

RNA Template Requirements for Target DNA-Primed Reverse Transcription by the R2 Retrotransposable Element

DONGMEI D. LUAN AND THOMAS H. EICKBUSH*

Department of Biology, University of Rochester, Rochester, New York

Received 21 February 1995/Returned for modification 6 April 1995/Accepted 28 April 1995

R2 is a non-long terminal repeat-retrotransposable element that inserts specifically in the 28S rRNA gene of most insects. The single protein encoded by R2 has been shown to contain both site-specific endonuclease and reverse transcriptase activities. Integration of the element involves cleavage of one strand of the 28S target DNA and the utilization of the exposed 3' hydroxyl group to prime the reverse transcription of the R2 RNA transcript. We have characterized the RNA requirement of this target DNA-primed reverse transcription reaction and found that the 250 nucleotides corresponding to the 3' untranslated region of the R2 transcript were necessary and sufficient for the reaction. To investigate the sequence requirements at the site of reverse transcription initiation, a series of RNA templates that contained substitutions and deletions at the extreme 3' end of the RNA were tested. The R2 templates used most efficiently had 3' ends which corresponded to the precise boundary of the R2 element with the 28S gene found in vivo. Transcripts containing short polyadenylated tails (8 nucleotides) were not utilized efficiently. R2 RNAs that were truncated at their 3' ends by 3 to 6 nucleotides were used less efficiently as templates and then only after the R2 reverse transcriptase had added extra, apparently nontemplated, nucleotides to the target DNA. The ability of the reverse transcriptase to add additional nucleotides to the target DNA before engaging the RNA template might be a mechanism for the generation of poly(A) or simple repeat sequences found at the 3' end of most non-long terminal repeat-retrotransposable elements.

Retrotransposable elements can be divided into two distinct groups on the basis of the presence or absence of terminal repeats and the nature of their encoded proteins (4, 12, 36, 39, 40). The first group to be identified consists of those elements that are flanked by long terminal repeats (LTR) and encode retrovirus-like *gag* and *pol* open reading frames. In agreement with their close structural similarity to retroviruses, all data on how this group of retrotransposable elements insert into eukaryotic genomes are consistent with a retrovirus-like mechanism (3, 30, 37). Indeed some LTR retrotransposable elements encode *env*-like genes and share certain transmission properties with retroviruses (23, 32).

The second group of retrotransposable elements to be identified are not flanked by terminal repeats. A reverse transcriptase domain is the only homology that can be consistently identified for these elements (15). We will use the term non-LTR retrotransposable elements to identify this second group of elements (39). It has been demonstrated that these elements undergo retrotransposition by the removal of intron sequences during the generation of new copies (16, 22, 28). However, the absence both of LTRs and of an encoded integrase domain strongly suggests that these elements use a fundamentally different mechanism for their retrotransposition than that of the LTR-containing retrotransposable elements.

Direct biochemical studies of the retrotransposition mechanism used by the non-LTR elements have been possible only with the R2 elements of insects (26, 38). These biochemical studies have been aided by the ability of the R2 element to insert into a unique sequence of the 28S rRNA gene (7, 8, 20, 21). By expressing the single open reading frame of the R2 element from *Bombyx mori* in *Escherichia coli*, the site speci-

ficity of R2 was found to be at least partly the result of an encoded endonuclease which generates a double-stranded cut at its target site in the 28S rRNA gene (38). The R2 protein was also found to have a reverse transcriptase activity (26). The mechanism used by R2 for the initial steps of its integration was revealed by a combined in vitro DNA cleavage and reverse transcription assay. The R2 protein was shown to first nick the 28S gene insertion site, use the 3' hydroxyl group exposed by this nick to prime reverse transcription of the R2 transcript, and then cleave the second strand of the target DNA (26). The mechanisms used for the attachment of the cDNA strand to the upstream target sequences, removal of the RNA template, and synthesis of the second DNA strand remain unknown. Similar models for the utilization of a 3' hydroxyl group on chromosomal DNA to prime reverse transcription have been suggested to explain features of the integrated copies of the *Cin4* element found in *Zea mays* (31), I and TART elements in *Drosophila melanogaster* (5, 24), and L1 elements in mammals (19).

In this study we have further characterized the target DNA-primed reverse transcription reaction of the R2 element. We find that the RNA requirement for this reaction is highly specific and involves the 250-nucleotide 3' untranslated region of the R2 element. Nucleotide substitutions or deletions at the extreme 3' end of the R2 RNA were used to identify those sequence requirements needed for initiation of reverse transcription at the 3' end of the RNA. These studies have revealed that the most efficiently utilized RNA templates were those that end at the precise R2-28S gene junction. For those templates shorter than this length, the R2 reverse transcriptase added additional nucleotides to the target DNA prior to reverse transcription of the R2 template.

MATERIALS AND METHODS

Protein purification. R2 protein was purified from *E. coli* JM109/pR260 with the following modifications of the previously described method (26). The starting

* Corresponding author. Mailing address: Department of Biology, Hutchison Hall, University of Rochester, Rochester, NY 14627. Phone: (716) 275-7247. Fax: (716) 275-2070. Electronic mail address: teickbush@the.biology.rochester.edu.

volume of the JM109/pR260 culture was increased from 0.8 to 1.2 liters. The volume of the Q-Sepharose column was correspondingly increased from 5 to 7.5 ml, but the volume of the DNA-cellulose column was retained at 1.0 ml. Because of a low level of RNase contamination, washing buffers were increased to 7.5 column volumes for the Q-Sepharose column and 12 column volumes for the DNA-cellulose column. Because the R2 protein is unstable after purification on the DNA-cellulose column, attempts were made to stabilize the purified R2 protein by adding RNA and bovine serum albumin and concentrating the protein on a Centrocon-10 column (Ameco). After an initial decrease of nearly 70% of the enzymatic activity, the remaining activity was stable for several days at 0°C. However, this initial loss may have changed the enzymatic properties of the protein; thus, all experiments reported in this article were conducted with R2 protein on the same day it was purified on the DNA-cellulose column.

DNA target-primed reverse transcription assays. All assays were performed as described by Luan et al. (26). Briefly, the assays were conducted in 20- μ l volumes containing 0.4 μ g of *Hind*III-*Eco*RI-predigested pB109 plasmid DNA, 0.2 μ g of RNA template with 15 U of RNasin (Pharmacia), and 5 μ l of the peak R2 fraction from the DNA-cellulose column (~100 ng of protein). Reaction conditions were 200 mM NaCl, 10 mM MgCl₂, 50 mM Tris-HCl (pH 8.0), 5 mM dithiothreitol, and 25 μ M deoxynucleoside triphosphates (dNTPs) including 125 nM ³⁵S-dATP. The amount of RNA in each assay (0.2 μ g) represented a three- to fivefold excess of that needed for maximum rates of integration (data not shown). The concentration of dNTP in the reactions could be varied from 5 to 25 μ M with no effect on the yield of reaction products. Unless otherwise indicated, the incubations lasted 2 h at 37°C, after which time 1 μ g of RNase A was added for an additional 25-min incubation. The final products were mixed with 5 \times loading buffer (0.02% bromophenol blue, 5% Sarkosyl, 0.1 M EDTA, 50% glycerol) and analyzed by electrophoresis in a 1% agarose gel. For relative quantitation of the 1.8- to 1.9-kb product band, autoradiographs of the dried agarose gel were scanned on a Hewlett Packard Scan Jet II scanner using Adobe Photoshop and NIH image programs. Determination of R2-28S junction sequences by PCR amplification, cloning into m13 vectors, and sequencing were carried out as described by Luan et al. (26).

Constructs for in vitro transcription of RNA. Constructs used to generate RNA containing various deletions of the R2 RNA were derived from the original pBluescript SK(-) clone HR4 (26). pBmR2-417 was generated by subcloning the *Eco*RV-*Eco*RI fragment of HR4 into pBluescript SK(-). pBmR2-204 was generated by subcloning the two *Hae*III-*Eco*RI fragments of HR4 into pBluescript SK(-). pBmR2-249 was constructed by PCR amplification of the HR4 plasmid using the universal primer and a primer (5'-TCAGTCGACGTTGGTTGAGC CTTGC-3') complementary to the beginning of the 3' untranslated region of R2 and ligating the *Hinc*II-*Eco*RI double-digested PCR products into pBluescript SK(-). pBmR2-417(D104-204) was generated by ligating the *Eco*RV-*Hae*III and *Hae*III-*Eco*RI fragments of HR4 into the *Eco*RV-*Eco*RI sites of pBluescript SK(-). pBmR2-417(R204-417) was generated as a rearrangement of the *Eco*RV-*Hae*III fragments from HR4. All of the above pBluescript constructs were predigested with *Xmn*I before being used as templates for in vitro transcription with T7 RNA polymerase. To generate the pBmR2-417(D0-113) RNA, clone pBmR2-417 was predigested with *Nae*I before in vitro transcription.

DNA templates used for the synthesis of HR4 RNAs with different 3' ends were generated by PCR amplification of HR4 using the universal -40 primer complementary to the pBluescript vector upstream of the T7 promoter and a second primer corresponding to the individual 3' ends of the RNA to be generated. The specific PCR primers used to generate RNAs were as follows: 5'-TTTTTTTTCATCGCCGGATCATCA-3' for HR4/8A, 5'-TCATCGCCG GATCATCATC-3' for HR4/1A, 5'-TCGCCGGATCATCATG-3' for HR4/D1, 5'-AAATCATCGCCGGATC-3' for HR4/V1, 5'-ATATCATCGCCGGATCA TC-3' for HR4/V2, 5'-AGAGCATCGCCGGATCATCAT-3' for HR4/V3, 5'-A TATCGCCGGATCATCATG-3' for HR4/V4, 5'-TATCGCCGGATCATCA TG-3' for HR4/V5, and 5'-TTATTATTAGATCATCATGCCATCG-3' for HR4/V6. All PCR products are treated with proteinase K (50 μ g/ml; 10 mM EDTA, pH 8), extracted with phenol-chloroform, precipitated by ethanol, and resuspended in 5 mM Tris-HCl (pH 8)-1 mM EDTA before in vitro transcription.

Identification of nucleotides added to the 28S rDNA target site. A 10 \times -volume DNA-primed reverse transcription assay was conducted with V4 RNA and pB109 target DNA cut with *Eco*RI. After the incubation, the DNA was treated at 0.1 μ g of RNase A per μ l for 20 min at 37°C and 50 μ g of proteinase K per ml at 55°C for 30 min, extracted with phenol-chloroform, and ethanol precipitated. The 3.7-kb DNA fragment which represented target DNA that had not undergone DNA-primed reverse transcription or second-strand cleavage was eluted from a 1% agarose gel, mixed with equal volume of denaturing-gel loading buffer (95% [wt/vol] formamide, 1 mM EDTA, 0.1% [wt/vol] xylene cyanol FF, 0.1% [wt/vol] bromophenol blue), and subjected to electrophoresis on a 5% polyacrylamide-7 M urea gel. The ~1,000-nucleotide single-stranded DNA generated by first-strand cleavage at the target site was eluted from the gel (29) and incubated with 17 U of terminal transferase (U.S. Biochemical) in the presence of 1 mM dCTP. The poly(C)-tailed single-stranded DNA was diluted 50 fold and used as template in "hot-start" PCR (34) using the anchor primer 5'-GGAATTC CGGATCCGGGGGGGGGGGGGGGG-3' and a second primer, 5'-AAAGAGCC GACATCGAAGGATC-3' complementary to the 28S rDNA sequence 631 bp downstream of the R2 insertion site. The PCR products were purified by agarose gel electrophoresis, eluted, digested with *Bam*HI, and cloned into the *Bam*HI-

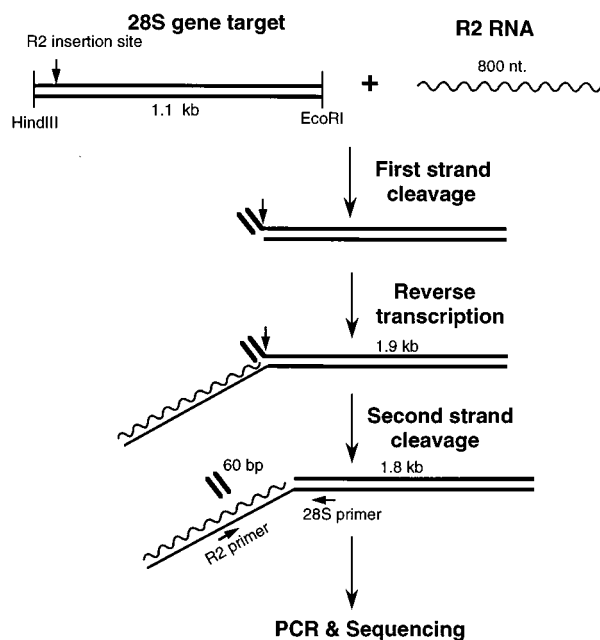


FIG. 1. Diagram of the in vitro assay for target DNA-primed reverse transcription. A 1.1-kb fragment containing the 28S gene target was incubated with an 800-nucleotide (nt.) R2 RNA transcript and the R2 protein. A nick at the target site creates a 3' OH group which is used to prime the reverse transcription of the R2 RNA. Second-strand cleavage of the target site is the slowest step in the reaction and occurs after reverse transcription. Cleaved and uncleaved products of the DNA-primed reverse transcription migrate at about 1.8 to 1.9 kb. Because of variation at the junction of the 28S gene and the cDNA sequence, determination of the nucleotide sequence of this junction requires PCR amplification of the products using one primer complementary to the 28S gene and the second primer complementary to the cDNA strand derived from the R2 RNA. The PCR products are then cloned into M13, and individual clones are sequenced.

*Hinc*II site of the mp19 vector for single-stranded sequencing of individual products.

RESULTS

The target DNA-primed reverse transcription assay. We have previously described purification of the 120-kDa R2 protein expressed in *E. coli* and the in vitro assay for the combined DNA cleavage and reverse transcription reactions (26). Figure 1 summarizes this target DNA-primed reverse transcription assay. A plasmid DNA substrate (pB109) containing a 1.1-kb segment of the 28S gene with the R2 insertion site 60 bp from one end is predigested with *Eco*RI and *Hind*III to release the 28S gene fragment from the 2.7-kb pUC vector. The DNA fragments are incubated at 37°C for up to 2 h with the purified R2 protein in the presence of 25 μ M each dNTP, 0.2 M NaCl, and 10 mM MgCl₂ (pH 8). An 800-nucleotide R2 RNA transcript, HR4, generated in vitro with T7 RNA polymerase serves as the template for reverse transcription. The first two steps of the reaction, cleavage of the first strand at the 28S target site and reverse transcription of the RNA template primed by the 3' OH generated by this cleavage, convert the 1.1-kb substrate DNA to a branched molecule with a molecular size of 1.9 kb. The third step, cleavage of the second DNA strand to generate a linear 1.8-kb fragment, is the slowest step in the reaction. Because the R2 insertion site is only 60 bp from the end of the cloned 28S gene fragment, second-strand cleavage has a minimal effect on the migration of the product molecules. After incubation the products of the DNA-primed

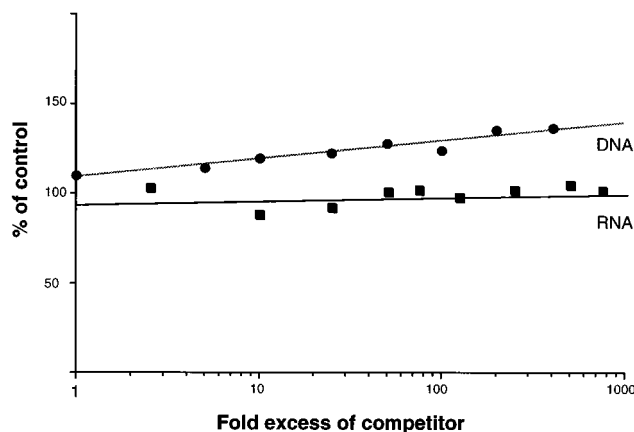


FIG. 2. Competition reactions demonstrating the RNA and DNA specificity of the DNA-primed reverse transcription reaction. For the RNA competition (squares), increasing amounts of total *E. coli* RNA (mean length, 500 nucleotides) were incubated with the R2 protein and target DNA for 10 min at 37°C prior to the addition of 0.2 μg of HR4 RNA. For the DNA competition (circles), increasing amounts of *Eco*RI-digested pUC19 were preincubated with the R2 protein and HR4 RNA for 10 min at 37°C prior to the addition of 0.01 μg of target DNA. This target DNA was a 150-bp fragment generated by PCR of the 28S target site using primers 85 bp upstream and 65 bp downstream of the R2 insertion site. Assays for both experiments were conducted with 200 mM NaCl–50 mM Tris-HCl (pH 8.0)–5 mM dithiothreitol–25 μM each dNTP (including 0.1 μM ³⁵S-dATP) at 37°C for 2 h. The yield of the 1.8- to 1.9-kb product band in the RNA competitions and the yield of the 1.0- to 0.9-kb product band in the DNA competitions were quantitated by scanning autoradiographs of dried agarose gels of the reaction products. For each competition, the protein concentration was adjusted so that in the absence of competitor ~50% of the target DNA was converted to product.

reverse transcription reaction are treated with Sarkosyl to remove protein from the DNA and separated on a 1% agarose gel. Quantitation of the reaction is obtained by direct visualization of the product bands with ethidium bromide staining or by autoradiography of the dried gel if labelled nucleotides have been added to the reaction mixture (26). Direct DNA sequencing of the cDNA-28S gene junction generated by this reaction revealed considerable sequence heterogeneity at the R2-28S gene junction (26). Therefore, to obtain individual examples of this junction sequence, the products of the reaction are PCR amplified with one primer complementary to the cDNA strand generated by reverse transcription of the R2 transcript and a second primer complementary to the 28S gene downstream of the insertion site. The PCR products are ligated into mp19 phage, and multiple clones are sequenced.

DNA and RNA specificity of the DNA-primed reverse transcription reaction. The endonuclease activity of the R2 protein has previously been shown to be specific for the 28S gene integration site (38), and DNA-primed reverse transcriptase activity has been shown to be specific for R2 RNA (26). However, the R2 protein has a high affinity for a variety of RNA and DNA sequences. Indeed, the ability of the R2 protein to tightly bind to total *E. coli* RNA in 0.4 M NaCl is used as the first step in the purification of the R2 protein from crude *E. coli* extracts, and the ability of the R2 protein to bind to DNA-cellulose columns in 0.6 M NaCl is used as the second step in the purification (26).

To test the ability of the R2 protein to locate and use as template a small amount of the R2 RNA in a large pool of RNA sequences, increasing amounts of total *E. coli* RNA were incubated with the R2 protein and target DNA for 10 min prior to the addition of 0.2 μg of HR4 RNA. The result of one such experiment is shown in Fig. 2. Even when the *E. coli* RNA in

the preincubation reaction was increased to 7.5 mg/ml (a 750-fold excess over the R2 RNA), no decrease occurred in the intensity of the 1.8- to 1.9-kb integration products nor were any other products of reverse transcription detected. The mean length of the *E. coli* RNA used in the competition reaction in Fig. 2 is 500 nucleotides. Similar results have been obtained with preparations of *E. coli* RNA degraded to greater or lesser extents, as well as specific T7 transcripts of pBluescript vector sequences (results not shown).

To test the ability of the R2 protein to find and cleave the 28S gene target in an excess of competing DNA, pUC19 DNA at concentrations up to 0.5 mg/ml was preincubated with the R2 protein and HR4 RNA for 10 min prior to the addition of 1.25 μg of a 150-bp DNA fragment containing the 28S target sequence per ml. No decrease in the final yield of integrated products was obtained even with a 400-fold excess of competitor DNA (Fig. 2). The addition of competitor DNA actually increased the yield of products by 25%. This increased level of product in the presence of competitor DNA was presumably a result of the greater stability of the R2 protein in the reverse transcription reaction (see Materials and Methods for further discussion of R2 stability). All subsequent reactions described in this report have been conducted in 20-μl mixtures with 0.2 μg of template RNA and 0.4 μg of pB109 DNA.

The 3' untranslated region of R2 RNA is required for DNA-primed reverse transcription. To identify which segments of the HR4 transcript are required for DNA-primed reverse transcription, various RNA templates that contain deletions of portions of this 800-nucleotide sequence have been tested. To generate each RNA, restriction fragments of the 800-nucleotide sequence were cloned into the pBluescript vector by using the sites shown in Fig. 3B and the RNA synthesized in vitro with T7 RNA polymerase. The abilities of the various RNAs to serve as templates in the reverse transcription assay are shown in Fig. 3A. Note that the sizes of the reverse-transcribed products are directly proportional to the size of the RNA template used in the reaction. Deletions of those RNA sequences encoding the carboxyl-terminal end of the open reading frame had no effect on the ability of the RNA to serve as a template for reverse transcription (constructs R2-417 and R2-249). Further deletion of the R2 sequences into the 3' untranslated region of the R2 RNA (R2-204) eliminated the ability of the RNA to support DNA-primed reverse transcription. Deletions of a large central region of the 3' untranslated region (nucleotides 104 to 204) or of the 3' end of the RNA (nucleotides 1 to 113) also eliminated the ability of the R2 RNA to serve as a template.

These results suggest that several segments within the 250-bp 3' untranslated region of R2 RNA are necessary and sufficient to serve as a template for the DNA-primed reverse transcription reaction. While only this 250-nucleotide 3' untranslated region is necessary for the reaction, we have continued to use the longer HR4 RNA because this template results in a larger separation of the reverse-transcribed products and the target DNA substrate on agarose gels.

Polyadenylated transcripts are not preferred in the reverse transcription reaction. Most integrated non-LTR retrotransposons contain short poly(A) or A-rich repeats at the 3' end of their mRNA synonymous strand, which have been assumed to be the result of posttranscriptional polyadenylation (4). In the case of R2, all elements isolated from species of the *Drosophila* genus end in a poly(A) tail from 10 to 25 nucleotides in length (13, 21), while all R2 elements isolated from *B. mori* end with exactly 4 A nucleotides (7). We have postulated that the integrated *B. mori* R2 elements could also have been derived from a polyadenylated transcript if reverse transcription initiated

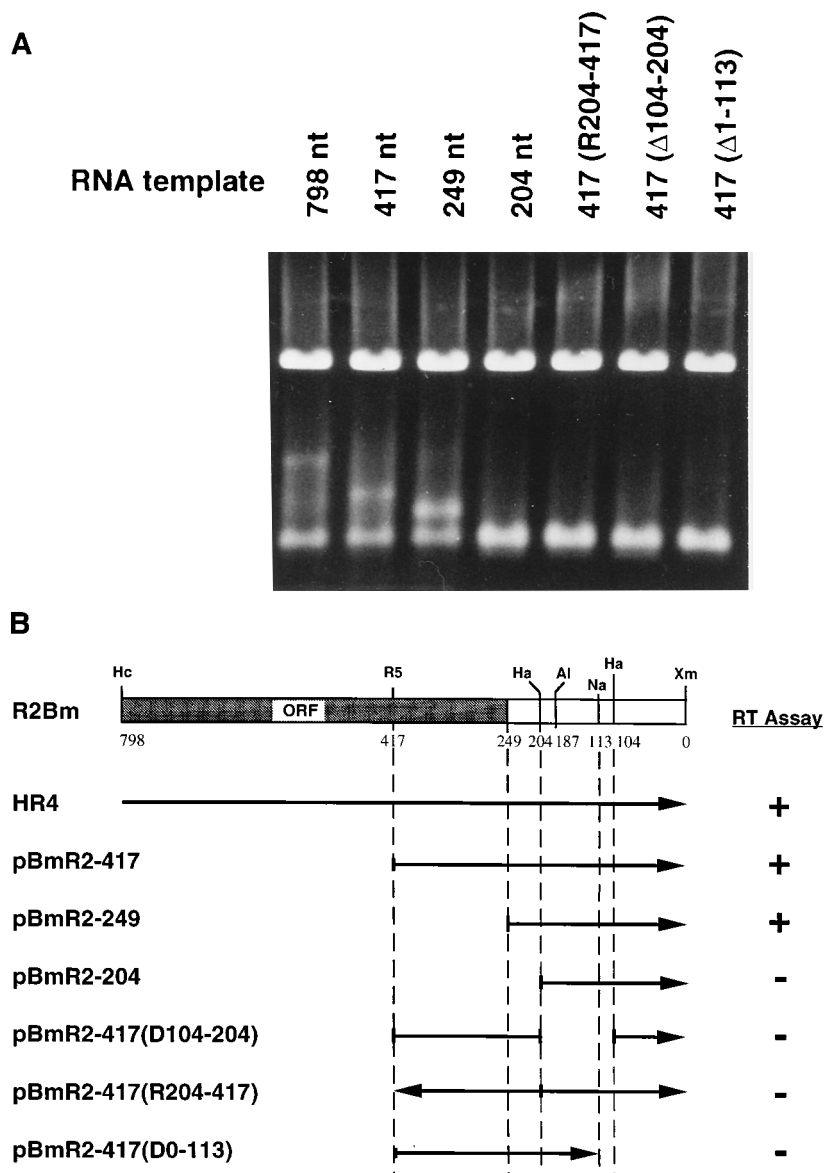


FIG. 3. Deletions of HR4 sequences to determine the regions required for the DNA-primed reverse transcription (RT) reaction. (A) Ethidium bromide-stained agarose gel of the reverse transcription reactions using the various deletion-containing R2 RNA templates. The R2 sequences present on each RNA are shown in panel B. Note that the product band is directly related to the length (in nucleotides [nt]) of the template RNA in the reaction. The amount of RNA used in each assay (0.2 μ g) represents a three- to fivefold excess of that required to give the maximum yield of product in the standard assay with HR4 RNA. (B) Restriction map of the 3' end of the R2 element from *B. mori*. All numbering is from the 3' end of the R2 sequence as defined by its junction with the 28S gene. The end of the open reading frame (ORF) of the element is located 249 bp from this 3' end. Enzyme abbreviations: Hc, *HincII*; R5, *EcoRV*; Ha, *HaeIII*; Al, *AluI*; Na, *NaeI*; Xm, *XmnI*. Arrows, R2 sequences present in the various pBluescript constructs used to generate the different RNA transcripts. The ability of each RNA to serve as a template for reverse transcription is summarized on the right and is based on at least two trials in addition to that in panel A. -, no 1.4- to 1.6-kb products could be detected by ethidium bromide staining or by ³⁵S-dATP incorporation.

precisely at the 4th nucleotide of a longer poly(A) tail (20). For our in vitro assays pBluescript constructs were designed to give rise to T7 runoff transcripts ending in 4 A nucleotides, the precise 3' end of the R2 elements seen in vivo (see Materials and Methods) (26). However, because T7 RNA polymerase can add extra nucleotides in such runoff reactions, HR4 templates containing 1 or more additional nucleotides at the 3' end were also generated (27). We have shown that, with such HR4 templates, most of the integrated products (8 of 11) had 4 or 5 T nucleotides at the junction of the cDNA strand with the 28S gene target sequence, suggesting that reverse transcription ini-

tiated opposite the 1st nucleotide at the 3' end of the transcript (26) (see Fig. 5). The three other sequenced products had junction sequences containing up to 12 T nucleotides. These longer T tails could have been added either by the R2 reverse transcriptase to begin the cDNA or by the T7 RNA polymerase adding additional A nucleotides to the RNA template.

To directly test the preference of the R2 reverse transcriptase for the length of the A tail at the 3' end of its transcript, HR4 RNA templates ending in a minimum of 1, 4, or 8 A nucleotides were compared in the DNA-primed reverse transcription reaction (Fig. 4A). Again, because the T7 RNA

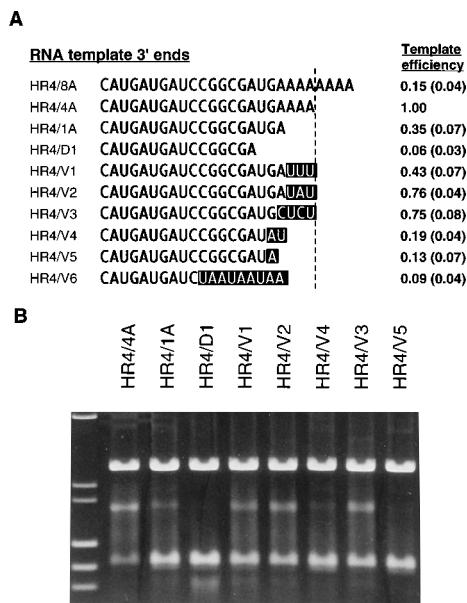


FIG. 5. Effects of nucleotide deletions and substitutions on the ability of the HR4 RNA to support DNA-primed reverse transcription. (A) Summary of the RNA templates used and the efficiencies with which they served as templates in the DNA target-primed reverse transcription reaction. Template efficiencies are given as the fraction of 1.8- to 1.9-kb products generated relative to that for the 4-A RNA template. These numbers are based on the average for three experiments, with standard deviations shown in parentheses. Dashed vertical line, *in vivo* junction of the R2 element with the 28S gene. Substitutions of the R2 RNA sequence are highlighted in black. (B) Agarose gel of the reverse transcription products obtained from most of the RNA templates in panel A.

reverse transcription reaction, a series of additional RNA templates which contain nucleotide substitutions and/or deletions at their 3' ends were made. Figure 5A shows the sequences at the 3' end of the various RNAs used in these experiments and summarizes their efficiency in supporting DNA-primed reverse transcription. The results of one such set of assays are shown in Figure 5B.

Of the various modified R2 RNAs tested, those that served as the best templates were V1, V2, and V3. These RNAs differ in the nucleotide sequence of their last 4 residues but are identical in length to the 4-A construct and will therefore be referred to as full length. Representative junction sequences obtained with the V1 and V3 RNAs are shown in Fig. 6A and B. The junction sequences obtained with the V2 RNA were similar to those obtained with the V3 RNA (data not shown). Most of the integrated products derived from these RNAs had initiated cDNA synthesis within 1 or 2 nucleotides of the 3' end of the RNA templates. The remaining products contained from 1 to 6 nucleotides located between those cDNA sequences clearly templated by the R2 RNA and the 28S junction.

All R2 RNA templates that were truncated at their 3' ends gave rise to considerably lower yields of the 1.8- to 1.9-kb product, and most of the sequenced products derived from these RNAs had a considerable number of additional nucleotides added between the R2 sequence and the 28S target DNA (Fig. 4C and 6C and D). With the previously described 1-A construct (truncated by 3 nucleotides), 80% of the integrated products had from 1 to 18 additional nucleotides that were of no identifiable pattern (Fig. 4C). In the case of the D1 construct (truncated by 6 nucleotides) the efficiency of the DNA-primed reverse transcription reaction was reduced until

it was difficult to detect on ethidium bromide-stained gels (Fig. 5B). The sequences of the integrated products that were PCR amplified, however, were surprisingly uniform (Fig. 6C). Thirteen of the sequenced products initiated reverse transcription at the precise 3' end of the RNA template, and each contained from 5 to 28 additional nucleotides. Most of these additional sequences were homopolymers of T.

In an attempt to examine the mechanism responsible for determining the sequence of these nontemplated nucleotides, three additional RNAs (V4, V5, and V6) that contained 3- or 4-nucleotide truncations and different sequences at their extreme 3' ends were tested. In the case of the V4 RNA template, most of the integrated products had nontemplated nucleotides that varied from 1 to 33 nucleotides in length (Fig. 6D). Unlike the nontemplated homopolymers seen with the 1-A or D1 RNA template, the nontemplated nucleotides generated with the V4 RNA contained abundant AG and TAG sequences. In the case of the V5 RNA, which was 1 nucleotide shorter than the V4 RNA, most integrated products had nontemplated nucleotides from 1 to 35 nucleotides in length that were similar in sequence to the V4 nontemplated nucleotides (data not shown). Finally, RNA construct V6 was tested because certain non-LTR retrotransposable elements end in TAA repeats (2, 17, 41). V6 RNA is truncated by 3 nucleotides and ends with three UAA repeats (Fig. 5A) and thus was predicted to promote the addition of TAA repeats before reverse transcription of the RNA. Of the 11 junctions sequenced, one did give rise to the addition of TAA and TAAA repeats. The other 10 sequenced junctions all represented internal initiations from 13 to 62 nucleotides within the R2 sequence (data not shown). The V6 RNA differed from the other RNAs in this experiment in that the GC-rich region 8 to 13 nucleotides from the 3' end of the RNA was changed to UAA repeats. The >90% internal initiations induced by the V6 RNA suggests that this GC-rich region plays an important role in the positioning of the RNA template in the initiation of reverse transcription.

One final point should be mentioned. A reverse-transcribed product band of approximately 0.8 kb was detected in many assays using suboptimal RNA templates (this band is most clearly visible in Fig. 5B with the D1 RNA). We have found that these 0.8-kb products represent heteroduplexes between the input RNA and its cDNA (data not shown). The mechanism used to prime synthesis of the cDNA in these products is currently under investigation.

Polymerization of nucleotides on the target DNA without R2 integration. If 3' truncated RNA templates result in the R2 reverse transcriptase adding nontemplated nucleotides to the target DNA before engaging the RNA template, then target DNA that contains only these nontemplated nucleotides should accumulate in the assay. On the other hand, if the extra nucleotides seen at the junction between the cDNA and the 28S target site had been added by the T7 RNA polymerase during the synthesis of the RNA, then target DNA with nontemplated nucleotides is unlikely to accumulate.

To differentiate between these two models, the experiment diagrammed in Fig. 7A was conducted. Plasmid pB109 was digested with *Eco*RI and used as the substrate in the DNA-primed reverse transcription reaction with V4 RNA as template. After incubation with the R2 protein, the 3.7-kb fragment was isolated. These 3.7-kb fragments contained the first strand cleaved at the target site (nearly 100% of the input target DNA is nicked within the first few minutes of the reaction [26]) but had not undergone the integration reaction or second-strand cleavage. The single-stranded 28S DNA fragment from the *Eco*RI site to the first-strand cleavage made by

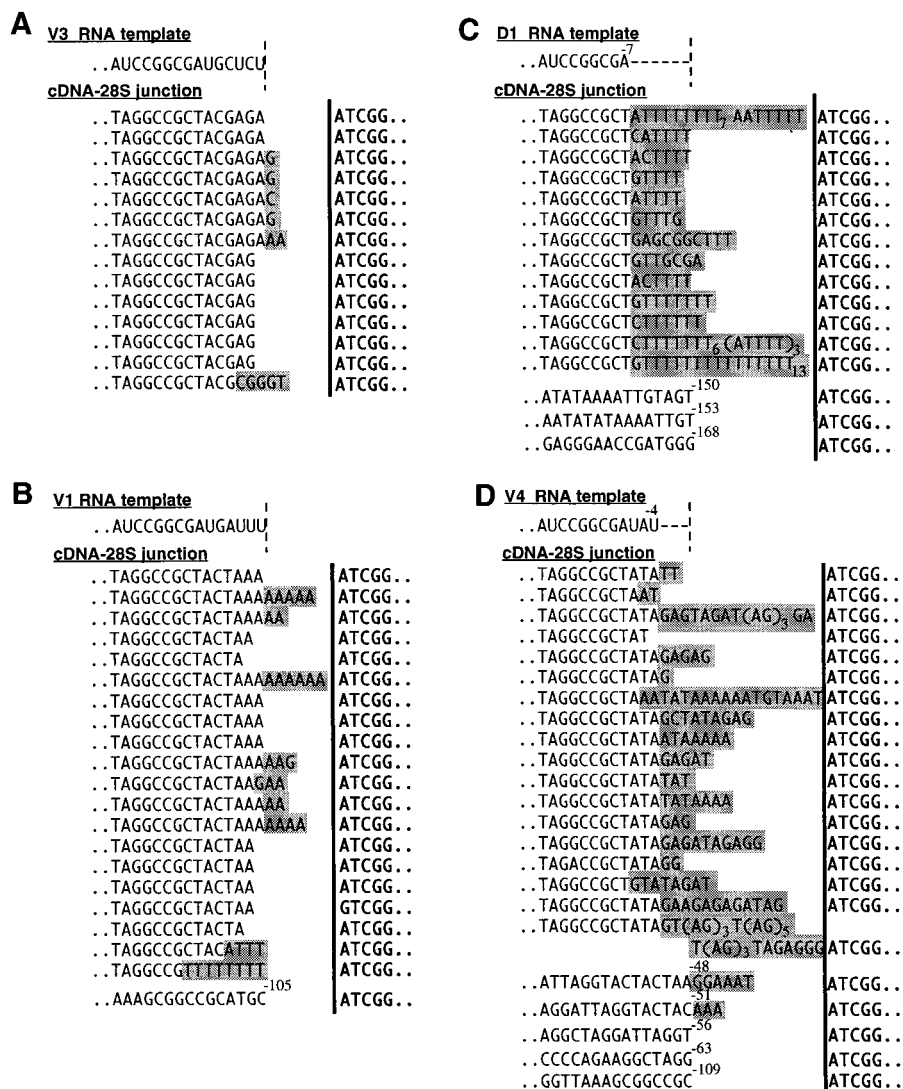


FIG. 6. 3' junctions obtained with HR4 RNA templates containing various deletions and substitutions at the 3' end. (A) Junctions obtained from the V3 RNA template; (B) junctions obtained from the V1 RNA template; (C) junctions obtained from the D1 RNA template; (D) junctions obtained from the V4 RNA template. Shown at the top of each panel is the 3' end of the RNA used in the reaction. Dashed vertical lines, *in vivo* junction of the R2 element with the 28S gene. Below are the junction sequences generated in the *in vitro* reaction. cDNA sequences are shown to the left of the solid vertical line, and 28S gene sequences are to the right of this line. Junctions derived from internal initiations are shown at the bottom of each panel. The distances of these internal initiation sites from the 3' end of the RNA are indicated. Extra nucleotides present at the junction that would not be predicted on the basis of simple reverse transcription of the RNA template are shaded.

the R2 protein was purified by denaturing polyacrylamide gel electrophoresis. Poly(dC) tails were added to the 3' end by terminal transferase, and the C-tail 28S DNA was PCR amplified by using a poly(dG)-containing anchor primer and a second primer complementary to the 28S sequence downstream of the R2 insertion site. The PCR product was gel purified and cloned into M13 phage, and individual clones were sequenced. Eight examples of the nucleotide sequences found associated with the nicked target site of the 28S gene are shown in Fig. 7B. The nontemplated sequences ranged from 1 to 68 nucleotides in length and were mostly long runs of A nucleotides interrupted by occasional T or G nucleotides. These results clearly argue that the nontemplated nucleotides are added by the R2 reverse transcriptase before the RNA template is engaged. It is unlikely that these additions are a result of an *E. coli* polymerase contaminating our preparation of the R2 protein. When ³⁵S-labelled nucleotides were added to the assays, only

the nicked site on the 28S gene became labelled by the addition of nontemplated nucleotides, and then only when R2 RNA transcripts with short truncations at their 3' ends were used as templates. When HR4 RNA or *E. coli* RNA was used as a template in the assay, we were not able to detect the labelling of the target site (data not shown).

DISCUSSION

The reverse transcriptase encoded by the R2 element of *B. mori* is remarkably specific for its own transcript, being capable of finding and reverse transcribing the transcript in the presence of 7.5 mg of nonspecific RNA per ml. The RNA sequences responsible for this specificity are limited to the 250-nucleotide 3' untranslated region of the element. The 3' untranslated regions of R2 elements from a variety of different insect species have been compared and have been shown to

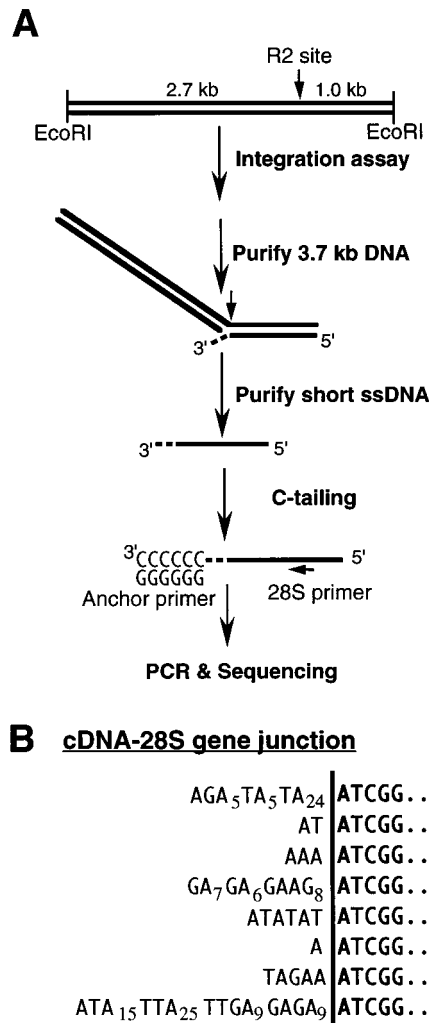


FIG. 7. Characterization of nucleotides added to the nicked target site without R2 integration. (A) Diagram of the anchor PCR approach used to obtain the sequences. The linearized 3.7-kb pB109 plasmid was incubated for 45 min with the V4 RNA template and the R2 protein in the standard DNA-primed reverse transcription reaction. The reaction products were resolved on a 1% agarose gel, the 3.7-kb DNA fragment was recovered, and the 1,000-nucleotide fragment generated by first-strand cleavage was isolated on a 5% polyacrylamide-urea gel. C tails were added to the end of this fragment with terminal transferase and PCR amplified with one primer specific to the 28S gene and a poly(G)-containing anchor primer. The PCR product was then inserted into M13, and individual clones were sequenced. Fewer than 10% of the cloned sequences contained extra nucleotides. ssDNA, single-stranded DNA. (B) Summary of the nontemplated nucleotides added to the target sequence. Nucleotides added by the reverse transcriptase are shown to the left of the vertical line, while 28S gene sequences are to the right of this vertical line.

vary from 150 to 550 bp in length (8). Comparisons of R2 elements from organisms within the genus *Drosophila* indicate that this region is under considerable selective constraint, evolving at a rate similar to that of the protein-coding regions (14). The highest level of sequence conservation within the 3' untranslated regions of the *Drosophila* elements was found in three regions spaced approximately 50 bp apart (14). Consistent with this phylogenetic approach, the deletion analysis of the *B. mori* R2 3' untranslated region described in this report indicated that multiple regions of the 3' untranslated region are required for specific recognition. Somewhat surprisingly, the sequence of the extreme 3' end of the RNA, where reverse

transcription initiates, is not critical for the reaction. We propose the model shown in Fig. 8A for the specific association of the R2 RNA with the target DNA during initiation of the DNA-primed reverse transcription reaction. When the RNA template is appropriately bound to the reverse transcriptase and the DNA target is bound to the endonuclease domain of the same or a second protein subunit, the 3' end of the R2 RNA is positioned opposite the nicked target site. This positioning enables the initiation of reverse transcription regardless of the sequence of the first few nucleotides. Further characterization of the RNA-protein interactions involved in this recognition by more detailed mutagenesis studies and direct RNA binding assays will require characterization of the secondary structure of this 250-nucleotide sequence.

Most integrated copies of non-LTR retrotransposable elements contain poly(A) tails at their 3' ends. This feature is sufficiently unusual among transposable elements for the non-LTR elements to be occasionally referred to as the poly(A) retrotransposable elements (4). We had previously assumed that the 4 A nucleotides which terminate the end of integrated R2 elements in *B. mori* were derived from a longer poly(A) tail of the RNA template (7, 26). In this article we report the first attempt to test whether polyadenylated transcripts are used in the integration reaction of a non-LTR retrotransposable element. R2 RNAs ending in at least 8 A nucleotides were utilized much less efficiently in the DNA-primed reverse transcription reaction than RNAs ending in either 4 or 1 A nucleotide. The 4 A nucleotides at the 3' end of the *B. mori* R2 element could even be substituted by the sequence CTCT with little reduction in the ability of the reverse transcriptase to utilize the RNA as template. These *in vitro* data strongly argue against the model that the R2 reverse transcriptase uses a polyadenylated transcript in the integration reaction.

It remains to be shown whether R2 RNA transcripts ending at the junction of the R2 element and the 28S gene are used by the R2 reverse transcriptase *in vivo*. It is unlikely that the RNA polymerase (either a Pol I read through from the 28S gene or a Pol II initiating at the 5' end of the element) would precisely end transcription at the 3' junction of the R2 element with the 28S gene. Therefore, an unknown RNA processing step would be required to release the free 3' end of the R2 transcript for use as an *in vivo* template. An alternative possibility is that no processing step occurs and the RNA templates used for integrations *in vivo* are ones that contain 28S gene sequences at the 3' end. We have tested such RNAs in our *in vitro* assay and have found that in some instances these RNAs can function as templates (25). The efficiency of these templates is inversely correlated with the length of the 28S sequences attached to the R2 sequences.

Nontemplated addition of nucleotides. Perhaps the most unexpected finding of this report was that the R2 reverse transcriptase had the ability to add nucleotides to the 28S target site before reverse transcription of the R2 template. Additional nucleotides were occasionally added with the various full-length RNA templates tested (i.e., transcripts which were sufficient in length to extend to the end of the R2 sequence as defined by the R2-28S gene junction seen *in vivo*). Additional nucleotides were added in nearly every instance with the various templates that were shorter than full length. For our discussions, we have described these additional nucleotides as nontemplated, although it is unlikely that they correspond to truly nontemplated additions in the same manner as terminal transferases are able to add any nucleotide to the end of a polynucleotide (10). The sequence of the extra nucleotides added by the R2 enzyme was dependent upon the nature of the RNA template used in the reaction. For example,

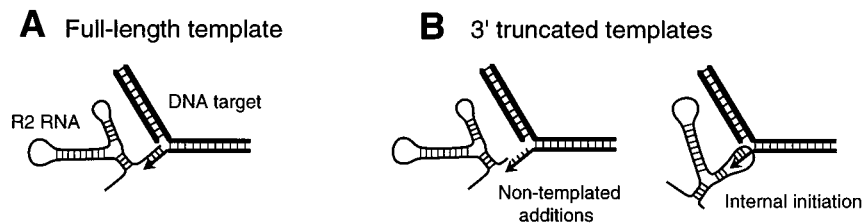


FIG. 8. Model for the initiation of synthesis in the DNA target-primed reverse transcription reaction. Because several regions of the R2 RNA appear to be required, it is assumed that the secondary structure of the RNA plays a role in its recognition by the R2 protein. The secondary structure drawn is stylistic and not intended to imply that this structure is known. Priming of reverse transcription is hypothesized to occur at a DNA bend because it occurs on DNA containing only one cleaved strand. (A) With full-length R2 RNA templates the RNA can be appropriately bound to the R2 protein such that the 3' end of the RNA is positioned opposite the nicked target site. (B) With truncated R2 templates the 3' end of the RNA cannot be positioned opposite the target site. As a result, either the reverse transcriptase adds extra nucleotides to the target DNA until it becomes capable of engaging the RNA template or the RNA-protein complex becomes repositioned such that an internal segment of RNA is opposite the target site.

if the RNA template ended in a short series of A nucleotides (4-A and 8-A RNAs), then the nontemplated additions on the cDNA strand were predominantly T nucleotides (Fig. 4B and 5) (26). If on the other hand, the RNA template ended in a series of U nucleotides (V1 RNA), then the nontemplated additions on the cDNA strand were predominately A nucleotides (Fig. 6B). A likely model to explain the addition of these extra nucleotides with full-length templates has multiple rounds of the reverse transcriptase engaging the 3' end of the RNA template, adding 1 or more templated nucleotides, and releasing from the RNA only to reengage again at the 3' end of the RNA.

Multiple rounds of aborted initiation at the 3' end of the RNA template do not appear to be the explanation for the addition of extra nucleotides with those RNA templates that were truncated by 3 to 6 nucleotides at their 3' ends. In the case of the 1-A RNA template, which is truncated by 3 nucleotides and ends with the sequence GA, the nontemplated nucleotides added to the cDNA strand were a mixture of simple G repeats and more complex sequences (Fig. 4C). In the case of the D1 RNA template, which is truncated by 6 nucleotides and also ends with the sequence GA, the nontemplated nucleotides added are predominantly T (Fig. 6C). Finally, in the case of the V4 RNA template, which is truncated by 3 nucleotides and ends in the sequence AU, the nontemplated nucleotides added are predominantly mixtures of AG or TAG sequences (Fig. 6D). We propose the model shown in Fig. 8B for the reverse transcription of these 3' truncated templates. The 3' ends of these RNAs are insufficient in length to extend to the active site of the reverse transcriptase. Nontemplated nucleotides are polymerized onto the target, enabling the enzyme to eventually engage the 3' end of the RNA template. What determines the nature of these additional nucleotides is not known. One likely model is that another segment of the same RNA template or a second RNA template may be positioned for multiple rounds of aborted initiation.

Finally, each of the 3' truncated RNAs also gave rise to internal initiations of reverse transcription. The 18 internal initiations observed occurred in three short regions of the RNA (43 to 64, 105 to 109, and 150 to 172 nucleotides from the 3' end). These internal initiations suggest that the 3'-truncated RNA templates are capable of binding to the reverse transcriptase in such a manner that other regions of the R2 template are positioned next to the active site (Fig. 8B). It is unlikely that these internal initiations occurred at the 3' ends of degraded RNA templates, because R2 RNA transcripts in which the 3' end of the untranslated region was deleted were not used as templates in the reaction (Fig. 3).

Relevance of nontemplated nucleotide additions to other non-LTR retrotransposable elements. We have previously summarized the evidence suggesting that the target DNA-primed reverse transcription which initiates the integration of R2 elements may be utilized by other non-LTR retrotransposable elements (26). Is the nontemplated addition of nucleotides by the R2 reverse transcriptase also relevant to the integration of other non-LTR elements? It is certainly possible that the unusual specificity of R2 elements for the 28S gene has led to the evolution of an unusual mechanism for the association of its reverse transcriptase with its RNA template. For example, if R2 elements are transcribed by a Pol I RNA polymerase, then it is unlikely that their RNA could be polyadenylated. In contrast, all "normal" non-LTR elements are probably transcribed by Pol II polymerases which would be expected to be polyadenylated. However, several features of the integrated copies of other non-LTR elements suggest that nontemplated additions of nucleotides may be involved in their retrotranspositions. First, while most non-LTR retrotransposable elements end in a short poly(A) sequence, very few of these elements contain the appropriately positioned AAU AAA (or AUUAAA) sequence upstream of this tail to promote the cleavage and polyadenylation steps (33). Second, some of the poly(A) tails found at the 3' end of integrated copies are interrupted by other nucleotides that are difficult to explain by the random mutation of a poly(A) tail after insertion (18). Third, some non-LTR elements end in repeating units other than poly(A). For example, I, Dong, and Q elements end in TAA repeats (2, 17, 41), R1 elements can end in TA or GTC repeats (13), T1 elements end in TGAAA repeats (1), and CR1 elements end in ATTCTGT repeats (6). Fourth, the reverse transcriptase from the human L1 element has been shown to be able to add nontemplated nucleotides in a yeast *in vivo* system using a Ty1/L1 hybrid element (11).

The ability of the reverse transcriptase from other non-LTR elements to add nontemplated nucleotides in a manner similar to that shown for the R2 element could explain the generation of simple repeats at their 3' ends. Those elements that end in simple A homopolymers may either encode a reverse transcriptase or use an RNA template that results in the preferential addition of T residues to the cDNA strand. It may be possible for us to differentiate between these two possibilities. All R2 elements isolated from *D. melanogaster* and related species end in a poly(A) tail from 10 to 25 nucleotides in length that is only rarely interrupted by another nucleotide (13). By using the approach we have applied to the R2 element of *B. mori* to analyze the corresponding components of the R2 element from *D. melanogaster*, we should be able to determine if

it is the enzyme or the template that is responsible for the uniform A tails.

A final argument to suggest that the ability to make non-templated additions is widespread in retroelements is the finding that the *Neurospora crassa* Mauriceville plasmid reverse transcriptase can add nontemplated nucleotides before the initiation of reverse transcription (9). On the basis of the sequence of its reverse transcriptase domain, the Mauriceville plasmid has been shown to be more closely related to the non-LTR retrotransposable elements than the LTR retrotransposable elements (40). Further evidence for the relationship of the Mauriceville reverse transcriptase to that of the non-LTR retrotransposable elements is the ability of the Mauriceville enzyme to use the 3' OH of DNA or RNA molecules to prime reverse transcription (34, 35). Clearly, the structures and mechanisms used by the reverse transcriptases encoded by the non-LTR retroelements need to be further compared with the enzymes encoded by retroviruses and LTR retrotransposable elements.

ACKNOWLEDGMENTS

This work was supported by a National Institutes of Health grant (GM42790) to T.H.E.

We thank Bill Burke, Danna Eickbush, and Janet George for comments on the manuscript. We also thank Bill Burke for generating the pBluescript constructs used to make the RNAs in Fig. 4.

REFERENCES

- Besansky, N. J. 1990. A retrotransposable element from the mosquito *Anopheles gambiae*. *Mol. Cell. Biol.* **10**:863–871.
- Besansky, N. J., J. A. Bedell, and O. Mukabayire. 1994. Q: a new retrotransposon from the mosquito *Anopheles gambiae*. *Insect Mol. Biol.* **3**:49–56.
- Boeke, J. D. 1989. Transposable elements in *Saccharomyces cerevisiae*, p. 335–374. *In* D. H. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society of Microbiology, Washington, D.C.
- Boeke, J. D., and V. G. Corces. 1989. Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43**:403–434.
- Bucheton, A. 1990. I transposable elements and I-R hybrid dysgenesis in *Drosophila*. *Trends Genet.* **6**:16–21.
- Burch, J. B. E., D. L. Davis, and N. B. Haas. 1993. Chicken repeat-1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. *Proc. Natl. Acad. Sci. USA* **90**:8199–8203.
- Burke, W. D., C. C. Calalang, and T. H. Eickbush. 1987. The site specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol. Cell. Biol.* **7**:2221–2230.
- Burke, W. D., D. G. Eickbush, Y. Xiong, J. Jakubczak, and T. H. Eickbush. 1993. Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. *Mol. Biol. Evol.* **10**:163–185.
- Chiang, C.-C., J. C. Kennell, L. A. Wanner, and A. M. Lambowitz. 1994. A mitochondrial retroplasmid integrates into mitochondrial DNA by a novel mechanism involving the synthesis of a hybrid cDNA and homologous recombination. *Mol. Cell. Biol.* **14**:6419–6432.
- Deng, G. R., and R. Wu. 1983. Terminal transferase: use in the tailing of DNA and for in vitro mutagenesis. *Methods Enzymol.* **100**:96–116.
- Dombroski, B. A., Q. H. Feng, S. L. Mathias, D. M. Sassaman, A. F. Scott, H. H. Kazazian, and J. D. Boeke. 1994. An in vivo assay for the reverse transcriptase of human retrotransposon L1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14**:4485–4492.
- Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure. 1989. Origin and evolutionary relationships of retroviruses. *Q. Rev. Biol.* **64**:1–30.
- Eickbush, D. G., and T. H. Eickbush. 1995. Vertical transmission of the retrotransposable elements R1 and R2 during the evolution of the *Drosophila melanogaster* species subgroup. *Genetics* **139**:671–684.
- Eickbush, D. G., W. C. Lathe III, M. P. Francino, and T. H. Eickbush. 1995. R1 and R2 retrotransposable elements of *Drosophila* evolve at rates similar to those of nuclear genes. *Genetics* **139**:685–695.
- Eickbush, T. H. 1994. Origin and evolutionary relationships of retroelements, p. 121–157. *In* S. S. Morse (ed.) *The evolutionary biology of viruses*. Raven Press, New York.
- Evans, J. P., and R. D. Palmiter. 1991. Retrotransposition of a mouse L1 element. *Proc. Natl. Acad. Sci. USA* **88**:8792–8795.
- Fawcett, D. H., C. K. Lister, E. Kellet, and D. J. Finnegan. 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINES. *Cell* **47**:1007–1015.
- Furano, A. V., C. C. Somerville, P. N. Tschlis, and E. D'Ambrosio. 1986. Target sites for the transposition of rat long interspersed repeated DNA elements (LINES) are not random. *Nucleic Acids Res.* **14**:3717–3727.
- Hutchison, C. A., S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell. 1989. LINES and related retrotransposons: long interspersed repeated sequences in the eukaryotic genome, p. 539–617. *In* D. H. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society of Microbiology, Washington, D.C.
- Jakubczak, J. L., W. D. Burke, and T. H. Eickbush. 1991. Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc. Natl. Acad. Sci. USA* **88**:3295–3299.
- Jakubczak, J. L., Y. Xiong, and T. H. Eickbush. 1990. Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *J. Mol. Biol.* **212**:37–52.
- Jensen, S., and T. Heidmann. 1991. An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE 1 element. *EMBO J.* **10**:1927–1937.
- Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prudhomme, and A. Bucheton. 1994. Retroviruses in invertebrates—the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**:1285–1289.
- Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar, and F. M. Sheen. 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75**:1083–1093.
- Luan, D. D. Unpublished data.
- Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**:595–605.
- Milligan, J. F., and O. C. Uhlenbeck. 1989. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.* **180**:51–62.
- Pelisson, A., D. J. Finnegan, and A. Bucheton. 1991. Evidence for retrotransposition of the I factor, a LINE element of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **88**:4907–4910.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sandmeyer, S. B., L. J. Hansen, and D. L. Chalker. 1990. Integration specificity of retrotransposons and retroviruses. *Annu. Rev. Genet.* **24**:491–518.
- Schwarz-Sommer, Z., L. Leclercq, E. Gobel, and H. Saedler. 1987. *Cin4*, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. *EMBO J.* **6**:3873–3880.
- Song, S. U., T. Gerasimova, M. Kurkulos, J. D. Boeke, and V. G. Corces. 1994. An Env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev.* **8**:2046–2057.
- Wahle, E., and W. Keller. 1992. The biochemistry of 3'-end cleavage and polyadenylation of messenger-RNA precursors. *Annu. Rev. Biochem.* **61**:419–440.
- Wang, H., J. C. Kennell, M. T. R. Kuiper, J. R. Sabourin, R. Saldanha, and A. M. Lambowitz. 1992. The Mauriceville plasmid of *Neurospora crassa*: characterization of a novel reverse transcriptase that begins cDNA synthesis at the 3' end of template RNA. *Mol. Cell. Biol.* **12**:5131–5144.
- Wang, H., and A. M. Lambowitz. 1993. The mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to reverse transcriptase and DNA polymerase progenitor. *Cell* **75**:1071–1081.
- Weiner, A. M., P. L. Deninger, and A. Efstratiadis. 1986. Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**:631–661.
- Whitcomb, J. M., and S. H. Hughes. 1992. Retroviral reverse transcription and integration: process and problems. *Annu. Rev. Cell Biol.* **8**:275–306.
- Xiong, Y., and T. H. Eickbush. 1988. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell* **55**:235–246.
- Xiong, Y., and T. H. Eickbush. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* **5**:675–690.
- Xiong, Y., and T. H. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**:3351–3362.
- Xiong, Y., and T. H. Eickbush. 1993. Dong, a non-long terminal repeat (non-LTR) retrotransposable element from *Bombyx mori*. *Nucleic Acids Res.* **21**:1318.