Vol. 202, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 15, 1994 Pages 111-119

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 DUPLICATIONS

Silke JENSEN, Marie-Pierre GASSAMA and Thierry HEIDMANN

Institut Gustave Roussy, CNRS URA147, 39 rue Camille Desmoulins
94805 Villejuif, FRANCE

Received May 17, 1994

Retrotransposition of the *Drosophila melanogaster* LINE I element normally generates target site duplications of variable length, as classically observed for most LINE elements. Using an I element "marked" with an indicator gene for *in vivo* detection of transposition that we previously developed, we show that deletion in the target DNA can also take place, as a direct consequence of I element transposition. We propose a simple model accounting for the generation of both target site duplications of variable length and target DNA deletions, which relies upon template switching of the LINE-encoded reverse transcriptase between single-strand DNA at the target site and the LINE template.

LINEs (Long Interspersed Nucleotidic Elements) are highly reiterated mobile elements found in many eukaryotic species including plants, insects and mammals [reviewed in (1, 2)]. LINEs are up to 7kb-long elements which commonly possess two long open reading frames, one (ORF1) with a putative coding domain also found in the nucleic acid binding domain of the *gag* polypeptide of retroviruses, and the other (ORF2) displaying a highly conserved region with homology to retroviral reverse transcriptase [reviewed in (3, 4, 5)]. LINEs possess an A-rich sequence at their 3' end, usually -but not always- preceded by a polyadenylation signal, and many of them are heterogeneously truncated at their 5' end. These features have led to the suggestion that LINEs transpose through an RNA intermediate and its reverse transcription. In a previous study, this assumption was tested for a *Drosophila* LINE -the I element [reviewed in (6, 7)]- that was tagged with an intron-containing indicator gene, developed to detect RNA-mediated transposition by selective methods (8, 9, 10, 11). *In vivo* transposition events of this marked element could be selected in transgenic *Drosophila*, and sequencing of the transposed copies disclosed precise splicing out of the intron thus unambiguously demonstrating "retrotransposition" of this *Drosophila* LINE [(12); see also (13)].

Although several full-length LINEs have now been entirely sequenced and characterized, still little is known on the refined molecular mechanisms of their retrotransposition. One of the open questions concerns their integration mechanism. LINEs are often -but not always- found associated with target site duplications, but unlike the related retrovirus-like transposable elements -which create target duplications of defined length- these duplications are of variable length. This observation, as well

0006-291X/94 $5.00

as the frequent occurrence of 5'-truncated transposed copies suggest that the integration mechanism of LINEs is fundamentally different from that of the retrovirus-like transposons. While R2Bm, a particular site-specific LINE from *Bombyx mori*, systematically induces target site deletion (14, 15), only one case of target deletion associated with a *de novo* insertion has been reported for other LINEs (16). We show here that the I element can induce target duplication as well as target deletion upon transposition, and a simple model for the mechanism of LINE integration accounting for this diversity is proposed.

MATERIALS AND METHODS

***Drosophila* strains and sequences of the transposed copies:** The *Drosophila* containing the transposed copies of the marked I element, the w^K (17) and *Charolles* (18) strains from which they originate, and the cloning and sequencing of copies (a), (b) and (c), are described in (12).
Nucleic acid purification and PCR amplification of target sites: Genomic DNA from w^K or *Charolles Drosophila* was extracted as described in (19), with in addition two phenol and one chloroform-isoamyl alcohol extractions prior to ethanol precipitation. 1 to 2 µg of genomic DNA were used for PCR amplification of the target sites. Target of copy (a) was amplified by inverted PCR amplification of w^K DNA with oligonucleotides a1 and a2, as described in (9); targets of copies (b) and (c) were amplified by direct PCR with primers b1 and b2 and primers c1 and c2, respectively. The oligonucleotides used were: a1=5'-AGTGAGAACTGAAGTTAGACACGAC-3'; a2=5'-CACCTTGAGT-ATGGCTTATTATGCC-3'; b1=5'-TATCGCTCAACGATAACTTTCGTGG-3'; b2=5'-CAAGTCTAGTGCCA-GCCAAACTGTTC-3'; c1=5'-GAGATGGTTGACTGTGATAGTTG-3'; c2=5'-GGTTTCGGTGTGTGGGCAT-CGCTG-3'. PCR amplifications were carried out in 50 µl containing 50 mM KCl, 1.5 mM $MgCl_2$, 0.001 % gelatin (w/w), 0.2 mM of each dNTP, 1 µM of each primer, 1.25 U of Taq Polymerase (Amersham), and DNA. After an initial step at 94°C (5 min.), 30 cycles of amplification were carried out: 30 sec. at 64°C, 1 min. 30 sec. at 72°C and 10 sec. at 94°C.
Cloning and sequencing of the PCR product: PCR-amplified DNA fragments were purified on agarose gels, Klenow-treated, kinased and cloned into the phosphatased *Eco*RV site of the Bluescript vector (Stratagene). DNA sequencing was performed on double-strand plasmid DNA, by the dideoxy chain-termination method (20) using primers a1 and a2 for copy (a), T3 and T7 for copy (b), and both c1 and c2, and T3 and T7, for copy (c).

RESULTS AND MODEL

In a previous publication (12), we reported a method for selecting retrotransposition events from transgenic *Drosophila* that contained an I element marked with an indicator gene for retrotransposition. The marked element could be tracked along its transposition, and 5' and 3' junctions of three transposed copies of this marked element were sequenced. All three copies disclosed short repeats (2-10 bp) at both ends, but the origin and precise length of these repeats could not be unambiguously assessed (except in one case) due to homologies between the ends of the I element and the putative target sites. We have therefore now cloned and sequenced the target sites of all three transposed elements, after PCR amplification from the initial genomic DNA (i.e. before transposition), using oligonucleotides in the flanking domains of the transposed copies. As illustrated in figure 1, two of the transposed copies (a, b) disclose target duplications [10 bp for (a), 7-10 bp for (b), the uncertainty resulting from a 3 bp homology between the I 5' end and the target site], but for the third copy (c) comparison of the target sequence and the 5' and 3' junctions reveals that transposition did not create a duplication but rather led to an 8 bp deletion in the target site. Since this transposed element was recovered from a cross between transgenic males derived from the w^K strain and *Charolles* females, we also sequenced the target site from the *Charolles* strain to ascertain that the deletion did not pre-exist

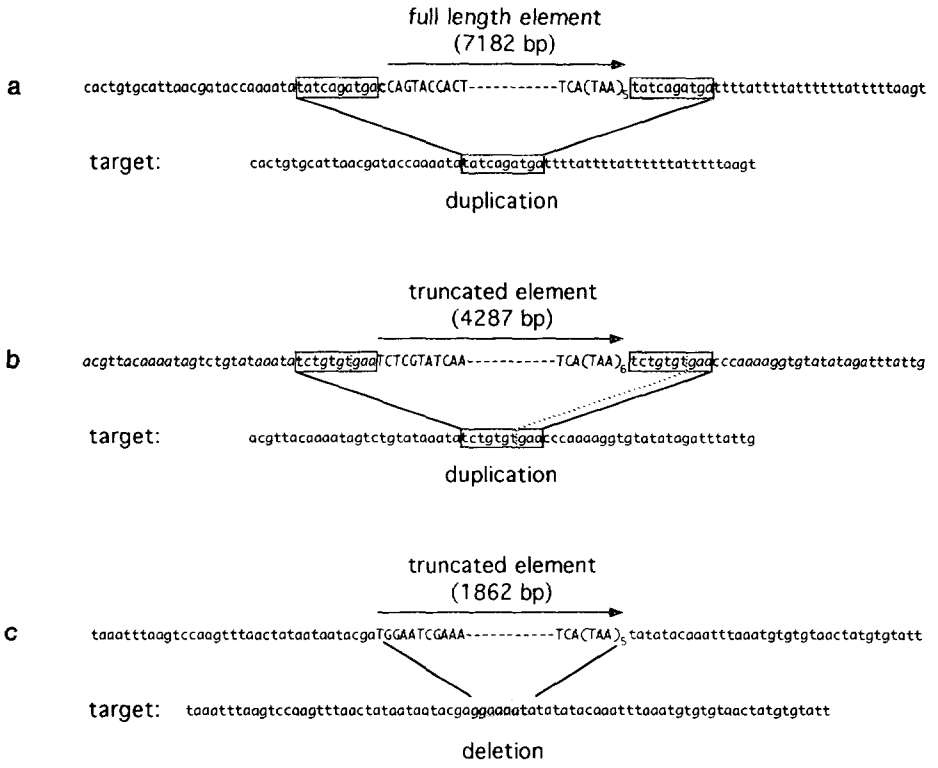


Figure 1. **Sequence of the 5' and 3' junctions of transposed marked I elements and of their target sites.**
Sequences of the transposed marked I copies (a), (b) and (c) are denoted with upper-case letters, target site sequences with lower-case letters; only a few nucleotides of the 5' and 3' ends (with the TAA *repeats) of the transposed I elements are shown. Arrows indicate 5'→3' orientation of the transposed* copies, with length given in brackets. Duplicated target site sequences are boxed [with an uncertainty *for (b), resulting from a 3 bp homology between the I 5' end and target site],* and the deleted sequence in the target site of copy (c) is shaded.

to transposition in either strain: both sequences, indeed, were found identical. Clearly then, deletion in the target DNA can result from the *de novo* transposition of the I element, furthermore under conditions where target site duplications -as normally observed- also take place.

Although this is the first reported case of target deletion unambiguously resulting from I element transposition (but actually sequencing of the target DNA before integration has only been rarely performed), this is not the only case reported for LINE elements: Narita *et al.* (16) reported a *de novo* transposition of a human L1 element into the dystrophin gene, most likely associated with a 2 bp target deletion. In addition, one out of six insertions of mouse L1 does not disclose any target duplication (1) and five out of thirteen cloned human full-length L1 sequences do not have defined target duplications (21, 22, 23, 24). In these last cases, since the corresponding target sites were not sequenced, it is even possible that target deletion took place. Finally, transposition of the LINE-like R2Bm element from *Bombyx mori* (which however is a rather atypical LINE because it inserts in a site-specific manner) systematically results in a target site deletion of fixed length (2 bp), without duplication (14, 15).

This apparently divergent set of data and the presented evidence that transposition of the LINE I element can both result in target site duplication and deletion, can actually be accounted for by a unique model for LINE integration (Figure 2). This model relies i) on already experimentally strongly supported assumptions concerning basic traits of LINE transposition (see below) and ii) on plausible properties of the LINE-encoded reverse transcriptases that can be postulated on the basis of the strong similarities between LINE and retroviral coding sequences for this essential enzymatic function. Concerning the former assumptions, it is now clearly demonstrated that LINE transposition is an RNA-mediated process which involves reverse transcription of an RNA transcript of the LINE element (see Introduction). Furthermore, it has been suggested that LINEs integrate into the host genome by an *"in situ"* reverse transcription taking place at the integration site (1, 6, 7, 25); this integration mechanism has been clearly demonstrated for the R2Bm element from *Bombyx mori* -and up to now only for this element-, by refined *in vitro* experiments disclosing initiation of R2Bm RNA reverse transcription at a 3'-OH end of a nick in the target DNA (15). It will be assumed that similar reverse transcription processes take place for other LINEs, and the model will be extended to account for the diversity of the observed target integration sites.

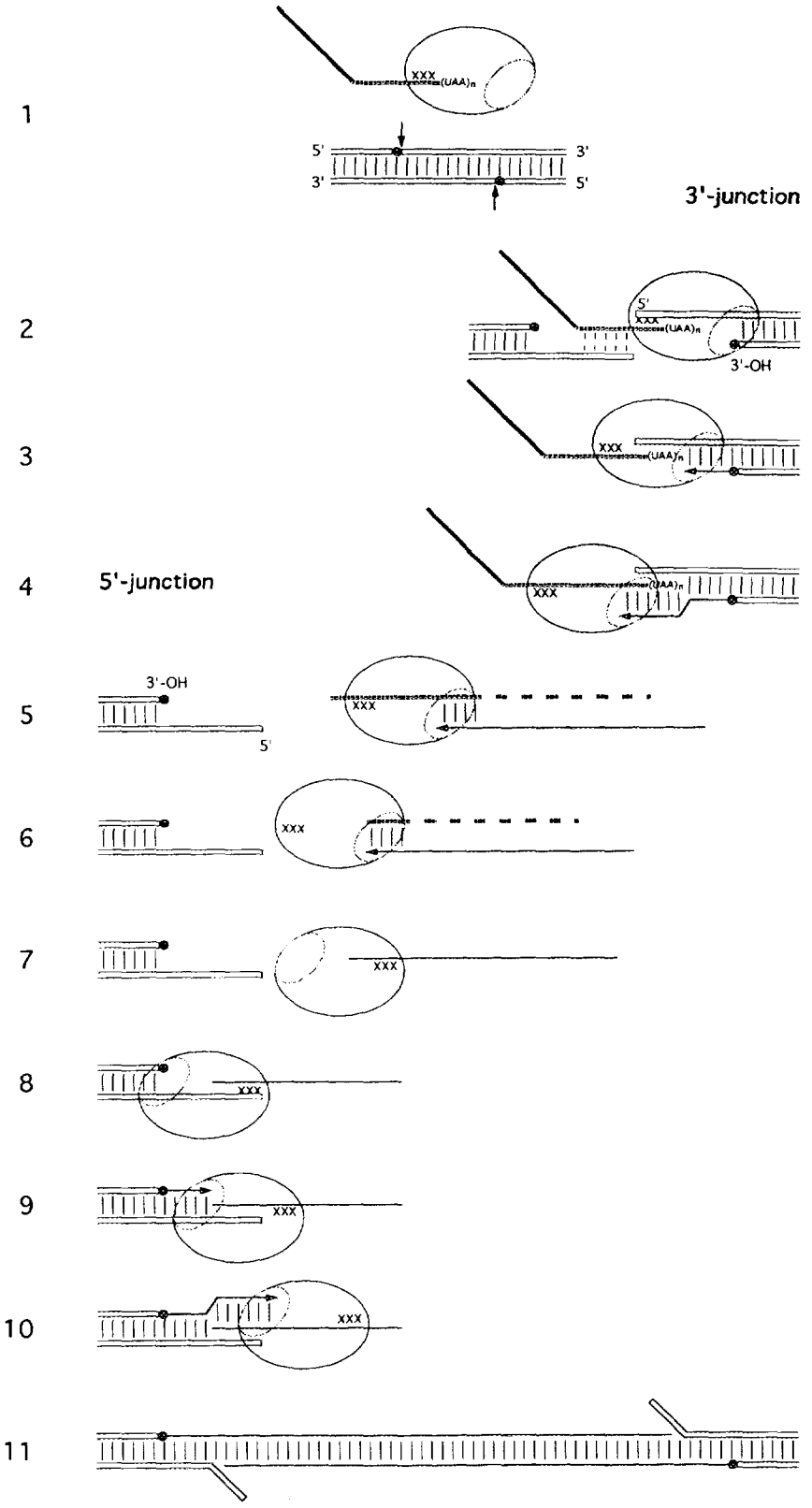
MODEL

1. Whatever the nature of the nick at the target site, i.e. single-strand nick or double-strand break with either 3' or 5' extensions, it is hypothesized that priming of reverse transcription takes place at a 3'-OH end of the target DNA (Figure 2).

2. It is assumed that LINE reverse transcriptases, as their retroviral homologues, i) can copy either DNA or RNA, and ii) can make "template switching". As a consequence, reverse transcription can initiate by copying either the target DNA overhanging strand (in case of nicks with 5' extensions), or the 3' end of the transposon RNA transcript; in the first case, reverse transcriptase will make either a "forced template

---

Figure 2. **A model for LINE retrotransposition involving reverse transcriptase template switching.**
[1] Reverse transcriptase is associated with the LINE transcript in a ribonucleoprotein complex, possibly via a direct interaction between a nucleic acid binding domain on the reverse transcriptase (xxx) and the 3' end of the RNA template [see (29)]. The LINE transcript is represented by a striped line -with the typical UAA repeats at the I element 3' end- and the reverse transcriptase active site by a shaded domain. The double-strand DNA target site is schematized, with the cleavage sites (either pre-existing or generated by a LINE-encoded endonuclease) indicated by arrows. [2] The ribonucleoprotein complex associates with a 3'-OH end of the target DNA (represented as a black circle) which acts as a primer for reverse transcription; pairing between the 5' overhang of the target site opposite to the reverse transcription initiation site and nucleotides within the LINE transcript is possible [see (25)]. [3] Reverse transcription initiates by replication of the target site DNA overhang (DNA-dependent DNA-polymerase activity) and then [4] the reverse transcriptase switches from target DNA to the LINE transcript (RNA-dependent DNA-polymerase activity). [5] *In situ* reverse transcription proceeds to the 5' end of the RNA template. Subsequent or concomitant to reverse transcription, the RNA template is degraded by an RNase H, which may be encoded by the reverse transcriptase itself [see (5)] or by a cellular gene [see (36)]. [6] The reverse transcriptase reaches the 5' end of the RNA template; truncated transposed copies could be due to premature reverse transcriptase-template dissociation or, alternatively, to a shortening of the RNA template resulting from RNase H digestion in case of pairing of the RNA transcript with the 5' overhang of the target site (see step 2, left). [7] After completion of reverse transcription, the reverse transcriptase can dissociate or remain attached to the single-strand cDNA (represented as a single line) via the nucleic acid binding domain. [8] The reverse transcriptase associates with the 3'-OH end of the target DNA to prime the second-strand DNA synthesis. [9, 10] Elongation proceeds using at first the target DNA overhang as a template and then the LINE cDNA, with a possible switching before reaching the 5' end of the target DNA. Alternatively second-strand synthesis may be achieved by a cellular DNA-polymerase (15). [11] The resulting reverse transcribed copy discloses protruding target DNA ends, which could be eliminated by repair enzymes or by structure-specific endonucleases (37) before ligation of the gaps.


1
XXX
(UAA)n
5'
3'
3'
5'
3'-junction
2
5'
XXX
(UAA)n
3'-OH
3
XXX
(UAA)n
4
5'-junction
XXX
(UAA)n
5
3'-OH
5'
XXX
6
XXX
7
XXX
8
XXX
9
XXX
10
XXX
11

switching" if reverse transcription goes through the 5' end of the overhanging DNA strand, or "copy choice" if the jump to the RNA template takes place prematurely; no template switching needs to take place if reverse transcription initiates directly with RNA as a template (necessarily the case for a nick with 3' extensions).

3. After complete or partial reverse transcription of the LINE transcript -and RNase H-mediated degradation of the RNA template- the reverse transcriptase machinery can act symmetrically at the other end of the nick, i.e. use the 3'-OH end of the left end DNA fragment as a primer for a DNA-dependent DNA polymerization, thus synthesizing a double-strand, target-DNA-linked, integrated DNA molecule. Again, "template switching" (in case of 5' extensions at the nick) can take place at any position within the target single-strand DNA overhang.

The first assumption is supported by the *in vitro* assays for reverse transcription using either retroviral or LINE-associated reverse transcriptases, which all disclose a strong requirement for a primer RNA or DNA sequence to initiate reverse transcription [(26, 27); reviewed in (28)] -although in some special cases priming can also take place using other pathways [see (29)]. The second and third assumptions are supported by the numerous evidences of template switching for retroviruses or viral-like retrotransposons, where both "forced template switching" -for instance after synthesis of the "strong stop" DNA- and "copy choice" -either homologous strand transfer or "illegitimate" transfer to nonhomologous templates- take place in the course of the replicative cycle [reviewed in (30, 31, 32)]. Retroviral reverse transcriptases further use either RNA or DNA as templates for polymerization, both *in vivo* [for the synthesis of the proviral structure, using RNA as a template for first-strand synthesis and the newly synthesized DNA for the complementary strand; reviewed in (33)], and *in vitro* [using either synthetic or natural templates; references in (33)]; it has also been reported that LINE reverse transcriptases can reverse transcribe RNA and DNA templates with similar efficiencies in an *in vitro* assay (26, 27). It should finally be emphasized that LINE reverse transcriptases -within identified ribonucleoprotein structures [(34, 35); see also (29)]-most likely specifically associate with the 3' end of the LINE RNA template, the spatial proximity of which would actually favor template switching after initiation of reverse transcription from the target DNA.

The basic prediction of the model, then, is that depending on the location where both template switchings take place during first-strand and second-strand synthesis, either duplications of varying length or even deletion will take place (see Figure 3): the model therefore very simply accounts for the previously presented data. It should be noted that if the nicks are with 3' extensions (see Figure 3), no template switching can take place and only target site duplication should be observed. This actually could account for what is observed with the *Bombyx mori* R1Bm element -which systematically generates a 14 bp target site duplication (14)-, in case an endonuclease for this site-specific LINE creates 14 bp-distant nicks with 3' overhangs.

It should also be emphasized that the occurrence of target site duplications of varying length has been repeatedly taken as a suggestion that LINE elements were inserting at pre-existing nicks in the target DNA (and therefore expectedly of varying length), but this assumption might not be correct. Clearly, the model can account for the occurrence of target deletions and/or duplications of varying length, even if a nick of fixed length is generated by an endonuclease encoded by the mobile element itself (as observed for all retroviruses and viral-like retrotransposons with the "integrase", as well as for

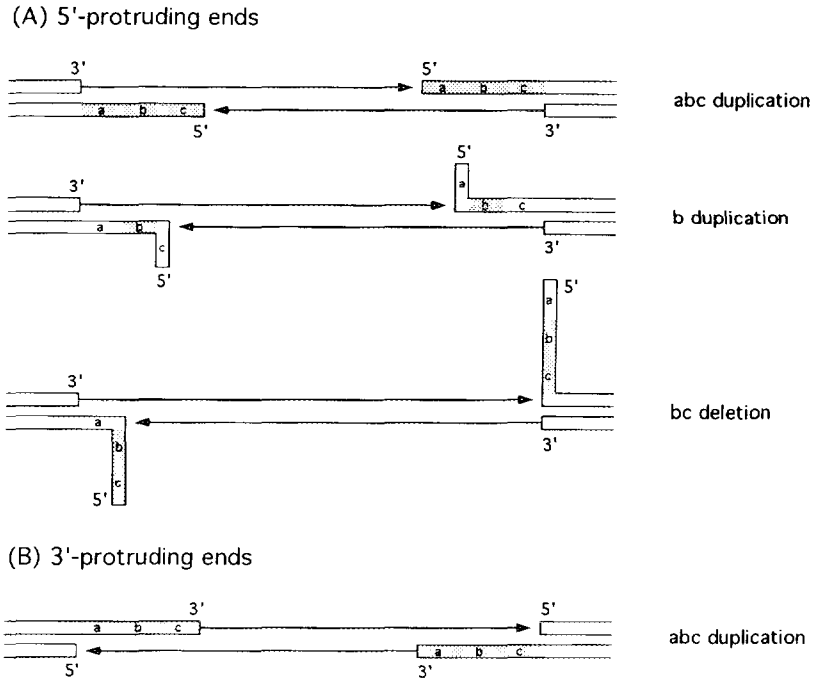


Figure 3. **Basic structural features of LINE target sites predicted by the model.** Nicks with either 5' or 3' extensions are considered for the possibility to create target site duplication or deletion of variable length, according to the model. Target DNA is divided into domains a, b and c, for clarity; arrows indicate direction of DNA synthesis.
(A) nicks with 5' extensions: reverse transcriptase template switching can create full target site duplication ("abc duplication"), partial duplication (for instance "b duplication") or target deletion (for instance "bc deletion").
(B) nicks with 3' extensions: only full target site duplication should be observed.

the R2Bm LINE-like element -which encodes an endonuclease which makes two successive single-strand nicks at a 2 bp distance-).

Finally, if one further assumes -as observed *in vitro* for the LINE-like R2Bm element- that LINEs only need for initiation of reverse transcription a single-strand nicks (which actually might be much less deleterious to the cell than double-strand nicks), the finally identified "target site" could even be very far from the single-strand nick, if reverse transcriptase "selects" DNA as a template rather than the associated LINE transcript, and "screens" for a region where template switching will take place efficiently (an AT-rich domain which would match better with the A-rich 3' end of the LINE RNA, for instance).

CONCLUSION

One "advantage" of the proposed model -besides its simplicity and its ability to account for the various structural features of LINE target sites (target site duplications of varying length, absence of

duplication, target site deletions)- is that it makes the LINE reverse transcriptase coding sequence closely "ressemble" the retroviral *pol* gene at the functional level, in addition to the close similarities already noticed at the sequence level [reviewed in (3, 4, 5)]: according to the model, both reverse transcriptases would have important properties in common in addition to that of reverse transcription, namely the non-specificity of the template (RNA or DNA) and the ability to make template switching. Finally, LINE integration into pre-existing nicks was conjectured as being at the origin of the variability observed for the length of the target duplications: actually, the model could still account for the data in the plausible case a LINE-encoded endonuclease -which has only been characterized up to now for the R2Bm LINE- would create defined nicks at a fixed distance in the target DNA, provided that these nicks are with 5' extensions.

ACKNOWLEDGMENTS

We thank Dr. C. Lavialle for critical reading of the manuscript. This work was financed by the Centre National de la Recherche scientifique (Unité associée 147), the Institut National de la Santé et de la Recherche Médicale (Unité 140) and the Association pour la Recherche sur le Cancer (contract 6552 to T.H., grant to S.J.).

REFERENCES

1. HutchinsonIII, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) In D. E. Berg and M. M. Howe (eds.), Mobile DNA, American Society for Microbiology, Washington, D.C., 593-617.
2. Martin, S.L. (1991) *Curr. Op. Genet. Dev.* **1**, 505-508.
3. Xiong, Y. and Eickbush, T.H. (1988b) *Mol. Biol. Evol.* **5**, 675-690.
4. Xiong, Y. and Eickbush, T. (1990) *EMBO J.* **9**, 3353-3362.
5. McClure, M.A. (1993) In A. M. Skalka and S. P. Goff (eds.), Reverse Transcriptase, Cold Spring Harbor Laboratory Press, 425-444.
6. Finnegan, D.J. (1989) In D. E. Berg and M. M. Howe (eds.), Mobile DNA, American Society for Microbiology, Washington, D.C., 503-517.
7. Bucheton, A. (1990) *Trends Genet.* **6**, 16-21.
8. Heidmann, T., Heidmann, O. and Nicolas, J.F. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2219-2223.
9. Heidmann, O. and Heidmann, T. (1991) *Cell* **64**, 159-170.
10. Tchénio, T. and Heidmann, T. (1991) *J. Virol.* **65**, 2113-2118.
11. Tchénio, T., Ségal-Bendirdjian, E. and Heidmann, T. (1993) *EMBO J.* **12**, 1487-1497.
12. Jensen, S. and Heidmann, T. (1991) *EMBO J.* **10**, 1927-1937.
13. Pélisson, A., Finnegan, D.J. and Bucheton, A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4907-4910.
14. Eickbush, T.H. and Robins, B. (1985) *EMBO J.* **4**, 2281-2285.
15. Luan, D.D., Korman, M.H., Jakubczak, J.L. and Eickbush, T.H. (1993) *Cell* **72**, 595-605.
16. Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Minami, R., Nakamura, H. and Matsuo, M. (1993) *J. Clin. Invest.* **91**, 1862-1867.
17. Lüning, K.G. (1981) *Hereditas* **95**, 181-188.
18. Pélisson, A. and Bregliano, J.C. (1987) *Mol. Gen. Genet.* **207**, 306-313.
19. Junakovic, N., Caneva, R. and Ballario, P. (1984) *Chromosoma* **90**, 378-382.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
21. Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. and Kazazian, H.H. (1991) *Science* **254**, 1805-1808.
22. Dombroski, B.A., Scott, A.F. and Kazazian, H.H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6513-6517.
23. Hohjoh, H., Minakami, R. and Sakaki, Y. (1990) *Nucleic Acids Res.* **18**, 4099-4104.
24. Crowther, P.J., Doherty, J.P., Linsenmeyer, M.E., Williamson, M.R. and Woodcock, D.M. (1991) *Nucleic Acids Res.* **19**, 2395-2401.
25. Schwarz-Sommer, Z., Leclercq, L., Göbel, E. and Saedler, H. (1987) *EMBO J.* **6**, 3873-3880.

26. Ivanov, V.A., Melnikov, A.A., Siunov, A.V., Fodor, I.I. and Ilyin, Y.V. (1991) *EMBO J.* **10**, 2489-2495.
27. Mathias, S.L., Scott, A.F., Kazazian, H.H., Boeke, J.D. and Gabriel, A. (1991) *Science* **254**, 1808-1810.
28. Finnegan, D.J. (1993) *Curr. Biol.* **3**, 468-470.
29. Wang, H. and Lambowitz, A.M. (1993) *Cell* **75**, 1071-1081.
30. Temin, H.M. (1991) *Trends Genet.* **7**, 71-74.
31. Hu, W.-S., Pathak, V.K. and Temin, H.M. (1993) In A. M. Skalka and S. P. Goff (eds.), Reverse Transcriptase, Cold Spring Harbor Laboratory Press, 251-274.
32. Gabriel, A. and Boeke, J.D. (1993) In A. M. Skalka and S. P. Goff (eds.), Reverse Transcriptase, Cold Spring Harbor Laboratory Press, 275-328.
33. Telesnitsky, A. and Goff, S.P. (1993) In A. M. Skalka and S. P. Goff (eds.), Reverse Transcriptase, Cold Spring Harbor Laboratory Press, 49-83.
34. Martin, S.L. (1991) *Mol. Cell. Biol.* **11**, 4804-4807.
35. Deragon, J.M., Sinnett, D. and Labuda, D. (1990) *EMBO J.* **9**, 3363-3368.
36. DiFrancesco, R.A. and Lehman, I.R. (1985) *J. Biol. Chem.* **260**, 14764-14770.
37. Harrington, J.J. and Lieber, M.R. (1994) *EMBO J.* **13**, 1235-1246.