Cell, Vol. 72, 595–605, February 26, 1993, Copyright © 1993 by Cell Press

Reverse Transcription of R2Bm RNA Is Primed by a Nick at the Chromosomal Target Site: A Mechanism for Non-LTR Retrotransposition

Dongmei D. Luan, Malka H. Korman, John L. Jakubczak,* and Thomas H. Eickbush
Department of Biology
University of Rochester
Rochester, New York 14627

Summary

R2 is a non-LTR retrotransposable element that inserts at a specific site in the 28S rRNA genes of most insects. We have expressed the open reading frame of the R2 element from Bombyx mori, R2Bm, in E. coli and shown that it encodes both sequence-specific endonuclease and reverse transcriptase activities. The R2 protein makes a specific nick in one of the DNA strands at the insertion site and uses the 3′ hydroxyl group exposed by this nick to prime reverse transcription of its RNA transcript. After reverse transcription, cleavage of the second DNA strand occurs. A similar mechanism of insertion may be used by other non-LTR retrotransposable elements as well as short interspersed nucleotide elements.

Introduction

Retrotransposable elements can be divided into two major classes: those with long terminal repeats (LTRs) and those without. Examples of the LTR class include copia and gypsy of Drosophila melanogaster, Ty1 and Ty3 of Saccharomyces cerevisiae, and Ta1 of Arabidopsis thaliana (Mount and Rubin, 1985; Marlor et al., 1986; Hansen et al., 1988; Voytas and Ausubel, 1988; Boeke, 1989). The structure and mechanism of retrotransposition of this class of elements are similar to those of retroviruses (reviewed by Varmus and Brown, 1989; Boeke and Corces, 1989; Sandmeyer et al., 1990). The LTRs of each element contain the start and stop signals for transcription of the RNA template. Reverse transcription of this template occurs in the cytoplasm with the first-strand DNA synthesis primed by a tRNA molecule and second-strand DNA synthesis primed by an oligopurine RNA molecule. Terminal redundancies of the LTR sequences on the RNA template enable strand transfer events to complete the synthesis of both strands. The linear double-stranded DNA intermediate produced by this synthesis migrates to the nucleus, where an integrase, bound to the LTRs, is capable of making a double-stranded break in the chromosome and covalently joining the ends of the intermediate to the chromosome (Brown et al., 1987; Bushman et al., 1990; Katz et al., 1990; Eichinger and Boeke, 1990).

In contrast, little is known about the mechanism of retrotransposition of the second class of retrotransposons, the non-LTR elements (Martin, 1991; Eickbush, 1992). Representatives of this class include the long interspersed nucleotide elements (LINEs) of mammals (L1), I of D. melanogaster, cin4 of Zea mays, and ingi of Trypanosoma brucei (Fawcett et al., 1986; Schwarz-Sommer et al., 1987; Kimmel et al., 1987; Hutchison et al., 1989). This class of elements has also been called the LINE-like elements (Singer and Skowronski, 1985), the nonviral elements (Weiner et al., 1986), poly(A) type retrotransposons (Boeke and Corces, 1989), and retroposons (Temin, 1989). These elements are missing many of the genes encoded by the LTR-containing retrotransposons. No protease domain has been detected in any of the non-LTR elements, most do not appear to have either the RNAase H or integrase domains, and some may be missing the *gag*-like gene. These elements clearly undergo retrotransposition, because intron sequences placed within the element are precisely removed during the generation of new copies (Pelisson et al., 1991; Jensen and Heidmann, 1991; Evans and Palmiter, 1991). Several non-LTR retrotransposons contain an internal promoter that is capable of initiating transcription upstream at the first nucleotide of the element (Mizrokhi et al., 1988; Swergold, 1990; Chaboissier et al., 1990; Minchiotti and Di Nocera, 1991). However, little or no data exist as to the mechanism by which this RNA transcript is reverse transcribed.

The non-LTR retrotransposon, R2 (Figure 1A), inserts into a specific sequence of the 28S ribosomal RNA (rRNA) genes of its host (Burke et al., 1987). A second non-LTR retrotransposon, R1, inserts 74 bp downstream of the R2 site (Xiong and Eickbush, 1988a). R1 and R2 elements occupy from a few percent to over half of the rDNA units of most insects (Jakubczak et al., 1991). We have previously shown that the single open reading frame (ORF) of the R2 element from Bombyx mori (R2Bm) encodes an endonuclease capable of specifically cleaving the 28S rRNA gene in an in vivo Escherichia coli expression system (Xiong and Eickbush, 1988b). As reported in this work, we have purified the 120 kd protein expressed from the R2Bm ORF and have shown that this protein also contains reverse transcriptase activity. Significantly, the R2 protein is able to use the 3′ hydroxyl group generated by a nick in one strand of the DNA target site to prime reverse transcription of its RNA template.

Results

Purification of the Protein Encoded by the R2 ORF

Previous work (Xiong and Eickbush, 1988b) demonstrated the functional expression in E. coli of the single ORF from R2Bm. The ORF was found to encode an endonuclease that could specifically cleave the 28S gene target site also located on the expression plasmid. The experiments reported here employed the pUC18 expression construct pR260, which contains the entire ORF of R2Bm starting from 18 bp upstream of the first methionine codon in-frame with the initiation codon of the *lacZ* gene (Figure 1B). DNA-

*Present address: National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892.

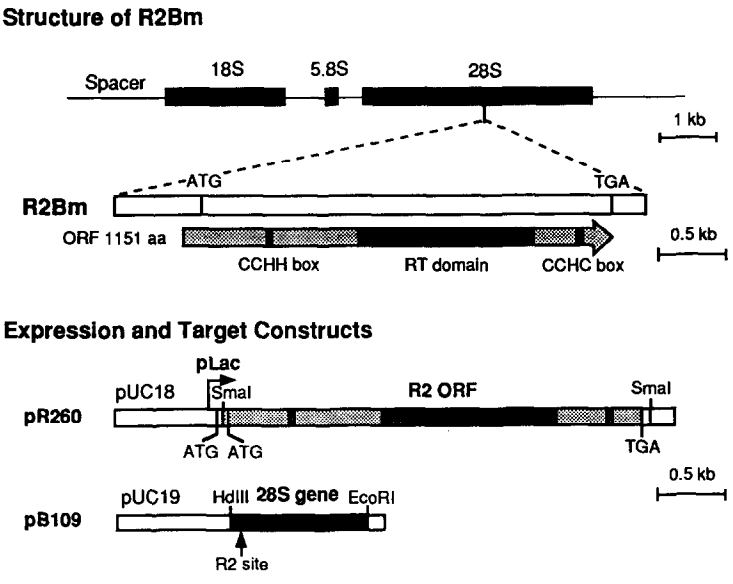

Figure 1. Structure of R2Bm and the Constructs Used for R2 Expression

(A) Diagram of an rDNA unit showing the location of R2Bm and the structure of its ORF. Closed boxes represent ribosomal coding regions. The first ATG in the 1151 amino acid ORF of R2Bm is codon 38. The reverse transcriptase domain (RT) and the two cysteine-histidine motifs within the ORF (CCHH and CCHC) are indicated by darker stippling.

(B) Construction of the expression plasmid pR260 and the target plasmid pB109. pR260 contains a 3.5 kb SmaI fragment with the entire R2 ORF subcloned into pUC18 under the control of the *lacZ* promoter and in-frame with the initiation codon of that gene (Xiong and Eickbush, 1988b). pB109 contains a 1 kb fragment of the 28S gene with the R2 insertion site subcloned into pUC19. Open boxes represent pUC sequences; pLac, *lacZ* promoter; R2 DNA encoding the ORF is stippled as in (A).

free extracts were made by gentle lysis with lysozyme and Triton X-100, followed by centrifugation in 1 M NaCl to sediment chromosomal and plasmid DNA. The R2 endonuclease activity was purified by column chromatography, taking advantage of the ability of the R2 protein to bind both RNA and DNA. The 1.0 M NaCl supernatant, after centrifugation, was diluted to 0.4 M NaCl and loaded onto a Q Sepharose column. More than 98% of the total protein in the crude extract was present in the flowthrough fraction of this column. After the column was washed in 0.4 M NaCl, the endonuclease activity was eluted with 0.6 M NaCl, along with all of the RNA present in the extract. The ability of the R2 protein to bind to the Q Sepharose column was dependent upon the presence of this RNA, because RNAase A added to the crude extract eliminated the binding of the R2 protein to the column. To separate the R2 endonuclease from the E. coli RNA, the Q Sepharose fraction was loaded onto a double-stranded DNA–cellulose column in 0.2 M NaCl. The column was washed in 0.4 M NaCl, and the endonuclease activity was eluted with 0.8 M NaCl. SDS–polyacrylamide gels of this DNA–cellulose-purified fraction revealed a single predominant protein band of approximately 120 kd (data not shown). The size of this protein corresponded to the 123 kd polypeptide expected from a full-length translation of the R2Bm ORF.

The R2 protein eluted from the DNA–cellulose column had only low levels of endonuclease activity. Adding back the flowthrough fraction from the DNA–cellulose column recovered endonuclease activity, suggesting the presence of a cofactor in this flowthrough. RNAase A digestion of the flowthrough fraction revealed that the cofactor was the E. coli RNA. Further evidence that RNA was a cofactor in the endonuclease reaction was obtained by demonstrating that RNAase A treatment also destroyed the double-stranded DNA cleavage activity of the crude supernatant and Q Sepharose preparations. Endonuclease activity could be restored to these fractions by the addition of excess RNA and RNAase inhibitors (data not shown).

Properties of the DNA Cleavage Reaction

Purification of the R2 endonuclease from contaminating E. coli nucleases and the discovery of its requirement for an RNA cofactor revealed an interesting property of the enzyme: it rapidly nicked one strand of the target site and then more slowly cleaved the second strand. The kinetics of this two-step cleavage reaction is shown in Figure 2A. In this experiment purified R2 endonuclease was incubated with supercoiled pB109 plasmid DNA containing the 28S gene target site (Figure 1B) in the presence of proteinase K–digested E. coli RNA. The endonuclease was able to convert most of the supercoiled molecules to open circular molecules within the first minute of digestion, while the generation of linear DNA gradually increased over the course of 2 hr. Both nicking and cleavage activity were dependent upon the 28S gene sequence. When supercoiled DNA lacking the target site was incubated with the endonuclease, open circular or linear DNA fragments were not generated. For easier visualization of the kinetics of this cleavage reaction, the agarose gel from a similar time course was scanned, and the percentage of nicked and cleaved DNA was plotted in Figure 2B. Different preparations of the R2 protein all showed that 100% of the pB109 DNA was rapidly nicked within the first few minutes of the reaction, while the total amount of cleaved DNA varied from 50% to 90%. Shown in Figure 2C is a third time course, but RNAase A was added to the digestion buffer to remove the trace amount of E. coli RNA that copurifies with the endonuclease through the DNA–cellulose column. In the absence of RNA, the open circular DNA again rapidly accumulated, but little second-strand cleavage of the DNA occurred. These experiments indicated that rapid nicking of one strand was the first step in the endonuclease reaction, while cleavage of the second strand occurred more slowly. An RNA cofactor was required for the slow cleavage of the second strand, but not for the initial rapid nicking activity. This two-step cleavage reaction occurs with both linear and supercoiled DNA sub-

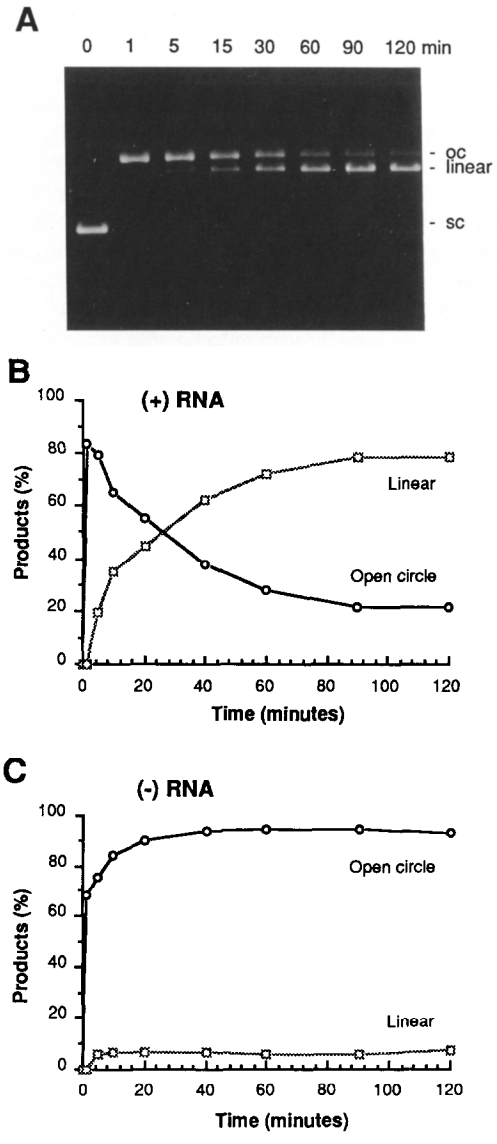

Figure 2. Effect of RNA on the Nicking and Cleavage Activity of the R2 Endonuclease

(A) Agarose gel of supercoiled pB109 DNA incubated with the R2 protein in the presence of RNA. For each time period 0.5 μg of pB109 was incubated with 3 μl of DNA–cellulose-purified R2 protein (60 ng) in 0.15 M NaCl, 10 mM $MgCl_2$, 50 mM Tris–HCl (pH 8), 1 mM DTT, and 1 μg total of E. coli RNA at 37°C. The reactions were electrophoresed on a 1% agarose gel and stained with ethidium bromide. oc, open circular DNA; sc, supercoiled DNA.

(B) Kinetics of the nicking and cleavage activity in the presence of E. coli RNA. The incubations were conducted as in (A). After electrophoresis, supercoiled, open circular, and linear DNA were quantified by laser densitometry. Nicked (open circular) and cleaved (linear) DNA are expressed as the percentage of total DNA in the reaction.

(C) Kinetics of the nicking and cleavage activity in the absence of E. coli RNA. The experiment was conducted as in (A), except that 0.1 μg of RNAase A was added to the reactions instead of E. coli RNA. Quantitation was conducted as in (B).

strates. A 3′ hydroxyl and 5′ phosphate appear to be generated by these cleavages, because the cleaved substrate can be over 90% resealed by T4 DNA ligase (data not shown).

The strand initially nicked in the cleavage reaction was determined using target DNA plasmids pR2603 and pR2604 (Xiong and Eickbush, 1988b). These plasmids, each containing a 160 bp segment of the 28S gene cloned in either orientation in the polylinker region of pUC18, were digested for 5 min with the DNA–cellulose-purified R2 protein, pretreated with RNAase A to allow only nicking activity. Primer extension experiments indicated that only the RNA coding strand of the 28S gene was nicked (data not shown). A modified primer extension–termination procedure (Brown and Smith, 1980) was used to determine precisely the location of the cleavages on both DNA strands. In this method, the 160 bp segment of the 28S gene containing the target site was subcloned in either orientation into the mp18 sequencing vector (see Experimental Procedures). For each orientation, Sequenase and the −40 sequencing primer were used to synthesize a ^{32}P-labeled complementary strand on the single-stranded phage DNA. The double-stranded products obtained from this synthesis were subsequently digested for 90 min with the R2 endonuclease in the presence of RNA. The digestion products were denatured and electrophoresed next to a sequencing ladder from each strand on an 8 M urea–polyacrylamide gel (Figure 3A). The results of these experiments indicated that the nick in the coding strand (lower strand of the insertion site as shown in Figure 3B) was between a C and an A residue. Cleavage of the upper strand was at a site 2 bp upstream of the nick site, resulting in free DNA ends that contained a 2 bp stagger with 5′ overhangs.

Our previous attempt to determine the location of the cleavage sites required digestion of the DNA in vivo, followed by runoff experiments using Sequenase (Xiong and Eickbush, 1988b). This approach, which suggested a 4 bp staggered cut, was in error, owing to the addition of an extra nucleotide by the polymerase in these runoff reactions. The location of the initial nick shown in Figure 3B is consistent with all available data on the location of R2 insertions in six species of insects (Jakubczak et al., 1991) and with the results of the DNA-primed reverse transcription experiments described below.

The R2 Protein Encodes a Reverse Transcriptase

The central region of the R2 ORF contains sequence similarity to reverse transcriptases (see Figure 1A). Reverse transcriptase activity using poly(rC) as template and oligo(dG) as primer has been previously shown for the non-LTR elements jockey (Ivanov et al., 1991), CRE1 (Gabriel and Boeke, 1991), and L1 (Mathias et al., 1991). Using this poly(rC)–oligo(dG) assay, no reverse transcriptase activity was detected in the crude E. coli extracts of the pR260 expression construct. Reverse transcriptase activity was detected in the Q Sepharose fraction and at even higher levels in the DNA–cellulose fractions, presumably because removal of endogenous E. coli RNA enabled more extensive binding of the poly(rC) template to the polymerase. As shown in Table 1, the reverse transcriptase activity detected with the DNA–cellulose fraction was more than 400-fold above background. Further characterization of the RNA-directed and DNA-directed DNA polymerase ac-

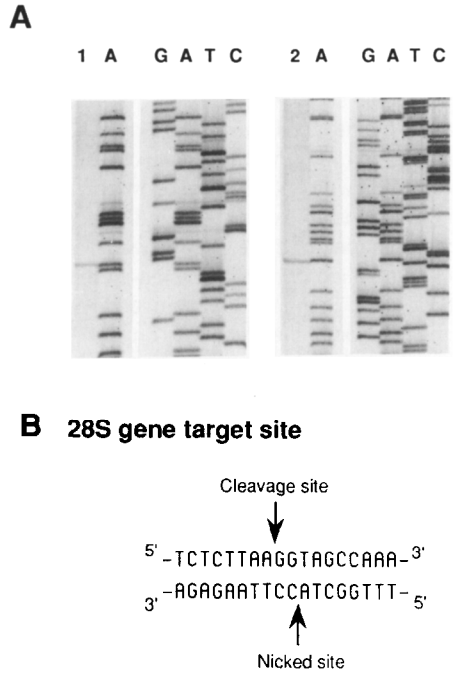

Figure 3. Location of the First-Strand Nick and Second-Strand Cleavage Site on the 28S Gene

(A) Determination of the cleavage sites on both DNA strands by the primer extension–termination method of Brown and Smith (1980). Lane 1, R2 cleavage of the noncoding strand (second-strand cleavage). Shown immediately adjacent to the cleavage lane is the A lane from a sequencing ladder. Also shown is a complete sequencing ladder of the noncoding strand (upper strand in [B]). Lane 2, R2 cleavage of the coding strand (first-strand nick). Shown immediately adjacent to the cleavage lane is again the A lane from a sequencing ladder and a complete sequencing ladder of the coding strand (lower strand in [B]). (B) Sequence of the 28S target site indicating the locations of the nick and cleavage sites.

tivities of this protein using various nucleic acid templates will be the subject of a separate report (J. L. J., D. D. L., and T. H. E., unpublished data).

Sedimentation analyses have also been conducted to determine if the reverse transcriptase and endonuclease activities would cosediment. Under all sedimentation conditions tested, endonuclease and reverse transcriptase activities cosedimented. The sedimentation value of the R2 protein in 1.0 M NaCl was 5.3S, consistent with both activities being associated with a spherical 120 kd protein (data not shown). At lower salt concentrations, endonuclease and reverse transcriptase activities were associated with more rapidly sedimenting complexes, the nature of which is currently under study.

The R2 Protein Can Use the 3′ OH Group Exposed at the DNA Nick to Prime Reverse Transcription of the R2 RNA Template

The presence of both endonuclease and reverse transcriptase activity in the same 120 kd protein, and the requirement of RNA for double-stranded cleavage at the target site, suggested that these two steps in retrotransposition of R2 might be directly related. Indeed, Schwarz-Sommer et al. (1987) have postulated that the non-LTR retrotransposon cin4 could use a 3′ hydroxyl present at a DNA break to prime reverse transcription. This method of priming reverse transcription has also been suggested for the I element (Finnegan, 1989; Bucheton, 1990) and L1 (Hutchison et al., 1989). We directly tested this model by adding RNA transcripts and dNTPs to the R2 cleavage reaction to permit reverse transcription if the cleaved pB109 DNA was able to serve as primer. In these proposed models the reverse transcriptases encoded by the non-LTR elements were predicted to be able to recognize sequences near the 3′ end of their RNA transcripts. Therefore, two RNA sequences were tested in our in vitro reaction. The first transcript corresponded to an 825 bp sequence synthesized from the pBluescript vector, referred to as vector RNA. The second transcript corresponded to the 802 bp at the 3′ end of R2Bm elements, referred to as R2 RNA. The pBluescript construct used to generate R2 RNA, clone pHR4, was formed by engineering an XmnI site at the 3′ R2Bm–28S gene junction. Cleavage of this plasmid with XmnI enabled the generation of runoff T7 transcripts that should end in the sequence GAAAA, the 3′ end of all R2Bm elements (Burke et al., 1987). However, direct observation of the length of this RNA on 8 M urea–8% polyacrylamide sequencing gels indicated that many of the R2 RNAs synthesized in vitro actually contained additional nucleotides at their 3′ ends. While the most abundant RNA species generated by the T7 polymerase were of the correct length to end in GAAAA or with one additional base, from 30% to 50% of the RNA species contained up to 10 additional nucleotides (data not shown). T7 polymerase has been shown to add additional nucleotides to the end of the transcript in such runoff reactions (Milligan and Uhlenbech, 1989). This addition appears to be more extensive when the DNA template ends in a series of A residues.

The integration assay was carried out in the presence of EcoRI–HindIII double-digested pB109 DNA, DNA–cellulose-purified 120 kd R2 protein, in vitro synthesized vector RNA or R2 RNA, and dNTPs (including [α-^{32}P]dATP). The results of these experiments are shown in Figure 4. An agarose gel of the reaction products, treated with Sarkosyl to remove protein from the DNA, is shown in Figure 4A, and an autoradiograph of this gel is shown in Figure 4B. It should be noted that endonuclease cleavage of the 28S target on the 1.0 kb fragment, which readily occurred under these conditions, was not monitored in this experi-

Table 1. Reverse Transcriptase Activity of the R2 Protein Using a Standard Homopolymer Assay

Protein	Poly(C)–Oligo(dG)	dGMP Incorporated (pmol)
pUC18	+	<0.03
pR260	–	<0.03
pR260	+	13.27

Protein (125 ng) eluted from the DNA–cellulose column from either the pR260 expression construct or pUC18 was used in each reaction.

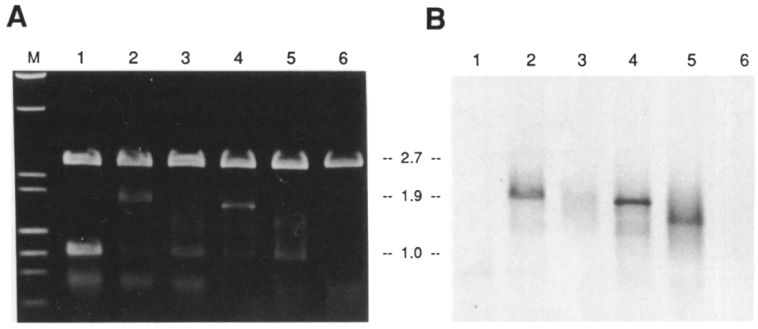

Figure 4. Assay for 28S Gene Target Site Priming of Reverse Transcription

For each reaction 0.4 μg of pB109 DNA, previously digested with EcoRI–HindIII, was incubated with 0.1 μg of DNA–cellulose-purified R2 protein, a 5 μM concentration of each dNTP (including 0.1 μM ^{35}S-dATP), and 0.2 μg of in vitro transcripts corresponding to a 825 nt pBluescript sequence (vector RNA) or a 802 nt transcript from the 3′ end of the R2Bm element (R2 RNA). Lane 1, vector RNA. Lane 2, R2 RNA; note the generation of a 1.9 kb labeled fragment. Lane 3, R2 RNA and 0.6 μM ddCTP to inhibit polymerization. Lane 4, R2 RNA; after polymerization the products were treated with 10 μg of RNAase A (Sigma). Lane 5, R2 RNA; after polymerization the products were treated with 2 U of E. coli RNAase H (Bethesda Research Laboratories). Lane 6, R2 RNA, but with EcoRI-digested pUC19 instead of the pB109 DNA.

(A) Ethidium bromide–stained agarose gel of the reaction products. M, DNA length markers: 6.6 kb, 4.4 kb, 2.3 kb, 2.0 kb, 1.4 kb, 1.1 kb, 0.9 kb, and 0.6 kb.

(B) Autoradiograph of the gel in (A).

ment because the R2 cleavage site was only 60 bp from the HindIII restriction site (see Figure 1B). In the presence of the vector RNA (Figure 4B, lane 1) the pB109 DNA fragments were not utilized as primers. Neither the 1.0 kb or 2.7 kb fragments became labeled, and there was no change in their mobility on the gel. Low level labeling of the vector RNA itself (~0.8 kb band) was detected in this reaction and is presumably reverse transcription of the RNA utilizing contaminating endogenous primers. In contrast with the vector RNA result, when R2 RNA corresponding to the 802 bp at the 3′ end of the element was added to the incubation (lane 2), high levels of reverse transcription occurred. A new 1.9 kb band was generated, while the 1.0 kb band containing the 28S target DNA was decreased by a corresponding amount. The 1.9 kb band was intensely labeled by the [^{32}P]dATP. When 0.6 μM dideoxyl cytosine was added to the reaction containing the R2 RNA (lane 3), a diffuse range of shorter products was labeled, consistent with the 1.9 kb band being the product of reverse transcription. To analyze the nature of the 1.9 kb product further, after incubation with the R2 protein the reaction was treated with RNAase A (lane 4) or RNAase H (lane 5) before the products were loaded on the gel. RNAase A reduced the 1.9 kb band to a sharper 1.8 kb band, the size predicted of a 0.8 kb RNA–DNA hybrid attached to the 1.0 kb target DNA fragment. RNAase H reduced the 1.9 kb band to a diffuse band with an average size of 1.5 kb, confirming that a portion of the 1.9 kb band was detected, indicating that at least one of the nucleotide strands was continuous DNA (data not shown).

vector band being completely unlabeled during incubation either in the presence (lanes 1–5) or absence (lane 6) of the target DNA. Further analysis of the 1.9 kb product generated by the addition of R2 RNA to the cleavage reaction indicated that this band was unaffected by SDS–proteinase K digestion. When the reaction products were run on alkaline gels, a labeled band of approximately 1.8 kb band was detected, indicating that at least one of the nucleotide strands was continuous DNA (data not shown).

The results in Figure 4 suggested that the 28S gene fragment was being used to prime reverse transcription of the R2 RNA. To determine if the 3′ OH of one of the two strands released at the cleavage site was priming the reverse transcription of the R2 RNA, the products of the reverse transcription reaction were isolated and subjected to polymerase chain reaction (PCR) amplification. One oligonucleotide primer was complementary to a region 250 bp from the end of the cDNA strand putatively made from the R2 RNA, while the second oligonucleotide primer was complementary to 28S gene sequences 40 bp either upstream or downstream of the R2 target site. Only the PCR using the primer downstream of the target site gave rise to a PCR product (300 bp), indicating that cDNA to the R2 RNA had been made and was covalently attached to the strand initially nicked by the endonuclease (data not shown). Direct sequencing of the PCR product revealed the integration of R2 sequences at the location of the nick in the 28S genes; however, there was considerable variability in the number of A nucleotides connecting the R2 element to the 28S gene at this site. Individual PCR products were therefore cloned into mp19, and 11 clones were sequenced (Figure 5B). All 11 clones indicated the insertion of R2 sequences at the nicked site. Four of the clones (1, 5, 6, and 8) contained a R2–28S junction identical to that seen in all naturally occurring R2Bm insertions (Burke et al., 1987), four clones contained an additional A at the junction, and three clones had short A-rich sequences at this junction. It is likely that these additional predominantly A nucleotides are derived from the extra nucleotides added to the end of the transcript by the T7 RNA polymerase in the runoff reaction. Because only a fraction of the input RNA is actually used as template in the reaction, it is not known whether initiation of reverse transcription begins at the 3′ end of each T7 transcript and the number of A nucleotides at the 28S junction corresponds to the length of the A-rich tail on the RNA template, or reverse transcription begins at internal locations within a longer A-rich tail present on the RNA transcripts.

Priming of Reverse Transcription Occurs on the Nicked Substrate

The results of the previous section indicated that the nicked strand was used to prime reverse transcription of the R2 RNA template. They did not indicate whether this

A 3' junction of R2Bm with the 28S gene

R2Bm	28S gene
GCATGATGATCCGGCGATGAAAA	TAGCCAAATGCCTCG
CGTACTACTAGGCCGCTACTTTT	ATCGGTTTACGGAGC

B 3' end of R2 RNA transcript

GCAUGAUGAUCCGGCGAUGAAAA$_n$

Clone no.

1.	CGTACTACTAGGCCGCTACTTTT	ATCGGTTTACGGAGC
2.	CGTACTACTAGGCCGCTACTTTTTCTTTTTT	ATCGGTTTACGGAGC
3.	CGTACTACTAGGCCGCTACTTTTTTTTTTT	ATCGGTTTACGGAGC
4.	CGTACTACTAGGCCGCTACTTTTT	ATCGGTTTACGGAGC
5.	CGTACTACTAGGCCGCTACTTTT	ATCGGTTTACGGAGC
6.	CGTACTACTAGGCCGCTACTTT	ATCGGTTTACGGAGC
7.	CGTACTACTAGGCCGCTACTTTTTTTTCTATT	ATCGGTTTACGGAGC
8.	CGTACTACTAGGCCGCTACTTTT	ATCGGTTTACGGAGC
9.	CGTACTACTAGGCCGCTACTTTTT	ATCGGTTTACGGAGC
10.	CGTACTACTAGGCCGCTACTTTTT	ATCGGTTTACGGAGC
11.	CGTACTACTAGGCCGCTACTTTTT	ATCGGTTTACGGAGC

Figure 5. 3' Junction Sequence of PCR Amplified Products from the DNA-Primed Reverse Transcription Reactions

The products of the reverse transcription reaction with R2 RNA and the pB109 substrate (reaction conditions identical to those in Figure 4, lane 1) were PCR amplified using one primer to the cDNA strand made from the R2 RNA and a second primer to the 28S gene downstream of the insertion site. The PCR product was cloned into mp19, and individual clones were isolated and sequenced across the R2–28S gene junction.

(A) The R2–28S gene 3' junction found in all cloned copies of R2Bm (Burke et al., 1987).

(B) Shown at the top is the 3' end of the R2 RNA transcript used in the reaction. Owing to slippage of the T7 RNA polymerase, additional A-rich sequences are added to the end of the transcript (Milligan and Uhlenbeck, 1989). Below this RNA sequence are the sequences of the cDNA–28S gene junctions obtained by PCR amplification and cloning of the reaction products. The cDNA strand (coding strand) is shown with 28S gene sequences enclosed within a box.

reverse transcription occurred before or after second-strand cleavage. To resolve this issue, a time course of the integration reaction was performed as shown in Figure 6. Figure 6A presents the agarose gel of the reaction products, and Figure 6B is an autoradiograph of this gel. The reaction conditions were the same as those in Figure 4, except that the pB109 substrate DNA was initially cleaved with only EcoRI, to resolve both cleavage and integration products. Shown in Figure 6C are the three types of products expected from the reactions. Reverse transcription before the second-strand cleavage yields a labeled 4.6 kb band. Because this molecule would be Y-shaped, it is expected to migrate at a higher apparent size under the conditions of the electrophoresis used in Figure 6 (Bell and Byers, 1983). Reverse transcription after double-stranded cleavage results in a 1.9 kb labeled band and a 2.7 kb unlabeled band. Finally, cleavage in the absence of reverse transcription results in 2.7 kb and 1.0 kb unlabeled bands.

The time course in Figure 6 indicates that most reverse transcription occurred on DNA molecules before second-strand cleavage. A labeled band, migrating with an apparent length of 5.4 kb, was detected at peak levels after 40 min of incubation, when the level of the 1.9 kb band was still quite low (this is easiest to see in the autoradiograph). With additional incubation time the 5.4 kb band decreased in intensity, and the 1.9 kb band increased in intensity, indicating that the 1.9 kb band was the final reaction product. Shown in the rightmost lane of Figures 6A and 6B is the 2 hr time point for the same reaction conducted in the absence of dNTPs, so reverse transcription could not occur. The higher levels of the cleaved 2.7 kb fragment detected in the presence of dNTPs suggested that second-

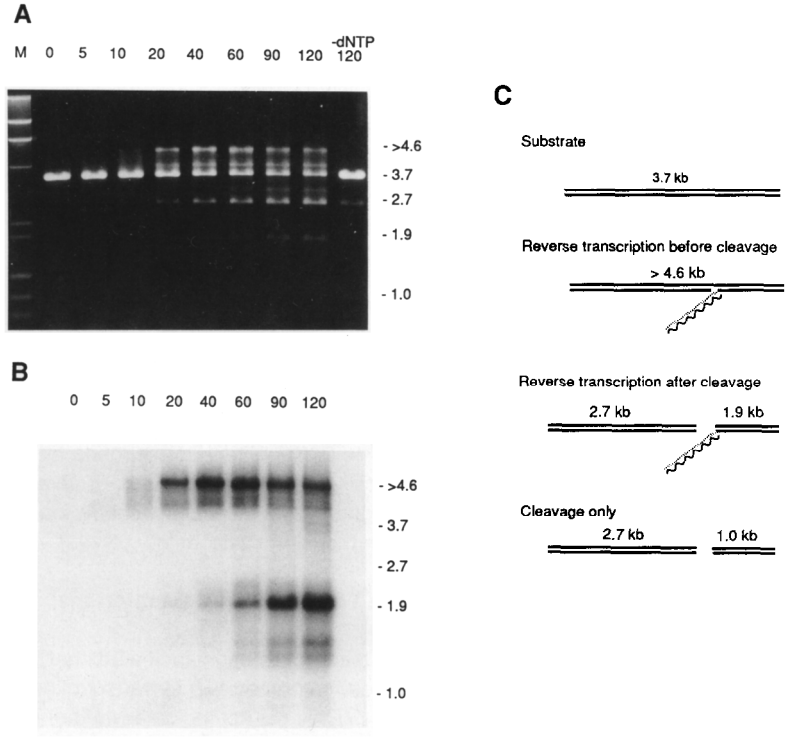

Figure 6. Time Course of the Cleavage and Reverse Transcription Reactions

Reaction conditions were the same as in Figure 4, except that the pB109 target DNA was predigested with only EcoRI (rather than EcoRI–HindIII) to enable resolution of the cleavage and insertion reactions.

(A) Ethidium bromide–stained agarose gel of the reaction products. The incubation time is indicated at the top of each lane. The reactions were digested with 10 μg of RNAase A before loading on the gel. The locations of the reaction products described in (C) are indicated at the right. The final lane corresponds to a 120 min reaction done in the absence of dNTPs to allow cleavage but no reverse transcription. M, DNA length markers as in Figure 4.

(B) Autoradiograph of the gel in (A). The nature of the labeled fragments below the >4.6 and 1.9 kb bands has not been characterized, but they are likely to represent incomplete reverse transcription.

(C) Diagrams of the various structures that are possible intermediates and end products in the reverse transcription–cleavage reaction. Solid lines, target site DNA strands; dotted line, cDNA; wavy line, RNA template.

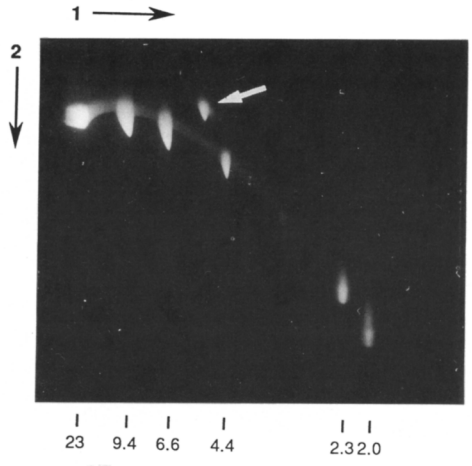

Figure 7. Two-Dimensional Gel Electrophoresis of the R2 Integration Intermediate

The 5.4 kb band used for the gel was isolated (Geneclean kit, BIO101) from a preparative 0.7% agarose gel similar to that in Figure 6 and mixed with a HindIII digest of λ DNA before loading. The two-dimensional gel was run as described by Bell and Byers (1983). The first dimension was 0.7% agarose at 0.75 V/cm for 26 hr. The second dimension was 1.5% agarose with 0.5 μg/ml ethidium bromide at 2 V/cm for 30 hr. The R2 integration intermediate is indicated by the arrow. The lengths of the linear DNA standards are shown at the bottom in kilobases.

strand cleavage was more likely to occur after reverse transcription than in its absence.

To determine if the 5.4 kb intermediate band had a conformation consistent with a branched molecule, this band was isolated from a preparative gel similar to that in Figure 6 and run on a two-dimensional gel designed to separate branched from linear DNA (Bell and Byers, 1983). The result of such an experiment is shown in Figure 7. Electrophoresis in the first dimension was conducted at low agarose concentrations and low voltage, so the migration of the DNA on the gel was primarily dependent upon its mass. Electrophoresis in the second dimension was at high agarose concentrations and high voltage, so branched molecules were preferentially slowed in the gel. The 5.4 kb intermediate band (arrow) clearly migrated at a position distinct from the diagonal arc of linear molecular weight standards, indicating that it is a branched molecule. A fraction of the 1.8–1.9 kb molecules seen in Figure 4 were also presumably branched. These branched molecules were not resolved from linear molecules, because the location of the branch was within 60 bp of one end of the molecule.

Discussion

The R2Bm Mechanism of Retrotransposition

The experiments in this report provide direct evidence for the mechanism used by a non-LTR retrotransposable element to prime reverse transcription of its RNA transcript. Because the RNA transcript from R2Bm is reverse transcribed into cDNA directly at the target site, we suggest the term RNA-mediated integration to refer to this method of retrotransposition. A summary of the critical steps involved in the reverse transcription of an R2 transcript is shown in Figure 8. The R2Bm protein recognizes sequences within 800 bp of the 3′ end of the RNA transcript. The first step in retrotransposition is the generation of a single-stranded nick at the target site. The 3′ end of the DNA strand exposed at the nick primes reverse transcription. Only R2 RNA containing the 3′ end of its presumed transcript was found capable of initiating this reaction in vitro. Vector RNA and total E. coli RNA were able to bind to the protein and serve as a cofactor for second-strand cleavage, but were unable to function as templates for reverse transcription primed at the cleavage site. The R2 RNA template used in these experiments ended with a short, variable-in-length A-rich sequence. About one-third of the resulting integrated sequences contained the identical R2Bm–28S gene 3′ junction sequence found in all genomic copies. Another third of the sequences contained an additional A at this junction, while the final third contained additional A-rich sequences at this junction.

We have previously suggested (Eickbush and Robins, 1985) that the four A residues at the 3′ end of all integrated R2Bm elements are part of a poly(A) tail, because R2 elements in D. melanogaster contain a true poly(A) tail from 13 to 22 nt in length (Dawid and Rebbert, 1981; Roiha et

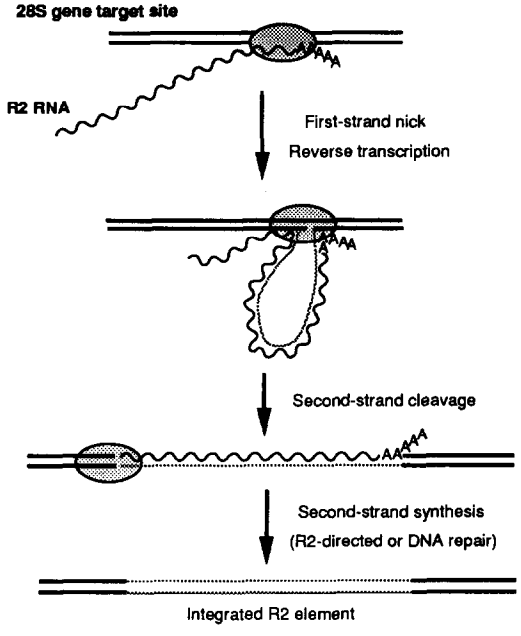

Figure 8. Model for R2Bm Retrotransposition

The 120 kd R2 protein is shown to associate with sequences near or at the 3′ end of the R2 transcript, including part of a poly(A) tail. The R2 protein first makes a specific nick in the lower (coding) strand of the 28S gene target site and uses the exposed 3′ end to prime reverse transcription. The R2 protein is suggested to remain bound to the 28S gene, involving sequences mostly upstream of the target site. After reverse transcription, cleavage of the upper DNA strand occurs. In our in vitro reaction this second-strand cleavage does not prime second-strand synthesis. Second-strand synthesis could be brought about by R2-encoded enzymatic activities that are not present in our protein expressed in E. coli, or may occur by cellular DNA repair mechanisms. Solid lines, original DNA strands; dotted line, cDNA; wavy line, RNA template.

al., 1981). The observation that all genomic R2Bm copies contain just four A residues could be explained if the accuracy of reverse transcription initiation is higher in vivo than the in vitro results reported here. It is also possible that in B. mori the R2 transcript is simply not polyadenylated.

The locations of the DNA-binding and RNA-binding domains within the 120 kd protein are currently unknown. The presence of two Cys-His motifs flanking the reverse transcriptase domain (Figure 1B) suggests a model that can be readily tested by in vitro mutagenesis. Located near the amino terminus of the protein is the motif Trp-X-Cys-X_2–Cys-X_3–Phe-X_5–Leu-X_2–His-X_4–His (X = any amino acid), which is identical to the consensus TFIIIA zinc finger motif known to bind DNA (reviewed in Berg, 1990). Located near the carboxyl terminal end of the protein is the motif Cys-X_2–Cys-X_7–His-X_4–Cys, which corresponds to a CCHC box known to be associated with the binding of single-stranded nucleic acids (Berg, 1990). These motifs suggest that the endonuclease domain is located near the amino terminus of the protein, while recognition of the RNA template for proper initiation of reverse transcription is located near the carboxyl terminus.

The time course of the integration reaction indicated that second-strand cleavage of the target site occurs after the reverse transcription step. Deletion studies of the 28S target site have indicated that most of the DNA sequences required for double-stranded cleavage are located upstream of the insertion site (Xiong and Eickbush, 1988b). We suggest (Figure 8) that the endonuclease remains bound to this upstream site throughout the reverse transcription process, with the RNA–cDNA heteroduplex able to spool through the R2 protein. The R2 protein component may be multimeric, because near the physiological ionic strength of the insertion reaction, the 120 kd R2 protein sediments at over 15S (D. D. L., unpublished data). It is possible that one subunit of the complex is responsible for the endonuclease activity and a separate subunit is responsible for the reverse transcriptase activity.

A major unresolved issue in the retrotransposition of R2 elements is what occurs after second-strand cleavage. We have used PCR primers that can bind within the 28S gene upstream of the target site and within the R2 sequences to amplify any continuous DNA strand at the 5′ end of the element (D. D. L., unpublished data). No PCR products were obtained, indicating that, under our in vitro conditions, cleavage of the second strand was unable to prime DNA synthesis even at extremely low levels. It is possible that specific sequences are required at the 5′ end of the RNA to prime this reaction. An alternative explanation for the failure of second-strand synthesis is that there is no RNAase H activity associated with the R2 ORF, providing no mechanism to remove the RNA template to allow synthesis of the second strand. No RNAase H domain has been found in sequence comparison of the R2 ORF from different insects (Burke et al., 1993), and no RNAase H activity has been detected in the purified 120 kd protein from R2Bm (J. L. J., D. D. L., and T. H. E., unpublished data).

It is possible that removal of the RNA template from the cDNA and even second-strand synthesis itself are controlled by cellular DNA repair mechanisms. Completion of R2Bm insertion by DNA repair mechanisms would explain why precise 3′ ends are associated with all integrated R2 elements while the 5′ ends are frequently truncated. Initiation of first-strand synthesis requires sequence-specific recognition of both the 3′ end of the RNA and the DNA target. However, once cDNA synthesis is initiated, if this process is aborted or if the RNA itself is not full length, the remaining insertion steps could still occur because they involve nonspecific repair functions. Even if the R2 protein is capable of second-strand synthesis in vivo, our model implies that the final step of the integration is probably accomplished by DNA repair. The 2 bp at the staggered cut made by the endonuclease are not complementary to the integrated sequences and thus must be removed.

The Use of RNA-Mediated Integration by LINEs and Short Interspersed Nucleotide Elements

Is RNA-mediated integration used by the many types of non-LTR retrotransposable elements (LINEs) widely distributed in animals, plants, fungi, and protozoans? The critical steps of the mechanism shown in Figure 8 are similar to models suggested for the retrotransposition of cin4 of Z. mays (Schwarz-Sommer et al., 1987), the I factor of D. melanogaster (Finnegan, 1989; Bucheton, 1990), and the L1 element of mammals (Hutchison et al., 1989). Several arguments favor the prediction that an RNA-mediated mechanism of integration similar to that of R2 is used by other non-LTR retrotransposable elements. First, most non-LTR elements have been shown to contain precise 3′ junctions with chromosomal DNA, usually involving a short A-rich tail. Second, most non-LTR elements are subject to the 5′ truncations that are readily explained by this model. Finally, the reverse transcriptase domains of all non-LTR retrotransposable elements are more closely related to each other than to any other retroelement, indicating that they represent a distinct family of elements with a common origin (Xiong and Eickbush, 1990).

On the other hand, the specific details of the retrotransposition of non-LTR elements are likely to be more variable than those of the LTR retrotransposons. R2 integrations result in a 2 bp deletion, while most non-LTR elements generate a target site duplication. In contrast with the LTR retrotransposons, which have highly uniform deduced protein domains, the ORF structures of different non-LTR retrotransposons appear to be quite variable. For example, the CCHC putative RNA-binding domain, located near the carboxyl terminus of the R2 protein, is found in only one-half of all non-LTR elements, while the amino-terminal TFIIIA zinc finger domain is not found in any other non-LTR element (Jakubczak et al., 1990).

Finally, short interspersed nucleotide elements (SINEs) are poly(A) tailed, highly abundant, repetitive sequences that move via an RNA intermediate (Rogers, 1985; Deininger, 1989). These elements have been studied most extensively in mammalian species but are now known to be present in other vertebrates (Okada, 1991), sea urchins (Nisson et al., 1988), and insects (Bradfield et al., 1985; Adams et al., 1986). SINEs do not encode their own reverse transcriptases, and it has been suggested that retro-

viruses or LTR retrotransposable elements could provide their integration machinery (Rogers, 1985; Deininger, 1989). However, with an LTR-based integration machinery, there is no obvious mechanism whereby first-strand and second-strand synthesis could be primed on the SINE RNA templates. Also, there is no mechanism by which this DNA copy, if formed, could be integrated into the genome, because it would not contain the LTR sequences recognized by the integrases of these elements.

Weiner et al. (1986) originally grouped SINEs, processed pseudogenes, and non-LTR retrotransposable elements as involving a similar, nonviral mechanism of retrotransposition. It can now be suggested that the RNA-mediated mechanism of retrotransposition described in this paper may also be the enzymatic machinery responsible for the retrotransposition of SINEs. The RNA-mediated mechanism requires no conventional sequence complementarity of the primer to the template RNA for the priming of first-strand synthesis. It is the reverse transcriptase complex, binding both the RNA and the DNA, that holds the primer and template together. The only requirement of the RNA template is that its 3′ end be able to mimic the correct sequences at the 3′ end of a non-LTR retrotransposable element, so it can be bound by the reverse transcriptase of that element.

Experimental Procedures

Strains and Constructs for In Vitro Expression

All plasmid constructions were maintained in E. coli strain JM101 grown at 37°C in L broth medium supplemented with 75 μg/ml ampicillin. All plasmid DNA was prepared by alkaline lysis and cesium chloride–ethidium bromide gradient centrifugation (Maniatis et al., 1982). The in vitro expression construct, pR260, was derived from construct pR250 (Xiong and Eickbush, 1988b). A 3.5 kb SmaI fragment of pR250 from 18 bp upstream of the first methionine codon to the 3′ untranslated region was subcloned into pUC18 in-frame with the *lacZ* gene. The 28S gene target site construct was derived from a 1.9 kb BamHI–XbaI subclone of lambda clone B108 (Eickbush and Robins, 1985). The 1055 bp HincII–SpeI fragment of this plasmid was subcloned into the HincII–XbaI sites of pUC19 to generate plasmid pB109.

Purification of the R2 Protein

E. coli strain JM109/pR260 was grown at 37°C in L broth until the optical density at 595 nm was 0.5–0.6. Isopropyl-β-D-thiogalactoside was then added to a final concentration of 0.2 mM, and the cultures were further incubated for 1 hr. Cells from 0.8 liter of culture were harvested by centrifugation, washed in cold 50 mM Tris–HCl (pH 8.0), and collected by centrifugation. All subsequent procedures were conducted at 0–4°C. The cell pellets were resuspended in 3.4 ml of buffer A (0.1 mM Tris–HCl [pH 7.5], 5 mM EDTA, 50% glycerol) and incubated for 30 min in 5 mM dithiothreitol (DTT), 2 mM benzamidine–HCl, and 2 mg/ml lysozyme. Sixteen milliliters of buffer B (0.1 M Tris–HCl [pH 7.5], 1 M NaCl, 5 mM DTT, 0.2% Triton X-100, 10 mM $MgCl_2$, 2 mM benzamidine) was then added, followed by an additional 30 min incubation. The lysate was centrifuged in an SW50.1 rotor at 33,000 rpm for 20 hr. The upper 1 ml of the supernatant contained no sequence-specific endonuclease activity and was discarded. The remaining 4 ml of supernatant from each tube was decanted and diluted with H_2O to lower the NaCl concentration to 0.4 M. The diluted crude extract was loaded onto a 5 ml Q Sepharose fast-flow column (Pharmacia) equilibrated in 0.4 M NaCl–buffer C (25 mM Tris–HCl [pH 7.5], 2 mM DTT). The column was washed with 20 ml of the 0.4 M NaCl–buffer C, and the R2 protein was eluted with 0.6 M NaCl–buffer C. Fractions containing the R2 endonuclease were pooled, dialyzed against 0.2 M NaCl–buffer D (25 mM Tris–HCl [pH 7.5], 2 mM DTT, 10% glycerol), and applied to a 1 ml DNA–cellulose column (Pharmacia) equilibrated with 0.2 M NaCl–buffer D. The column was washed with 9 ml of 0.4 M NaCl–buffer D and eluted with 0.8 M NaCl–buffer D. The endonuclease activity eluted from the DNA–cellulose column was unstable (losing 50% of its activity within 12 hr); therefore, all characterization of the enzymatic activity of the R2 protein was performed immediately after the DNA–cellulose purification step. One unit of R2 endonuclease activity was defined as the enzyme required to generate 1 μg of cleaved pB109 DNA product in 1 hr at 37°C. Approximately 80 U of activity was typically recovered from the DNA–cellulose column, representing a 50% recovery of the activity present in the crude supernatant.

Assays for Endonuclease Activities

R2 endonuclease activity of the crude extract was assayed by incubating 0.5 μg of supercoiled pB109 plasmid DNA with 30 μl of the extract in 50 mM Tris–HCl (pH 8.0), 150 mM NaCl, and 1 mM DTT in a total reaction volume of 300 μl for 45 min at 37°C. After incubation the plasmid DNA was purified by proteinase K digestion, phenol–chloroform extraction, and ethanol precipitation. The DNA was resuspended in 15 μl of EcoRI digestion buffer, digested with EcoRI, and analyzed by agarose gel electrophoresis. Assay of R2 endonuclease activity in fractions from the Q Sepharose and DNA–cellulose columns was conducted with linear DNA. pB109 DNA (0.5 μg), previously digested with EcoRI, was incubated with 5 μl aliquots of each fraction in 50 mM Tris–HCl (pH 8.0), 10 mM $MgCl_2$, and 1 mM DTT in a total reaction volume of 20 μl at 37°C for 30 min. Total E. coli RNA, added to some reactions, was obtained by proteinase K–phenol extraction of the flowthrough fractions from the DNA–cellulose column. To assay for double-stranded cleavage, the reactions were stopped by the addition of 5 × loading buffer (0.02% bromophenol blue, 5% Sarkosyl, 0.1 M EDTA, 50% glycerol) and analyzed by agarose gel electrophoresis. To assay for nicking activity on linear substrates, the reactions were precipitated with ethanol and resuspended in 10 μl of 50 mM NaOH, 1 mM EDTA. The samples were mixed with 2 μl of formamide loading dye (0.3 N NaOH, 6 mM EDTA, 18% Ficoll, 0.15% bromophenol blue, 0.25% xylene cyanole FF) and analyzed on 1% agarose gels in alkaline buffer (Maniatis et al., 1982).

Quantitation of nicking and cleavage activities was carried out by densitometry. After electrophoresis, gels were stained in a solution of 0.1 μg/ml ethidium bromide in H_2O at room temperature for 30 min and destained in H_2O for another 30 min. Ethidium bromide fluorescence at 254 nm was photographed on Polaroid 55 positive/negative films. Relative amounts of DNA in the different bands of each sample were determined by scanning this film using an LKB Instruments UltraScan densitometer.

Determination of the Nick and Cleavage Sites

The nicked site was determined using the "termination" method described by Brown and Smith (1980) with the following modifications: A HindIII–BamHI fragment containing the R2 target site from plasmid pR2603 (Xiong and Eickbush, 1988b) was subcloned into mp18 in both orientations, and single-stranded DNA templates were prepared. For each orientation the −40 sequencing primer (US Biochemicals) was annealed to the template and divided into five samples, four of which were used for routine sequencing (Sanger et al., 1977). The fifth sample was extended in the absence of ddNTP and the presence of excess nucleotides (15 μM) to drive the primer extension reaction to completion. The products from the extension reaction were recovered by ethanol precipitation and incubated with 5 μl of Q Sepharose–purified R2 protein in a 50 μl reaction containing 50 mM Tris–HCl (pH 8.0), 100 mM NaCl, 2 mM $MgCl_2$, and 1 mM DTT at 37°C for 1.5 hr. The digested DNA was recovered by ethanol precipitation, denatured by boiling in 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanole FF, and analyzed by electrophoresis next to the sequence ladder on a standard 6% acrylamide–urea sequencing gel.

Reverse Transcriptase Assays

The RNA-dependent DNA polymerase activity was assayed in a 20 μl reaction as follows: 100 ng of the DNA–cellulose-purified protein was incubated in 50 mM Tris–HCl (pH 8.0), 6 mM $MgCl_2$, 200 mM NaCl, 1 mM DTT, containing 0.3 μg of substrate poly(C)–oligo$(dG)_{12-18}$ (Pharmacia), 2.6 μM [α-^{32}P]dGTP (40 Ci/mmol) at 37°C for 15 min. Ten microliter aliquots of each reaction were then spotted onto DE81 chromatography paper and allowed to air dry. The filters were washed three

times in 2 x SSC for 15 min each, briefly washed in 70% ethanol, and allowed to air dry. Radioactive counts per minute were determined in a liquid scintillation counter by Cerenkov counting.

Construction of Vectors and Synthesis of Specific RNA

Clone HR4 was constructed by PCR amplification of the 3′ end of R2Bm from B131 (Burke et al., 1987), using one oligonucleotide (5′-GCTGGAGAATTCCGCAG-3′) to prime synthesis within the R2 element (starting at position 2186) and the second primer (5′-CGAGGAATTCGAATATTTTCATCGCCGG-3′) to prime synthesis at the R2Bm–28S gene 3′ junction. The primer complementary to the 3′ junction contained an engineered XmnI site to enable restriction at the precise 3′ junction (GAAAATAGCCAAATGC in the element became GAAAATATTCGAATTC in the product), as well as an EcoRI site for cloning. The 0.8 kb 3′ terminal HincII–EcoRI fragment of the resultant PCR product was subcloned into HincII–EcoRI-digested pBSIISK– (Stratagene).

For in vitro transcription of R2-specific RNA, plasmid HR4 was linearized with XmnI. Vector RNA was synthesized from pBSIISK– linearized with ApaLI. After digestion the plasmid DNA was treated with proteinase K, extracted with phenol–chloroform, and ethanol precipitated. About 1 μg of template DNA was used in the in vitro transcription reaction in 40 μl of transcription buffer (Promega), 1 mM DTT, 0.4 mM NTP, 5 U of RNasin (Sigma), and 10 U of T7 RNA polymerase (Pharmacia) at 37°C for 1.5 hr. The reaction was then incubated with 10 U of DNAase I (Pharmacia) for 45 min. The synthesized RNA was recovered after phenol extraction by ethanol precipitation and resuspended in 10 μl of 5 mM Tris–HCl (pH 8), 1 mM EDTA with 5 U of RNasin.

DNA-Primed Reverse Transcription Assays

Five microliter aliquots from the DNA–cellulose column (100 ng of protein) in 0.8 M NaCl were diluted to 0.2 M NaCl in a buffer containing 0.4 μg of pB109 predigested with HindIII and EcoRI, 0.2 μg of RNA, 10 mM $MgCl_2$, 50 mM Tris (pH 8.0), 1 mM DTT, dNTPs (5 μM each), 5 U of RNasin, 125 nM ^{35}S-dATP. The total reaction volume was 20 μl. Incubations were performed for 1.5 hr at 37°C, stopped by the addition of 5 x Sarkosyl loading buffer, and analyzed by electrophoresis in 1% agarose gel. After photography, the gel was dried, and autoradiography was performed.

Determination of the 3′ Junction of Inserted R2 and 28S DNA

After the DNA-primed reverse transcription assay, the products were amplified by PCR, using one oligonucleotide (5′-TTGGTTGAGCCTTGCACAG-3′) to prime synthesis within the R2 sequence 256 nt from the 3′ end and a second oligonucleotide (5′-CGCGCATGAATGGATTAACG-3′) to prime synthesis within the 28S gene 40 nt downstream of the R2 insertion site. The PCR products were purified (Geneclean, BIO101), denatured by alkali, and sequenced on both strands by the dideoxy chain termination method, primed by the oligonucleotides used for the PCR amplifications (Sanger et al., 1977). The PCR products were also cloned into ddT-tailed mp18 vectors (Holton and Graham, 1991) for single-stranded sequencing of individual products.

Acknowledgments

We thank D. Hinkle for invaluable advice throughout the course of this work and for comments on the manuscript. We thank W. Burke for help in the sequencing of PCR products and Y. Xiong for help in the initial attempts to monitor the activity of the R2 protein in vitro. This study was supported by National Institutes of Health Grant GM42790 and American Cancer Society Grant NP-691 to T. H. E.

Received October 28, 1992; revised December 23, 1992.

References

Adams, D. S., Eickbush, T. H., Herrera, R. J., and Lizardi, P. M. (1986). A highly reiterated family of transcribed oligo(A)-terminated, interspersed DNA elements in the genome of *Bombyx mori*. J. Mol. Biol. *187*, 465–478.

Bell, L., and Byers, B. (1983). Separation of branched from linear DNA by two-dimensional gel electrophoresis. Anal. Biochem. *130*, 527–535.

Berg, J. M. (1990). Zinc fingers and other metal-binding domains. J. Biol. Chem. *265*, 6513–6516.

Boeke, J. D. (1989). Transposable elements in *Saccharomyces cerevisiae*. In Mobile DNA, D. H. Berg and M. M. Howe, eds. (Washington, DC: American Society of Microbiology), pp. 335–374.

Boeke, J. D., and Corces, V. G. (1989). Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. *43*, 403–434.

Bradfield, J. Y., Locke, J., and Wyatt, G. R. (1985). An ubiquitous interspersed DNA sequence family in an insect. DNA *4*, 357–363.

Brown, N. L., and Smith, M. (1980). A general method for defining restriction enzyme cleavage and recognition sites. Meth. Enzymol. *65*, 391–404.

Brown, P. O., Bowerman, B., Varmus, H. E., and Bishop, J. M. (1987). Correct integration of retroviral DNA in vitro. Cell *49*, 347–356.

Bucheton, A. (1990). I transposable elements and I–R hybrid dysgenesis in *Drosophila*. Trends Genet. *6*, 16–21.

Burke, W. D., Calalang, C. C., and Eickbush, T. H. (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., Eickbush, D. G., Xiong, Y., Jakubczak, J., and Eickbush, T. H. (1993). Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. Mol. Biol. Evol. *10*, in press.

Bushman, F. D., Fujiwara, T., and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. Science *249*, 1555–1558.

Chaboissier, M. C., Busseau, I., Prosser, J., Finnegan, D. J., and Bucheton, A. (1990). Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in *Drosophila melanogaster*. EMBO J. *9*, 3557–3563.

Dawid, I. B., and Rebbert, M. L. (1981). Nucleotide sequence at the boundaries between gene and insertion regions in the rDNA of *Drosophila melanogaster*. Nucl. Acids Res. *9*, 5011–5020.

Deininger, P. (1989). SINEs: short interspersed repeated DNA elements in higher eukaryotes. In Mobile DNA, D. H. Berg and M. M. Howe, eds. (Washington, DC: American Society of Microbiology), pp. 619–636.

Eichinger, D. J., and Boeke, J. D. (1990). A specific terminal structure is required for Ty1 transposition. Genes Dev. *4*, 324–330.

Eickbush, T. H. (1992). Transposing without ends: the non-LTR retrotransposable elements. New Biol. *4*, 430–440.

Eickbush, T. H., and Robins, B. (1985). *Bombyx mori* 28S genes contain insertion elements similar to the type I and type II elements of *Drosophila melanogaster*. EMBO J. *4*, 2281–2285.

Evans, J. P., and Palmiter, R. D. (1991). Retrotransposition of a mouse L1 element. Proc. Natl. Acad. Sci. USA *88*, 8792–8795.

Fawcett, D. H., Lister, C. K., Kellett, E., and Finnegan, D. J. (1986). Transposable elements controlling I–R hybrid dysgenesis in D. melanogaster are similar to mammalian LINEs. Cell *47*, 1007–1015.

Finnegan, D. J. (1989). The I factor and I–R hybrid dysgenesis in *Drosophila melanogaster*. In Mobile DNA, D. H. Berg and M. M. Howe, eds. (Washington, DC: American Society of Microbiology), pp. 503–517.

Gabriel, A., and Boeke, J. D. (1991). Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. Proc. Natl. Acad. Sci. USA *88*, 9794–9798.

Hansen, L. J., Chalker, D. L., and Sandmeyer, S. B. (1988). Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol. Cell. Biol. *8*, 5245–5256.

Holton, T. A., and Graham, M. W. (1991). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucl. Acids Res. *19*, 1156.

Hutchison, C. A., Hardies, S. C., Loeb, D. D., Shehee, W. R., and Edgell, M. H. (1989). LINEs and related retroposons: long interspersed repeated sequences in the eukaryotic genome. In Mobile DNA, D. H. Berg and M. M. Howe, eds. (Washington, DC: American Society of Microbiology), pp. 593–617.

Ivanov, V. A., Melnikov, A. A., Siunov, A. V., Fodor, I. I., and Illyin, Y. V. (1991). Authentic reverse transcriptase is coded by jockey, a mobile *Drosophila* element related to mammalian LINEs. EMBO J. *10*, 2489–2495.

Jakubczak, J. L., Xiong, Y., and Eickbush, T. H. (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.

Jakubczak, J. L., Burke, W. D., and Eickbush, T. H. (1991). Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc. Natl. Acad. Sci. USA *88*, 3295–3299.

Jensen, S., and Heidmann, T. (1991). An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. EMBO J. *10*, 1927–1937.

Katz, R. A., Merkel, G., Kulkosky, J., Leis, J., and Skalka, A. M. (1990). The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. Cell *63*, 87–95.

Kimmel, B. E., ole-MoiYoi, O. K., and Young, J. R. (1987). Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half a smaller mobile element at their end, has homology with mammalian LINEs. Mol. Cell. Biol. *7*, 1465–1475.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Marlor, R., Parkhurst, S., and Corces, V. (1986). The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. *6*, 1129–1134.

Martin, S. L. (1991). LINEs. Curr. Opin. Genet. Dev. *1*, 505–508.

Mathias, S. L., Scott, A. F., Kazazian, H. H., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. Science *254*, 1808–1810.

Milligan, J. F., and Uhlenbeck, O. C. (1989). Synthesis of small RNAs using T7 RNA polymerase. Meth. Enzymol. *180*, 51–62.

Minchiotti, G., and Di Nocera, P. P. (1991). Convergent transcription initiates from oppositely oriented promoters within the 5′ end region of *Drosophila melanogaster* F elements. Mol. Cell. Biol. *11*, 5171–5180.

Mizrokhi, L. J., Georgieva, S. G., and Ilyin, Y. V. (1988). *jockey*, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell *54*, 685–691.

Mount, S. M., and Rubin, G. M. (1985). Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. *5*, 1630–1638.

Nisson, P. E., Hickey, R. J., Boshar, M. F., and Crain, W. R. (1988). Identification of a repeated sequence in the genome of the sea urchin which is transcribed by RNA polymerase III and contains the features of a retroposon. Nucl. Acids Res. *16*, 1431–1452.

Okada, N. (1991). SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol. Evol. *6*, 358–361.

Pelisson, A., Finnegan, D. J., and Bucheton, A. (1991). Evidence for retrotransposition of the I factor, a LINE element of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA *88*, 4907–4910.

Rogers, J. (1985). The origin and evolution of retroposons. Int. Rev. Cytol. *93*, 187–279.

Roiha, H., Miller, J. R., Woods, L. C., and Glover, D. M. (1981). Arrangements and rearrangements of sequence flanking the two types of rDNA insertion in *D. melanogaster*. Nature *290*, 749–753.

Sandmeyer, S. B., Hansen, L. J., and Chalker, D. L. (1990). Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. *24*, 491–518.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schwarz-Sommer, Z., Leclercq, L., Gobel, E., and Saedler, H. (1987). Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of non-viral retrotransposons. EMBO J. *6*, 3873–3880.

Singer, M. F., and Skowronski, J. (1985). Making sense out of LINEs: long interspersed repeat sequences in mammalian genomes. Trends Biochem. Sci. *10*, 119–122.

Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol. Cell. Biol. *10*, 6718–6729.

Temin, H. M. (1989). Retrons in bacteria. Nature *339*, 254–255.

Varmus, H., and Brown, P. (1989). Retroviruses. In Mobile DNA, D. H. Berg and M. M. Howe, eds. (Washington, DC: American Society of Microbiology), pp. 53–108.

Voytas, D. F., and Ausubel, F. M. (1988). A copia-like transposable element family in *Arabidopsis thaliana*. Nature *336*, 242–244.

Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. *55*, 631–661.

Xiong, Y., and Eickbush, T. H. (1988a). The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. *8*, 114–123.

Xiong, Y., and Eickbush, T. H. (1988b). Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell *55*, 235–246.

Xiong, Y., and Eickbush, T. H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. *9*, 3353–3362.