRNAs can serve at some frequency to prime DNA synthesis, even though these tRNAs are only poorly homologous to the tRNA primer-binding site. It is likely that reverse transcriptase is responsible and that the enzyme can bind and misutilize a variety of tRNAs for priming. Our previous results showed that the Moloney murine leukemia virus enzyme can utilize tRNAs other than the wild-type proline tRNA, at least when the primer-binding site is altered (2a); thus, it is not unreasonable that the enzyme might occasionally misutilize a tRNA other than proline tRNA, and it is possible that the mutation could promote such errors. Glycine tRNA, in particular, is used by several viruslike 30S (VL30) elements (9, 10, 14) and so must be compatible with the proteins of at least some murine viral genomes. Second, the retention of the tRNA sequence in the clone suggests that the aberrant tRNA may have been poorly removed from the minus strand. Preliminary screening of other clones of this mutant suggested that retention of tRNA sequences is not infrequent (data not shown). RNase H is normally responsible for the removal of the tRNA while it is paired with the 3' end of plus-strand strong-stop DNA (15). The poor removal could be accounted for by either of two possibilities. If the minus strand did serve as template for the formation of a plus-strand strong stop, the structure near the 3' end may have been a poor substrate; the mutant sequence of the DNA joined to the tRNA or the altered tRNA itself may be poorly recognized by the enzyme. A second alternative is that the aberrant minus strand may not have served as the template for plus-strand strong-stop DNA synthesis, but that tranfer in *trans* from another minus-strand template was involved in the formation of the plus strand. In this case, the aberrant

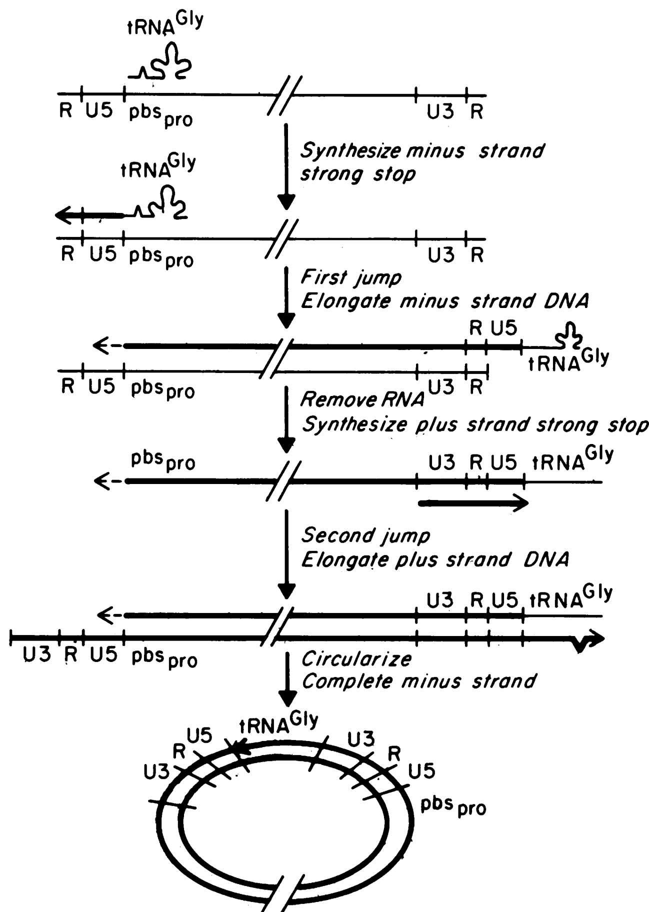

FIG. 2. Proposed model for the generation of the aberrant clone bearing the glycine tRNA sequence. Thin lines, RNA; thick lines, DNA. The U3, R, and U5 regions of the LTR are indicated; the central portion of the genome has been shortened at the position of the slashed lines. For a detailed explanation of the model, see the text.

tRNA would not have been in RNA-DNA duplex form and for this reason could not be removed by RNase action.

Finally, it is interesting that reverse transcriptase was able to copy nearly all of a tRNA, including a number of modified bases found in mature glycine tRNA. If the fully modified tRNA was utilized, four 5-methylcytidine bases were copied and paired with guanine; surprisingly, a 1-methyladenine and a pseudouridine base in the TΨC arm of the tRNA were copied and in each case apparently paired with adenine bases. These last two modified bases cannot form normal Watson-Crick base pairs and should pose problems for reverse transcriptase. It should be noted that the enzyme has been shown to exhibit a preference for adenine incorporation opposite apurinic and apyrimidinic sites (11, 19); thus, the enzyme may insert adenines at any position which cannot form a good base pair. Two other modified bases, 2′-O-methyluridine and dihydrouridine, were located in the regions of the tRNA that were not retained in the extra sequences (Fig. 1). The fact that these bases were not retained may reflect difficulties in copying them. Overall, the results suggest that reverse transcriptase is surprisingly versatile at copying aberrant modified bases in RNA. The only alternative, which we consider unlikely, is that an unmodified precursor form of the tRNA acted as the primer.

This work was supported by Public Health Service grant CA30488 from the National Cancer Institute and by an award from the Chicago Community Trust/Searle Scholars Program. Salary support for S.P.G. was provided by the Irma T. Hirschl Trust.

LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening lambda gt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
2. **Colicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell. **42**:573–580.
2a. **Colicelli, J., and S. P. Goff.** 1986. Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. J. Virol. **57**:37–45.
3. **Gauss, D. H., and M. Sprinzl.** 1983. Compilation of tRNA sequences. Nucleic Acids Res. **11**:1–53.
4. **Gilboa, E., S. W. Mitra, S. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and tests of crucial aspects. Cell **18**:93–100.
5. **Gupta, R. C., B. A. Roe, and K. Randerath.** 1979. The nucleotide sequence of human $tRNA^{Gly}$ (anticodon GCC). Nucleic Acids Res. **7**:959–970.
6. **Harada, F., G. G. Peters, and J. E. Dahlberg.** 1979. The primer tRNA for Moloney murine leukemia virus DNA synthesis. J. Biol. Chem. **254**:10979–10985.
7. **Haseltine, W. A., D. G. Klein, A. Panet, E. Rothenberg, and D. Baltimore.** 1976. Ordered transcription of RNA tumor virus genomes. J. Mol. Biol. **106**:109–131.
8. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–371.
9. **Hodgson, C. P., P. K. Elder, T. Ono, D. N. Foster, and M. J. Getz.** 1983. Structure and expression of mouse VL30 genes. Mol. Cell. Biol. **3**:2221–2231.
10. **Itin, A., and E. Keshet.** 1985. Primer binding sites corresponding to several tRNA species are present in DNAs of different members of the same retrovirus-like gene family (VL30). J. Virol. **54**:236–239.
11. **Kunkel, T. A., R. M. Schaaper, and L. A. Loeb.** 1983. Depurination-induced infidelity of deoxyribonucleic acid synthesis with purified deoxyribonucleic acid replication proteins in vitro. Biochemistry **22**:2378–2384.
12. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–599.
13. **Mitra, S. W., M. Chow, J. Champoux, and D. Baltimore.** 1982. Synthesis of murine leukemia virus plus strand strong stop DNA initiates at a unique site. J. Biol. Chem. **257**:5983–5986.
14. **Norton, J. D., J. Conner, and R. J. Avery.** 1984. Unusual long terminal repeat sequence of a retrovirus transmissible mouse (VL30) genetic element: identification of functional domains. Nucleic Acids Res. **12**:3445–3459.
15. **Omer, C. A., and A. J. Faras.** 1982. Mechanism of release of the avian retrovirus $tRNA^{trp}$ primer molecule from viral DNA by ribonuclease H during reverse transcription. Cell **30**:797–805.
16. **Panganiban, A. T., and H. M. Temin.** 1984. Circles with two LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.
17. **Peters, G., F. Harada, J. E. Dahlberg, A. Panet, W. A. Haseltine, and D. Baltimore.** 1977. Low-molecular-weight RNAs of Moloney murine leukemia virus: identification of the primer for RNA-directed DNA synthesis. J. Virol. **21**:1031–1041.
18. **Rimm, D. L., D. Hornes, J. Kucera, and F. R. Blattner.** 1980. Construction of coliphage lambda Charon vectors with Bam HI cloning sites. Gene **12**:301–309.
19. **Sagher, D., and B. Strauss.** 1983. Insertion of nucleotides opposite apurinic/apyrimidinic sites in deoxyribonucleic acid during in vitro synthesis: uniqueness of adenine nucleotides. Biochemistry **22**:4518–4526.
20. **Shank, P. R., and H. E. Varmus.** 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. J.

Virol. **25**:104–114.

21. **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. USA **77**:3932–3936.
22. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
23. **Taylor, J. M., and T. W. Hsu.** 1980. Reverse transcription of avian sarcoma virus RNA into DNA might involve copying of the tRNA primer. J. Virol. **33**:531–534.
24. **Wigler, M., R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell **16**:777–785.
25. **Yoshimura, F. K., and R. A. Weinberg.** 1979. Restriction endonuclease cleavage of linear and closed murine leukemia viral DNAs: discovery of a smaller circular form. Cell **16**:323–332.