

## Mini-Review Article

# The ORF1 Protein Encoded by LINE-1: Structure and Function During L1 Retrotransposition

**Sandra L. Martin**

*Department of Cell and Developmental Biology, School of Medicine, University of Colorado, Fitzsimons Campus, PO Box 6511, Mail Stop 8108, Aurora, CO 80045, USA*

Received 17 September 2005; Revised 10 December 2005; Accepted 13 December 2005

LINE-1 or L1 is an autonomous non-LTR retrotransposon in mammals. Retrotransposition requires the function of the two L1-encoded polypeptides, ORF1p and ORF2p. Early recognition of regions of homology between the predicted amino acid sequence of ORF2 and known endonuclease and reverse transcriptase enzymes led to testable hypotheses regarding the function of ORF2p in retrotransposition. As predicted, ORF2p has been demonstrated to have both endonuclease and reverse transcriptase activities. In contrast, no homologs of known function have contributed to our understanding of the function of ORF1p during retrotransposition. Nevertheless, significant advances have been made such that we now know that ORF1p is a high-affinity RNA-binding protein that forms a ribonucleoprotein particle together with L1 RNA. Furthermore, ORF1p is a nucleic acid chaperone and this nucleic acid chaperone activity is required for L1 retrotransposition.

Copyright © 2006 Sandra L. Martin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## LINE-1 BACKGROUND

L1 is an interspersed repeated DNA found in mammalian genomes that attained its high copy number by retrotransposition. It belongs to a large family of mobile elements that replicate via reverse transcription of an RNA intermediate. These elements, non-LTR retrotransposons, are distinct from retroviruses and LTR-containing retrotransposons which also replicate via reverse transcription of an RNA intermediate. Non-LTR retrotransposons are widely distributed throughout eukaryotes and likely all share the novel mechanism of replication known as target-site-primed reverse transcription, or TPRT, whereby reverse transcription of the L1 RNA template is primed from a 3' hydroxyl at the genomic insertion site (reviewed in [1]).

Close inspection of sequences of the > 500,000 copies of either mouse or human L1 reveals a multigene family comprised mainly of truncated, mutated, or rearranged copies of a small number of functional, full-length elements; only a subset of the full-length elements is capable of retrotransposition (see [2, 3], for recent reviews). The functional retrotransposons are 6-7 kb in length and contain two long open reading frames (ORFs), both of which encode proteins that are required for retrotransposition [3]. ORF2 encodes a 146 kd polypeptide which provides the two known enzymatic activities that are required to convert the RNA

intermediate into a new genomic DNA copy of L1 during TPRT: endonuclease [4] and reverse transcriptase [5]. These two activities of ORF2p were predicted from sequence similarity between L1 and known apurinic/aprimidinic endonuclease [4] and reverse transcriptase [6, 7] proteins, then verified biochemically [4, 8]. In contrast, the role of ORF1p during retrotransposition has remained far more elusive because the amino acid sequence predicted by ORF1 lacks homology with any protein of known function (see [9], and results of October 2005 NCBI protein and nucleotide searches).

## ORF1 AND RELATED SEQUENCES

The first intact coding sequence for an ORF1 protein was found by sequence analysis of a mouse L1 element called L1Md-A2 [7]. Comparison of the theoretical translation of this 41.2 kd protein from L1Md-A2 to that of a consensus primate L1 sequence revealed that the C-terminal half of ORF1 was evolving under selective pressure, whereas the N-terminal half was not. This early analysis also noted that the predicted ORF1 protein is quite basic, a common feature of nucleic acid-binding proteins [7]. Subsequently, many ORF1-like sequences have been determined from the L1 elements of different mammals, and from related elements found in fish [16]. The C-terminal, homologous, basic

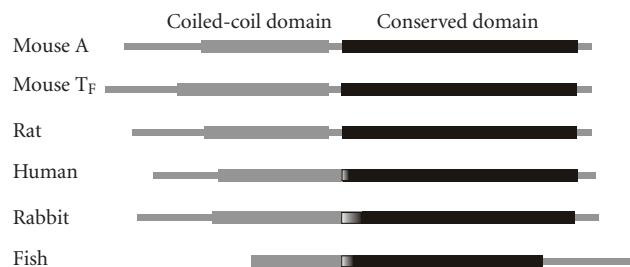


FIGURE 1: Schematic representation of the ORF1 protein. The thin bar for each species represents the entire length of the protein. Thicker bars represent the coiled-coil (gray, based upon coils analysis [10]) and conserved domains (black, based upon multiple-sequence alignments using T-Coffee [11]). These two domains overlap in the human, rabbit, and fish, but not the two mouse or rat ORF1 protein sequences, as indicated. Sequences used were mouse A101 (Q91V68, [12]), mouse L1spa (O54849, [13]), rat (Q63303), human L1rp (AAD39214, [14]), rabbit ([15], not in GenBank), and fish swimmer (AAD02927, [16]). The two mouse and the human ORF1 protein sequences are from retrotransposition-competent elements; the other sequences are from untested elements. A model for the trimeric structure of mouse L1 and its role in TPRT appeared in [17] by Martin et al.

domain is a general feature of the ORF1 proteins from all of these elements (the conserved domain in Figure 1).

A second predicted feature of all of these ORF1 amino acid sequences is the presence of a long coiled-coil domain upstream of the conserved domain (Figure 1). In human L1, this coiled-coil domain encompasses a leucine zipper [18]. In rabbit ORF1, this coiled-coil region appears similar to keratin [15]. The most likely explanation for the poor sequence similarity among the different ORF1 sequences in this region with one another and with other coiled-coil-containing proteins (eg, keratin) is that all share a coiled-coil domain with distinct evolutionary origins, probably brought into proximity with the conserved, basic C-terminal domain via recombination. Recombination to create novel sequence variants is often evident in L1 lineages [19–23]. With this scenario, the constraints imposed by a requirement for protein-protein interaction via a coiled-coil domain in ORF1 protein forces a small degree of apparent similarity in the absence of homology among these diverse sequences. Conversely, it is possible that all of these ORF1 sequences evolved from a common ancestor, but extremely rapid divergence of the sequences towards the N-terminus of the protein has obscured the evidence for this homology. Interestingly, sequence variation in this N-terminal region of ORF1p is particularly great within subtypes of human [24], rat (see [2], and references therein), and mouse (see [25, 26], and references therein) L1. Positive selection acting within this portion of the ORF1 protein is associated with the evolutionary success/extinction of human L1 lineages, perhaps reflecting drive for ORF1p to either attract or avoid an interacting factor [24]. Additional sequences from L1 elements in other species may shed light on whether the amino acid sequence of the N-terminal region is undergoing strong selective pressure for rapid sequence

divergence by accumulating replacement substitutions, or if novel sequences are often acquired from nonhomologous sources.

An unexpected feature of the L1 ORF1 sequence is revealed when its amino acid sequence is used as the query in a BLASTP search. The program reports that it has detected a putative conserved domain. This conserved domain is essentially the entire mammalian ORF1 protein sequence and has been annotated “transposase 22.” Given that transposase is the enzyme responsible for the DNA breaking-joining reactions that occur during transposition of a wide variety of DNA elements [27], it seems likely to be a misnomer to call this domain a transposase for several reasons. Most significantly, the TPRT reaction used by L1 and the other non-LTR retrotransposons does not require a transposase enzymatic activity because cDNA is synthesized *in situ* using chromosomal primers [28]. Thus, L1 replication lacks any intermediate equivalent to the double-strand DNA substrate of transposases and the related integrases [27]. Furthermore, biochemical and mutational analyses demonstrate that the endonuclease activity of L1 ORF2p is responsible for the DNA cleavages that occur during TPRT [4, 29, 30]. Finally, the conserved domain in ORF1p of known functional mouse and human L1 elements lacks an apparent DDE motif [31], which is conserved in the active sites of transposases and integrases. Due to vast sequence divergence among members of the transposase/integrase superfamily of proteins, their DDE motifs are best recognized in structure [32] rather than sequence alignments; hence absolute resolution of the question of whether L1 ORF1p should be annotated “transposase 22” awaits atomic-level resolution of its structure.

### ORF1p IS REQUIRED FOR RETROTRANSPOSITION

Even the relatively conserved C-terminal domain of ORF1 is more divergent than ORF2 when the sequences of human and mouse L1s are compared [7]. Hence, when it was finally possible to measure L1 retrotransposition activity using an autonomous retrotransposition assay [5, 33], it was unexpected to learn that mutations in ORF1 were at least as severe, if not more so, than those that abolish reverse transcriptase activity. No retrotransposition events were detected in human L1 mutants in which either the serine at position 119 of ORF1p was replaced with a stop codon, or a highly conserved diarginine at 261/262 was replaced by dialanine; in both of these cases, the frequency of retrotransposition was less than 0.06% of the wild-type parental element. In contrast, mutation of a critical active-site residue in the reverse transcriptase domain of ORF2 (D702Y), which abolishes *in vitro* enzymatic activity [8], knocked retrotransposition down to 0.15% of wild type [5]. The other known enzymatic activity of ORF2 in L1 is endonuclease, which is also required for TPRT [34]. Several mutations that eliminate detectable endonuclease activity in an *in vitro* nicking assay again knock retrotransposition down to 0.2–1%, but do not eliminate it [34]. We observe similar effects of mutations in the ORF1 conserved domain compared to the endonuclease and reverse transcriptase domains in mouse L1.

Thus, to date, the most stringent mutations of L1 are those in ORF1. As noted when the leakiness of the ORF2 mutations was originally observed, it is likely that ORF2p is more readily supplied *in trans* (albeit with substantially reduced efficiency, [5]), whereas ORF1p appears to be more stringently required *in cis* with the L1 RNA. These findings imply that ORF1 is critical for an earlier step in the retrotransposition cycle than reverse transcription itself, for example, regulating expression of ORF2 [35] or recruitment of ORF2p into the L1 ribonucleoprotein complex, and/or delivering the L1 RNP to the chromosomal DNA and facilitating the strand exchanges that are required during TPRT [17, 36].

In light of the stringent *cis*-requirement for ORF1p during L1 retrotransposition, it is interesting that ORF1p appears to be dispensable when the L1 machinery provided by ORF2p is usurped by the human SINE, Alu, for its amplification—this surprising finding may be explained if the SRP9/14 protein can replace ORF1p function [37]. In contrast, ORF1p is required along with ORF2p for processed pseudogene formation by L1 [38, 39].

#### **FUNCTIONAL ANALYSIS OF ORF1p: PROTEIN-PROTEIN INTERACTION**

Leucine zippers and coiled-coil domains are typically associated with protein-protein interactions. In cytoplasmic extracts from human cells that express high levels of L1, NTERa2D1, the ORF1p (also called p40) partitions into a 160,000 x g pellet. Treatment of this pellet with increasing concentrations of glutaraldehyde to cross-link the protein shifts increasing amounts of the 40 kd ORF1 protein into complexes that run at 78, 89, 100, and 200 kd on SDS-PAGE, suggesting that the ORF1ps in these cytoplasmic particles are interacting closely with one another, or with other cellular proteins. This study also examined full-length p40 and various truncations expressed in *E coli* for protein-protein interactions, thereby mapping the multimerization domain to the N-terminal half of the protein, in the region of the predicted coiled-coil [9].

Similar findings were obtained with mouse L1 ORF1p using somewhat different experimental approaches. Recombinant protein purified from *E coli* coimmunoprecipitated <sup>35</sup>S-labeled protein synthesized *in vitro* in rabbit reticulocyte lysate, demonstrating that mouse ORF1p is able to self-associate [40]. A combination of yeast 2-hybrid and GST pull-down assays were later used to map the region in mouse ORF1p responsible for multimerization; the predicted coiled-coil domain is both necessary and sufficient for protein-protein interaction [41]. More recently, overexpression of soluble ORF1p in baculovirus permitted analysis of its multimerization state by analytical ultracentrifugation. These studies revealed that the full-length protein forms a highly stable homotrimer, whereas a truncated ORF1p containing just the carboxy-terminal C-1/3 does not self-associate, even at relatively high protein concentrations [17]. All of the above findings consistently support the conclusion that the coiled-coil domain is wholly responsible for multimerization in both mouse and human L1 ORF1ps, with

the trimer being the biologically relevant form of mouse ORF1p [17].

#### **FUNCTIONAL ANALYSIS OF ORF1p: NUCLEIC ACID BINDING**

The bias towards highly basic amino acids in ORF1p led to the hypothesis that this protein interacts with nucleic acids [7]. Early evidence for such interaction was provided by cosedimentation of ORF1p with L1 RNA in sucrose gradients loaded with cytoplasmic extracts prepared from the mouse embryonal carcinoma cell line F9. The heavy complexes that formed were termed L1 ribonucleoprotein particles, or L1 RNP. L1 RNPs are not sensitive to disruption by EDTA, but are sensitive to proteolysis [42]. Exposure of the RNPs to UV light rapidly cross-links the RNA to protein, indicating a close association between L1 RNA and protein [21]. The human ORF1p (p40) also associates with L1 RNA based upon a series of cosedimentation experiments. p40 remains in the supernatant upon centrifugation at 800 and 12,000 x g, but pellets at 160,000 x g. Treatment of the cytoplasmic extract (800 x g supernatant) with RNase but not DNase prior to centrifugation shifts p40 from the 160,000 x g pellet to the supernatant, indicating that the protein is pelleting because it is in a large complex with RNA, or an RNP. The human L1 RNPs are not dependent on divalent cations or disturbed by 10 mM EDTA, thus it appears that human ORF1p is bound to RNA in an RNP that is quite similar to the mouse L1 RNP. Further experiments indicated that the RNA present in these RNP was L1 RNA and not actin or G3PDH RNA [9]. The presence of ORF1p in RNP was found to be sensitive to high concentrations of monovalent cations as well as RNase treatment [43, 44], leading to an enrichment procedure for RNA-free ORF1p from human cells [43], which was then used to provide evidence for one or two relatively high-affinity binding sites for ORF1p in L1 RNA [35]. It is important to note that all of the above experiments examined the interaction of L1 RNA with ORF1p in extracts from animal cells where L1 RNA and ORF1p were present as minority components of a complex mixture.

A more direct assessment of the nucleic acid-binding properties of the ORF1 protein is provided by studies of highly purified protein prepared after overexpression in either *E coli* or baculovirus-infected insect cells. As with ORF1p from mammalian cells, it is critical to take precautions against copurification of RNA with the protein—when protein is purified in standard, nondenaturing conditions without high concentrations of monovalent cation, RNA is coenriched and the protein is heavily contaminated with nucleic acid. This is readily apparent on a wavelength scan, or by examining purified protein by ethidium bromide staining after electrophoresis through agarose gels [41]. Our earliest experiments with protein expressed in *E coli* used denaturing conditions (8 M urea) to purify the protein from the insoluble inclusion body fraction which simultaneously removed bound nucleic acid. That protein was used for UV cross-linking and electrophoretic mobility-shift assays (EMSAs), which demonstrated that ORF1p binds RNA and

single-stranded DNA [40]. The affinities observed in those experiments, however, were lower than those obtained with subsequent experiments that were done using protein purified from the soluble fraction rather than the refolded denatured protein from inclusion bodies, probably because most of the protein was not correctly refolded to its native form after the denaturation. Interestingly, the RNA-binding region of the full-length ORF1p was mapped by simply examining various GST-ORF1p fusion constructs (containing full-length and a variety of truncated regions of ORF1p) for the presence of copurifying RNA. As long as the *E coli* extracts and affinity purification steps were kept in physiological concentrations of monovalent cation, RNA copurified with the protein if it contained the RNA-binding domain. All deletions containing the C-1/3 basic domain were contaminated with RNA and those that lacked it were free of RNA contamination. This same region of mouse ORF1p was found to be both necessary and sufficient for binding nucleic acid based upon transfer of  $^{32}\text{P}$  from RNA to protein by UV cross-linking [41].

The RNA-binding properties of the full-length mouse ORF1 protein purified from baculovirus were further assessed using coimmunoprecipitation and filter-binding assays. These experiments examined the affinity of ORF1p for a variety of transcripts, and tested whether a specific *cis*-acting sequence in mouse L1 RNA recruits ORF1p. The presence of a high-affinity site in human L1 RNA was suggested based upon preferential coimmunoprecipitation of a 41 nt T1 nuclease-resistant fragment with ORF1 antibody [35]. The mouse L1 RNA coimmunoprecipitation experiments revealed that efficient recovery of the  $^{32}\text{P}$ -labeled RNA required at least 38 nt, suggesting a length effect rather than a sequence requirement. All longer RNAs tested precipitated efficiently, independent of sequence. Further evidence that ORF1p is a nonsequence-specific RNA-binding protein was provided by results of nitrocellulose filter-binding assays using highly purified mouse ORF1p expressed in baculovirus. Transcripts that contained specifically the 38 nt sequence in either the sense or antisense orientation both bound with high affinity. Although there is a slight increase in the apparent binding affinity of ORF1p to RNA containing the sense 38 nt sequence compared to the same sequence in antisense orientation, it is only 4- to 7-fold and therefore too small to be considered specific binding for sense versus antisense L1 RNA [45].

This discrepancy between the results with mouse and human L1 ORF1ps regarding the existence of a high-affinity binding site within L1 RNA has not been resolved. Possibly, it is due to differences between mouse and human L1, or, more likely, between the reagents used for the assays. For example, it is possible that another protein that is critical for the site-specificity was present in the partially purified preparation from human cells, but missing when the protein was purified from baculovirus-infected insect cells or *E coli*. The question of whether L1 RNA contains a specific, high-affinity binding site for ORF1p is important for L1 biology because it offers an attractive explanation for the *cis*-preference of ORF1p for L1 RNA that is evident from both the evolution-

ary pattern of L1 as well as experimental evidence from the autonomous retrotransposition assay (see [38, 39], and references therein).

### FUNCTIONAL ANALYSIS OF ORF1p: NUCLEIC ACID CHAPERONE ACTIVITY

Non-LTR retrotransposons are present throughout Eukaryota, but diverged long ago into five groups based upon the phylogenetic relationships of their reverse transcriptase region (the only sequence feature conserved among all non-LTR retrotransposons) and the type and organization of their protein domains [1]. Three of these groups, L1, I, and Jockey, each named for the first element of that group described, have a separate ORF upstream of their reverse transcriptase-containing ORF. In two of these three groups, I and Jockey, the upstream ORFs (ie, ORF1s), both contain zinc-finger domains, making their ORF1 proteins reminiscent of retroviral gag proteins. An important function associated with the zinc-finger-containing, nucleocapsid domain of gag is that of nucleic acid chaperone, which is critical for retroviral replication [46]. Nucleic acid chaperones are proteins that facilitate rearrangements of nucleic acids to their thermodynamically most stable form. A combination of at least three protein features contribute to nucleic acid chaperone activity: charge neutralization due to an excess of basic amino acids, a higher affinity for single-stranded than for double-stranded nucleic acids, and the ability to lower the cooperativity of the helix: coil transition [47]. These properties must be exquisitely balanced so that the chaperone can promote both melting and annealing of nucleic acids. The ORF1 protein from the non-LTR retrotransposon, *I* factor, shares several biochemical properties with retroviral nucleocapsid proteins, including the ability to accelerate annealing of complementary single-strand DNA sequences; these observations led to the suggestion that the *I* factor ORF1 protein functions as a nucleic acid chaperone during replication [48].

Mouse L1 ORF1 protein also accelerates annealing of complementary oligonucleotides. In addition, it lowers the  $T_m$  of mispaired duplex DNA, accelerates a strand displacement reaction if an imperfect duplex is challenged by the addition of the perfect complement, and alters the force required for the helix: coil transition in single-molecule studies using optical forceps [36, 49]. Significantly, the nucleic acid chaperone activity of ORF1p is required for retrotransposition. A single-point mutation that destroys effective chaperone activity (R297K) without affecting RNA or single-stranded DNA binding affinity also destroys retrotransposition activity [49]. Consistent with this observation, the analogous mutation in human L1 also destroys retrotransposition, but not RNP formation [44].

### SUMMARY

L1 is arguably the most significant dynamic force currently operating upon the mammalian genome. Retrotransposition is just one of many facets of L1's contribution to genetic plasticity and diversity [50], although it lies at the root of

all of the others. Retrotransposition requires the proteins encoded by both of the two open reading frames in L1. The two known functions of the protein encoded by ORF2, endonuclease and reverse transcriptase, were readily predicted based upon sequence homology, whereas homology has so far failed to provide clues regarding the function of the ORF1 protein. In spite of this disadvantage, however, several significant advances have been made in establishing the structure and function of this critical retrotransposition protein through a series of *in vivo* and *in vitro* experiments. The protein binds both RNA and DNA, with a higher affinity for single-stranded than double-stranded nucleic acids. The RNA-binding function leads to RNP formation and safe delivery of the RNP to genomic DNA so that it can undergo TPRT. The nucleic acid chaperone activity of ORF1p likely contributes more directly to reverse transcription by TPRT, perhaps by facilitating the strand exchanges that place the DNA primer onto the RNA or cDNA template, or by melting secondary structure in the RNA, or both.

## ACKNOWLEDGMENTS

I thank D Branciforte, E Epperson, P W-l Li, C Rose, and A Walker for helpful comments on the manuscript and NIH GM40367 for support.

## REFERENCES

- [1] Eickbush TH, Malik HS. Origins and evolution of retrotransposons. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington, DC: American Society of Microbiology Press; 2002:1111–1144.
- [2] Furano AV. The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Progress in Nucleic Acid Research and Molecular Biology*. 2000;64:255–294.
- [3] Moran JV, Gilbert N. Mammalian LINE-1 retrotransposons and related elements. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington, DC: American Society of Microbiology Press; 2002:836–869.
- [4] Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*. 1996;87(5):905–916.
- [5] Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. High frequency retrotransposition in cultured mammalian cells. *Cell*. 1996;87(5):917–927.
- [6] Hattori M, Kuhara S, Takenaka O, Sakaki Y. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. *Nature*. 1986;321(6070):625–628.
- [7] Loeb DD, Padgett RW, Hardies SC, et al. The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons. *Molecular and Cellular Biology*. 1986;6(1):168–182.
- [8] Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science*. 1991;254(5039):1808–1810.
- [9] Hohjoh H, Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *The EMBO Journal*. 1996;15(3):630–639.
- [10] Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. *Science*. 1991;252(5010):1162–1164.
- [11] Notredame C, Higgins DG, Heringa J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*. 2000;302(1):205–217.
- [12] DeBerardinis RJ, Goodier JL, Ostertag EM, Kazazian HH Jr. Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nature Genetics*. 1998;20(3):288–290.
- [13] Naas TP, DeBerardinis RJ, Moran JV, et al. An actively retrotransposing, novel subfamily of mouse L1 elements. *The EMBO Journal*. 1998;17(2):590–597.
- [14] Schwahn U, Lenzner S, Dong J, et al. Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nature Genetics*. 1998;19(4):327–332.
- [15] Demers GW, Matunis MJ, Hardison RC. The L1 family of long interspersed repetitive DNA in rabbits: sequence, copy number, conserved open reading frames, and similarity to keratin. *Journal of Molecular Evolution*. 1989;29(1):3–19.
- [16] Duvernell DD, Turner BJ. Swimmer 1, a new low-copy-number LINE family in teleost genomes with sequence similarity to mammalian L1. *Molecular Biology and Evolution*. 1998;15(12):1791–1793.
- [17] Martin SL, Branciforte D, Keller D, Bain DL. Trimeric structure for an essential protein in L1 retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(24):13815–13820.
- [18] Holmes SE, Singer MF, Swergold GD. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *The Journal of Biological Chemistry*. 1992;267(28):19765–19768.
- [19] Adey NB, Schichman SA, Hutchison CA III, Edgell MH. Composite of A and F-type 5' terminal sequences defines a subfamily of mouse LINE-1 elements. *Journal of Molecular Biology*. 1991;221(2):367–373.
- [20] Adey NB, Schichman SA, Graham DK, Peterson SN, Edgell MH, Hutchison CA III. Rodent L1 evolution has been driven by a single dominant lineage that has repeatedly acquired new transcriptional regulatory sequences. *Molecular Biology and Evolution*. 1994;11(5):778–789.
- [21] Saxton JA, Martin SL. Recombination between subtypes creates a mosaic lineage of LINE-1 that is expressed and actively retrotransposing in the mouse genome. *Journal of Molecular Biology*. 1998;280(4):611–622.
- [22] Hayward BE, Zavanelli M, Furano AV. Recombination creates novel L1 (LINE-1) elements in *Rattus norvegicus*. *Genetics*. 1997;146(2):641–654.
- [23] Cabot EL, Angeletti B, Usdin K, Furano AV. Rapid evolution of a young L1 (LINE-1) clade in recently speciated *Rattus* taxa. *Journal of Molecular Evolution*. 1997;45(4):412–423.
- [24] Boissinot S, Furano AV. Adaptive evolution in LINE-1 retrotransposons. *Molecular Biology and Evolution*. 2001;18(12):2186–2194.
- [25] Goodier JL, Ostertag EM, Du K, Kazazian HH Jr. A novel active L1 retrotransposon subfamily in the mouse. *Genome Research*. 2001;11(10):1677–1685.
- [26] Mears ML, Hutchison CA III. The evolution of modern lineages of mouse L1 elements. *Journal of Molecular Evolution*. 2001;52(1):51–62.
- [27] Craig NL. Unity in transposition reactions. *Science*. 1995; 270(5234):253–254.
- [28] 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(4):595–605.

- [29] Cost GJ, Boeke JD. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry*. 1998;37(51):18081–18093.
- [30] Morrish TA, Gilbert N, Myers JS, et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genetics*. 2002;31(2):159–165.
- [31] Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Molecular and Cellular Biology*. 1992;12(5):2331–2338.
- [32] Davies DR, Mahnke Braam L, Reznikoff WS, Rayment I. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-Å resolution. *The Journal of Biological Chemistry*. 1999;274(17):11904–11913.
- [33] Ostertag EM, Prak ETL, DeBerardinis RJ, Moran JV, Kazazian HH Jr. Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Research*. 2000;28(6):1418–1423.
- [34] Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. *The EMBO Journal*. 2002;21(21):5899–5910.
- [35] Hohjoh H, Singer MF. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *The EMBO Journal*. 1997;16(19):6034–6043.
- [36] Martin SL, Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Molecular and Cellular Biology*. 2001;21(2):467–475.
- [37] Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genetics*. 2003;35(1):41–48.
- [38] Wei W, Gilbert N, Ooi SL, et al. Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Molecular and Cellular Biology*. 2001;21(4):1429–1439.
- [39] Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genetics*. 2000;24(4):363–367.
- [40] Kolosha VO, Martin SL. In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proceedings of the National Academy of Sciences of the United States of America*. 1997;94(19):10155–10160.
- [41] Martin SL, Li J, Weisz JA. Deletion analysis defines distinct functional domains for protein-protein and nucleic acid interactions in the ORF1 protein of mouse LINE-1. *Journal of Molecular Biology*. 2000;304(1):11–20.
- [42] Martin SL. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Molecular and Cellular Biology*. 1991;11(9):4804–4807.
- [43] Hohjoh H, Singer MF. Ribonuclease and high salt sensitivity of the ribonucleoprotein complex formed by the human LINE-1 retrotransposon. *Journal of Molecular Biology*. 1997;271(1):7–12.
- [44] Kulpa DA, Moran JV. Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Human Molecular Genetics*. 2005;14(21):3237–3248.
- [45] Kolosha VO, Martin SL. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *The Journal of Biological Chemistry*. 2003;278(10):8112–8117.
- [46] Rein A, Henderson LE, Levin JG. Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. *Trends in Biochemical Sciences*. 1998;23(8):297–301.
- [47] Williams MC, Rouzina I, Wenner JR, Gorelick RJ, Musier-Forsyth K, Bloomfield VA. Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(11): 6121–6126.
- [48] Dawson A, Hartwood 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. *The EMBO Journal*. 1997;16(14):4448–4455.
- [49] Martin SL, Cruceanu M, Branciforte D, et al. LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *Journal of Molecular Biology*. 2005;348(3):549–561.
- [50] Han JS, Boeke JD. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *BioEssays*. 2005;27(8):775–784.