Cell, Vol. 87, 905–916, November 29, 1996, Copyright ©1996 by Cell Press

# Human L1 Retrotransposon Encodes a Conserved Endonuclease Required for Retrotransposition

Qinghua Feng,* John V. Moran,†
Haig H. Kazazian, Jr.,† and Jef D. Boeke*
*Department of Molecular Biology and Genetics
School of Medicine
Johns Hopkins University
Baltimore, Maryland 21205
†Department of Genetics
School of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania 19104

## Summary

**Human L1 elements are highly abundant poly(A) (non-LTR) retrotransposons whose second open reading frame (ORF2) encodes a reverse transcriptase (RT). We have identified an endonuclease (EN) domain at the L1 ORF2 N-terminus that is highly conserved among poly(A) retrotransposons and resembles the apurinic/apyrimidinic (AP) endonucleases. Purified L1 EN protein (L1 ENp) makes 5′-$PO_4$, 3′-OH nicks in supercoiled plasmids, shows no preference for AP sites, and preferentially cleaves sequences resembling L1 in vivo target sequences. Mutations in conserved amino acid residues of L1 EN abolish its nicking activity and eliminate L1 retrotransposition. We propose that L1 EN cleaves the target site for L1 insertion and primes reverse transcription.**

## Introduction

Retrotransposons can be grouped into two classes, the retrovirus-like long terminal repeat (LTR) retrotransposons, and the poly(A) elements such as human L1 elements, Neurospora crassa TAD elements (Kinsey, 1990), Drosophila melanogaster I factors (Bucheton et al., 1984), and Bombyx mori R2Bm (Luan et al., 1993). The two types of retrotransposon have different types of structural organization and use radically different mechanisms for their transposition. Unlike the LTR retrotransposons, poly(A) elements (also called non-LTR elements) lack LTRs and end with poly(A) or A-rich sequences. The LTR retrotransposition mechanism is relatively well understood, thanks to its parallels with the retroviral life cycle (Boeke and Stoye, 1996); in contrast, the mechanism of poly(A) elements is still being elucidated (Luan et al., 1993).

Poly(A) elements can be subdivided into sequence-specific and non-sequence-specific types. The human L1 element (also known as LINE, for long interspersed nuclear element) is of the latter type, with insertions scattered among all chromosomes. L1 has been extremely successful at colonizing the human genome; the ~100,000 copies are estimated to comprise 5% of nuclear DNA (Fanning and Singer, 1987). Most L1s are 5′ truncated and presumed defective. Like full-length elements, the truncated copies are often flanked by short target site duplications (TSDs), suggesting that specific nucleotide sequences at the L1 5′ end are not required for insertion in *cis*, and that transactions involving the 3′ poly(A) may be critical. Retrotransposition of L1 can cause human disease (Dombroski et al., 1991; Holmes et al., 1994; Kazazian et al., 1988; Miki et al., 1992; Narita et al., 1993). The full-length L1 consensus sequence is about 6 kb long and unlike most genomic L1 elements, which contain many mutations, has two intact open reading frames (ORFs) (Scott et al., 1987). ORF1 is not homologous to known cellular protein sequences; there is evidence that it is an RNA-binding protein, but its precise function is unknown (Hohjoh and Singer, 1996; Martin, 1991). ORF2 encodes reverse transcriptase (RT) (Dombroski et al., 1994; Mathias et al., 1991). L1 and other poly(A) elements lack recognizable homologs of retroviral integrase, protease, and RNase H. These findings suggest that L1s employ a fundamentally different transposition mechanism than the LTR elements.

The best model for the transpositional movement of poly(A) elements comes from studies of the R2Bm element of B. mori (Luan et al., 1993). R2Bm encodes a single protein with both sequence-specific endonuclease activity and RT activity (Luan et al., 1993). The protein nicks target DNA, generating a 3′-OH that is subsequently used by the R2Bm protein to reverse transcribe specifically the 3′ end of the R2Bm RNA in vitro (Luan and Eickbush, 1995). This mechanism of target DNA–primed reverse transcription may apply to other poly(A) elements. However, since many of the other poly(A) elements are not exquisitely target sequence–specific, this is by no means certain.

Recent studies on mobile group II introns support the above model (Zimmerly et al., 1995a). Such introns can sequence-specifically home from intron-containing alleles into intronless alleles (Lazowska et al., 1994; Meunier et al., 1990; Moran et al., 1995). Group II intron aI2 encodes both a maturase that mediates its own splicing (Moran et al., 1994) and an RT that is closely related to poly(A) element RTs (Michel and Lang, 1985; Kennell et al., 1993). Further studies showed that aI2 retrotransposition is initiated by specific endonucleolytic cleavage of the intronless allele. The 3′-OH is then used by the aI2 RT activity to prime intron-specific reverse transcription (Zimmerly et al., 1995a), analogous to the DNA-primed reverse transcription of R2Bm. Remarkably, aI2 intron homing involves partial or complete reverse splicing of the aI2 intron. The specific DNA endonuclease activity resides in both aI2 protein and aI2 RNA (Zimmerly et al., 1995b).

In contrast, L1s are non-sequence-specific elements. Until now, we know of no evidence for an endonuclease activity associated with L1. Therefore, it is unclear whether L1 employs a similar retrotransposition mechanism to R2Bm. Instead, it has been proposed that L1 and related elements use nicks generated by cellular endonucleases to prime reverse transcription (Branciforte and Martin, 1994; Schwarz-Sommer et al., 1987).

Table 1. Many Sequence-Specific Retroelements Lack the EN Domain

| Element | Specific for | ORFs | EN Domain? |
|---|---|---|---|
| R2Bm | sequence in rDNA array | 1 | − |
| CRE-1 | sequence in mini-exon array | 1 | − |
| CRE-2 | sequence in mini-exon array | 1 | − |
| SLACS | sequence in mini-exon array | 1 | − |
| Group II intron | intronless target site | 1 | − |
| R1Bm | sequence in rDNA array | 2 | + |
| TART | telomeres and heterochromatin | 2 | + |
| TRAS-1 | telomeric repeat unit | 2 | + |
| DRE-1 | tRNA upstream regions | 2 | + |

Here, we identify an apurinic/apyrimidinic (AP) endonuclease homologous domain (EN) that is conserved among non-sequence-specific poly(A) elements. We have expressed and purified an L1 ENp and demonstrated its endonuclease activity. Mutations in conserved putative active site residues abolish its nicking activity in vitro, as well as L1 retrotransposition in vivo. We propose that the EN domain of L1 ORF2 functions to identify and cleave the target site for L1 retrotransposition, generating a reverse transcription primer. As the EN domain is conserved among poly(A) elements, the same is likely to be true for all these elements.

## Results

### PolyA Elements Encode an AP Endonuclease-like Domain

Recently, a novel poly(A) element, L1Tc, was identified from Trypanosoma cruzi (Martín et al., 1995). It encodes three ORFs; ORF1 resembles AP endonucleases. We investigated whether this AP endonuclease homology exists in other poly(A) elements.

An amino acid sequence alignment of the AP endonucleases was generated with the program PILEUP (Genetics Computer Group, University of Wisconsin) and hand-edited. Conserved residues and the putative catalytic active site residues were identified on the basis of the structure of Exo III (Mol et al., 1995), the major AP endonuclease from Escherichia coli. Next, an amino acid sequence alignment between an AP endonuclease and several poly(A) element ORFs was generated (Figure 1). Several conserved AP endonuclease motifs were present in the poly(A) elements in the alignment. In particular, all conserved residues are equivalent to Exo III residues in or near the active site (Mol et al., 1995). These motifs are only conserved in the non-sequence-specific poly(A) elements, and in a subset of the sequence-specific poly(A) elements, but not in group II introns or LTR retrotransposons (Table 1). Spacing of the catalytic motifs was also conserved between the transposons and the AP endonucleases. We refer to this conserved domain of poly(A) elements as the EN domain.

The AP endonucleases are important enzymes with known roles in DNA repair. They also have 3′ exonuclease, 3′ phosphatase, and RNase H activities specified by a single active site (Barzilay et al., 1995b; Weiss, 1976). RNase H or 3′–5′ exonuclease activities could play roles in retrotransposition. Several conserved AP endonuclease motifs are also shared with DNase I, a nonspecific nicking endonuclease, notably the proposed catalytic active site residues (Figure 1). The overall fold in the DNase I structure resembles that of Exo III (Mol et al., 1995); thus, the EN domain might be a target site nickase.

### EN Domain at L1 ORF2 N-Terminus Encodes a Nicking Endonuclease

To test whether the observed EN domain in L1 was functional, we expressed and purified the N-terminal 26 kDa domain of L1 ORF2 (ending at residue 239). We tagged the protein with six His residues to facilitate purification on nickel–agarose. A single protein band was observed on SDS–polyacrylamide gel electrophoresis (PAGE) (Figure 2A, lane 7).

Endonuclease activity was assayed by measuring the ability to convert supercoiled plasmids into open circles. L1 EN nicking activity depends on divalent cations, with Mg greatly preferred over Mn; L1 EN activity was optimized with regard to buffer, pH, and salt (data not shown). When 2.6 ng (approximately 81 fmol) of purified L1 ENp was incubated with 0.2 μg of supercoiled substrate DNA (approximately 100 fmol), 50% of the supercoiled DNA was converted to open-circle DNA in 20 min (Figure 2B, lane 3); when 26 ng of L1 ENp was used, 100% of the supercoiled DNA was converted to a mixture of open-circle DNA, as well as some linear plasmid DNA (Figure 2B, lane 4).

The nicking activity was specific to expression of L1 ENp, because no activity was detected in mock-purified protein from a vector containing strain. The specific activity of L1 ENp in the nicking assay is approximately 20,000-fold lower than that of DNase I on a molar basis (data not shown; Price, 1975). The calculated turnover number for L1 EN is approximately two phosphodiester bond cleavages per hour on a supercoiled pBS substrate under optimal conditions. A time course in which 1.7 ng (50 fmol) L1 ENp digested 1 μg (500 fmol) of DNA shows that L1 ENp turns over and is therefore an enzyme (Figure 2C). Although the expressed fragment of ORF2 was designed to match AP endonucleases in length, longer (or shorter) segments of ORF2 might behave differently.

### Mutations in Conserved Residues of L1 EN Abolish Endonuclease Activity

Since many endonucleases are found in extracts, it was possible that an E. coli activity copurified with L1 ENp. Therefore, we made missense mutations in the L1 ENp expression construct and determined whether nicking activity was affected. Certain residues are conserved among all poly(A) elements and AP endonucleases (see Figure 1). Three of these are thought to be especially critical for catalysis, including E43 (numbering refers to L1 ORF2), believed to bind the essential divalent cation, and the D205 and H230 residues, believed to effect catalysis in both Exo III (Mol et al., 1995) and DNase I (Oefner and Suck, 1986; Suck et al., 1988). We mutagenized five conserved L1 residues, including the above three. We tagged, expressed, and purified (Figure 2A, lanes 2–6) the mutant proteins in parallel with wild-type L1 ENp. All five mutants have greatly reduced nicking

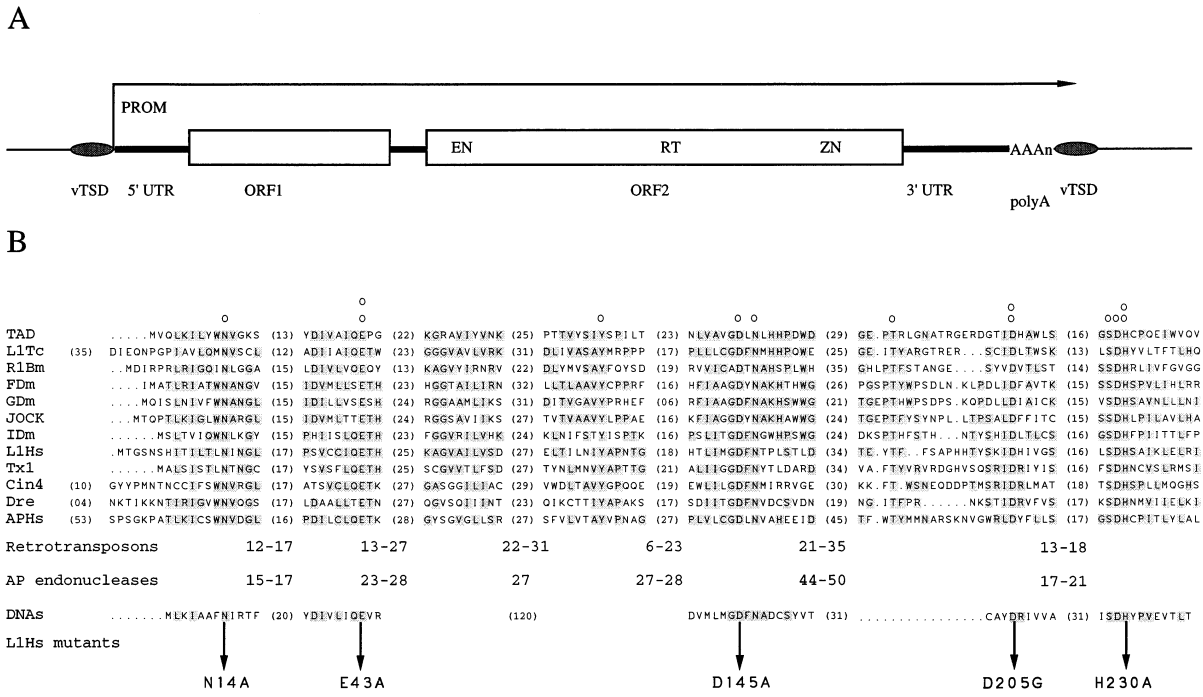



Figure 1. Poly(A) Element EN Domain

(A) Structure of the human L1 element. PROM, L1 internal promoter; vTSD, variable target site duplication; EN, endonuclease domain; RT, reverse transcriptase domain; ZN, putative zinc finger–like domain.

(B) Amino acid sequence alignment of poly(A) elements and human AP endonuclease. The sequences are TAD, from Neurospora crassa; L1Tc, from T. cruzi; R1Bm, from B. mori; FDM, GDM, and IDM, which are F, G, and I elements from Drosophila; Jock, *jockey* from D. melanogaster; L1Hs, human L1; Tx1, from Xenopus laevis; Cin4, from Zea mays; and DRE, from Dictyostelium discoideum. APHs is the human AP endonuclease, DNase I from bovine pancreas. The EN domain was also identified in the following elements: CR1 (chicken, J. Burch, personal communication), *ingi* (trypanosome), L1Md (mouse, and other mammalian L1s), Ta11 (Arrabidopsis thaliana), TART (D. melanogaster), TRAS (B. mori), and T1 (mosquito). Conserved (>2 identities) residues are stippled; residues conserved among all poly(A) elements and the human AP endonuclease as shown as a single circle; putative active site residues, as double circles. Numbers refer to the residues between two conserved blocks. Residues mutated in L1 ENp are indicated by arrows.

activities relative to that of wild-type L1 EN (Figure 2B, lanes 5–9). The E43A mutant is slightly leaky (Figure 2B, lane 5). Since this residue binds $Mg^{2+}$ in Exo III (Mol et al., 1995), human AP endonuclease (Barzilay et al., 1995a), and DNase I (Oefner and Suck, 1986; Suck et al., 1988) the excess $Mg^{2+}$ provided in vitro may partially suppress this mutant. Nevertheless, the E34A mutant protein still has at least 20-fold less activity than wild-type L1 ENp. Thus L1 EN has the nicking activity.

### L1 ENp Leaves 5′-$PO_4$ and 3′-OH Residues

Nucleases can leave either 5′-$PO_4$, 3′-OH or 5′-OH, 3′-$PO_4$ termini. The L1 EN domain resembles class II AP endonucleases and DNase I, both of which leave 5′-$PO_4$, 3′-OH termini. To examine the L1 ENp products, we tested whether the products of L1 ENp were substrates for T4 DNA ligase (Lehman, 1974). Nicked circles generated by L1 ENp were incubated with T4 DNA ligase, and the ends were efficiently ligated (Figure 3A, lanes 2–5). Thus, L1 ENp generates 5′-$PO_4$, 3′-OH termini. We observed that the closed relaxed circle DNA product generated is itself a substrate for L1 EN (Figure 3A, lane 5). Thus, L1 EN can cleave both supercoiled and relaxed DNAs.

### L1 ENp Is Not Specific for AP-DNA

Since L1 ENp cleaves native DNA (pBS plasmid DNA), and L1 ENp is closely related to AP endonucleases, we investigated whether apurinic DNA was a preferred substrate for L1 ENp. Native pBS DNA and pBS containing one to two AP sites per molecule (AP-DNA) were tested for cleavage by L1 ENp and Exo III. L1 ENp cleaved both substrates equally (Figure 3B, lanes 3 and 6), whereas Exo III only cleaved AP-DNA (lanes 4 and 7). Titration experiments (data not shown) showed that L1 ENp cleaved both substrates with the same kinetics. Thus, L1 ENp is not specific for AP-DNA.

### L1 ENp Preferentially Cleaves Supercoiled DNA

Bacterial chromosomal DNA is apparently supercoiled in vivo (Pettijohn and Pfenninger, 1980; Sinden et al., 1980), whereas virtually all supercoiling of eukaryotic DNA results from nucleosome wrapping (Sinden and Pettijohn, 1981). However, chromosomal DNA targets in eukaryotic cells are likely to be supercoiled transiently as the result of moving polymerases and transient nucleosome removal (Drolet et al., 1994; Liu and Wang, 1987; Tsao et al., 1989). We asked whether supercoiled or closed relaxed circle DNAs were preferred targets for L1 ENp in vitro. We repeated the ligation experiment with a limiting amount of

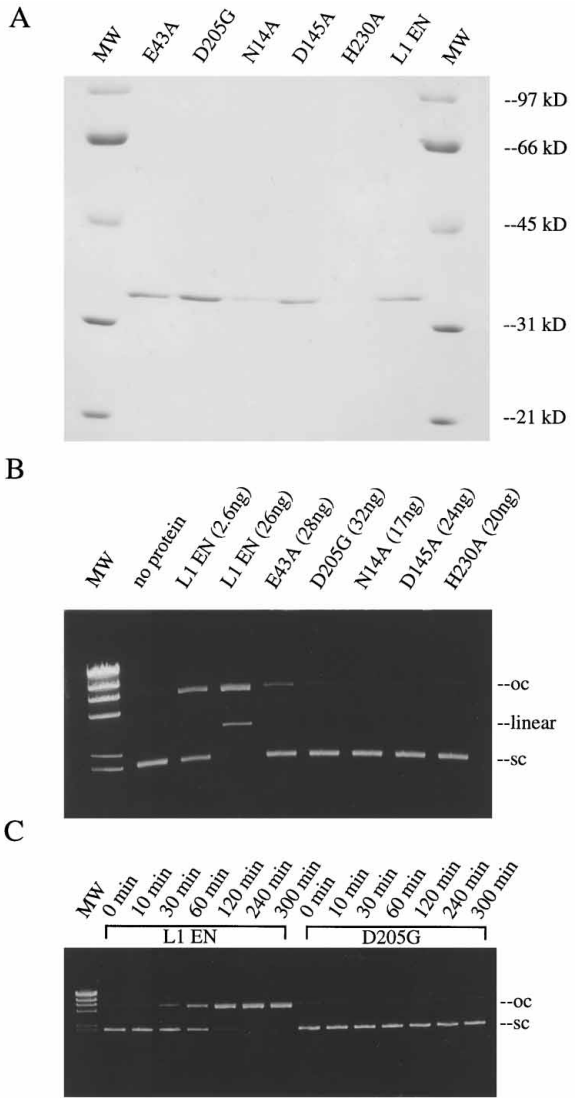



Figure 2. Purification and Nicking Activities of L1 ENp and Mutant Proteins

(A) Purification. Purified proteins were separated on a 10% SDS–PAGE gel and stained with Coomassie blue. Approximately equal amounts of protein were loaded, except that for H230A, 10-fold less protein was loaded. MW, molecular mass standards.

(B) Nicking activities. Lane 1 (lanes are numbered left to right), phage λ HindIII digest MW marker; lane 2, substrate pBS DNA, no protein added; lane 3, with 2.6 ng of wild-type L1 ENp; lane 4, with 26 ng of wild-type L1 ENp; lane 5, E43A mutant; lane 6, D205G; lane 7, N14A; lane 8, D145A; lane 9, H230A. Sc, supercoiled plasmid; oc, open (nicked) circular plasmid; l, linear plasmid.

(C) Time course, 50 fmol of L1 ENp (or D205G mutant) was used to digest 500 fmol of pBS.

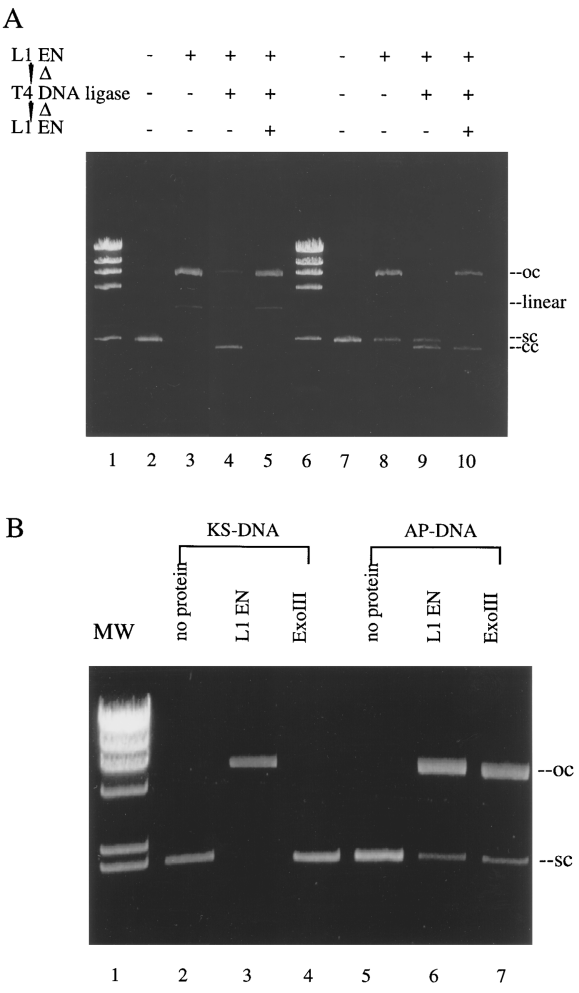



Figure 3. Enzymatic Properties

(A) Structure of nick and preference for supercoiled substrate. Supercoiled pBS DNA (0.2 μg) (lane 2) was incubated with L1 ENp to generate open circle DNA (lane 3). Subsequently, L1 ENp was heat-inactivated, and T4 DNA ligase was added (lane 4). After ligation, T4 DNA ligase was heat-inactivated, and product was again incubated with L1 ENp (lane 5). Lanes 7–10 are similar, except that 10-fold less L1 ENp was added initially. Cc, closed relaxed circle DNA.

(B) L1 ENp cleaves native DNA and apurinic DNA equally well. DNA substrate was either native DNA or apurinic DNA (prepared as described in Experimental Procedures). KS-DNA, native pBS KS(−) DNA; AP-DNA, apurinic DNA; sc, supercoiled DNA; oc, open circle DNA; MW, λ HindIII digest.

L1 ENp (Figure 3A, lanes 7–10), i.e., such that about half the supercoiled DNA was converted into open circle DNA (Figure 3A, lane 8). The products were then converted to closed relaxed circle DNA by treatment with T4 DNA ligase (Figure 3A, lane 9), generating an approximately 60/40 mixture of supercoiled and closed relaxed circle substrates. L1 ENp preferred the supercoiled DNA substrate (Figure 3A, lane 10).

## Supercoiled pBS DNA Has Cleavage Hotspots for L1 ENp

The ease with which linear DNA was generated by L1 ENp cleavage suggested that L1 ENp cleavage was not random. L1 ENp double-strand breaks presumably result from closely spaced nicks on opposite strands. We investigated whether breaks made by L1 ENp were randomly distributed by linearizing supercoiled pBS DNA with L1 ENp and then digesting the linears with seven different restriction enzymes, either single or double cutters of the plasmid. We observed discrete bands (two or three major bands, respectively) instead of the smears expected of random cleavage (Figure 4A; data not shown), placing

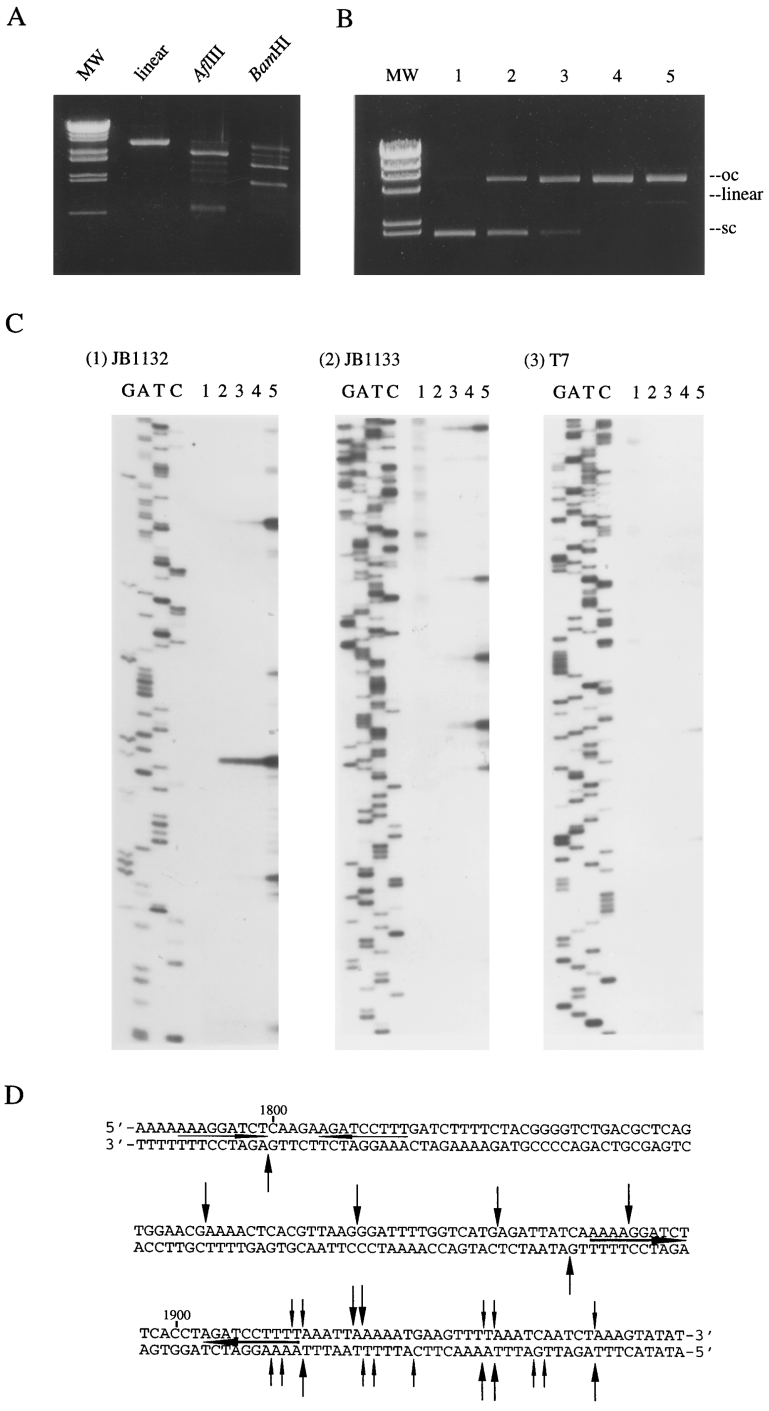



Figure 4. Cleavage Hotspots in pBS Plasmid

(A) L1 ENp double-strand break hotspot. Linear pBS DNA products were electroeluted, digested with retriction enzymes, and run on agarose gels.

(B) L1 ENp cleavage reaction. Lane 1, supercoiled DNA substrate; lanes 2–5, 13 ng, 26 ng, 65 ng, and 130 ng of L1 ENp added to 3.2 μg of DNA, respectively; 5% of this was run on the gel.

(C) Primer extension on uncleaved substrate and L1 ENp products as in (B). A sequence ladder generated with the indicated kinased primer was included for each reaction. Primers JB1132 and JB1133 are specific for each strand flanking the cleavage hotspot region of pBS.

(D) Cleavage hotspots in pBS. Major cleavage sites, large vertical arrows; minor ones, smaller vertical arrows; horizontal arrows, inverted repeats (heavy arrows, pBR322 minor; thin arrows, pBR322 subminor [Lilley, 1981]).

the major double-strand break made by L1 ENp at about position 1900 in pBS.

Primer extension was used to define precisely the cleavage sites on each strand in this region. Using $^{32}$P-labeled oligonucleotide primers flanking this region, we mapped the linear pBS DNA ends generated by L1 ENp. Surprisingly, a cluster of six major cleavage sites was seen on each strand (Figure 4D). All major sites identified had a purine immediately 3′ to the point of cleavage, usually an A, and most sites had several purines in a row. In addition, about half the sites had a run of pyrimidine residues just 5′ to the point of cleavage, suggesting a consensus sequence of $(Py)_n\downarrow(Pu)_n$ for L1 ENp cleavage. All the L1 ENp cleavages observed fall in a very AT-rich segment of pBS.

Since a cluster of cleavage hotspots had been mapped in this region, we investigated the kinetics of cleavage. L1 ENp cleavage was titrated, ranging from ~40% conversion to open circles to 90% conversion to open circles and 10% conversion to linears (Figure 4B). Primer extension was done on these DNAs with primers

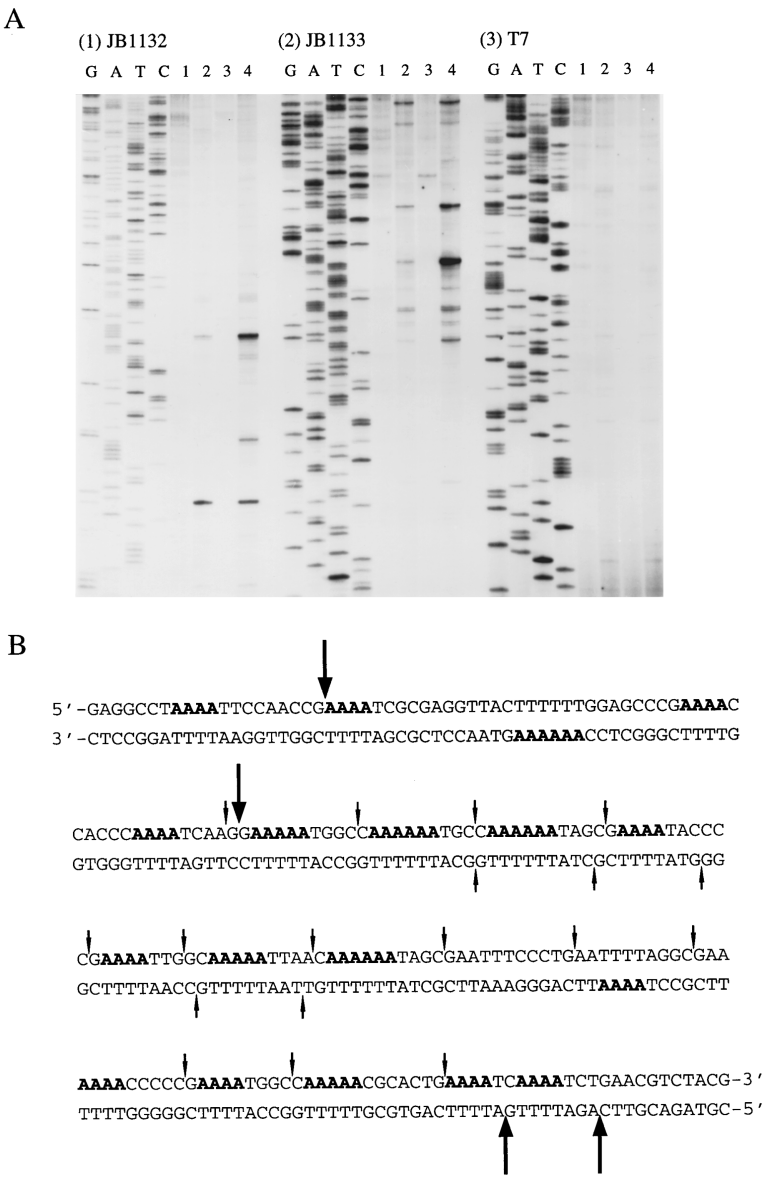



Figure 5. Cleavage Specificity

(A) Cleavage specificity does not require supercoiling. DNAs were treated with L1 ENp and used as templates for primer extension experiments as in Figure 4. Lanes 1, supercoiled DNA, no L1ENp; lanes 2, supercoiled DNA plus 20 ng of L1 ENp; lanes 3, relaxed closed circular DNA, no L1 ENp; lanes 4, relaxed closed circular DNA plus 80 ng of L1 ENp. GATC lanes indicate sequencing reactions primed with indicated kinased oligonucleotide.

(B) K-DNA contains a hotspot for L1 ENp cleavage (indicated by bold arrow); cleavage sites determined as in Figure 4B but using primer SP6. Sites of enhanced cleavage by hydroxyl radical determined by Burkhoff and Tullius (1987) are indicated by small vertical arrows. Bold letters indicate phased A tracts.

flanking the site (Figure 4C, JB1132 and JB1133) and a control T7 primer. One highly preferred site on the arbitrarily defined top strand was cleaved first, and overall, sites on this strand were cleaved faster than bottom strand sites. When the T7 primer was used, little cleavage was observed. Thus, cleavage of pBS DNA by L1 ENp is highly nonrandom. Interestingly, this region of pBS (derived from pBR322) migrates more slowly than expected on polyacrylamide gels (Stellwagen, 1983), raising the possibility that it may adopt an unusual structure, such as bent or cruciform DNA.

### Cleavage Specificity Is Not Affected by Supercoiling

We investigated whether supercoiling was necessary for specific recognition and cleavage of hotspot sites, or whether it only affected cleavage rate. This aids in defining the enzyme specificity, because it addresses whether specificity is intrinsic to sequence or whether a higher order structure (e.g., cruciform) is required. We nicked pBS DNA with HpaII in the presence of ethidium bromide and ligated the DNA to form relaxed closed circular substrates. As expected, four times more L1 ENp had to be added to the relaxed substrate than to the supercoiled substrate for it to cleave to similar extents. However, cleavage specificity of supercoiled and relaxed DNAs was identical (Figure 5A).

### L1 ENp Cleavage of Other Targets

We tested whether cruciform or bent DNAs were preferred substrates for L1 ENp cleavage. Cruciform sequences tested included endogenous cruciforms mapping within the pBS hotspot region (Lilley, 1981). Cruciform sequences and their boundaries with normal DNA occasionally contained sites of preferred cleavage, but usually did not; in particular, the so-called major pBR322 cruciform (Lilley, 1981) lacked such sites (data not shown). The bent DNA we investigated was from

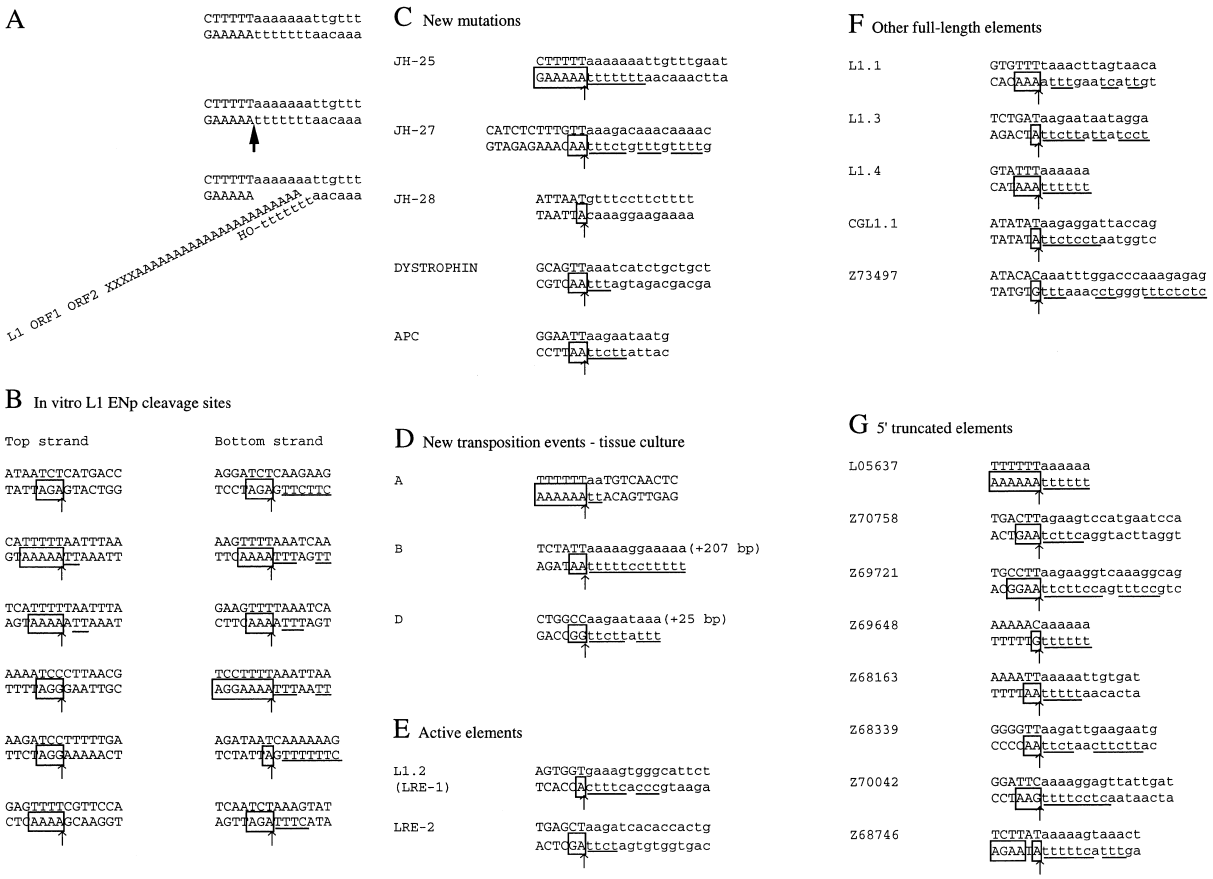



Figure 6. Similarity of In Vitro Cleavage Sites for L1 ENp and Predicted Sites of Priming of Reverse Transcription

(A) A model, based on the JH-25 sequence, for concerted target DNA nicking and reverse transcription of the 3′ poly(A) end of L1 RNA. The specificity of L1 ENp for $(Py)_n|(Pu)_n$ generates a polypyrimidine 3′ terminus that can in principle base pair to the 3′ poly(A) of L1 RNA. Such complementarity might stabilize a reverse transcription priming complex.

(B–G) Comparison of cleavage sites determined in vitro (B) to various in vivo inferred priming sites involved in L1 retrotransposition. Note that the nucleotide 3′ to the cleavage site is always a purine, is usually an A, and is usually part of an oligopurine run (boxed residues). In many cases, there is a symmetrically placed oligopyrimidine tract 5′ of the cleavage site or inferred priming site (underlined residues). For parts (C–G), letters in lower case represent the TSD. Note that the runs of As at the 5′ end of many of the TSDs represent an area of microhomology with the 3′ poly(A) tract of the L1 insertion; these are assumed to represent part of the TSD here.

(B) pBS targets; top strand is arbitrarily defined as the strand cleaved first.

(C) New mutations caused by L1 insertion include three hemophilia A mutations (Kazazian et al., 1988; Woods-Samuels et al., 1989) and a dystrophin mutation (Holmes et al., 1994), and a somatic insertion into the APC tumor suppressor gene associated with cancer (Miki et al., 1992).

(D) New L1-*neo* transposition events that occurred in HeLa cells are described by Moran et al. (1996).

(E) Active transposon copies discovered as progenitor elements for the JH-27 insertion (L1.2) and the dystrophin insertion (LRE-2).

(F) Other full-length elements cloned intentionally in searches to find active elements (L1.1–1.4 [Dombroski et al., 1991], CGL1.1 [Hohjoh et al., 1990]) or discovered by searching for element copies in GenBank (Z73497).

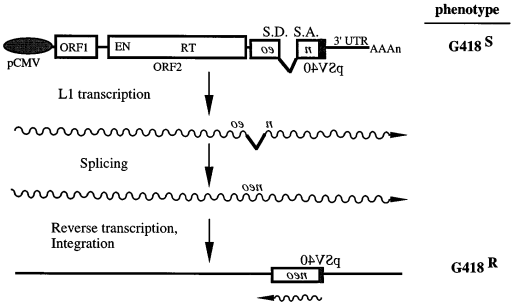
(G) Genbank was searched using BLASTN with the 3′ UTR sequence of L1.2 and the top 34 hits were studied. Approximately half of the truncated elements had a precise TSD. These are all listed on this figure, identified by accession number.

kinetoplast K-DNA, which contains a severely bent fragment (Kitchin et al., 1986). Primer extension mapping of the preferred site(s) of nicking was performed with pPK201/CAT, which consists of the K-DNA fragment in a vector also containing the hotspot region we previously mapped in pBS. Two approximately equally utilized hotspots for double-strand cleavage were observed in pPK201/CAT. One of these was the previously mapped hotspot; the other was in a subset of the oligo(A) tracts of the K-DNA (Figure 5B). However, not all bent segments in the K-DNA were nicked, indicating that some but not all bent DNA segments are hotspots for L1 ENp cleavage.

## L1 In Vivo Target Sites Resemble L1 ENp Cleavage Sites

We examined the sequences of new human mutations caused by L1 insertion and the termini of full-length L1s, including several elements known to be active in transposition (Moran et al., 1996 [this issue of *Cell*]). From the TSDs, we inferred the site of priming of minus-strand reverse transcription, assuming that the reverse transcription of L1 3′-end RNA had been primed by chromosomal nicks. Finally, we collected a number of TSDs from truncated human L1 sequences present in GenBank. These sequences indicate that like L1 ENp, the activity that cleaves the target for L1 insertion recog-

A




B

| construct | transposition frequency ($10^{-6}$cell$^{-1}$) |
|---|---|
| wildtype L1 | 335 |
| D703Y (RT$^-$) | 0.5 |
| N14A (EN$^-$) | 3.4 |
| D145A (EN$^-$) | 1.0 |
| D205G (EN$^-$) | 0.7 |
| H230A (EN$^-$) | 1.3 |

Figure 7. L1 EN Domain Is Required for Transposition in HeLa Cells

(A) Diagram of *L1.2mneoI* retrotransposition assay. A *neo* marker gene with a backward intron (*mneoI*) is inserted upstream of the L1 3′UTR such that *neo* and L1 are convergently transcribed. L1 transcription from the CMV promoter leads to the splicing of the intron and reconstruction of the *neo* coding region. Reverse transcription and integration lead to expression of *neo* from its SV40 promoter. pCMV, cytomegalovirus early promoter; S. D., splicing donor; S. A., splicing acceptor; wavy line, RNA; V, intron sequence.

(B) L1 retrotransposition frequencies. D703Y is the RT active site mutant; others are EN domain mutants.

nizes one or more purines just 3′ to the site of cleavage, and these often involve short runs of As. These are usually symmetrically juxtaposed to a run of pyrimidines 5′ to the site of cleavage (Figure 6). All the sequences are very AT-rich; this specificity is completely consistent with that observed for L1 ENp in vitro.

### Mutations in L1 EN Domain Kill L1 Retrotransposition

A transposition-competent L1 bearing a genetic marker (*L1.2mneoI*; described in the accompanying paper by Moran et al., 1996) was used to evaluate the importance of the EN domain in retrotransposition. The construct consists of a functional L1 element, L1.2A, driven by the cytomegalovirus (CMV) early promoter–enhancer. The element is marked with the *mneoI* gene, driven by the simian virus 40 early promoter, and inserted downstream of L1 ORF2, within the L1 3′ untranslated region (3′ UTR) and in the opposite transcriptional orientation relative to L1. The marker gene is also disrupted with an intron oriented such that it can only be spliced out of L1 RNA (Figure 7). G418$^R$ cells arise only when the *L1.2mneoI* retrotransposes, resulting in generation of a functional *neo* gene (Moran et al., 1996). The number of G418$^R$ colonies gives a readout of transposition frequency. The wild-type L1–*neoI* element gave rise to G418$^R$ colonies at a frequency of $3.4 \times 10^{-4}$ per cell, but a control mutation in the L1 ORF2 RT domain (D702Y) reduced transposition frequency about 600-fold. Four conserved residue mutations in the L1 EN domain similarly reduced the G418$^R$ frequency 100- to 500-fold (Figure 7), as do two additional conserved residue mutations (R. DeBerardinis, J. V. M., and H. H. K. unpublished data). These data demonstrate that the EN domain is essential for L1 retrotransposition.

## Discussion

The presence of a conserved amino acid sequence motif in the N-termini of poly(A) element RTs suggests that it represents a function essential for the transposition of the elements. The limited homology to AP endonucleases, perfectly maintained in the putative active site residues, suggests that this function is an AP-like endonuclease. The human L1 element indeed encodes an endonuclease, but it apparently lacks specificity for AP sites. Rather, L1 ENp recognizes some feature intrinsic to certain sequences in native DNA. L1 EN is essential for retrotransposition of L1s in human tissue culture. We propose that similar endonucleases are encoded by homologous poly(A) elements. The R2Bm element encodes a sequence-specific endonuclease as part of a large multifunctional reverse transcriptase protein, but the identity of the endonuclease domain is unclear. L1 EN is the first retrotransposon endonuclease defined biochemically as an isolated domain.

### Recognition Specificity of L1 EN

The homology of L1 EN to AP endonucleases suggests that there may be a shared aspect of how these two nucleases recognize their substrates. But both enzymes share a degree of structural homology with the relatively nonspecific nickase DNase I. L1 ENp does not prefer apurinic DNA over native DNA, so it is unlikely to be an AP endonuclease. However, it is formally possible that some special form of AP-DNA might represent a preferred substrate. L1 ENp cleaves a specific subset of sequences in native DNA, including at least 12 high affinity sites in pBS. However, most phosphodiester bonds are attacked very weakly, if at all. Aside from a requirement for a purine or short run of purines immediately 3′ to the cleavage site, and a preference for a symmetrical run of pyrimidines 5′ to it in a subset of the sites, there is no unambiguous, highly specific consensus sequence for cleavage. Although we observe that supercoils are cleaved at an increased rate, specificity is unaffected by superhelicity. Thus, the information for cleavage specificity in vitro resides in the DNA sequence.

The mechanism by which AP endonucleases recognize AP sites is not yet clear. These enzymes have three activities: AP endonuclease, 3′–5′ exonuclease, and RNaseH, utilizing a common active site (Barzilay et al., 1995a, 1995b; Weiss, 1976). The substrates for all three Exo III–mediated reactions contain DNA backbones that are either deformed relative to B-DNA structure or, in the case of the terminal sequences digested by the exonuclease activity, can be readily deformed through breathing. Thus, one hypothesis for how AP endonucleases might recognize their target phosphodiester bonds is that a unique phosphodiester backbone conformation is required for cleavage. Thus, the retrotransposon endonucleases may also

recognize an altered backbone conformation, but one different from that recognized by AP endonucleases. Interestingly, the DNA segment containing the L1 EN cleavage hotspot migrates anomalously slowly on polyacrylamide gels, suggesting that its conformation differs from B-DNA. We have also investigated a severely bent DNA, the kinetoplast DNA, as a target for L1 EN and find that it is a very good target for cleavage. However, the preferred cleavages are limited to a few phosphodiester bonds within this sequence, even though many more sequences in this DNA are acutely bent. Thus, severe DNA bending is compatible with, but not sufficient for, defining an L1 ENp hotspot.

An alternative hypothesis for recognition is based on the possibility that a severe structural distortion might be recognized. A cocrystal of Exo III with dCMP shows that the nucleotide is within a cleft in the enzyme, considerably distant from the proposed helix-binding surface (Mol et al., 1995). This ability to bind a nucleotide could be interpreted as evidence that Exo III recognizes AP sites by what could be called a nucleotide flipping mode, similar to the recognition mechanism utilized by the DNA modification methylases (Klimasauskas et al., 1994; Roberts, 1994). In the latter enzymes, the base to be modified is flipped or rotated completely out of the double-helical substrate so that the base can be accessed by the catalytic site(s) of the enzyme. In any case, physical study of L1 ENp complexed to DNA will be useful in answering these questions as well as possibly providing insights into how AP endonucleases recognize their targets.

The consensus sequence cleaved by L1 ENp is consistent with a preferred integration site for L1 elements. Cleavage at this sequence within AT-rich DNA, followed by limited breathing of the 3′ end, would generate a structure that could readily hybridize to the 3′ poly(A) tail of L1 RNA, stabilizing what might otherwise be a rather unstable structure (Figure 6A). The sequence we have identified in L1 EN targets is similar to that defined by Jurka (submitted) as a target signal for Alu, B1, B2, and ID elements, all of which, like L1 elements, have 3′ poly(A) tails. This and other evidence (Smit et al., 1995) suggest that L1 EN and RT may be responsible for the transposition of such elements, which do not encode proteins.

### Essential Function of the EN Domain in Retrotransposition

Three models for the function of L1 EN can be imagined on the basis of the known activities of AP endonucleases: 3′–5′ exonuclease proofreading; RNase H activity; or target site definition (the equivalent of the retroviral integrase function). The proofreading hypothesis is probably incorrect, because true 3′–5′ exonuclease activity (releasing 5′ dNMPs) is not detected (Q. F. and J. D. B., unpublished data), and, importantly, proofreading is not expected to be an essential function for the transposon, but only to increase its fidelity. It seems unlikely a priori that an element that uses RNA polymerase to make its first strand would require a proofreading function for second-strand synthesis.

The second model, the possibility of an RNase H activity, is consistent with an essential transposition function, but we have been unable to detect any RNase H activity in L1 ENp in vitro (Q. F. and J. D. B., unpublished data). It is unknown whether an element-encoded RNase H is required for L1, especially if reverse transcription occurs in the nucleus, where cellular RNase H should be available. Certain retroelements utilize cellular RNases H (Shimamoto et al., 1995; Wang and Lambowitz, 1993b). Finally, homologies to RNase H have been reported for the Drosophila I factor, and these are located downstream of the RT domain in ORF2 of the I factor (Abad et al., 1989).

According to the third model, L1 EN is likely to be responsible for target DNA cleavage in the L1 retrotransposition mechanism. Obviously, target cleavage is an essential step. As predicted by this model, L1 ENp produces the 5′-$PO_4$, 3′-OH termini required for priming reverse transcription. Its preference for supercoiled DNA might reflect the state of preferred chromosomal DNA targets. Its relatively sluggish specific activity might reflect the fact that it needs to make only one (or two) cleavages to effect a complete retrotransposition event. Interestingly, retroviral integrases have comparably low turnover numbers, which is likely to be significant evolutionarily but not mechanistically. Finally, L1 EN cleavage sites resemble inferred sites of L1 reverse transcription priming in human cells. The biochemical properties of the L1 EN activity are most consistent with a role in target site cleavage.

The current transposition model for poly(A) elements is based on in vitro data on R2Bm. In R2Bm, both sequence-specific endonuclease activity and RT activity are encoded by a single transposon ORF. It has been proposed that non-sequence-specific elements such as L1 might employ a similar mechanism, but since no endonuclease activity was known from these elements, it was speculated that cellular endonucleases mediated transposition. Several lines of evidence suggest that non-sequence-specific retroelements employ the same modus operandi as the sequence-specific ones: first, a nuclease-homologous sequence is conserved in the non-sequence-specific poly(A) elements; second, the N-terminus of human L1 ORF2 encodes an endonuclease activity; third, the in vitro specificity of L1 EN resembles L1 in vivo target sequences; and fourth, EN domain mutations kill retrotransposition. We propose that for all poly(A) elements, an encoded endonuclease activity generates target site nicks; subsequently, the target nick is used as a primer for reverse transcription of the element RNA 3′ end. In contrast with retroviruses, target site recognition and reverse transcription are coupled in these elements.

### Distribution of EN Domain among Retroelements

The EN domain is found in a diverse collection of retrotransposons from fungi to plants to mammals. However, it is absent from many elements (Table 1). The R2Bm element, which encodes a single protein with both sequence-specific endonuclease and RT activities, lacks it. R2Bm presumably contains a distinct class of EN domain or perhaps has diverged so radically that it is no longer recognizable. Among the poly(A) elements, only about half of the sequence-specific ones encode

an EN domain. The phylogenetically related group II intron ORFs, which mediate the homing of those introns to intronless target sites (and, like R2Bm, are sequence-specific), lack the EN domain. These remarkable elements contain two components to their nucleolytic mechanism: the RNA lariat mediates cleavage of the sense strand of the target via a full or partial reverse splicing mechanism, and a zinc finger–like nuclease domain located C-terminal to the RT domain cleaves the antisense strand (Yang et al., 1996; Zimmerly et al., 1995b). All but one of the non-sequence-specific class of poly(A) elements we have examined contain the EN domain, and in nearly all cases, the domain defines the N-terminus of the ORF2 conceptual translation product. The exceptions to the latter are the L1Tc and *ingi* elements, which encode the EN domain in a separate ORF. The single poly(A) element that is non-sequence-specific and lacks the domain is the Dong element, which contains the first motif (QET) at the N-terminal of ORF2 but lacks the remaining motifs, suggesting that the sequenced Dong copy represents an internal deletion mutant of ORF2.

## Evolutionary Considerations

The immense diversity of retroelement types can be categorized into two major branches by comparing RT sequences (Xiong and Eickbush, 1990). These are the poly(A) element branch and the LTR element branch, including the retroviruses. The poly(A) branch includes L1 and other poly(A) elements, the group II introns, the retroplasmids, and the prokaryotic msDNAs. The latter three classes inhabit prokaryotic genomes (or their endosymbiotic organellar genome descendants). Lambowitz and colleagues have suggested that the Mauriceville retroplasmid RT might represent a very primitive form of RT, because it shares with RNA-directed RNA polymerases the ability to carry out unprimed synthesis. As such, it may be descended from a missing link between the RNA world and the DNA world (Wang and Lambowitz, 1993a). Presumably, a gene encoding such an RT acquired new coding sequences at some point, enabling its mobility. Such acquisitions occurred multiple times in evolution; the poly(A) elements acquired the EN domain, perhaps others (the progenitors of R2Bm?) acquired a different type of endonuclease, and yet others, the progenitors of infectious group II introns, became associated with catalytic RNA components, as well as a zinc finger–like nuclease. Debate is sure to rage about the relative timing of such acquisitions, but it seems likely that they occurred early in retroelement evolution, because the EN domain is widely represented and highly successful in the eukaryotic lineage, whereas the group II intron/zinc domain acquisition appears restricted to the organellar lineage. It will be of great interest to know whether these elements are found in the archaeal lineage.

Both efficiency and regulation are offered by self-encoding an endonuclease activity rather than utilizing a cellular endonuclease. Coupling reverse transcription to target site recognition may increase retrotransposition efficiency. Endonuclease could be deleterious to host cells and might have to be regulated. Thus, nicking activity in full-length L1 ORF2 protein may be activated only when reverse transcription is possible (i.e., when both template RNA and target DNA are properly engaged). Finally, since the specificity of target site definition appears to be provided by L1 EN, it may be possible to modulate the specificity of L1 insertion by manipulating L1 EN cleavage specificity.

## Experimental Procedures

### Plasmids and Strains

The L1 EN domain was PCR-amplified with primers JB1073 5′-CCT CATGACAGGATCAAATTCACAC-3′ and JB1083 5′-GCCCATGGCA ATCCTGAGTTCTAGTTTG-3′ from the pL1.2 A DNA plasmid and cloned into the pCR(II) vector (Invitrogen), resulting in pQF218. Five different point mutations (Figure 1) were introduced by site-directed mutagenesis. Each mutation was PCR-amplified and cloned as above and verified by DNA sequencing. For expression of L1 ENp and mutant proteins in E. coli, the BspHI–NcoI fragments from pQF218 and its mutant derivatives were cloned into the NcoI site of pET15b (Novagen) and transformed into strain BL21(DE3) for protein production.

### Expression and Purification of L1 ENp

Induction and purification of $His_6$-tagged protein followed the protocol from Qiagen (catalog number 30210). Cells were grown at 37°C in Luria broth containing 50 μg/ml ampicillin to an $A_{600}$ of 0.8, IPTG was added to 1 mM for another 3 hr, and cells were pelleted and stored at −20°C. Cells from a 10 ml culture were thawed at 0°C for 30 min, resuspended in 0.3 ml of buffer (50 mM sodium phosphate [pH 8.0], 300 mM NaCl), and disrupted by sonication for 30 s. The clarified supernatant (12,000 rpm, 20 min) was mixed with 0.1 ml of preequilibrated nickel–agarose (Qiagen) at 4°C for 1 hr. The nickel–agarose was centrifuged and washed twice with 0.5 ml of sonication buffer, twice with 0.5 ml of washing buffer (50 mM sodium phosphate, 300 mM NaCl, 10% glycerol [pH 6.0]), twice with 0.5 ml of washing buffer containing 0.7 M NaCl, and twice with 0.5 ml of washing buffer containing 30 mM imidazole. Finally, the protein was eluted with 0.25 ml of washing buffer containing 100 mM imidazole and 0.25 ml of washing buffer containing 150 mM imidazole; most protein was eluted in the 100 mM imidazole fraction. Either eluted fraction was directly used in the endonuclease nicking assay.

### Nicking Assay

Supercoiled pBS (Bluescript KS(−), Stratagene) DNA was prepared by double banding in cesium chloride–ethidium bromide (Maniatis et al., 1982). Partially depurinated DNA (AP-DNA) preparations of pBS was prepared by incubation of purified supercoiled DNA in 8 mM Tris–HCl (pH 7.5), 0.8 mM EDTA, 20 mM sodium citrate (pH 5.0), and 0.2 M NaCl for 20 min at 70°C, followed by chilling on ice (Johnson and Demple, 1988). The extent of AP-DNA generated was determined by the minimum amount of depurination leading to quantitative conversion of supercoiled DNA to nicked circles upon treatment with Exo III.

The L1 EN reaction mix contained 50 mM HEPES–KOH (pH 7.6), 50 mM KCl, 5 mM $MgCl_2$, 100 μg/ml BSA, 0.2 μg of supercoiled DNA, and 20 ng of purified protein in a total reaction volume of 25 μl at 37°C for 20 min or as indicated. The reaction was stopped by 25 mM EDTA or heating to 69°C for 10 min. Half the reaction mix was loaded on a 1% agarose gel in TTE buffer containing 0.5 μg/ml ethidium bromide. For the ligation experiment, after the supercoiled DNA was incubated with L1 ENp, it was heated at 69°C for 10 min to inactivate the residual L1 ENp activity. T4 DNA ligase and 1 mM ATP were added and incubated at 16°C overnight. Subsequently the T4 DNA ligase was heat inactivated, and L1 EN was added to the reaction as indicated.

The AP-endonuclease activity of Exo III was assayed as follows: the reaction mix contained 50 mM Tris–HCl (pH 8.0), 5 mM $CaCl_2$, 1 mM β-mercaptoethanol, 0.2 μg of AP-DNA, and 1 μl of Exo III (100 U/μl, NEB) in 25 μl. The reaction was performed at 37°C for 20 min.

### Determination of Specificity of Cleavage Sites

L1 ENp-digested supercoiled DNA was run on gels, and linear products were electroeluted and digested with various restriction enzymes. Primer extension with Taq polymerase was performed on nicked and linear DNA generated from L1 ENp cleavage by using kinased primers JB1132 5′-TCTTTTCTACGGGGTCTG-3′, and JB1133 5′-CAGGCAACTATGGATGAA-3′, and the T7 primer, 5′-AAT ACGACTCACTATAG-3′. The reaction mix was loaded side by side with a sequencing reaction on supercoiled DNA carried out using the same $^{32}$P-labeled primers; cleavage sites were determined by comparison with these standards. For the K-DNA, primer SP6 5′-AGCTATTTAGGTGACACTATAG-3′ was used.

### Tissue Culture Analysis of Transposition

Wild-type and mutant derivatives of the pL1.2mneo-I plasmid were introduced into HeLa cells by lipofection, and hygromycin-resistant cell populations were obtained as described (Moran et al., 1996). The efficiency of plating of these cells on G418-containing medium was then measured to obtain the reported transposition frequencies.

### Acknowledgments

Correspondence should be addressed to J. D. B. We thank David Wilson, Bernard Weiss, Rick Cunningham, John Tainer, and Bruce Demple for helpful discussions on AP endonucleases. We thank Lys Guilbride, Dwight Nissley, Jeff Smith, Ralph DeBerardinis, and Greg Cost for providing plasmids and helpful comments. This work supported in part by a Damon Runyon Postdoctoral Fellowship DRG-1332 (J. V. M.) and National Institutes of Health grants GM-45398 (H. H. K.) and CA-16519 (J. D. B.).

Received August 6, 1996; revised September 19, 1996.

### References

Abad, P., Vaury, C., Pélisson, A., Chaboissier, M.C., Busseau, I., and Bucheton, A. (1989). A long interspersed repetitive element—the I factor of *Drosophila teissieri*—is able to transpose in different *Drosophila* species. Proc. Natl. Acad. Sci. USA *86*, 8887–8891.

Barzilay, G., Mol, C.D., Robson, C.N., Walker, L.J., Cunningham, R. P., Tainer, J.A., and Hickson, I.D. (1995a). Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAP1. Nature Struct. Biol. *2*, 561–568.

Barzilay, G., Walker, L.J., Robson, C.N., and Hickson, I.D. (1995b). Site-directed mutagenesis of the human DNA repair enzyme HAP1: identification of residues important for AP endonuclease and RNase H activity. Nucl. Acids Res. *23*, 1544–1550.

Boeke, J.D., and Stoye, J.P. (1996). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In Retroviruses, H. Varmus, S. Hughes and J. Coffin, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), in press.

Branciforte, D., and Martin, S.L. (1994). Developmental and cell-type specificity of LINE-1 expression in mouse testis: implications for transposition. Mol. Cell. Biol. *14*, 2584–2592.

Bucheton, A., Paro, R., Sang, H.M., Pélisson, A., and Finnegan, D.J. (1984). The molecular basis of I–R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. Cell *38*, 153–163.

Burkhoff, A.M., and Tullius, T.D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. Cell *48*, 935–943.

Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F., and Kazazian, H.H., Jr. (1991). Isolation of an active human transposable element. Science *254*, 1805–1808.

Dombroski, B., Feng, Q., Mathias, S.L., Sassaman, D.M., Scott, A.F., Kazazian, H.H., and Boeke, J.D. (1994). An *in vivo* assay for the reverse transcriptase of human retrotransposon L1 in Saccharomyces cerevisiae. Mol. Cell Biol. *14*, 4485–4492.

Drolet, M., Bi, X., and Liu, L.F. (1994). Hypernegative supercoiling of the DNA template during transcription elongation *in vitro*. J. Biol. Chem. *269*, 2068–2074.

Fanning, T.G., and Singer, M.F. (1987). LINE-1: a mammalian transposable element. Biochim. Biophys. Acta. *910*, 203–212.

Hohjoh, H., and Singer, M.F. (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J. *15*, 630–639.

Hohjoh, H., Minakami, R., and Sakaki, Y. (1990). Selective cloning and sequence analysis of the human L1 (Line-1) sequences which transposed in the relatively recent past. Nucl. Acids Res. *18*, 4099–4104.

Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. Nature Genet. *7*, 143–148.

Johnson, A.W., and Demple, B. (1988). Yeast DNA 3′-repair diesterase is the major cellular apurinic/apyrimidinic endonuclease: substrate specificity and kinetics. J. Biol. Chem. *263*, 18017–18022.

Kazazian, H.H., Jr., Wong, C., Youssoufian, H., Scott, A.F., Phillips, D.G., and Antonarakis, S.E. (1988). Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. Nature *332*, 164–166.

Kennell, J.C., Moran, J.V., Perlman, P.S., Butow, R.A., and Lambowitz, A.M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. Cell *73*, 133–146.

Kinsey, J.A. (1990). *Tad*, a LINE-like element of *Neurospora*, can transpose between nuclei in heterokaryons. Genetics *126*, 317–326.

Kitchin, P.A., Klein, V.A., Ryan, K.A., Gann, K.L., Rauch, C.A., Kang, D.S., Wells, R.D., and Englund, P.T. (1986). A highly bent fragment of *Crithidia fasciculata* kinetoplast DNA. J. Biol. Chem. *261*, 11302–11309.

Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357–369.

Lazowska, J., Meunier, B., and Macadre, C. (1994). Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. EMBO J. *13*, 4963–4972.

Lehman, I.R. (1974). DNA ligase: structure, mechanism, and function. Science *186*, 790–797.

Lilley, D. (1981). Hairpin-loop formation by inverted repeats in supercoiled DNA is a local and transmissible property. Nucl. Acids Res. *9*, 1271–1288.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA *84*, 7024–7027.

Luan, D., and Eickbush, T.H. (1995). RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol. Cell. Biol. *15*, 3882–3891.

Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell *72*, 595–605.

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

Martín, F., Marañon, C., Olivares, M., Alonso, C., and Lopez, M.C. (1995). Characterization of a non–long terminal repeat retrotransposon cDNA (L1Tc) from *Trypanosoma cruzi*: homology of the first ORF with the Ape family of DNA repair enzymes. J. Mol. Biol. *247*, 49–59.

Martin, S.L. (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. Mol. Cell. Biol. *11*, 4804–4807.

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

Meunier, B., Tian, G.L., Macadre, C., Slonimski, P.P., and Lazowska,

J. (1990). Group II introns transpose in yeast mitochondria. In Structure, Function and Biogenesis of Energy Transfer Systems, E. Quagliariello, S. Papa, F. Palmieri, and C. Saccones, eds. (Amsterdam: Elsevier), pp. 169–174.

Michel, F., and Lang, B.F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature *316*, 641–643.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B., and Nakamura, Y. (1992). Disruption of the *APC* gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res. *52*, 643–645.

Mol, C.D., Kuo, C., Thayer, M.M., Cunningham, R.P., and Talner, J.A. (1995). Structure and function of the multifunctional DNA-repair enzyme exonuclease III. Nature *374*, 381–386.

Moran, J.V., Mecklenburg, K.L., Sass, P., Belcher, S.M., Mahnke, D., Lewin, A., and Perlman, P.S. (1994). Splicing defective mutants of the *COX1* gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. Nucl. Acids Res. *22*, 2057–2064.

Moran, J.V., Zimmerly, S., Eskes, R., Kennell, J.C., Lambowitz, A.M., Butow, R.A., and Perlman, P.S. (1995). Mobile group II introns of yeast mtDNA are novel site-specific retroelements. Mol. Cell. Biol. *15*, 2828–2838.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. Cell *87*, this issue.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R., Nakamura, H., and Matsuo, M. (1993). Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J. Clin. Invest. *91*, 1862–1867.

Oefner, C., and Suck, D. (1986). Crystallographic refinement and structure of DNase I at 2 Å resolution. J. Mol. Biol. *192*, 605–632.

Pettijohn, D.E., and Pfenninger, O. (1980). Supercoils in prokaryotic DNA restrained *in vivo.* Proc. Natl. Acad. Sci. USA *77*, 1331–1335.

Price, P. (1975). The essential role of $Ca^{2+}$ in the activity of bovine pancreatic deoxyribonuclease. J. Biol. Chem. *250*, 1981–1986.

Roberts, R.J. (1994). An amazing distortion in DNA induced by a methyltransferase. Biosci. Rep. *14*, 103–117.

Schwarz-Sommer, Z., Leclercq, L., Gobel, E., and Saedler, H. (1987). Cin 4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J. *6*, 3873–3880.

Scott, A.F., Schmeckpeper, B.J., Abdelrazik, M., Comey, C.T., O'Hara, B., Rossiter, J.P., Cooley, T., Heath, P., Smith, K.D., and Margolet, L. (1987). Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics *1*, 113–125.

Shimamoto, T., Shimada, M., Inouye, M., and Inouye, S. (1995). The role of ribonuclease H in multicopy single-stranded DNA synthesis in retron-Ec73 and retron-Ec107 of *Escherichia coli*. J. Bacteriol. *177*, 264–267.

Sinden, R.R., and Pettijohn, D.E. (1981). Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. Proc. Natl. Acad. Sci. USA. *78*, 224–228.

Sinden, R.R., Carlson, J.O., and Pettijohn, D.E. (1980). Torsional tension in the DNA double helix measured with trimethylpsoralen in living E. coli cells: analogous measurements in insect and human cells. Cell *21*, 773–783.

Smit, A.F.A., Toth, G., Riggs, A.D., and Jurka, J. (1995). Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J. Mol. Biol. *246*, 401–417.

Stellwagen, N.C. (1983). Anomalous electrophoresis of deoxyribonucleic acid restriction fragments on polyacrylamide gels. Biochemistry *22*, 6186–6193.

Suck, D., Lahm, A., and Oefner, C. (1988). Structure refined to 2 Å of a nicked DNA octanucleotide complex with DNase I. Nature *332*, 464–468.

Tsao, Y.P., Wu, H.Y., and Liu, L.F. (1989). Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. Cell *56*, 111–118.

Wang, H., and Lambowitz, A.M. (1993a). The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to the progenitor of reverse transcriptases and DNA polymerases. Cell *75*, 1071–1081.

Wang, H., and Lambowitz, A.M. (1993b). Reverse transcription of the Mauriceville plasmid of *Neurospora*. J. Biol. Chem. *268*, 18951–18959.

Weiss, B. (1976). Endonuclease II of *Escherichia coli* is exonuclease III. J. Biol. Chem. *251*, 1896–1901.

Woods-Samuels, P., Wong, C., Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., and Antonarakis, S.E. (1989). Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. Genomics *4*, 290–296.

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

Yang, J., Zimmerly, S., Perlman, P.S., and Lambowitz, A.M. (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. Nature *381*, 332–335.

Zimmerly, S., Guo, H., Perlman, P.S., and Lambowitz, A.M. (1995a). Group II intron mobility occurs by target DNA-primed reverse transcription. Cell *82*, 545–554.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., and Lambowitz, A.M. (1995b). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell *83*, 529–538.