Rous sarcoma virus genome is terminally redundant: The 5′ sequence
(reverse transcription/DNA sequence/tumor virus replication/ribosome binding sites/RNA dimers)

WILLIAM A. HASELTINE*, ALLAN M. MAXAM†, and WALTER GILBERT‡

*Sidney Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115; and †Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contribution by Walter Gilbert, January 6, 1977

ABSTRACT When Rous sarcoma virus RNA is transcribed into DNA by the reverse transcriptase, a tRNA primer is elongated into DNA. The primer is near the 5′ end of the virus genome; the first major DNA made is a “run-off” product extending 101 bases from the primer to the 5′ end of the template. We have studied this DNA molecule to determine the sequence of the first 101 bases at the 5′ end of the Rous sarcoma virus genome (virus strain, subgroup C). Twenty-one bases at the extreme 5′ end are also at the 3′ end of the virus genome (see D. E. Schwartz, P. C. Zamecnik, and H. L. Weith, this issue, pp. 994–998), and thus this virus is terminally redundant. The existence of this sequence repetition immediately suggests mechanisms by which the growing DNA copy can jump from the 5′ end to a 3′ end of the template and become circular. The sequence also displays a possible ribosome binding site and enough secondary structure to permit a possible 5′-3′ linkage of viral RNA molecules.

During infection, RNA tumors make a DNA copy of their genome. This DNA molecule becomes a covalently closed double-stranded circle and then integrates into the host DNA before the life cycle of the virus can continue. The virion carries into the cell two copies of a linear viral RNA, several copies of an RNA-dependent DNA polymerase (the reverse transcriptase), and, annealed to the RNA, a primer for transcription, a host tRNA molecule [tRNA<sup>32</sup>P for Rous sarcoma virus (RSV)](1). The reverse transcriptase can elongate the primer tRNA to make a single-stranded DNA copy of the viral RNA, as an RNA-DNA hybrid. An RNase H activity of the reverse transcriptase can digest the RNA template in such an RNA-DNA hybrid to reveal a template for the synthesis of a second strand. (For reviews see refs. 2–5.) But the original template viral RNA is linear. How then is a circular DNA molecule formed?

In vitro, when reverse transcriptase copies the RNA of RSV, DNA products of various discrete lengths are synthesized. At sufficient deoxyribonucleoside triphosphate concentrations, the longest of these chains covers a major portion of the genome [1000–4000 (6–9) up to full length, 1000 nucleotides (E. Rothenberg and D. Baltimore, personal communication)] A tRNA primer initiates these chains (1, 10); the first eight deoxynucleotides added to its 3′ end are AATGAAGC (9, 11, 12); and the longer fragments are extensions of this sequence (9, 12). Thus, in vitro, the reverse transcriptase initiates DNA synthesis at a unique site and terminates at a series of defined sites along the genome.

This would be a natural, expected behavior, except for the curious finding by Taylor and Illmensee (13) and Staskus et al. (14) that the tRNA primer is bound very near the 5′ end of the template RNA. Thus, as the reverse transcriptase copies (in the 5′ to 3′ direction), it soon will come to the 5′ end of the template and fall off. How is this to be reconciled with the existence of ordered long DNA transcripts?

With RSV, under most in vitro conditions, about 20% of all initiation events terminate at a specific site about 100 bases from the primer (9, 10, 12, 15). This product, strong-stop DNA, obviously is a candidate for the run-off product: a DNA fragment running from the primer to the 5′ end of the template. We have sequenced this strong-stop DNA and argue that it is, in fact, a transcript of the 101 nucleotides at the 5′ end of the RSV genome.

MATERIALS AND METHODS

Rous Sarcoma Virus. RSV (Prague strain, subgroup C) was obtained from University Laboratories via the Office of Programs and Logistics of the National Cancer Institute.

DNA. This was synthesized in a 10-ml reaction containing RSV virion protein at 2 mg/ml, 0.05 M Tris-HCl (pH 8.3), 0.06 M KCl, 0.02% (wt/vol) Nonidet P-40, 0.006 M MgCl<sub>2</sub>, 0.005 M dithiothreitol, 1 mM each dATP, dCTP, and dGTP, 0.2 mM [3H]dTP (specific activity, 2 Ci/mmol), and actinomycin D at 100 µg/ml. After a 90-min incubation at 43°, we added 10 ml of 0.2 M NaOAc/0.01 M Tris-HCl, pH 7.5/1 M EDTA (buffer A) and 20 ml of a 1:1 (vol/vol) mixture of phenol/chloroform. After vigorous mixing for 1 min, centrifugation for 5 min at 25° and 20,000 × g separated the phases. The phenol/chloroform was reextracted with buffer A; the aqueous phases were combined and reextracted with an equal volume of chloroform. Then, 2 volumes of ethanol was added, and the nucleic acids were collected by centrifugation at 20° for 5 hr at 30,000 × g. The pellet was resuspended in water, 2 volumes of ethanol was added, and the suspension was centrifuged before.

The final pellet was dissolved in 50 µl of 0.3 M NaOH and held at 37° for 15 hr to hydrolyze primer and template RNA; then the DNA was precipitated with 100 µl of 1.0 M NaOAc and 500 µl of 95% ethanol by chilling in a Dry ice-ethanol bath and collected by centrifuging at 12,000 × g for 10 min. The pellet was rinsed with 95% ethanol, dried, and resuspended in 25 µl of distilled H<sub>2</sub>O.

DNA Phosphorylation. In this procedure (16) we used [γ-<sup>32</sup>P]ATP (1200 Ci/mmol). After the reaction with polynucleotide kinase, 2 M NH<sub>4</sub>Ac and tRNA carrier were added, and the DNA and RNA were precipitated with 3 volumes of ethanol at −70° and sedimented. The ethanol-rinsed and dried pellet was dissolved in 25 µl of 5 mM Tris-borate, pH 8.3/0.1 mM EDTA/10% (vol/vol) glycerol/0.02% xylene cyanol/0.02% bromophenol blue, layered on a slab gel polymerized from 8% acrylamide/0.27% bisacrylamide/50 mM Tris-borate, pH 8.3/1 mM EDTA, and subjected to electrophoresis at 300 V (regulated) and 20 mA (average). After autoradiography, labeled DNA was located, extracted, and ethanol-precipitated (16).

Nucleotide Sequence Analysis. For analysis, 5′-[<sup>32</sup>P]-labeled strong-stop DNA was partially cleaved by the four base-specific reactions described by Maxam and Gilbert (16). For cleavage

Abbreviation: RSV, Rous sarcoma virus.

† To whom reprint requests should be addressed.
at adenine (and to some extent, guanine) the DNA was treated with 50 mM dimethyl sulfate at 20° for 20 min, the methylated adenines were preferentially released at pH 1, and the strands were broken at depurination sites by treatment with 0.1 M NaOH at 90° for 30 min. Two cleavages at guanine were used after the same reaction with dimethyl sulfate: either heat depurination for a strong C/weak A pattern after NaOH cleavage, or 1.0 M piperidine at 90° for 30 min for a G-specific cleavage. Reaction with 18 M hydrazine in 2 M NaCl at 20° for 30 min, followed by displacement and strand scission with 0.5 M piperidine at 90° for 30 min, cleaved at cytosine. The same reaction with hydrazine alone for only 20 min, followed by piperidine, cleaved at both cytosine and thymine. The end-labeled partial cleavage products from the four reactions were fractionated on a 0.15 × 35 × 40 cm slab gel polymerized from 20% acrylamide/0.67% bisacrylamide/7 M urea/50 mM Tris-borate, pH 8.3/1 mM EDTA as described (16). The wrapped gel on one glass plate was then exposed to Kodak XR-5 x-ray film at −20°.

RESULTS

We made strong-stop DNA [a DNA product about 100 bases long covalently linked to a 75-base tRNA primer (1, 9, 10)] in a reaction mixture containing detergent-disrupted virions, actinomycin D [to suppress the formation of double-stranded DNA (17)], and high concentrations of deoxynucleoside triphosphates. At the end of the reaction, after a phenol/chloroform extraction, alkaline hydrolysis of the primer released a free DNA 5'-hydroxyl. We then labeled the 5' ends of the DNA products with polynucleotide kinase and [γ-32P]ATP and resolved the labeled molecules on a preparative polyacrylamide gel. Fig. 1 shows that about 80% of the synthesis product was unique, strong-stop DNA.

We excised strong-stop DNA from the gel and sequenced it by partial, base-specific, chemical cleavages and separation of the end-labeled products by size on polyacrylamide gels (16). Figs. 2 and 3 show autoradiographs of the sequencing gels. Fig. 3 displays the first 10 partial cleavage products beginning from the labeled end, and Fig. 2, displays the next 90. Taken together, the four patterns in the two figures determine, in order, 101 deoxynucleotides added to the primer in strong-stop DNA, with the exception of the very first one. That one is known to be A from analysis by others (9, 11, 12, 18) who found the first eight deoxynucleotides to be AATGAAGC, overlapping with our first seven, NATGAAGC . . . . The sequence of all 101 deoxynucleotides in strong-stop DNA, joined to the last 17 ribonucleo-

FIG. 1. Isolation of 5'-32P-labeled RSV endogenous reverse transcripts. One-fifth of the DNA synthesized in endogenous reactions containing 20 μg of virion protein was phosphorylated with 32P (1200 Ci/mmol) as described in Materials and Methods, layered on an 8% polyacrylamide slab gel, and subjected to electrophoresis until the bromphenol blue marker dye had moved about 30 cm. The gel was then exposed to Kodak XR-5 x-ray film for 30 min, and the labeled strong-stop (SS) DNA was extracted for sequence analysis.

FIG. 2. Base-specific partial cleavage products 10–101 from end-labeled strong-stop DNA separated by size on a polyacrylamide gel. Equal portions of [5'-32P]DNA from gel region SS, see Fig. 1, were subjected to four limited chemical degradations (specificities: A>G, G, C, and C+T) and electrophoresed in parallel on a denaturing polyacrylamide slab gel. Degradation products from the four reactions were repeatedly layered on the gel during electrophoresis to expand different regions of the sequence: at 0 hr in the center, at 13 hr on the left, and at 24 hr on the right, after which electrophoresis was continued for another 12 hr. The sequence is derived by beginning with the smallest partial product at the lower right, a T cleavage product, reading upward, and then switching to the left and finally to the middle patterns whenever bands are not well resolved.
FIG. 3. Base-specific cleavage products 2-14 from end-labeled strong-stop DNA separated by size on a polyacrylamide gel. Equal portions of [5'-32P]DNA were subjected to chemical degradations of specificity A> G, G< A, C, and C+ T and electrophoresed in parallel for 10 hr on a denaturing polyacrylamide slab gel. The band at the bottom is from cleavage at the second nucleotide in from the labeled end of the strong-stop fragment, an A, and bands above it are from nucleotides increasingly distant from the end up to the 14th, a G.

FIG. 4. Nucleotide sequence of the strong-stop reverse transcript and the 5' end of the RSV genome (Prague C). Top. That portion (underlined) of the host tRNA primer that specifically binds to the viral template (19) and the first 101 deoxynucleotides added to its 3' end by the reverse transcriptase. The sequence of this strong-stop DNA was derived from the band patterns of Fig. 2 and 3. Middle. The complementary viral RNA sequence inferred from the DNA transcripts, the primer binding sequence (19), and the cap structure (21, 22). Bottom. The putative proviral DNA sequence encompassing this region, with potential cleavage sites for restriction endonucleases (36, 37).

DISCUSSION

Strong-Stop DNA Is a Run-Off Product from the 5' End of the Template. The 5' end of RSV RNA is capped by the structure

\[ G^m7pppG^2OmC \ldots \]

which can be released by RNase A. Although the sequence at this end is not known, RNase T1 releases a 25-base-long fragment carrying the cap, whose partial sequence is

\[ G^m7pppG^2OmCCACUUUUA, U(C, U), 2CCA, CA]UUG \]

We deduce that the strong-stop DNA is complementary to the

FIG. 5. Terminal redundancy and possible replication of the RSV genome. Top. The first 21 nucleotides in RSV viral RNA are identical in sequence and polarity with the last 21 (see Fig. 4, and Schwartz, Zamecnik, and Weith, this issue, pp. 994-998). Middle. Strong-stop DNA complementary to the 5' end of viral RNA has been synthesized. Bottom. The redundant portion of this template has been displaced, allowing the same sequence from the 3' end to hydrogen bond with the DNA. This allows the strong-stop reverse transcript, synthesized from the extreme 5' end of the viral template, to serve as a primer for more extensive DNA synthesis from the extreme 5' end. The figure arbitrarily shows all of this occurring on one template; synthesis could just as well continue on the 3' end of another molecule—for instance, the second viral RNA contained in each virion.
et al. (19) have argued that the last A in the tRNA is not base-paired to the template. Thus, we do not know base 102.

Terminal Redundancy of the RSV Genome Can Lead to Circularization. Schwartz, Zamecnik, and Weith (this issue, pp. 994-998) have determined the sequence at the 3' end of the RSV genome by priming DNA synthesis off the region near the poly(A) tail using specially constructed primers containing oligo-dT. They concluded that the RNA sequence at the 3' end is

\[
\ldots \ldots \text{GCCAUUUUACCAAUCACCA}\ [\text{poly(A)}].
\]

This sequence is identical, for 21 bases, to the one we find just after the cap. Thus, the first 21 nucleotides of RSV are identical to the last 21: the RSV genome (excluding the poly(A)) is terminally redundant.

How is the strong-stop DNA to be elongated into an extensive copy of the genome? Because the initiation occurs at the 5' end of the RNA, the growing DNA chain must jump to the 3' end of the RNA if the entire genome is to be transcribed. Clearly, the terminally redundant sequence could be used to elongate the DNA chain past the 5' end of the genome. Fig. 5 shows how this redundant information might be used. The nascent DNA chain could float away from its template at the 3' end and pair with a new template at the 3' end for further elongation, or the RNase H activity of the reverse transcriptase may digest some of the RNA from the RNA-DNA hybrid containing the strong-stop DNA, revealing that RNA so that it can pair with the repeated sequence at the 3' end.

The terminal redundancy, used in this jumping mechanism, will also serve to create DNA circles. If the nascent chain jumps to the other RNA molecule in the virion, the ultimate single-stranded DNA product could be a greater-than-genome-length strand; if it becomes double-stranded, there is enough information for circularization by recombination. Alternatively, the nascent chain might pick up at the 3' end of the same RNA. Then, as the DNA elongates, the growing point may finally come to the tRNA primer and, on displacing it, arrive at the 5' end of itself, to make a DNA circle. This displacement and sealing problem, to make a single-stranded DNA circle, is analogous to the displacement of RNA primers and sealing of DNA chains that occurs elsewhere in DNA synthesis and could be done by host or virion-encoded enzymes. Since these DNA circles have only one copy of the originally repeated sequence, the RNA redundancy must be recreated during the viral life cycle. One way this could occur would be through transcription of an integrated proviral DNA from an external promoter, if the integration is at the repeated sequence.

There Is a Possible Translation-Initiation Site Near the 5' End of the Genome. The virion RNA is a messenger, at least for those viral proteins that it expresses in an in vitro translation (24, 25). There are six translation-initiation codons in the 118-nucleotide sequence at the 5' end of the genome: AUG codons centered on positions 42 and 83 and GUG codons at positions 26, 28, 105, and 116 (Fig. 4). Only the AUG at 83 and the GUG at 116 are not in phase with termination sequences; all the others are in phase with UGA codons. Fig. 6 shows that there are sequences partially complementary to the 3' end of chicken fibroblast 18S ribosomal RNA (26) near AUG 83, supporting the suggestion that this may be a functional initiation codon in a ribosome binding site (27). This AUG is immediately preceded by the termination codon UGA.

If AUG 83 is an initiation codon, then that protein should begin Met-Lys-Gln-Lys-Ala-...; we suggest that this may be the NH2-terminal sequence of the precursor to the group-specific antigen (gag) proteins, the gene for which is near the 5' end of the genome (24, 25, 28, 29).

The ribosome binding site is close to the tRNA primer binding site. This proximity may allow bound tRNA to interfere with ribosome attachment and thus to control translation.

5'-to-3' Subunit Linkage? Each virion contains two viral genomes and two tRNA primers tightly bound together in a
complex that sediments at 70 S (30–32). Electron microscopy has often shown, especially for primate tumor viruses, the two RNA subunits held together at their 5' ends (33, 34). The sequence of RSV permits such an interaction. Fig. 7 shows two such hypothetical dimer linkages. A palindromic sequence from bases 64 to 116 would allow 36 to 40 base pairs to connect the subunits. If the primer tRNAs also base-pair to the subunits of the 70S structure, then they may help link the 35S RNAs together, because sequences complementary to the tRNA occur in apposed regions of the two viral RNAs. Curiously, the RNase T1 and T2 cutting sites in the primer-viral RNA hybrid (19, 35) fall in convenient places in this speculative model. The strong-stop DNA fragment is part of a tumor virus replication intermediate. It is a short reverse transcript, polymerized onto an RNA primer bound at a unique place on the viral RNA, which ultimately primes the synthesis of a complete transcript of the viral genome. Sequence analysis of replication intermediates is worthwhile in its own right, but fortuitously, the RSV strong-stop intermediate is also a DNA copy of the 5' end of the RNA genome, so sequencing this DNA yields information about an essential region of viral RNA, suggesting speculations on how the viral genome is translated, linked to itself, and replicated.

W.G. is an American Cancer Society Professor of Molecular Biology. This work was supported by National Institutes of Health Grants CA 19341-01 to W.A.H. and GM 09541-15 to W.G.