

## Retrotransposon R1Bm endonuclease cleaves the target sequence

QINGHUA FENG<sup>\*†</sup>, GERALD SCHUMANN<sup>\*</sup>, AND JEF D. BOEKE<sup>‡</sup>

Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore MD 21205

Communicated by Thomas J. Kelly, The Johns Hopkins University, Baltimore, MD, December 23, 1997 (received for review July 4, 1997)

**ABSTRACT** The R1Bm element, found in the silkworm *Bombyx mori*, is a member of a group of widely distributed retrotransposons that lack long terminal repeats. Some of these elements are highly sequence-specific and others, like the human L1 sequence, are less so. The majority of R1Bm elements are associated with ribosomal DNA (rDNA). R1Bm inserts into 28S rDNA at a specific sequence; after insertion it is flanked by a specific 14-bp target site duplication of the 28S rDNA. The basis for this sequence specificity is unknown. We show that R1Bm encodes an enzyme related to the endonuclease found in the human L1 retrotransposon and also to the apurinic/aprimidinic endonucleases. We expressed and purified the enzyme from bacteria and showed that it cleaves *in vitro* precisely at the positions in rDNA corresponding to the boundaries of the 14-bp target site duplication. We conclude that the function of the retrotransposon endonucleases is to define and cleave target site DNA.

Retrotransposons that lack long terminal repeats are very diverse in structure and can insert into a wide variety of different types of DNA targets. Some of these elements, such as the human L1 element, insert into a relatively wide array of targets distributed on all host chromosomes. In contrast, related retroelements from insects and trypanosomes integrate at very specific sequences. The basis for this extreme specificity is in most cases unknown. We recently showed that all of the elements that lack sequence specificity as well as a subset of those that are sequence-specific encode an endonuclease (EN) domain, usually at the N terminus of the second ORF (1). ORF2 also encodes reverse transcriptase (RT), and an attractive model is that the bifunctional ORF2 protein nicks target DNA and then primes reverse transcription of transposon RNA from the target DNA nick. The EN domain resembles apurinic/aprimidinic (AP) ENs, which are important for DNA repair (2), but the human L1 EN does not cleave at AP sites.

The silkworm *Bombyx mori* genome contains two sequence-specific retroelements, each of which is inserted at a specific position in 28S rDNA; R1Bm encodes an EN domain and R2Bm does not. The R2Bm element has a single ORF that nevertheless encodes both an EN activity and an RT activity and has served as the best model system for the target-primed reverse transcription (TPRT) model of transposition in these elements (3). In contrast, R1Bm has two ORFs. ORF1 encodes a protein with certain similarities to retroviral Gag proteins, and ORF2 encodes a protein with homology to both RT and an EN (1, 4) (Fig. 1). The retrotransposition of R1Bm has not been studied. The basis for its sequence specificity is unknown; in principle, it could be either specified by the EN itself, by the RT, or by host factors. We wished to determine the function of the R1Bm EN and specifically to test the hypothesis that the R1Bm EN specified target sequence cleavage.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/952083-6\$2.00/0  
PNAS is available online at <http://www.pnas.org>.

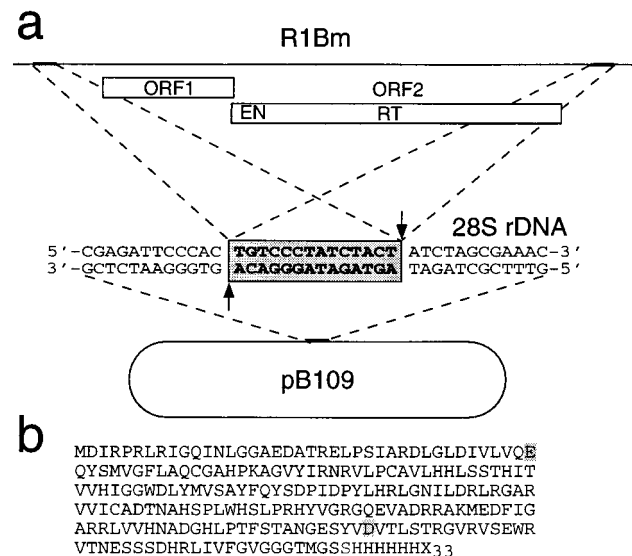


FIG. 1. Structure of R1Bm element and its target site. (a) The genome organization of R1Bm is indicated; the ORFs are shown as boxes, and the positions of the EN and RT domains in ORF2 are indicated. tsd is shown as bold lines (top) or nucleotide sequence (shaded box, middle) in target plasmid pB109. (b) The sequence of the R1Bm EN domain expressed is indicated, with the putative active site residues mutagenized to alanine codons shaded. The expressed protein contained 43 amino acids derived from the pET15b vector at the C terminus, including 6 His residues (the remainder are designated X<sub>33</sub>).

The R1 elements are widely distributed among insect orders and are inserted in precisely the same site in the rDNA of these diverse species (5). Remarkably, in some strains of *Drosophila*, these insertions interrupt as many as 50–70% of the copies of rDNA (6, 7), but insertion into other genomic sites can also occur (8, 9). However, the basis for the exquisite target specificity of this element was unknown until now.

We expressed the R1 EN domain in bacteria, purified the protein, and showed that it encodes a sequence-specific EN. It specifically cleaves *Bombyx* rDNA at both boundaries of the 14-bp target site (Fig. 1), indicating that this R1Bm EN defines and cleaves the DNA target for R1Bm. Because R1Bm ORF2 also encodes an RT activity, we propose a TPRT model to explain the R1Bm retrotransposition mechanism.

The existence of a conserved AP EN-like domain in diverse retrotransposons that lack long terminal repeats raised questions about its function. AP ENs have three biochemical activities: endonucleolytic cleavage, RNase H, and 3' → 5'-exonuclease. Any related activities in the transposon en-

Abbreviations: EN, endonuclease; RT, reverse transcriptase; AP, apurinic/aprimidinic; TPRT, target-primed reverse transcription; tsd, target site duplication.

\*Q.F. and G.S. contributed equally to this work.

<sup>†</sup>Present address: Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142.

<sup>‡</sup>To whom reprint requests should be addressed. e-mail: [jef.boeke@qmail.bs.jhu.edu](mailto:jef.boeke@qmail.bs.jhu.edu).

zymes could potentially play a role in retrotransposition. The endonucleolytic cleavage activity suggests a role in target site definition and cleavage, RNase H activity could be required for degradation of an RNA/DNA hybrid, and exonuclease could play a role in proofreading. We have shown that the EN encoded by the human L1 retrotransposon is not an APEN but rather a simple nicking enzyme; we found no evidence for RNase H or 3' → 5'-exonuclease activity. We used a functional assay for L1 retrotransposition (10) to demonstrate that the L1 EN was required for retrotransposition, suggesting an essential function for L1 EN in retrotransposition and arguing against a proofreading function (1). *In vitro* assays failed to provide any evidence for an RNase H activity in L1 EN (Q.F. and J.D.B., data not shown). The cleavage specificity of the L1 EN is consistent with target site definition and cleavage, but as L1 can insert at thousands of different genomic sites, we have no direct positive evidence for the function of the EN domain in retrotransposition. To provide such evidence, we studied the R1Bm element, which inserts at a specific sequence. If the role of the EN domain is to recognize and cleave the target DNA, then R1Bm EN should cleave specifically at the boundaries of the R1Bm target site duplication (tsd). Also, the TPRT model of Luan *et al.* (3) predicts that the bottom strand of the target site should be cleaved before the top strand as bottom strand cleavage generates the primer for reverse transcription of the RNA. We show here that R1Bm EN protein indeed recognizes and cleaves the target sequence *in vitro*.

## MATERIALS AND METHODS

**Plasmids, Protein Expression, and Purification.** The R1Bm EN domain was amplified by PCR with primers JB1158 (5'-TACCATGGATATTAGCCCCGAC) and JB1159 (5'-GCCCATGGTACCGCCCCCACCCC) and p78.Xho-1.9kb (4) as template. Primers contain *Nco*I sites (underlined) at their 5'-ends. The resulting 670-bp fragment was cloned into pCRII (Invitrogen) and confirmed by DNA sequencing to contain no unwanted mutations. pGS405 was constructed by inserting the 670-bp *Nco*I fragment from this construct into the *Nco*I site of expression vector pET-15b (Novagen). Point mutants were created by oligonucleotide-directed mutagenesis with a Quickchange site-directed mutagenesis kit (Stratagene). For E40A, PCR was performed with primers JB1476 (5'-GTTCTTGTACAGGCCCAATATTCATG-3') and JB1477 (5'-CATGGAATATTGGCCTGTACAAGAAC-3') and *Pfu* Polymerase on a pGS405 template. The PCR product was digested with *Dpn*I and transformed into *Escherichia coli*. The mutant construct pQF371 was confirmed by DNA sequencing. Similarly, the D186A mutant was made with primers JB1478 (5'-GAATCTTATGTGCTGTCACGCTGTCT-3') and JB1479 (5'-AGACAGCGTGACAGCGA-CATAAGATTC-3') to generate pQF372. The rDNA target plasmid, pB109, kindly provided by T. Eickbush, consists of a *Hinc*II-*Spe*I fragment (1055 bp) of *B. mori* cloned into the *Hinc*II-*Xba*I site of pUC19 (3).

Proteins were expressed and purified exactly as previously described (1) except that the proteins were eluted with 0.25 ml of washing buffer containing 150 or 300 mM imidazole. Most of the protein eluted in the 300 mM imidazole fraction. Both fractions were pooled, dialyzed, and concentrated against storage buffer (50 mM Tris·HCl, pH 7.6/300 mM NaCl/10% glycerol/10 mM 2-mercaptoethanol). Protein aliquots of 5–10 μg in storage buffer at 0.2–0.5 μg/μl were stored at –70°C. Freeze-thaw had no apparent effect on protein activity.

**Activity Assay.** Optimal salt, divalent cation, temperature, and pH conditions were defined for the R1Bm EN activity based on its ability to nick pB109 (data not shown). The EN reaction mix contained 50 mM Pipes·HCl at pH 6.0, 30 mM NaCl, 1 mM CoCl<sub>2</sub>, 0.2 μg of supercoiled plasmid DNA, and 0.5 μg of purified protein in a total volume of 25 μl. MgCl<sub>2</sub> was

later shown to substitute for CoCl<sub>2</sub>. Incubation was at 25°C for 1 h. The reaction was stopped by adding EDTA to a final concentration of 25 mM. Half of the reaction mix was loaded on a 1% agarose gel in TTE (Tris-taurine EDTA) buffer containing 0.5 μg/ml ethidium bromide.

**Cleavage Site Mapping.** End-labeled DNA molecules containing the R1Bm target site were created by PCR with pB109 as a template and a combination of one kinased primer and one unlabeled primer. The sequences of the primers used were: JB1291, 5'-TCCTTACAATGCCAGACTAG-3'; JB1296, 5'-CTTAAGGTAGCCAAATGC-3'; JB1531, 5'-AACGTGAA-GAAATTCAAGC-3'; JB1534, 5'-GTTTTTCAGCGAC-GATCG-3'.

R1 ENp (1–2 μg) was used to digest approximately 100 ng of PCR product in the presence of MgCl<sub>2</sub>. The protein was inactivated by adding EDTA to a final concentration of 25 mM. Initially, formamide was added to a final concentration of 35%, but this resulted in incomplete denaturation; in subsequent experiments the samples were precipitated and resuspended in 95% formamide and boiled for 10 min. The products were run on 6% polyacrylamide DNA sequencing gels together with dideoxy sequencing reactions performed by using the same radiolabeled primers and pB109 template as size standards. The double-strand cleavage reactions were done exactly as above except that the products were not mixed with formamide and were run on a non-denaturing polyacrylamide gel.

**Gel Filtration.** 24 μg of purified EN protein in 200 μl (4.1 μM) of storage buffer were applied to a Superose-12HR 10/30 column with a Pharmacia FPLC system. The column was equilibrated with storage buffer, eluted at a flow rate of 0.4 ml/min, and monitored for protein by absorbance at 280 nm. For calibration, catalase, aldolase, ovalbumin, and ribonuclease A were used as standards. Values of  $[-\log(K_{av})]^{1/2}$  were plotted against the corresponding Stokes radii of the standards. The partition coefficient  $K_{av}$  of each protein was calculated by using the equation  $K_{av} = V_e - V_0/V_t - V_0$ , where  $V_e$  is elution volume of the protein,  $V_0$  is column void volume determined by blue dextran 2000, and  $V_t$  is total bed volume of the column. Fractions were assayed for R1Bm EN protein by immunoblotting with antibody G-18 (Santa Cruz Biotechnology), which recognizes the C-terminal tag.

**Sedimentation Equilibrium.** The molecular mass of R1Bm EN was examined by analytical ultracentrifugation to determine the multimeric state of EN in solution. The experiments were conducted on a Beckman Optima XL-A analytical ultracentrifuge with an An-60Ti rotor and a standard six-sector cell. A 100-μl sample containing 1.5 nmol of 15 μM EN in storage buffer was centrifuged at 4°C, with rotor speeds of 12,000 and 15,000 rpm, and equilibrium data were collected at a wavelength of 280 nm. Equilibrium was checked by comparing scans at various times up to 24 h. Data were analyzed with NONLIN, a program that performs a global nonlinear least-squares fit of sedimentation equilibrium data (11). The extinction coefficient at 280 nm used was estimated at 20,400 M<sup>-1</sup>·cm<sup>-1</sup> based on the amino acid composition of R1Bm EN.

## RESULTS

**Expression and Purification of R1Bm EN.** We expressed the R1Bm EN domain (Fig. 1*b*) in bacteria with a C-terminal His<sub>6</sub> tag and purified the protein by nickel chelate affinity chromatography (Fig. 2*a*). By expressing the EN domain separately from the rest of the ORF2 protein, we could test whether specificity of retrotransposition was conferred by the EN domain itself. We used a very sensitive plasmid nicking assay to detect endonucleolytic activity of the R1Bm EN protein. We found that it had weak nicking activity on supercoiled plasmids and defined conditions under which supercoiled plasmids bearing the *B. mori* rDNA (the *in vivo* target of the R1Bm transposon) were nicked (Fig. 2*b*). We optimized the cleavage

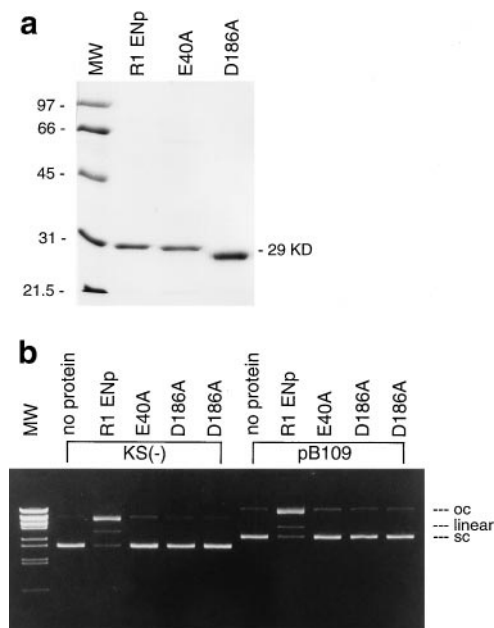


FIG. 2. Expression, purification, and activity of R1Bm EN protein. (a) R1Bm EN protein and its mutant versions E40A and D186A were expressed from the phage T7 promoter. The proteins were purified by nickel chelate chromatography and run on SDS/PAGE; the predicted  $M_r$  is 29,000. The unusual electrophoretic mobility of the D186A mutant was observed for two independent isolates. (b) The activity of the R1Bm EN protein was optimized and tested on two plasmid substrates, pBluescript ( $KS^-$ ) vector and pB109. The positions of open circle (oc), linear, and supercoiled (sc) plasmid controls are indicated on the right. Note that the mutant proteins are inactive. MW, molecular weight standards.

of an rDNA plasmid substrate, pB109, for R1Bm EN relative to monovalent cation, pH, buffer, and temperature conditions and showed that cleavage absolutely required the divalent cations  $Mg^{2+}$  or  $Co^{2+}$  (data not shown). Cleavage of a control plasmid lacking the target site was also observed, so the presence of an R1Bm target site was not required for cleavage. The specific activity of this protein was extremely low, about 1% of the specific activity of L1 EN (1). The R1Bm enzyme, like L1 EN protein, was inactive on apurinic DNA (data not shown). The L1 EN is itself about 20,000-fold less active than DNase I, a distantly related nicking EN (12). Thus the specific activity of the isolated R1Bm EN domain is about  $2 \times 10^6$ -fold lower than that of the digestive enzyme DNase I. We attempted to increase the specific activity of R1Bm EN by expressing lengthened versions (containing amino acid residues 1–230, 1–239, and 1–259), but these proteins instead showed modest decreases in specific activity (data not shown), suggesting that we had not omitted critical amino acid sequences in the design of our original construct. It is formally possible that the low specific activity of the R1Bm EN is because of a low fraction of properly folded molecules rather than an intrinsic property of the enzyme. Because of the extremely low levels of enzyme activity detected in our sensitive nicking assay, it was necessary to demonstrate that the detected activity was encoded by R1Bm ORF2. We mutated the highly conserved E40 and D186 residues in our R1Bm EN construct; the corresponding residues are known to be critical for catalysis in pancreatic DNase, exonuclease III (12), human AP EN (13), and the human L1 EN domain (1). Multiple sequence alignments (1, 14) show that these residues are absolutely conserved among retrotransposon ENs as well as in AP ENs. We purified the E40A and D186A mutant proteins and found that they were inactive in our nicking assay (Fig. 2). Thus the observed activity is indeed encoded by R1Bm ORF2.

**Cleavage Specificity.** We tested whether cleavage occurred at a specific sequence(s). To test this, we prepared end-labeled 175-bp fragments of the *B. mori* rDNA containing the R1Bm insertion site and incubated them with or without wild-type R1Bm enzyme. The cleavage products were then separated on polyacrylamide DNA sequencing gels with sequence standards. On the bottom strand, the most prominent cleavage product corresponded precisely to the left boundary of the 14-bp *tsd* that R1Bm generates *in vivo*, whereas on the top strand, a prominent cleavage product corresponded precisely to the left boundary of the *tsd* (Fig. 3a). Additional cleavage products were observed on the top strand, indicating that cleavage by R1Bm EN protein is not absolutely sequence-specific *in vitro*. In time course experiments (e.g., Fig. 3b), bottom strand cleavage was faster than top strand cleavage, consistent with the TPRT model (3) in which bottom strand cleavage defines the initial target site primer for reverse transcription. Top strand cleavage occurs subsequently, when it may be required for priming of the second strand.

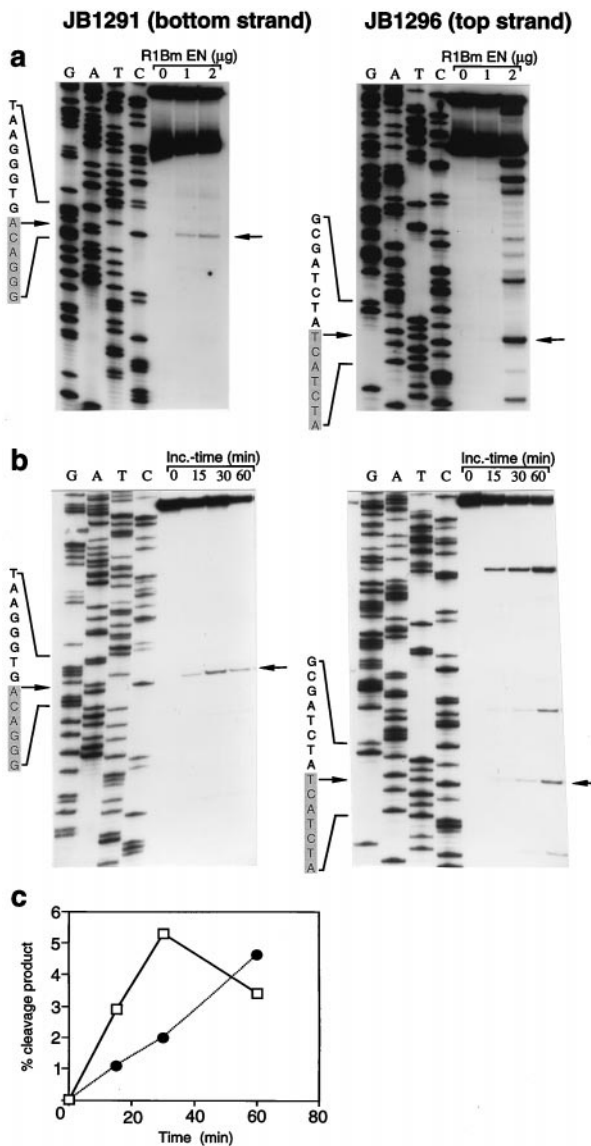
We have also investigated the specificity of cleavage on other substrates, including a fragment just 100 bp longer than the above substrate generated using primers JB1531 and JB1534. Obviously, sequences other than the preferred target can also be cleaved, as vector lacking the rDNA substrate is nicked to about the same extent as the plasmid containing the rDNA target site. We found that the relative efficiency of cleavage of the R1Bm *in vivo* target site was not consistently the highest efficiency cleavage site in every fragment tested, and other sites of cleavage (lacking obvious sequence similarity to the target site) were sometimes equally prominent, as in the 300-bp substrate mentioned above (data not shown). Thus it appears that the nature of flanking sequences can affect the relative rates of cleavage at various sites. We conclude that R1Bm EN is a sequence-specific nuclease but that its specificity can be altered by the effects of flanking sequences.

#### Double-Strand Cleavage and Evidence for Multimerization.

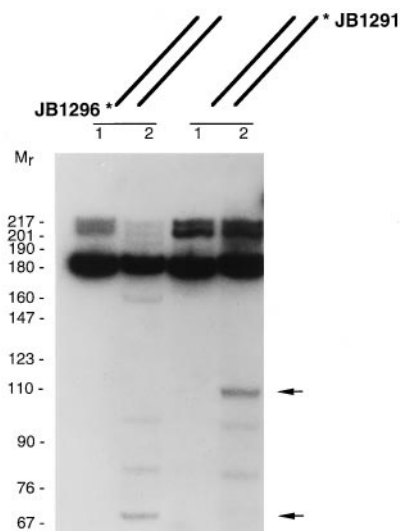
We next examined whether double-strand cleavage could be mediated by R1Bm EN. Only a small fraction of the substrate was nicked on each strand under the conditions used in Fig. 3. Similarly digested products were run on a nondenaturing gel, and products with electrophoretic mobilities consistent with double-strand cleavage at the *tsd* boundaries were observed (Fig. 4). The ability of the enzyme to make a double-strand break suggested two possibilities: (i) a monomer might make both cleavages or (ii) the enzyme might be multimeric. We examined the latter possibility by gel-filtration chromatography and equilibrium sedimentation and found that bulk R1Bm EN protein indeed behaves as a multimer, probably a tetramer (Fig. 5). Thus the fragment of the ORF2 protein we purified contains both an EN active site and a multimerization domain. Because of the low activity of the enzyme, it was not possible to determine unambiguously whether the tetramer form represented the active form of the enzyme. Therefore, it is formally possible that the active species is monomeric. Nevertheless, the fact that this retrotransposon EN shows evidence of tetramerization is interesting, because other enzymes involved in integration, including retroviral integrases and Mu transposase, have multimeric active forms (15–17).

## DISCUSSION

We recently showed that the human retrotransposon L1, which inserts into many sites in human DNA, encodes an EN with nicking activity (1). L1 EN also nicks specific DNA sites but with less specificity than R1Bm EN. Essentially, L1 EN prefers to nick at sequences that conform to the sequence  $Y_n \downarrow R_n$  that are A+T-rich, a rather degenerate consensus sequence. Although L1 EN *in vitro* cleavage sites resemble the sites of TPRT inferred from the sequences of various L1 *in vivo* transposition events, the similarities were restricted to runs of



**FIG. 3.** Sequence-specific cleavage of the R1Bm target DNA. (a) A 175-bp fragment centered around the R1Bm insertion site was synthesized by PCR with individually 5' end-labeled primers JB1291 (bottom strand) or 1296 (top strand) and pB109 template. The radioactive full-length double-stranded target DNAs were incubated without or with 1 or 2  $\mu\text{g}$  of R1Bm EN protein. Arrows indicate the major sites of cleavage that correspond to the tsd boundaries indicated in Fig. 1a; tsd sequences are shaded. The molecular weight standards were dideoxy sequencing reactions primed with the same radioactive oligonucleotides on pB109 template. The exact positions of the major cleavages were confirmed by mixing experiments in which the products were mixed with the molecular weight standards (not shown). The faster moving band in the substrate in part a was shown to correspond to incompletely denatured double-stranded DNA and could be eliminated from subsequent experiments by denaturing by boiling for 10 min in 95% formamide. (b) The cleavage of the bottom strand at lower enzyme concentrations than the top strand (part a above) suggested that bottom strand cleavage might precede top strand cleavage kinetically. This was confirmed in a time course experiment. The experiment was performed as above except with 2  $\mu\text{g}$  of R1Bm EN, and samples were removed at 0, 15, 30, and 60 min. Note that cleavage at the target sequence boundary (arrow) on the bottom strand (JB1291) peaks at 30 min whereas top strand cleavage (JB1296) peaks later. (c) The radioactivity in the bands representing the target sequence boundary cleavages was measured by PhosphorImager analysis, expressed as percent of initial substrate, and plotted as a function of time. Late in the reaction, the signal begins to decrease, suggesting the presence of a small amount of a contaminating random nuclease activity. Solid line, bottom strand; dotted line, top strand.



**FIG. 4.** Double-strand cleavage of target DNA by R1Bm EN. The 175-bp substrates described in Fig. 3a were incubated without (lanes 1) or with 2  $\mu\text{g}$  of R1Bm EN enzyme (lanes 2). The arrows indicate the size of the double-stranded DNA products resulting from cleavage on both strands at the R1Bm insertion site. The expected cleavage products would be 67 and 94 bp long with 3' overhangs of 14 nt, the effect of which on electrophoretic mobility is uncertain. Molecular weight standards ( $M_r$ ) are radiolabeled pBR322 DNA *MspI* fragments.

one to a few consecutive purines immediately 3' to the cleavage site. Because ENs of this general class (such as *E. coli* Exo III) have ribonuclease H activity (13) and because ribonuclease H activity could in principle be important for retrotransposition, Barzilay and Hickson (2) actually proposed that these ENs are ribonucleases rather than target site definition ENs. Our work showing that R1Bm EN cleaves with sequence specificity precisely at the boundaries of the R1Bm tsd provides the strongest evidence yet that the critical role of the retrotransposon ENs is instead to define and cleave the target DNA. Furthermore, the sluggish specific activity of R1Bm EN is consistent with target site cleavage, which requires only two cleavages. Also, because retrotransposition is predicted to take place in the nucleus, cellular RNase H could carry out this degradative role, unlike the case for retroviruses and retrotransposons, which carry out reverse transcription in the cytoplasm. However, it remains a formal possibility that the R1 EN has multiple functions in retrotransposition. Finally, recent studies of retrotransposon Tx1, a *Xenopus* element proposed to be specific for a sequence within a small DNA transposon (18), have provided independent evidence for our findings on R1Bm EN. The Tx1 EN domain was shown to have a precise bottom strand sequence-specific nicking activity *in vitro* on a substrate containing the putative Tx1 target DNA (S. Christensen and D. Carroll, University of Utah, personal communication).

The fact that R1Bm EN can make paired cleavages at each end of the tsd suggests a double TPRT model for the complete R1Bm retrotransposition process (Fig. 6). The multimeric state of the EN enzyme we expressed supports the possibility that the bottom and top strand cleavages might be made by independent subunits of a multimer of ORF2 protein. In this model, a multimer of bifunctional ORF2 proteins (each monomer containing both EN and RT activities) initially nicks target site DNA on the bottom strand and then uses the 3' end generated to prime reverse transcription on R1Bm RNA. In support of this model, we have expressed a larger fragment of R1Bm ORF2 including the region homologous to RT in *E. coli* and shown that this protein has RT activity *in vitro* (Q.F. and J.D.B., data not shown). The detailed structure of R1Bm RNA is unknown, but a low level of cotranscription (readthrough) with rRNA (corresponding to less than one such transcript per

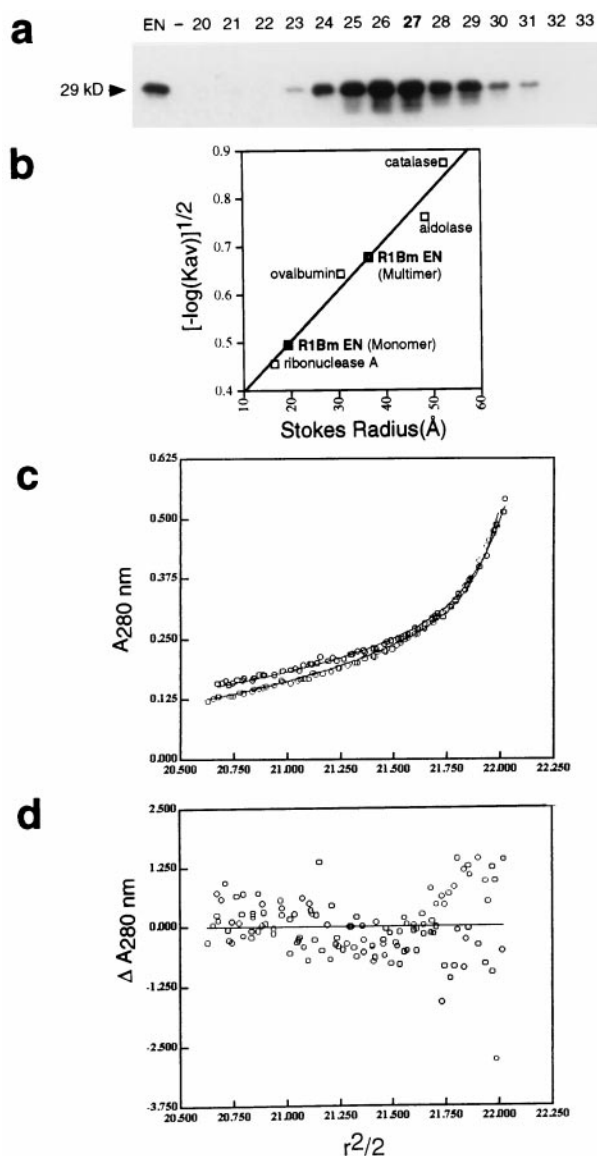


FIG. 5. R1Bm EN behaves as a multimer. (a) Purified R1Bm EN protein (24  $\mu$ g) was applied to a sizing column with molecular weight standards, and the fractions were immunoblotted with the anti-tag antibody to detect EN protein. Nearly all of the EN protein (as well as bulk absorbance) peaked at fractions 23–31, but a second small peak was observed on a long exposure at fraction 42 (not shown) and is assumed to represent monomeric enzyme. (b) Determination of Stokes radius of R1Bm EN monomer and multimer by gel filtration chromatography at 4°C with the indicated standards. The EN monomer peak runs with a Stokes radius of 19 Å corresponding to a globular protein of 19 kDa even though the actual mass is 29 kDa, suggesting that the EN has weak affinity for the column matrix. The Stokes radius of the multimer (36 Å) corresponds to a globular protein of 81 kDa and thus is consistent with a spherical tetramer with weak affinity for the column matrix. (c and d) Sedimentation equilibrium data for R1Bm EN at 12,000 and 15,000 rpm at 4°C. c represents the actual data (open dots) and the results of a global fit of these two data sets to a monomer–tetramer–dodecamer model (lines; see *Materials and Methods*). d presents a composite residual plot for the global fit in c; r, radius. A random distribution of actual data points (dots) about the predicted value (line) indicates very good agreement with the model. The data rule out a pure monomeric state for the R1Bm EN. A monomer–tetramer–dodecamer model with a tetramer as the predominant species (92%) fit these data best. Thus both gel filtration and sedimentation equilibrium methods are consistent with a predominantly tetrameric enzyme.

cell) has been reported (19). The readthrough transcripts would have rRNA sequences flanking the transposon se-

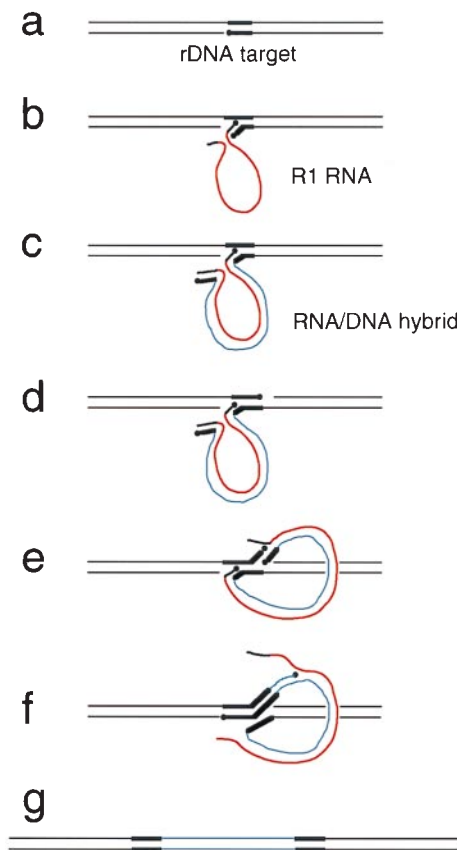


FIG. 6. Proposed model for R1Bm retrotransposition. The model described here is based on the TPRT model developed for the R2Bm element (3), an element that like R1 is sequence-specific for rDNA but unlike R1Bm has no clearly definable EN domain. Also, unlike R2Bm, which deletes a few base pairs of target DNA as part of the retrotransposition process, R1Bm creates a 14-bp tsd. The bifunctional EN/RT protein encoded by ORF2 (Fig. 1a) is presumably required for retrotransposition of R1, as is known to be the case for the human L1 element (1). For clarity, the protein has been omitted from the diagram; in principle, the ORF2 protein could carry out all of the diagrammed steps. As the protein is predominantly multimeric, we propose that the two target nicks could be made by separate subunits. (a) The rDNA target for R1 is symbolized by thin black lines with the 14-bp target site recognized by R1 (Fig. 1a) shown as bold black lines. The R1Bm EN domain creates a nick in the rDNA bottom strand; 3'-hydroxyls are indicated by black dots. (b) R1 RNA sequences (thick red line) are expressed as rDNA cotranscripts (RNA segments derived from rDNA are shown in black). Complementary base pairing between the cotranscribed rRNA sequence and the rDNA target allows formation of a primer–template complex, which is then (c) extended by the R1 RT activity to form an RNA/DNA hybrid R1 intermediate; the R1 DNA strand is shown as a thin blue line. (d) Nicking of the target top strand is carried out by the R1 EN, generating a primer for R1 second strand synthesis. (e) The newly synthesized rDNA sequences to the left of the R1 sequences can serve as a template for such priming; the R1 RNA could be displaced during continued polymerization, presumably by the R1 RT activity (f), or could be degraded by host RNase H; in either case this would result in completion of the R1 element insertion (g).

quences. The fact that this element's RNA has such target sequences flanking its own sequences will serve to increase the precision of a TPRT mechanism, as has been shown for R2Bm (20). Sequence complementarity between the cotranscript RNA and target DNA should increase the precision of retrotransposition even in the absence of cleavage precisely at the top strand tsd boundary and could explain how sequence-specific insertion can be effected in the absence of precise cleavage on the top strand (D. Carroll, personal communication). A simple series of nicking and priming events and a

strand-transfer event could readily explain the observed R1Bm structure (Fig. 6). Experiments are under way to more completely understand the mechanism of this unique retrotransposition pathway.

We are especially grateful to Tom Eickbush for freely providing clones and useful information and an anonymous reviewer for insightful observations. We thank Shawn Christensen and Dana Carroll for communicating unpublished data. We thank Shani Waninger, Dyche Mullins, David Symer, and Cynthia Wolberger for assistance with analytic ultracentrifugation and Tom Kelly for helpful comments on the manuscript. This work was funded in part by National Institutes of Health Grant CA16519.

1. Feng, Q., Moran, J., Kazazian, H. & Boeke, J. D. (1996) *Cell* **87**, 905–916.
2. Barzilay, G. & Hickson, I. D. (1995) *BioEssays* **17**, 713–719.
3. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. (1993) *Cell* **72**, 595–605.
4. Xiong, Y. & Eickbush, T. H. (1988) *Mol. Cell. Biol.* **8**, 114–123.
5. Jakubczak, J. L., Burke, W. D. & Eickbush, T. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3295–3299.
6. Wellauer, P. K. & Dawid, I. B. (1977) *Cell* **10**, 193–212.
7. Hollocher, H., Templeton, A. R., DeSalle, R. & Johnston, J. S. (1992) *Genetics* **130**, 355–366.
8. Dawid, I. B. & Botchan, P. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4233–4237.
9. Xiong, Y., Burke, W. D., Jakubczak, J. L. & Eickbush, T. H. (1988) *Nucleic Acids Res.* **16**, 10561–10573.
10. Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D. & Kazazian, H. H., Jr. (1996) *Cell* **87**, 917–927.
11. Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981) *Biophys. J.* **36**, 575–588.
12. Mol, C. D., Kuo, C.-F., Thayer, M. M., Cunningham, R. P. & Tainer, J. A. (1995) *Nature (London)* **374**, 381–386.
13. Barzilay, G., Walker, L. J., Robson, C. N. & Hickson, I. D. (1995) *Nucleic Acids Res.* **23**, 1544–1550.
14. Martín, F., Marañón, C., Olivares, M., Alonso, C. & Lopez, M. C. (1995) *J. Mol. Biol.* **247**, 49–59.
15. Engelman, A., Bushman, F. D. & Craigie, R. (1993) *EMBO J.* **12**, 3269–3275.
16. Aldaz, H., Schuster, E. & Baker, T. A. (1996) *Cell* **85**, 257–269.
17. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M. & Skalka, A. M. (1992) *J. Biol. Chem.* **267**, 16037–16040.
18. Garrett, J. E., Knutzon, D. S. & Carroll, D. (1989) *Mol. Cell. Biol.* **9**, 3018–3027.
19. Long, E. O. & Dawid, I. B. (1979) *Cell* **18**, 1185–1196.
20. Luan, D. & Eickbush, T. H. (1995) *Mol. Cell. Biol.* **15**, 3882–3891.