Transposable elements: **How non-LTR retrotransposons do it**

D.J. Finnegan

The source of the enzyme activity responsible for the transposition of retrotransposons of the type that lack terminal repeats has at last been identified: in *L1Hs* elements, it is encoded by the second open reading frame and is a nuclease related to the apurinic repair endonucleases.

Address: Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh, EH9 3JR, UK.

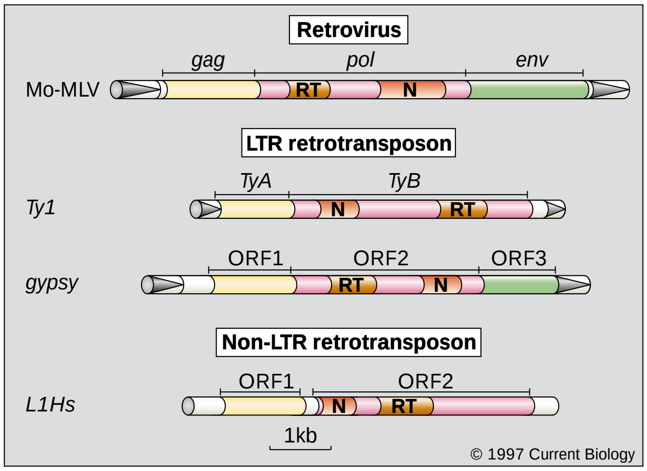
Electronic identifier: 0960-9822-007-R0245

Current Biology 1997, 7:R245–R248

Transposable elements make up a substantial proportion of the total DNA in most, if not all, eukaryotic genomes. These elements fall into two classes: the retrotransposons that transpose by a process involving reverse transcription, and the transposons that move by an excision-insertion mechanism. During the transposition of a retrotransposon, an RNA copy of the element is converted into DNA by a reverse transcriptase, and this DNA is inserted at a new site in the genome. Surprisingly, until recently no element-encoded enzyme had been identified to carry out this last step for the most abundant transposable element in human DNA —the *L1Hs* elements that are members of the class of retrotransposons known as long interspersed nucleotide elements (LINEs). Two papers [1,2] have now shown that this reaction is carried out by a novel nuclease encoded by the *pol*-like gene of these elements.

Retrotransposons are of two types, the best known of which resemble retroviral proviruses in having long terminal repeats (LTRs) and open reading frames equivalent to the *gag*, *pol* and, in some cases, *env* genes of retroviruses (Fig. 1). Elements of this type include the *Ty* elements of the yeast *Saccharomyces cerevisiae*, and the *copia* and *gypsy* elements of *Drosophila melanogaster*. During a cycle of retroviral infection, the corresponding provirus is transcribed into full-length RNAs that are first packaged into viral core particles and then into enveloped virions that are released from one cell to infect another. After infection, the genomic RNA is copied into linear extrachromosomal DNA molecules that integrate into the genome of the host cell. The enzymes responsible for these reactions, reverse transcriptase and integrase, are encoded by different domains of the *pol* gene of the virus. Reverse transcriptase, like all DNA polymerases, requires a primer to initiate DNA synthesis, and this is provided by a tRNA hybridized to the viral RNA near its 5′ end.

Figure 1



The organization of the genomes of the retrovirus Moloney murine leukaemia virus (Mo-MLV), the LTR-containing retrotransposons *Ty1* of *S. cerevisiae* and *gypsy* of *D. melanogaster*, and the non-LTR retrotransposon *L1Hs* of humans. The positions of the open reading frames within these elements are shown. The LTRs of those elements that have them are indicated by the arrowheads. Within the second open reading frame of each element, the regions coding for the reverse transcriptase (RT) and nuclease (N) domains are indicated.

The structural similarities between retroviruses and LTR-containing retrotransposons suggest that the latter transpose by a mechanism similar to this cycle, but without the formation of infectious particles. There is evidence to support this for many LTR retrotransposons, and they all have recognizable reverse transcriptase and integrase domains encoded by their *pol*-like open reading frames. The clearest demonstration of the relationship between retroviral infection and retrotransposition comes from *in vitro* and *in vivo* experiments with *Ty1* elements. These studies have shown that *Ty1* RNA is reverse transcribed into linear DNA within virus-like particles [3] using the initiator methionine tRNA as primer [4]. These extrachromosomal molecules are then inserted into target DNA in a reaction requiring the putative *Ty1* integrase [3].

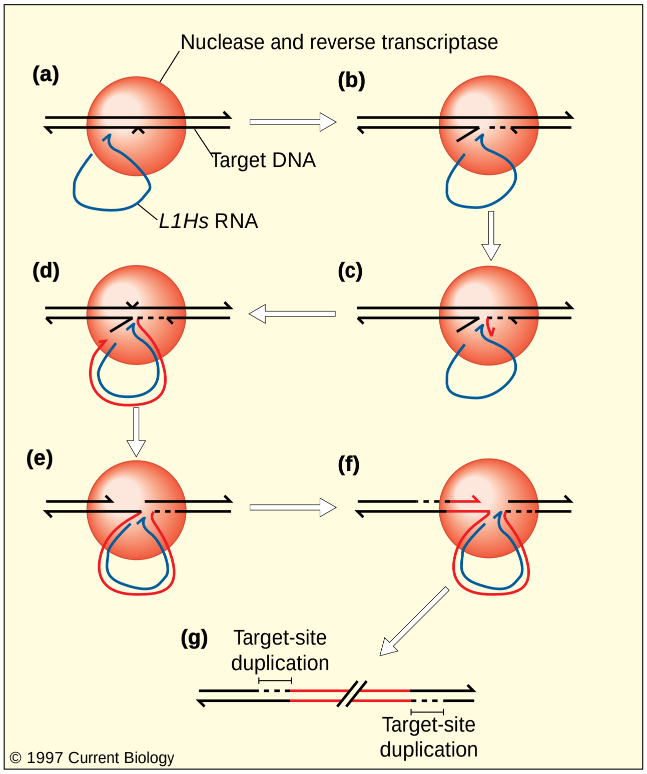
Retrotransposons of the second type, the LINES or ‘non-LTR’ elements, have no terminal repeats. These elements, originally discovered in mammalian genomes [5], have now been detected in a wide range of species from protozoa to fungi, plants and animals, and they generally have two open reading frames. There is little, if any, sequence similarity between the proteins encoded by the first of these open reading frames, although they may all be nucleic-acid-binding proteins that are able to form ribonucleoprotein particles during transposition [6,7]. The proteins encoded by the second open reading frames of these elements, however, are related in sequence to reverse transcriptases, and elements of this type have been shown to transpose by a process involving reverse transcription [8–10]. To this extent, transposition of non-LTR retrotransposons is similar to that of the retrovirus-like elements. There must be some differences, however, as they have no obvious mechanism for priming reverse transcription. Furthermore, there is no indication as to how the products of reverse transcription integrate into the host genome, as the proteins that they encode do not include recognizable integrase domains.

Most non-LTR retrotransposons can be found inserted at many sites within a genome, and the target-site duplications flanking them vary in both length and sequence. This has led several authors to suggest that integration takes place at sites at which chromosomal DNA has been nicked by host-encoded products, as might be expected to occur during DNA repair or recombination. The idea is that the 3′ hydroxyl groups present at such sites could serve as primers for reverse transcription of element-encoded RNA. Synthesis of the second strand might be primed in a similar way at an adjacent nick in the opposite strand, the nucleotides between the two nicks giving rise to the target-site duplication (Fig. 2).

Although integration at ‘random’ breaks in chromosomal DNA may be reasonable for elements that can insert at many sites, this cannot account for the integration of the subset of non-LTR retrotransposons that insert at specific sequences. This subset includes the elements found only at specific sites within ribosomal RNA genes of insects, and elements inserted within the mini-exon genes of some trypanosomes. Elements of this kind presumably require the activity of a nuclease that cleaves the insertion site specifically, whether or not this is used to prime reverse transcription. In this case, one might expect an element to encode a nuclease able to recognize and cleave its own integration site.

This is indeed the case, at least for the *R2Bm* element that inserts within the 28S rRNA genes of *Bombyx mori*. This element encodes a single protein that has a sequence-specific nuclease activity that recognizes the sequence of the *R2Bm* integration site, where it makes staggered nicks two base pairs apart [11]. RNA is a cofactor for this reaction and, in the presence of an RNA corresponding to the 3′ end of a full-length *R2Bm* transcript, the two strands of the insertion site are nicked at different rates, with the 3′ hydroxyl at the first nick being used to prime DNA synthesis using *R2Bm* RNA as template. Most DNA synthesis takes place before the second strand is cleaved, as would be expected if its 3′ hydroxyl is to be used as the primer for the second strand, as it would not be required until synthesis of the first strand had taken place. So far, there is no direct evidence that synthesis of the second strand takes place in this way, and second-strand synthesis has not been detected *in vitro*. The region of *R2Bm* responsible for this nuclease activity has not been identified but it is not obviously related to a retroviral integrase.

Figure 2



A possible mechanism for reverse transcription and integration of a new copy of *L1Hs*. **(a)** The ORF2 product, with reverse transcriptase and nuclease domains, binds to both the target DNA and full-length *L1Hs* RNA. The nuclease domain then cleaves one strand of the target at the site of integration. **(b)** DNA synthesis is initiated using the 3′ hydroxyl of the broken strand of the target DNA as primer and a short region of the opposite strand as template. This generates the first copy of the target-site duplication. **(c)** DNA synthesis continues using the 5′ end of *L1Hs* RNA as template. **(d)** When DNA synthesis reaches the 3′ end of *L1Hs* RNA, the second strand of the target DNA is cut. **(e)** The 3′ end of the first strand *L1Hs* DNA is joined to the 5′ end of the target DNA. **(f)** Synthesis of the second strand of L1Hs is initiated using the new 3′ hydroxyl of target DNA as primer. The short exposed strand of target DNA is used as template to generate the second copy of the target-site duplication. Synthesis then proceeds using the first strand of *L1Hs* DNA as template. **(g)** Synthesis of the new copy of *L1Hs* is complete and is flanked by the target-site duplication. An arrowhead indiactes the 5′ end of each strand of nucleic acid.

These results strongly support the idea that site-specific LINE-like elements insert by a mechanism in which integration and reverse transcription take place simultaneously. Is this also true of the related elements that insert throughout the genome, and do they encode a nuclease that can initiate the process? The first suggestion that this might be so came with the realization [12] that the second open reading frames of several such elements potentially encode polypeptides with regions of similarity to members of the apurinic (AP) family of nucleases (Fig. 1), so-called because of their ability to cleave abasic sites in DNA. These include exonuclease III of *Escherichia coli*, Rrp1 of *Drosophila melanogaster* and Ap1 of humans. This sequence was detected in proteins encoded by several elements that insert at many sites, raising the possibility that a nuclease activity may be associated with most, if not all, such elements and that this might play a role in reverse transcription and integration.

This has been tested recently for the human LINE element, *L1Hs*, in experiments reported in two papers by Jef Boeke and Haig Kazazian and their colleagues [1,2]. The protein encoded by the second open reading frame of *L1Hs* contains a putative AP nuclease domain near its amino terminus. This has been expressed in *E. coli* and its nuclease activity tested on supercoiled and relaxed, covalently closed plasmid DNAs. Both substrates were cleaved to give linear molecules, although the supercoiled DNA was the preferred substrate. This activity is associated with the AP-related sequence, as it was greatly reduced when mutations were introduced that changed some of the amino-acid residues common to the *L1Hs* protein and members of the AP family. Abasic sites were not required for nuclease activity, however, as the *L1Hs* protein cut both apurinic and native plasmid DNA whereas exonuclease III only cut the apurinic substrate [1].

If the AP-like nuclease is responsible for cleaving insertion sites then, if its amino acid sequence were altered, *L1Hs* transposition should be reduced. This has been tested in tissue culture cells, using an *L1Hs* element marked so as to allow selection of cells in which the element has transposed [2]. A modified neomycin resistance gene containing an intron was inserted at the 3′ end of an active *L1Hs* element in the opposite transcriptional orientation. The intron was oriented so that it should be spliced from an *L1Hs* transcript, but not from a transcript of the neomycin resistance gene itself. As a result, colonies resistant to the neomycin analogue G418 could only be recovered if the marked element had been transcribed into an RNA that was processed to remove the intron and then reverse transcribed and integrated at a new site. The marked element was placed under the control of a cytomegalovirus promoter to ensure efficient transcription in HeLa cells. Cells resistant to G418 and containing transposed copies of the marked element were readily obtained when the sequence of the AP-like nuclease domain was intact, but were reduced over 100-fold if the conserved residues were altered. This clearly indicates that the *L1Hs* nuclease is important for transposition.

There are few biological events, if any, that are truly random, and this is certainly true of transposition. Whenever new insertion sites for a particular element have been mapped, even for an element than can transpose throughout the host genome, ‘hot’ and ‘cold’ spots are found at which insertions are more or less frequent. This can be seen both on a large scale, some loci being preferred over others, and at the nucleotide level, with some sequences within a preferred locus having more insertions than others. This presumably reflects some form of target-site selection by the transposition machinery. In the case of *L1Hs* this might be due to a degree of sequence specificity of the AP-like nuclease, as this enzyme appears to initiate integration. If so, the sequences cut by the enzyme *in vitro* should be similar to the sites at which *L1Hs* inserts *in vivo*.

The sequence at which the *L1Hs* endonuclease cleaved supercoiled plasmid DNA were mapped and found to be confined to a small region of the molecule [1]. The breaks in each strand were located precisely by primer extension experiments and found to be confined to a region of about 150 base pairs, with six major cleavage sites in each strand. Each cleavage site had a short run of purines, usually adenosines, to the 3′ side of the break, and about half had a short run of pyrimidines to the 5′ side. A similar distribution of bases was found at sites at which the marked *L1Hs* element had inserted in HeLa cells and in DNA insertions flanking wild-type *L1Hs* elements recovered previously by others.

These results support the notion that non-LTR retrotransposons that integrate at many sites also transpose by a mechanism in which integration and reverse transcription are coupled (Fig. 2). There is no direct evidence that the 3′ hydroxyl group generated by this event is used to prime reverse transcription, although this appears to be very likely. This will have to be investigated by further *in vitro* experiments in which transcripts from *L1Hs*, or a similar element, are mixed with target DNA and the endonuclease and reverse transcriptase that the element encodes, either as a single protein or as two separate molecules.

Most non-LTR retrotransposons have a deoxyadenosine-rich sequence at the 3′ end of the coding strand, often as a poly(dA) sequence. This is also true of the highly repeated short interspersed nucleotide elements (SINEs) which occur in several hundred thousand copies in the genomes of many higher eukaryotes [13]. The SINEs are related in sequence to the transcripts of polymerase III transcribed genes and appear to have been produced by reverse transcription and integration. They have no coding capacity themselves and must rely on enzymes from elsewhere, an obvious source being non-LTR retrotransposons. SINE transcripts might occasionally be incorporated into the ribonucleoprotein particles that are thought to be transposition intermediates for non-LTR retrotransposons, and in this way be carried into the nucleus to be reverse transcribed and integrated at new sites. Processed pseudogenes might have a similar origin, although there is some evidence that retroviruses ma be involved in their formation [14] .

The fact that non-LTR retrotransposons code for an enzyme that cleaves target sites during transposition does not necessarily mean that integration cannot take place at breaks in chromosomal DNA introduced by other means. In recombination-deficient *S. cerevisiae* mutants, transposition of sequences mediated by *Ty1* elements can heal double-stranded breaks introduced at the *MAT* locus by HO endonuclease [15,16]. This was seen with strains containing marked *Ty1* elements, and with similar elements in which sequences coding for the *Ty1* reverse transcriptase had been replaced by those coding for the reverse transcriptases of the non-LTR retrotransposons *L1Hs* or *cre* [15]. This suggests that retrotransposons may play a role in healing breaks in chromosomal DNA produced in one way other another. All in all, there are many ways in which these elements may influence genome structure and evolution.

References

1. Feng O, Moran JV, Kazazian HH, Boeke JD: Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 1996, 87:905–916.
2. Moran JV, Holmes SE, Naase TP, DeBerardinis RJ, Boeke JD, Kazazian HH: High frequency retrotransposition in cultured mammamlian cells. *Cell* 1996, 87:917–927.
3. Eichinger DJ, Boeke JD: The DNA intermediate in yeast ***Ty1*** element transposition copurifies with virus-like particles: cell-free ***Ty1*** transposition. *Cell* 1988, 54:955–966.
4. Chapman KB, Bystrom AS, Boeke JD: Initiator methionine tRNA is essential for ***Ty1*** transposition in yeast. *Proc Natl Acad Sci USA* 1992, 89:3236–3240.
5. Hutchison CA III, Hardies SC, Loeb DD, Sehee WR, Edgell MH: LINEs and related retrotransposons: long interspersed repeated sequences in the eukaryotic genome. In *Mobile DNA.* Edited by Berg DE, Howe MM. Washington, DC: American Society for Microbiology; 1989:593–617.
6. Dawson A, Hartswood E, Paterson T, Finnegan DJ: A LINE-like transposable element in ***Drosophila***, the ***I*** factor, encodes a protein with properties similar to those of retroviral nucleocapsids. *EMBO J* 1997, in press.
7. Hohjoh H, Singer MF: Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J* 1996, 3:630–639.
8. Jensen S, Gassama MP, Heidmann T: Retrotransposition of the ***Drosophila*** LINE I element can induce deletion in the target DNA: a simple model also accounting for the variability of the normally observed target site duplication. *Biochem Biophys Res Comm* 1994, 202:111–119.
9. Pelisson A, Finnegan DJ, Bucheton A: Evidence for retrotransposition of the ***I*** factor, a ***LINE*** element of ***Drosophila melanogaster***. *Proc Natl Acad Sci USA* 1991, 88:4907–4910.
10. Evans JP, Palmiter RD: Retrotransposition of a mouse ***L1*** element. *Proc Natl Acad Sci USA* 1991, 88:8792–8795.
11. Luan DD, Korman MH, Jakubczak JL, Eickbush TH: Reverse transcription of ***R2Bm*** RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 1993, 72:595–605.
12. Martin F, Maranon C, Olivares M, Alonso C, Lopez MC: 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* 1995, 247:49–59.
13. Deininger PL: SINEs: short interspersed repeated nucleotide elements in higher eukaryotes. In *Mobile DNA.* Edited by Berg DE, Howe MM. Washington: American Society for Microbiology; 1989:619–636.
14. Tchenio T, Segal-Bendirdjian E, Heidmann T: Generation of processed pseudogenes in murine cells. *EMBO J* 1993, 12:1487–1497
15. Teng S-C, Kim B, Gabriel A: Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* 1996, 383:641–644.
16. Moore JK, Haber JE: Capture of retrotransposon DNA at the sites of chromosomal double-stranded breaks. *Nature* 1996, 383:644–646.