L1 retrotransposition
The snap-velcro model and its consequences

Sébastien Viollet,1,2,3 Clément Monot,1,2,3 and Gael Cristofari1,2,3,*
1INSERM; U1081; Institute for Research on Cancer and Aging of Nice (IRCAN); Nice, France; 2CNRS; UMR 7284; Institute for Research on Cancer and Aging of Nice (IRCAN); Nice, France; 3University of Nice-Sophia Antipolis; Faculty of Medicine; Nice, France

Keywords: reverse transcription, reverse transcriptase, target-primed reverse transcription, TPRt, non-LTR retrotransposon, LINE-1, L1, poly(A) tail, endonuclease

Submitted: 03/07/2014
Revised: 04/14/2014
Accepted: 04/15/2014
Published online: 04/24/2014
Citation: Viollet S, Monot C, Cristofari G. L1 retrotransposition: The snap-velcro model and its consequences. Mobile Genetic Elements 2014; 4:e28907; http://dx.doi.org/10.4161/mge.28907
*Correspondence to: Gael Cristofari; Email: Gael.Cristofari@unice.fr

LINE-1 (L1) elements are the only active and autonomous transposable elements in humans. The core retrotransposition machinery is a ribonucleoprotein particle (RNP) containing the L1 mRNA, with endonuclease and reverse transcriptase activities. It initiates reverse transcription directly at genomic target sites upon endonuclease cleavage. Recently, using a direct L1 extension assay (DLEA), we systematically tested the ability of native L1 RNPs to extend DNA substrates of various sequences and structures. We deduced from these experiments the general rules guiding the initiation of L1 reverse transcription, referred to as the snap-velcro model. In this model, L1 target choice is not only mediated by the sequence specificity of the endonuclease, but also through base-pairing between the L1 mRNA and the target site, which permits the subsequent L1 reverse transcription step. In addition, L1 reverse transcriptase efficiently primes L1 DNA synthesis only when the 3’ end of the DNA substrate is single-stranded, suggesting so-far unrecognized DNA processing steps at the integration site.

L1 Retrotransposition can be Initiated through Two Different Pathways

The L1 replication cycle starts with the synthesis of a bicistronic mRNA coding for the two L1 proteins, ORF1p and ORF2p (Fig. 1). ORF1p is a 40 kDa RNA-binding protein able to form trimers.11-13 It exhibits nucleic acid chaperone activity.14

L1 Elements are Endogenous Mutagens in the Human Genome

Transposable elements account for half to two-thirds of the human genome.1 Among them, LINE-1 (L1) non-LTR retrotransposons form the only autonomous and active family and are the most abundant, representing 17% of our DNA. Each individual genome contains hundreds of potentially active L1 copies, and hundreds of thousands of defective copies, which are truncated, fragmented, and/or mutated.2 The active copies can proliferate via an RNA-mediated copy-and-paste mechanism, called retrotransposition. L1 insertions are intrinsically mutagenic, however their actual impact on gene expression depends on their specific site of integration. Intergenic or deep intronic insertions can often have no detectable effects on genes. In contrast, insertions in exons or regulatory sequences have the potential to profoundly alter gene expression/function, by disrupting coding- or cis-regulating sequences, or by carrying cis-regulating sequences (transcription factor binding sites, cryptic splicing and polyadenylation sites, etc.).2,3 Hence, germline L1 insertions sporadically cause de novo genetic diseases,2,4 and somatic L1 retrotransposition in cancer has been shown to contribute to tumor genome dynamics, including driver mutations.5-10 Therefore, exploring the mechanisms that influence L1 target choice is crucial to our understanding of L1-driven genome plasticity.

Commentary to: Monot C, Kuciak M, Viollet S, Mir AA, Gabus C, Darlix J-L, Cristofari G. The specificity and flexibility of l1 reverse transcription priming at imperfect T-tracts. PLoS Genet 2013; 9:e1003499; PMID:23675310; http://dx.doi.org/10.1371/journal.pgen.1003499
the function of which has not been elucidated. ORF2p is a large, 149 kDa, protein with endonuclease (EN) and reverse transcriptase (RT) activities. Both ORF1p and ORF2p bind the L1 RNA to form a stable ribonucleoprotein particle (RNP), the core of the L1 retrotransposition machinery. The L1 RNP can mediate two different integration processes. In the canonical pathway, called target-primed reverse transcription (TPRT), the L1 EN activity produces a nick at the recognized target site in the chromosomal DNA. A particular case of this pathway is retrotransposition at telomeres, the natural extremity of chromosomes. Regardless of whether a particular retrotransposition event is initiated by EN or not, the subsequent steps of the reaction, such as second strand DNA synthesis or ligation of the 3’ ends of the neo-synthesized DNA to the target DNA, have not been explored yet.

The Snap-Velcro Model of L1 Reverse Transcription Initiation

Recently, we explored the mechanism of L1 reverse transcription initiation using a direct L1 extension assay (DLEA). In an alternative pathway, named endonuclease-independent (ENi) retrotransposition or non-classical L1 insertion (NCLI), reverse transcription is initiated at pre-existing DNA lesions, without the need for endonuclease cleavage. A particular case of this pathway is retrotransposition at telomeres, the natural extremity of chromosomes. Regardless of whether a particular retrotransposition event is initiated by EN or not, the subsequent steps of the reaction, such as second strand DNA synthesis or ligation of the 3’ ends of the neo-synthesized DNA to the target DNA, have not been explored yet.

The Snap-Velcro Model of L1 Reverse Transcription Initiation

Recently, we explored the mechanism of L1 reverse transcription initiation using a direct L1 extension assay (DLEA). In this approach, native L1 RNP expressed in—and enriched from—human cells are incubated with oligonucleotide primers of various sequences or structures, and with radioactive dTTP only, for a very short time (less than 5 min). The products are then resolved on sequencing gels to directly visualize the extension of the primer. Due to the short incubation time and the use of dTTP only, it focuses on the initiation of reverse transcription. Advantages of this method, compared with previously PCR-based techniques, include its versatility with regards to the primers that can be used and its quantitative nature. A limitation of DLEA is the absence of sequence information and its lower sensitivity.

One of the unresolved questions related to L1 reverse transcription priming was whether—or to which degree—the 3’ end of the nicked genomic DNA needs to be accessible and to base-pair with the
poly(A) tail of the L1 RNA. Indeed, R2, a related non-LTR retrotransposon which has been used to establish the basis of the TPRT model, does not require such a complementarity.\textsuperscript{29,30} Although the consensus sequence released upon L1 EN cleavage (5\textsuperscript{'}-TTTT-3\textsuperscript{'}) could in principle anneal to the poly(A) tail of the L1 RNA, it is extremely short for maintaining a stable interaction and the actual sequences cleaved by the L1 EN can significantly differ from the consensus sequence.\textsuperscript{22-24} To directly address this question, we quantified the efficiency of extension of more than 65 primers by DLEA. Based on the results of these experiments, and on additional analyses of the distribution of polymorphic L1 insertions in the human genome, we proposed the snap-velcro model for L1 reverse transcription initiation.\textsuperscript{28} This model is detailed in Figure 2. The efficiency of reverse transcription initiation is influenced by the 10 last nucleotides of the target DNA. The 4 last nucleotides (the snap) contribute the most to this process. Reverse transcription priming is the most efficient when the snap corresponds to 4 Ts (snap closed). However suboptimal sequences with terminal or internal mismatches can be tolerated (snap open). For terminal mismatches, the efficiency of extension depends on the nature of the base ending the primer (T > C > A > G). These suboptimal sequences can be more efficiently extended if mismatches are compensated by an increased number of matching Ts in the upstream 6 nt (velcro strap fastened). Finally an important aspect of our results is that priming only occurs if the DNA substrate is single-stranded. Indeed, double-stranded DNA substrates are extended only if they end with a 3\textsuperscript{'} overhang. If the 3\textsuperscript{'} extremity of the target DNA is embedded in duplex DNA, either as a blunt- or as a 3\textsuperscript{'}-recessed end, no extension could be detected under the DLEA conditions employed.

**Consequences of the Snap-Velcro Model Related to Primer-Template Sequence Match**

The snap-velcro model indicates that complementarity between the L1 poly(A)
tail and the last 10 nucleotides of the target DNA is important for priming of L1 reverse transcription, yet allows sufficient flexibility to accommodate a wide range of potential target sites. It explains a long-standing observation that L1 elements are often flanked by imperfect T-rich sequences significantly longer than expected for the recognition site of the L1 endonuclease.23,24,31,32 This model also implies that L1 target-site selection relies not only on the sequence specificity of the EN nicking reaction, but also on the subsequent ability of the RT to efficiently extend the cleaved product. This observation has practical and technological implications. Non-LTR retrotransposons fall into two main classes: i) the stringent elements, which always insert into discrete and defined genomic locations with high sequence specificity (such as R1, R2, Tx1L, SART1, Tras1), and ii) more promiscuous elements, which insert into multiple locations within a very short and degenerate sequence (such as L1). EN domain swapping between two stringent retrotransposons is sufficient to exchange their respective target site selectivity both in vitro and in vivo.33 Similarly, structure-driven domain swapping between a stringent EN and the L1 EN moiety has succeeded in modifying its target selectivity in vitro.34 However, such EN variants, when reintroduced into a complete L1 element and tested in vivo, are unable to redirect L1 insertions to altered target sites. Instead, they continue to insert in T-rich stretches,35 indicating that other determinants downstream of the initial EN cleavage contribute to L1 target site selection. Our results suggest that L1 RT priming specificity could be one of these determinants. Therefore engineering L1 to achieve site-specific integration in vivo might require to take into account both EN and RT specificities.

Consequences of the Snap-Velcro Model Related to the Accessibility of the Target DNA

A second important feature of the snap-velcro model is the requirement...
for a single-stranded 3′ overhang to initiate reverse transcription at the target site. The original TPRT model stipulates that retrotransposition is initiated by a nick in the target DNA.\(^1\)\(^2\)\(^3\)\(^4\) How does the 3′-OH extremity of the substrate, which is embedded in the duplex DNA, become available for L1 RT? We can envisage several possibilities. First, the L1 EN produces double-stranded staggered cuts rather than nicks. In vitro, plasmid DNA can indeed be linearized upon prolonged incubation with an isolated recombinant EN domain.\(^5\) Whether L1 ORF2p acts as a monomer or a multimer in the context of the L1 RNP is unknown. However, many other reverse transcriptases form dimers, including the R2 RT or human telomerase.\(^6\)\(^7\)\(^8\) Dimerization of ORF2p could lead to concomitant cleavage of bottom and top strands, while maintaining the two target site DNA extremities together (Fig. 3A). Of note, the average length of the target-site duplication, which reflects the distance between the top and bottom strand cuts, is 15 nucleotides,\(^9\) a length compatible with the minimal size of the single-stranded region (6 nt). A second hypothetical mechanism involves a strand-transfer or a DNA helicase activity (Fig. 3B). Although ORF1p has been proposed to perform this task through its nucleic acid chaperone activity,\(^10\) the presence of this protein in native L1 RNPs was not sufficient under our experimental conditions to prime reverse transcription with duplex DNA substrates. Recent efforts have identified a number of cellular proteins associated with the L1 retrotransposition machinery.\(^11\)\(^12\) Among them, Utp1 is an RNA- and DNA-dependent 5′-3′ helicase implicated in nonsense-mediated mRNA decay, and also in telomeric DNA replication,\(^13\) a serious candidate. It is noteworthy that our attempts to extend duplex DNA by L1 RNP were performed in the absence of ATP, an essential cofactor of helicase activities. Thus, to explore the mechanism allowing annealing of the target site DNA to the L1 poly(A) tail, future follow-up DLEA experiments could be performed with the additional inclusion of ATP. Finally, unlike a TPRT reaction, DLEA decouples EN cleavage and RT priming. Thus we cannot exclude that, in vivo, the product of the nicking reaction would somehow be channeled to the RT active site, being unwound in the process.

In this perspective, DLEA reactions are more similar to reverse transcription taking place in the ENi pathway.\(^14\)\(^15\) The nature of the preexisting lesions used by L1 RT to prime reverse transcription is still unclear, in particular whether they are single- or double-stranded. Telomeres are frequent integration sites in this pathway,\(^16\) and notably they end with a 3′ overhang, which is also the preferred substrate of telomerase\(^2\) (Fig. 3D). Another interesting observation is that ENi-retrotransposition only occurs efficiently in cells defective for both p53 and non-homologous end-joining (NHEJ).\(^17\)\(^18\) Impairing NHEJ results in end-resection and extremities with 3′ overhangs, which ultimately leads to homologous recombination-mediated repair.\(^19\) We speculate that NHEJ inhibition could be necessary to permit the generation of these 3′ overhangs and thereby reverse transcription priming in the ENi pathway (Fig. 3C).

Possible end-processing factors include the MRN complex (Mre11, Rad50, Nbs1), CtIP, Exo1, BLM or Dna2. Poly(ADP-ribose) polymerase 1 (PARP-1) was shown to interact with both ORF2p\(^20\) and with Mre11.\(^21\) Therefore it could potentially recruit both end-resection factors and the L1 machinery at sites of DNA damage.

Because DLEA uses native L1 RNP enriched from cells, it can be combined with shRNA-mediated knock-down of specific cellular factors. This approach will help in the future to elucidate the role of cellular factors in retrotransposition.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We are grateful to Dominic van Essen for critical reading. This work was funded by INSERM and INCa (Avenir program, agreement #2009-340), the European Research Council (Retrogenomics, #243312) and the French Government (National Research Agency, ANR) through the “Investments for the Future” (LABEX SIGNALIFE, #ANR-11-LABX-0028-01). CM was the recipient of a PhD fellowship from the Fondation ARC.

References

1. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DB. Repetitive elements may comprise two-thirds of the human genome. PLoS Genet 2011; 7:e1002384; PMID:22144907; http://dx.doi.org/10.1371/journal.pgen.1002384

2. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease.Curr Opin Genet Dev 2012; 22:191-203; PMID:22406018; http://dx.doi.org/10.1016/j.gde.2012.02.006

3. Kaer K, Speck M. Retroelements in human disease. Genen 2013; 51B:231-41; PMID:23356077; http://dx.doi.org/10.1016/j.gene.2013.01.008

4. Kazazian HH Jr., Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 1988; 332:164-6; PMID:383458; http://dx.doi.org/10.1038/332164a0

5. Shukla R, Upson KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 2013; 155:101:11; PMID:23540693; http://dx.doi.org/10.1016/j.cell.2013.02.032

6. Solovyov S, Ewing AD, Rahmann EP, Doucet T, Nelson HH, Burns MB, Harris RS, Sgomon DF, Casella A, Erlander M, et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res 2012; 22:2328-38; PMID:22968929; http://dx.doi.org/10.1101/gr.145235.112

7. Lee E, Iskow R, Yang L, Gokcenen O, Haseley P, Luquette LJ 3rd, Lohr JG, Harris CC, Ding L, Wilson RK, et al.; Cancer Genome Atlas Research Network. Landscape of somatic retrotransposition in human cancers. Science 2012; 337:967-71; PMID:22745252; http://dx.doi.org/10.1126/science.1227077

8. Iskow RC, McCabe MT, Mills RE, Torene S, Pirart J, Neuwald AF, Van Meir EG, Vertino PM, Devine WS. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 2010; 141:1253-61; PMID:20603005; http://dx.doi.org/10.1016/j.cell.2010.05.020

9. MiKi Y, Nishio A, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y. Disruption of the APC gene by a retrotransposonal insertion of L1 sequence in a colon cancer. Cancer Res 1992; 52:643-5; PMID:1500668

10. Rodič N, Burns KH. Long, interspersed element-1 (LINE-1): passenger or driver in human neoplasms? PLoS Genet 2013; 9:e1003402; PMID:23555307; http://dx.doi.org/10.1371/journal.pgen.1003402

11. Huhjoh H, Singer M. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. EMBO J 1996; 15:630-9; PMID:8599946

12. Martin SL, Branciforte D, Keller D, Bain DL. Trimeric structure for an essential protein in L1 retrotransposition. Proc Natl Acad Sci U S A 2003; 100:13815-20; PMID:14615577; http://dx.doi.org/10.1073/pnas.2336221100

13. Khazina E, Truffault V, Bärtner R, Schmidt S, Coles M, Weichenrieder O. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. Nat Struct Mol Biol 2011; 18:1006-14; PMID:21822284; http://dx.doi.org/10.1038/nsmb.2097
