Nucleic Acids Research, 2006, Vol. 34, No. 3 ***853–864***
doi:10.1093/nar/gkj490

The dicistronic RNA from the mouse LINE-1 retrotransposon contains an internal ribosome entry site upstream of each ORF: implications for retrotransposition

Patrick Wai-Lun Li[1,2], Jinfang Li[1], Stephanie L. Timmerman[4], Les A. Krushel[3–5] and Sandra L. Martin[1–3,*]

[1]Cell and Developmental Biology, [2]Human Medical Genetics Program, [3]Program in Molecular Biology, [4]Biochemistry and Molecular Genetics and [5]Department of Pharmacology, University of Colorado School of Medicine, 12801 E. 17th Avenue, Aurora, CO 80010, USA

Received November 29, 2005; Revised and Accepted January 18, 2006

ABSTRACT

Most eukaryotic mRNAs are monocistronic and translated by cap-dependent initiation. LINE-1 RNA is exceptional because it is naturally dicistronic, encoding two proteins essential for retrotransposition, ORF1p and ORF2p. Here, we show that sequences upstream of ORF1 and ORF2 in mouse L1 function as internal ribosome entry sites (IRESes). Deletion analysis of the ORF1 IRES indicates that RNA structure is critical for its function. Conversely, the ORF2 IRES localizes to 53 nt near the 3′ end of ORF1, and appears to depend upon sequence rather than structure. The 40 nt intergenic region (IGR) is not essential for ORF2 IRES function or retrotransposition. Because of strong *cis*-preference for both proteins during L1 retrotransposition, correct stoichiometry of the two proteins can only be achieved post-transcriptionally. Although the precise stoichiometry is unknown, the retrotransposition intermediate likely contains hundreds of ORF1ps for every ORF2p, together with one L1 RNA. IRES-mediated translation initiation is a well-established mechanism of message-specific regulation, hence, unique mechanisms for the recognition and control of these two IRESes in the L1 RNA could explain differences in translational efficiency of ORF1 and ORF2. In addition, translational regulation may provide an additional layer of control on L1 retrotransposition efficiency, thereby protecting the integrity of the genome.

INTRODUCTION

LINE-1, or L1, is an autonomous retrotransposon that has amplified to high copy number in mammalian genomes. There are at least 599 000 copies of mouse L1 that are interspersed in all chromosomes, and together comprise 19% of the genomic DNA (1). Most individual L1 insertions are the truncated, rearranged or otherwise inactivated progeny of a much smaller number of retrotransposition-competent, active copies of L1 (2). Recent L1 insertions causing mutant phenotypes, such as human hemophilia and the spastic mouse document that L1 is replicating in present-day mammals [reviewed in (3)].

Retrotransposition-competent L1s are ~6500 nt in length and encode two proteins (Figure 1A). L1 retrotransposition begins with transcription of one of the full-length active copies. This natural dicistronic transcript acts first as the mRNA for translation of the two L1-encoded proteins, ORF1p and ORF2p, and later as the template for RT. Both ORF1p and ORF2p are required for retrotransposition. ORF1p is a nucleic acid binding protein (4–7) with essential nucleic acid chaperone activity (8,9). ORF2p provides the enzymatic activities of endonuclease (10) and reverse transcriptase (11) for target-site primed reverse transcription [TPRT (12)]; TPRT is the mechanism used by L1 and other non-LTR retrotransposons to insert a new copy into genomic DNA (13). Because both L1-encoded proteins are required *in*

*To whom correspondence should be addressed. Tel: +1 303 724 3467; Fax: +1 303 724 3420; Email: sandy.martin@uchsc.edu
Present address:
Jinfang Li, Replidyne, 1450 Infinite Drive, Louisville, CO 80027, USA

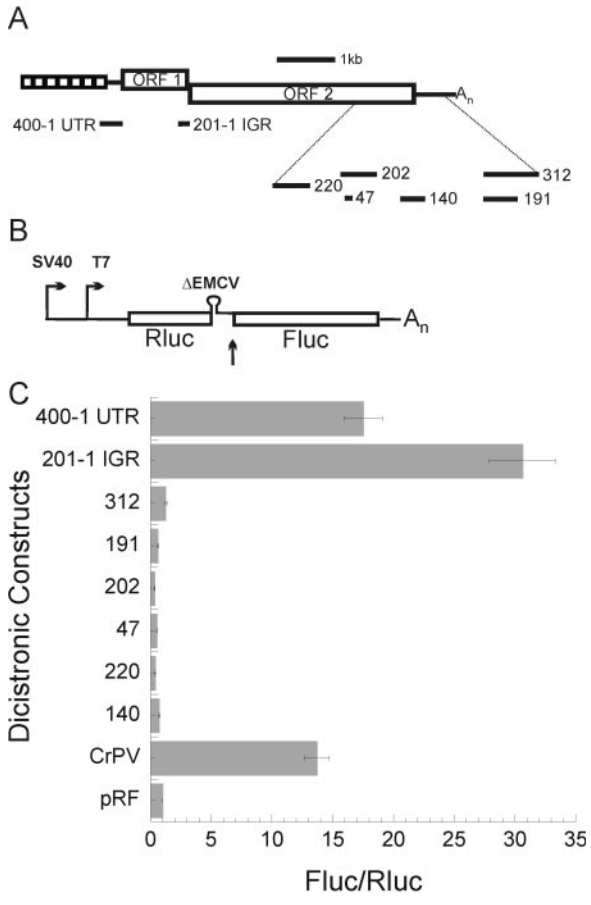

Figure 1. An IRES upstream of ORF1 and ORF2 in mouse L1. (**A**) Schematic of the RNA that created $L1_{spa}$: tandem boxes on the left represent 7.2 copies of a 212 nt repeated motif known as the T_F monomer which contains a promoter for transcription (18); these are followed by a 257 nt 5′-UTR (line); ORF1, which overlaps with ORF2 by 14 nt, although the UAA termination codon of ORF1 is separated from the first AUG in ORF2 by a 40 nt IGR; ORF2; a 3′-UTR (line) and a polyA tail (An). Note that the 5′ end of each L1 mRNA depends upon which monomer was used to initiate transcription, thus the sequences upstream of the ORF1 AUG can differ dramatically in length and base composition (monomer sequences are 65% GC, whereas the remaining 5′-UTR is 47% GC). Regions of $L1_{spa}$ tested for IRES activity using the dicistronic reporter gene assay are delineated. (**B**) Schematic of the pRF dicistronic reporter vector containing firefly (Fluc) and Renilla (Rluc) luciferase genes. The dicistronic mRNA is transcribed by an SV40 promoter/enhancer sequence in mouse cells and by a T7 promoter *in vitro*. Insertion of ΔEMCV upstream of the test sequence greatly reduces translation of Fluc by read-through or reinitiation from Rluc (28). (**C**) Relative lucificerase activity (to pRF without an insert ± SD) obtained from pRF containing 400 or 201 nt upstream of the ORF1 (400-1 UTR) or ORF2 (201-1 IGR) AUG or various sequences further 3′ in L1. These are named by their lengths and are from the 3′ end of ORF2 (202, 47, 220, 140) or the 3′-UTR (312, 191); their GC contents vary between 38 and 49%. The known Cricket Paralysis Virus intergenic IRES [CrPV (28)] was included as a positive control.

cis for efficient retrotransposition (14,15), both proteins must be translated from the primary distronic transcript. This raises questions about how translation of the L1 proteins is initiated.

Routine translation of the majority of mRNAs in eukaryotic cells is initiated by a cap-dependent mechanism. This involves recognition and binding of the cap structure (m^7GpppN) on the 5′ ends of mRNAs by the eukaryotic translation initiation factor, eIF4F. Upon binding an mRNA, eIF4F recruits the small ribosomal subunit and additional initiation factors, and then this 43S complex scans 5′–3′ until the first AUG initiation codon is encountered. The 60S subunit is then recruited and elongation begins (16).

Although ORF1 is proximal to the 5′ end of the RNA and hence the presumed cap, initiation of its translation by a cap-dependent mechanism is likely to be problematic. The first AUG lies at least 300 nt downstream of the 5′ end of L1 mRNA in the T_F-type element studied here because the transcriptional promoter lies in a repeated region known as the T_F monomer (17). In T_{Fspa} (18) any one of the 7.2 monomers may theoretically be used to initiate transcription, and T_{Fspa} itself retrotransposed successfully with a 5′-untranslated region (5′-UTR) of at least 1786 nt (18,19). The variability of the length of the 5′-UTR and the highly stable secondary structures associated with even the shortest 5′-UTR would be expected to lead to dramatic fluctuations in the efficiency of ribosome scanning and hence the initiation of translation of ORF1. Highly structured 5′-UTRs are known to represent a significant barrier to scanning ribosomes. One way around this difficulty of scanning through long, highly structured 5′-UTRs is to recruit the 40S subunit directly, using internal ribosome entry sites (IRES) [reviewed in (20,21)].

ORF2, on the other hand, is the second cistron in the dicistronic L1 mRNA, and its AUG is separated from the termination codon of ORF1 by a 40 nt intergenic region (IGR). This arrangement alone makes it unlikely that ORF2 is translated by a classical cap-dependent mechanism; the stringent *cis*-requirement for ORF2p during L1 retrotransposition (14,15) eliminates the possibility that a truncated L1 RNA is the normal mechanism for ORF2 translation. Several other mechanisms could explain translational initiation of ORF2, including: a ribosomal frameshift at the end of ORF1, since the two open reading frames (ORFs) in mouse L1 overlap by 14 nt (22); ribosome reinitiation, which follows translation of an upstream ORF (uORF), in this case ORF1 (23); and ribosomal shunting, as used for translation of the adenovirus late mRNAs (24). All of these mechanisms rely on recognition of the 5′ cap by eIF4F. Alternatively, ORF2 could be initiated using an IRES, which is cap-independent.

Here we test sequences upstream of both ORF1 and ORF2 in mouse L1 for IRES activity. The results of dicistronic reporter gene assays provide evidence for an IRES upstream of both ORF1 and ORF2. Neither IRES activity is explained by artifactual cryptic promoter or splicing activities. Dicistronic assays were also used to map the functional IRES for each ORF, revealing that the ORF1 IRES encompasses ~300 nt, whereas the ORF2 IRES localizes to a 53 nt region of the ORF1 coding sequence. Additionally, we show that the IGR is not required for ORF2 IRES function or L1 retrotransposition and eliminate ribosomal frameshifting as the mechanism of translation of ORF2 in mouse L1. These findings are placed into the context of L1 retrotransposition and have implications for the evolution of L1 and its replicative success in mammals.

MATERIALS AND METHODS

Plasmid constructions

L1 sequences were tested for IRES activity using the dicistronic reporter vector pRF (25), which was obtained from

Dr P. Sarnow. PCR products were generated from $L1_{spa}$ (18) using the oligonucleotide primers listed in Supplementary Table 1, then cloned into pGEM-TA (Promega). The inserts were recovered using EcoRI and NcoI, and then moved into the IGR of pRF (Figure 1B) to generate the plasmids pRF-1 to pRF-18. For pRF-19 to pRF-22, multiple mutations were introduced into the pGEM-TA clone carrying the 138-1 IGR insert using QuikChange® Multi Site-Directed Mutagenesis Kit (Stratagene), and the mutagenized inserts were moved as restriction fragments into pRF as described above. The TrkB construct contains the full-length 5′-UTR from the first promoter in mouse TrkB (26).

Plasmids for the promoterless reporter assays were based on RF_{A30} (27). Inserts were either made by direct amplification of $L1_{spa}$ (18) using the oligonucleotide primers listed in Supplementary Table 1, or by excising inserts from pGEM-TA intermediate clones. All inserts were moved into the promoterless vector using EcoRI and NcoI.

Plasmids for *in vitro* studies using rabbit reticulocyte lysates, T7-EMCV/Fluc-L1 5′-UTR /Rluc and T7-EMCV/Fluc-L1 200IGR/Rluc, were created by moving the EcoRI/SalI fragment from pRF-2 or pRF-13, containing the 299-1 UTR or the 201-1 IGR from L1, respectively, together with the adjacent Fluc gene into pGEM 3Z, to make the intermediate vectors, pGEM-5′-UTR Fluc and pGEM-200IGR Fluc. A PCR fragment containing the encephalomyocarditis virus (EMCV) IRES and the Rluc gene was amplified from T7-EMCV/Fluc-CrPV (Cricket Paralysis Virus) 5′ nc/Rluc (28) with EMCVEX.for and Rluc.rev primers (Supplementary Table 1). The product was digested with EcoRI, then cloned into the EcoRI sites of pGEM -5′-UTR Fluc and pGEM-200IGR Fluc. To create T7-ΔEMCV/Fluc-L1 5′-UTR /Rluc and T7-ΔEMCV/Fluc-L1 200IGR/Rluc constructs, the EMCV IRES was replaced with an inactive form of the EMCV IRES, ΔEMCV. ΔEMCV was amplified by PCR from T7 ΔEMCV/Fluc-CrPV 5′ nc/Rluc using EMCVEX.for and Rluc.rev as the primers, digested with EcoRI and KpnI, then cloned into T7-EMCV/Fluc-L1 5′-UTR /Rluc and T7-EMCV/Fluc-L1 200IGR/Rluc.

T_{FC}-containing plasmids for autonomous retrotransposition (9) were mutagenized in the vicinity of the ORF2 AUG to test for effects on L1 retrotransposition. These mutations were first made in pTN201 (18) and later moved as restriction fragments into T_{FC}. The PCR products were digested with NotI/NsiI, and then ligated into an intermediate vector, pDB25, which contains the NotI/BclI fragment of pTN201 in pET28A, for ease of cloning. The entire NotI/BclI fragment with the mutation was then used to replace the NotI/BclI fragment of pTN201. In order to examine the effects of mutations in the vicinity of the first AUG of ORF2 on retrotransposition, a unique HpaI site was introduced into pTN201 by making a silent substitution of G for the T at nt nine (if A of the first AUG in ORF1 is nt one). For this, the PstI/SspI fragment from pTN201 was subcloned into pGEM 3Z, creating pGEM-int. Separate PCR amplifications were used to amplify the partial ORF1 and ORF2 fragments with an HpaI site by using the primer pairs listed in Supplementary Table 1. The PCR fragments were digested with PstI/HpaI, or HpaI/KpnI for partial ORF1 and partial ORF2, respectively, then those fragments were ligated into PstI/KpnI-digested pGEM-int. The PstI/StuI fragment from this subclone was recovered, then ligated into pDB25 to obtain pET-int Hpa. Various mutants in the IGR were PCR amplified, digested with PstI/HpaI and cloned into pET-int Hpa. Those mutations, as well as mut1 made by site-directed mutagenesis of pDB44 as described for pRF-19 were then moved into L1 T_{FC} as an NsiI–EcoRV restriction fragment first into the intermediate pDB46 and then recovered with NheI and NotI into pCEP-puro (9) for retrotransposition analysis because pTN201 retrotransposition frequencies were too low to give reliable differences among mutants in our hands.

All of the constructions were verified by restriction mapping as well as by DNA sequencing across junctions between vector and insert. All mutations were verified by DNA sequencing. Plasmid names and the primers used to generate them are listed in Supplementary Table 2.

DNA transfection

An aliquot of 1×10^5 LTK^- cells were seeded in each well of a 24-well plate the day before transfection. Cells were transfected with 0.8 μg DNA using LipoFectamine 2000 as recommended (Invitrogen). Cells were harvested 48 h post-transfection, washed with phosphate-buffered saline (PBS), then lysed in 100 μl of Passive Lysis Buffer (Promega).

In vitro transcription and RNA transfection

A total of 1 μg of pRF-1 and pRF-13 were linearized with BamHI, and transcribed *in vitro* using the mMESSAGE mMACHINE T7 Ultra® kit (Ambion). An aliquot of 1×10^5 LTK^- cells were seeded in each well of a 24-well plate one day before transfection. A total of 1.6 μg of RNA were transfected using TransMessenger™ Transfection Reagent (Qiagen). After 3 h, cells were fed with normal medium and then harvested 4 h later by washing with PBS and lysing in 100 μl of Passive Lysis Buffer (Promega).

Luciferase assays

The activities of firefly (Fluc) and Renilla (Rluc) luciferase were determined from cell lysates following transient transfection. Assays were performed using 40 μl of cell lysate per assay in the Dual-Luciferase Reporter Assay System (Promega) with a Thermal Labs Luminoskan Ascent luminometer. For the dicistronic reporter assay, the relative ratio was calculated by dividing the firefly luciferase activity by the Renilla luciferase activity, then normalizing to the ratio calculated for the negative control, pRF with no insert, whose value was set to 1.0. For the promoterless reporter assay, the relative luciferase ratio was normalized to the negative control, in this case the IRES with no promoter activity, CrPV. Each reported value is the mean from three independent transfections (28). The absolute values for relative luciferase activities varied from experiment to experiment, but the relative values were consistent. Therefore the data shown in each figure were collected together in one experiment.

Northern analysis

A total of 4 μg of polyA + RNA (Oligotex Direct, Qiagen, Inc.) isolated 48 h after DNA transfection were separated on 1% formaldehyde-agarose gels and transferred to Hybond N+ (Amersham Biosciences). A gel-purified Fluc fragment (NcoI to XbaI of pRF) was radiolabeled with ^{32}P (Prime-a-probe, Amersham Biosciences). Hybridization and washing

were carried out using Rapid-hyb (Amersham Biosciences) as recommended by the manufacturer.

In vitro translation

RNA was transcribed *in vitro* using T7 polymerase (MAXIscript, Ambion, Inc.) with templates made from dicistronic luciferase constructs after linearization with XbaI. These uncapped transcripts were translated in the presence of ^{35}S-methionine (Amersham Biosciences) using rabbit reticulocyte lysate (Promega). Translation products were separated by SDS–PAGE (12% acrylamide); phosphorimages were analyzed with ImageQuant (Molecular Dynamics).

Autonomous retrotransposition assay

Retrotransposition assays were essentially as described (29) using 143B osteosarcoma cells. Briefly, 2.5×10^5 143B cells were seeded in each well of a 6-well plate the day before transfection. Cells were transfected with 2 μg of DNA using LipoFectamine 2000 as recommended (Invitrogen). Cells were selected for transfection with 10 μg/ml puromycin 24 h post-transfection. Cells were then harvested after 96 h of puromycin selection and flow cytometry was performed using a FACS Calibur flow cytometer (Becton Dickenson).

RESULTS

Evidence for an IRES upstream of both ORF1 and ORF2 in mouse L1

L1 sequences were tested in pRF, a dicistronic reporter vector. This vector contains Rluc cloned downstream of a composite SV40/T7 promoter as the first cistron, and Fluc as the second cistron. A highly structured element, ΔEMCV, between the two reporter genes but upstream of the cloning site for the IRES candidate sequences from L1, provides an effective barrier to expression of firefly luciferase by read-through translation (Figure 1B). Stimulation of Fluc expression is indicative of IRES activity (25). The known CrPV intergenic IRES serves as a positive control.

Mouse L1 sequences extending 5′ of the initial AUGs of both ORF1 and ORF2 (Figure 1A), as well as several sequences taken from the 3′ end of ORF2 or the 3′-UTR that vary in their length and GC-content, were examined for IRES activity. To test whether an IRES is present upstream of ORF1, a construct containing 400 nt upstream of the first L1 AUG was cloned into pRF (400-1 UTR). This region was chosen as a starting point in an attempt to avoid the known promoter activity of the monomer region; promoter activity is minimal with one or less monomer, and increases with increasing copies of monomers (17). Briefly, 400-1 UTR activated translation of the Fluc at least as well as that of the positive control, CrPV [(28), Figure 1C]. Likewise, a robust stimulation of Fluc expression was detected when an L1 fragment extending 201 nt upstream of the ORF2 AUG was tested (201-1 IGR, Figure 1C). In contrast, all of the other regions of L1 tested showed expression of Fluc similar to the empty vector (Figure 1C), pRF, demonstrating that the ability to function as an IRES is not a general feature of L1 sequence.

Four additional assays were used to test whether translation of monocistronic transcripts, rather than IRES activity on a dicistronic transcript, is responsible for Fluc expression. First, RNA was isolated from the cells transfected with the dicistronic vector containing 400-1 UTR or 201-1 IGR in order to determine whether a shorter monocistronic RNA was detectable that could account for expression of Fluc. The expected 4.4 kb dicistronic transcript, but no smaller transcript, was detected by northern hybridization using the firefly luciferase cDNA probe (Figure 2A).

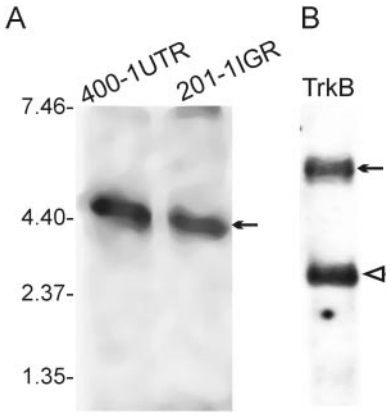

Figure 2. Northern blot analysis of Fluc RNA. Phosphorimage of northern blot of RNA recovered from DNA transfection of the dicistronic constructs containing L1 (**A**) or TrkB (**B**) sequences as indicated. Blots were hybridized to Fluc. The positions of size standards run in the adjacent lane are indicated. Full-length dicistronic transcripts are indicated by the arrows; the truncated TrkB Fluc transcript from the cryptic promoter is indicated by the open arrowhead. The coding sequence for Fluc is 1656 nt.

Second, these regions of L1 were tested in a promoterless dicistronic reporter vector to evaluate whether they contain cryptic promoters. The creation of truncated, monocistronic RNAs by cryptic promoters in the test sequences is an alternative mechanism that would elevate Fluc expression in these constructs (27). The known CrPV IRES was used as a negative control because this sequence is known to function as an IRES rather than a promoter (28). The 312 nt from the 3′-UTR of L1, which had no detectable IRES activity in pRF (Figure 1), was also inactive in the context of the promoterless construct (Figure 3), with Fluc values equivalent to CrPV (Table 1). Briefly 400-1 UTR and 201-1 IGR from L1 were both more active than CrPV and 312 in this assay (Figure 3 and Table 1), suggesting that they both contain weak cryptic promoters.

It appears that some of the Fluc expression driven by either 400-1 UTR or 201-1 IGR from L1 is accounted for by a cryptic promoter in each of these sequences, with the greatest promoter activity residing within the 400-1 UTR. However, these activities are very weak compared with the 5′ end of mouse TrkB (containing exon 1 and 2), which contains a cryptic promoter that is strong enough to produce a truncated mRNA that can be readily detected by northern blot hybridization (Figure 2B). In addition, if Fluc expression in the promoter-containing dicistronic vector is due to a combination of translation by IRES activity from a dicistronic mRNA and promoter activity from a truncated monocistronic RNA, then the Fluc activity seen in the promoterless assay (where dicistronic mRNA is eliminated as seen by the Rluc values in Table 1) must represent the activity due to the cryptic promoter. Since this accounts for only a fraction of the total Fluc

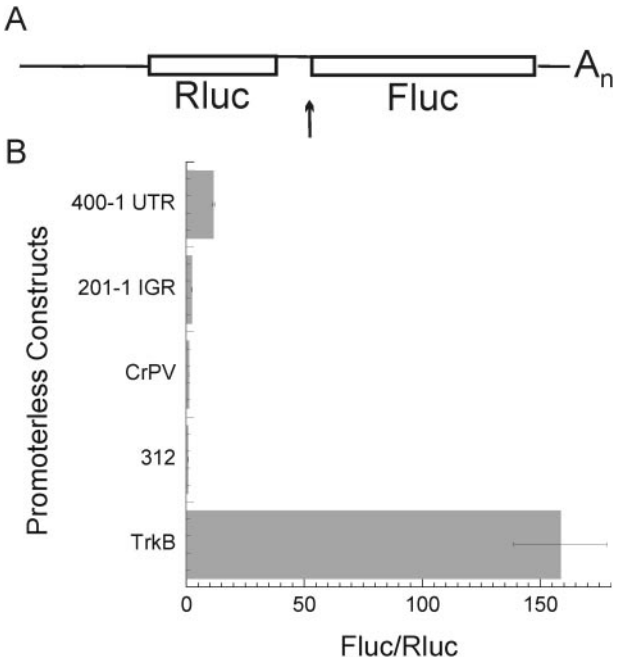

Figure 3. Promoterless vector assays. (**A**) Schematic of promoterless vector, the arrow indicates the insertion site of the test sequences. (**B**) Relative luciferase activities (to CrPV ± SD) of the 400 nt (400-1 UTR) or 201 nt (201-1 IGR) upstream of ORF1 and ORF2 compared with 312 (L1 3′-UTR) and mouse TrkB (exon 1 and 2) are plotted below the schematic of the promoterless vector. This region of TrkB contains a cryptic promoter (Figure 2). 312 is also devoid of activity in the dicistronic assay of Figure 1.

Table 1. Rluc and Fluc activities from dicistronic reporter constructs

	With promoter		Without promoter	
	Rluc	Fluc	Rluc	Fluc
400-1 UTR	208.74 ± 18.43	26.59 ± 0.81	0.06 ± 0.00	4.47 ± 0.31
201-1 IGR	324.29 ± 21.67	72.84 ± 11.08	0.06 ± 0.02	0.77 ± 0.18
CrPV	282.54 ± 35.88	28.26 ± 4.16	0.06 ± 0.00	0.36 ± 0.05
312	309.49 ± 28.51	2.82 ± 0.08	0.08 ± 0.01	0.32 ± 0.02
pRF	240.77 ± 56.95	1.75 ± 0.37	ND	ND
TrkB	ND	ND	0.08 ± 0.01	78.59 ± 5.04
Mock	0.06 ± 0.01	0.07 ± 0.01	0.06 ± 0.01	0.07 ± 0.01
Blank	0.07	0.08	0.07	0.08

Average and SD luminometer readings for triplicate samples after transfection with promoter-containing versus promoterless dicistronic reporter constructs (Figures 1 and 3). Mock, untransfected cells; blank, no cells; ND, not determined.

expression (Table 1), the majority of Fluc expression appears to be due to IRES activity in both of the L1 sequences tested. It is not unexpected that the 400-1 UTR had weak promoter activity, because this region exhibits weak promoter activity (10% of full-length TFspa, which is 2-fold over background) in a monocistronic reporter gene assay (17).

Third, dicistronic mRNAs containing 400-1 UTR and 201-1 IGR were transcribed *in vitro* and then transfected into cells to assay for Fluc and Rluc activity. The dicistronic reporter assay with *in vitro* transcribed RNA is effective for identifying IRESes because it bypasses the possibility of Fluc expression from monocistronic RNAs created by cryptic promoter and splicing activities. Both of these L1-containing constructs showed similar Fluc expression as the CrPV IGR IRES

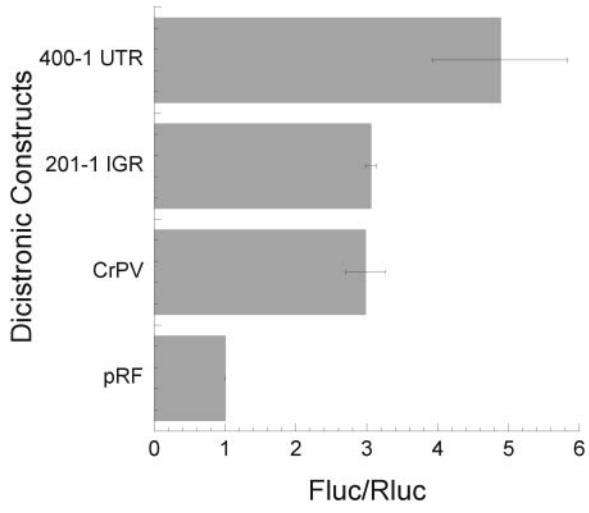

Figure 4. RNA transfection. Relative luciferase activity (to pRF ± SD) following transfection of RNA transcribed *in vitro* from pRF constructs as depicted in Figure 1.

(Figure 4), further supporting the presence of an IRES upstream of both ORF1 and ORF2 in mouse L1 RNA. RNA transfections of the same constructs have been reported to show lower activities than DNA transfections for some cellular IRESes. To explain this anomaly, it was proposed that the mRNA undergoes 'nuclear events' that are important for IRES function by virtue of being transcribed in the nucleus (30–32). Such a discrepancy between DNA and RNA transfection for the dicistronic reporter assay was also observed for the ORF2 IRES; 201-1 IGR was twice as active as CrPV when DNA was transfected (Figure 1C), but was at the same level as CrPV when RNA was transfected (Figure 4). This discrepancy is even more dramatic if one considers the fact that 400-1 UTR has some cryptic promoter activity, which is negligible in 201-1 IGR (Figure 3). It is possible that 201-1 IGR requires a 'nuclear event' to function optimally as an IRES, which did not occur when the *in vitro* transcript was transfected.

Finally, if the sequences between Rluc and Fluc are truly acting to recruit ribosomes by internal initiation, then translation of the second cistron, Fluc, should be independent of the translation of Rluc (28,33). To test this, we used *in vitro* translation in rabbit reticulocyte lysates programmed with RNA made *in vitro* using T7 polymerase. The L1 sequences under investigation were again cloned between the two luciferases. In contrast to the transfection assays above, however, Rluc was preceded by either an active IRES from EMCV, or the inactive ΔEMCV sequence, rendering its translation either robust or nearly undetectable, respectively. In spite of the dramatic reduction in the amount of Rluc protein (compare the Rluc signals in Figure 5A and B), the amounts of Fluc remained the same in constructs containing either of the two L1 IRESes or the CrPV IRES (Figure 5A and B).

Collectively, the results of these five assays argue strongly that mouse L1 sequences upstream of ORF1 and ORF2 promote internal initiation of translation on dicistronic mRNAs, rather than the creation of monocistronic mRNAs that are then translated by a classic cap-dependent mechanism.

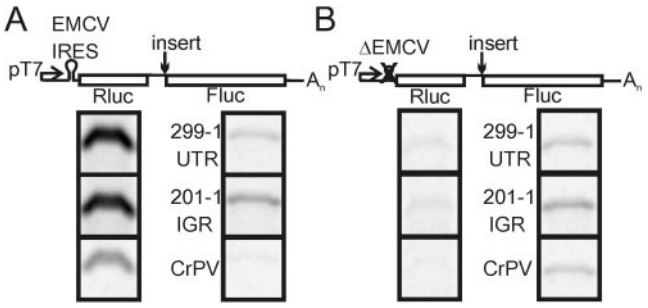

Figure 5. Test of translational independence of the second cistron in the dicistronic reporter construct. Phosphorimages of SDS–PAGE gels containing proteins labeled with [^{35}S]methionine during *in vitro* translation. Rabbit reticulocyte lysates were programmed with 20 ng/μl of dicistronic RNA transcribed *in vitro* by T7 polymerase as indicated. Rluc translation is controlled by either the EMCV IRES (**A**), or ΔEMCV (**B**), Fluc translation is controlled by the three indicated sequences. Results identical to these were obtained using 10 ng/μl of RNA in separate reactions.

Mapping of the ORF1 IRES

pRF constructs containing progressive truncations from the 5′ or 3′ end, or both, of 400-1 UTR were tested in the dicistronic and promoterless reporter assay in an attempt to localize the sequences important for the ORF1 IRES. Figure 6 shows the relative luciferase activities obtained from these fragments in both the dicistronic and promoterless constructs. Removing 100 nt from the 5′ end partially reduced the IRES activity, but the additional deletion of 100 or 200 nt did not lead to significant loss of IRES activity. The shortest 44 nt sequence tested was significantly reduced, but still retained 5-fold IRES activity over background (Figure 6A, middle panel). Although the longer constructs did have minimal cryptic promoter activity, neither of the shorter two constructs did (Figure 6A, right panel). A similar series of 3′ truncations were also tested. Removing 100 nt from the 3′ end had a similar effect to removing 100 nt from the 5′ end. Unlike the 5′ end truncations, however, removing an additional 100 nt, or more, from the 3′ end eliminated IRES activity (Figure 6B, middle panel). These data suggest that the essential sequence lies within the 200 nt immediately upstream of the ORF1 AUG. Consistent with this, when 100 nt fragments across the region were tested, the two 3′-most fragments (101-1 and 199-102) had IRES activity, but the 5′-most fragment (401-310) did not. Surprisingly, fragment 302-202 also had robust IRES activity (Figure 6C, middle panel). This was unexpected because deleting this sequence from the 5′ end of 299-1 had no effect on the activity (compare 299-1 and 199-1 in Figure 6A). It was also intriguing to see that both 302-202 and 199-102 were as active as 400-1 (Figure 6C, middle panel). This could be due to the presence of an inhibitory sequence in 400-1 that reduces the combined activity of 302-202 and 199-102 to the level of 400-1. If this is true, then 302-102 would be expected to have even higher IRES activity than 400-1. Remarkably, 302-102 is at best a weak IRES. The activities of the other two 200 nt constructs were also reduced relative to the 100 nt segments that comprised them (Figure 6C and D, middle panel).

Taken together, these data indicate that the IRES activity is largely confined to the 300 nt upstream of ORF1. The data further suggest that the activity depends upon RNA secondary structure rather than the sequences *per se*. It is likely that the RNA secondary structure is altered by these various sequence juxtapositions, causing the IRES function to appear or disappear. Indeed, Mfold analysis (34) predicted that 302-102 forms a structure very different than those formed by either 302-202 or 199-102 alone.

Mapping of the ORF2 IRES

A series of deletions in the 201-1 IGR construct were also tested in the dicistronic and promoterless reporter systems to delineate the region with IRES activity upstream of ORF2 (Figure 7A, middle and right panel). These deletions were based on the results of an Mfold analysis of 201-1 (34), which predicted very little stable secondary structure in this region of mouse L1. However, one significant stem–loop was consistently found in the most stable thermodynamically equivalent structures even when 201-1 was analyzed in the context of much longer surrounding L1 sequences (data not shown). This predicted stem–loop was the specific target of the deletion constructs. A total of 63 nt were deleted from the 5′ end of 201-1 (138-1) to just 5′ of the putative stem–loop. This deletion had reduced, but still significant IRES activity (Figure 7A, middle panel) and also eliminated all evidence of cryptic promoter activity (Figure 7A, right panel). A further deletion of 52 nt, removing the predicted stem–loop, reduced Fluc expression to background levels (86-1), whereas the 53 nt alone (138-86) had activity similar to 138-1 and CrPV (Figure 7A, middle panel). To directly test the importance of the L1 IGR for ORF2 IRES activity, those 40 nt were removed from 201-1 to make 201-41, which retained the IRES activity of 138-1 and 138-86. The 40 nt IGR alone, 40-1, like 86-1, was inactive (Figure 7A, middle panel). These results demonstrate that the IGR is not required for successful translation of the second cistron and place important sequences for translational initiation of ORF2 within the coding sequence of ORF1 (Figure 7A).

To further assess the significance of the putative stem–loop, we introduced three point mutations designed to disrupt base pairing and hence its formation. This modified sequence actually increased the apparent translation of Fluc (Figure 7B, middle panel, compare 138-1 with mut1), strongly suggesting that formation of this stem–loop is not important for IRES function. Consistent with this conclusion, introduction of three additional point mutations that restored complementarity to the stem–loop of mut1, to make mut2, effectively destroyed IRES activity (Figure 7B, middle panel, compare 138-1 and mut2). The 5′-most point mutation in mut2 inadvertently introduced an upstream AUG, which was followed by two in frame stop codons 27 nt downstream (Figure 7B, left panel). Because this small uORF would likely affect translation of Fluc, 1 bp was exchanged to eliminate the introduced AUG, to create mut4. The equivalent base pair exchange was also made in mut1, creating mut3, to allow direct comparison of the disrupted and restored stem–loop structures without the complication of the uORF (Figure 7B, middle panel). Both mut3 and mut4 had higher apparent IRES activity than the original mut1 and mut2 (Figure 7B, middle panel). It is difficult to discern, however, whether this is a real increase in IRES activity, or if it is caused by the slight increase in cryptic promoter activity (Figure 7B, right panel). Nonetheless, one can conclude that

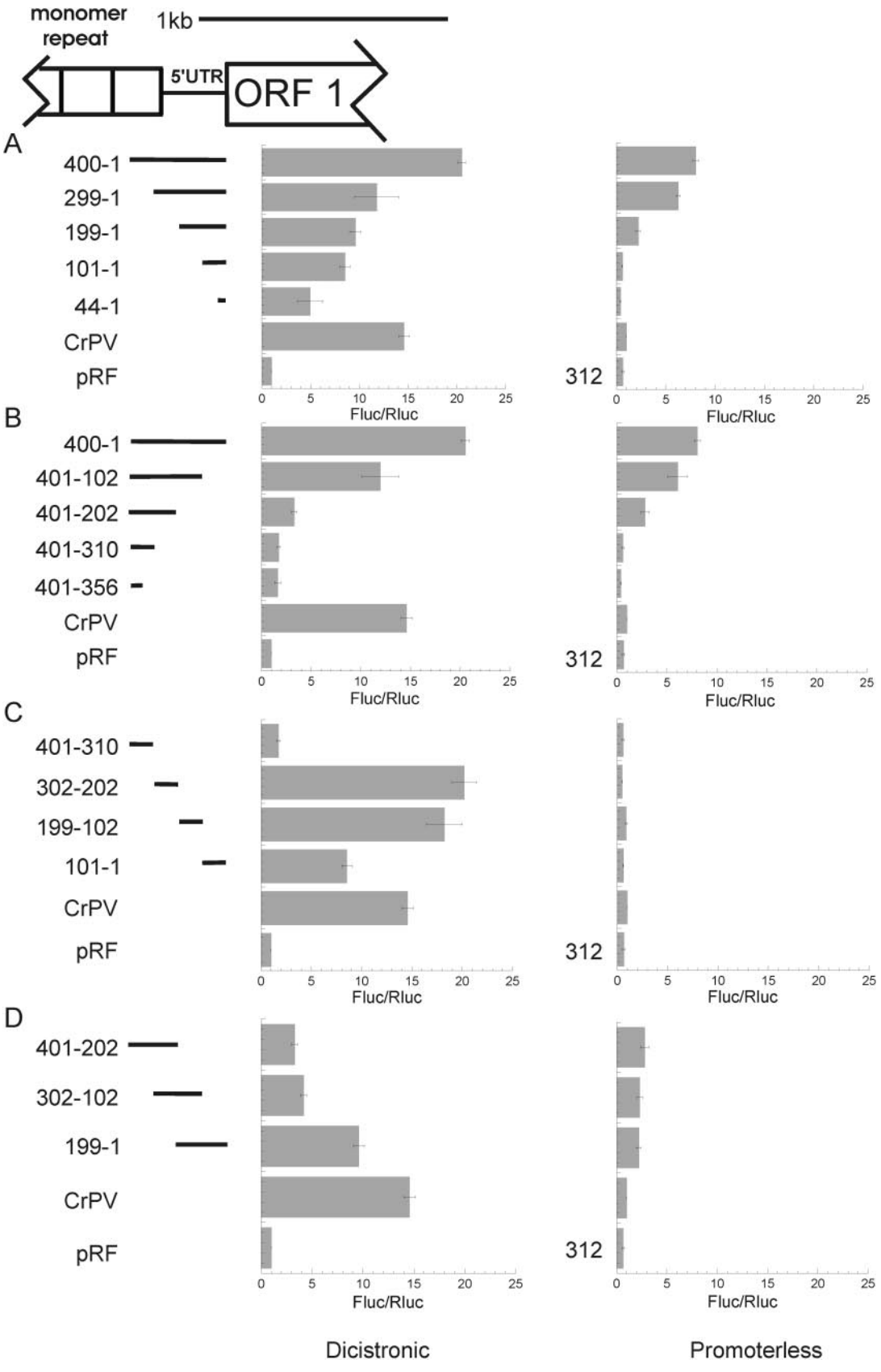

Figure 6. Mapping the ORF1 IRES. Schematics of deletion constructs upstream of ORF1 are shown on the left. These regions of L1 were cloned into either pRF or the promoterless vector and tested for luciferase expression; relative luciferase activities are plotted after normalization to pRF (dicistrionic, middle) or to the CrPV IRES (promoterless, right). (**A**) 5′ truncation series, (**B**) 3′ truncation series, (**C**) 100 nt fragments and (**D**) 200 nt fragments as indicated.

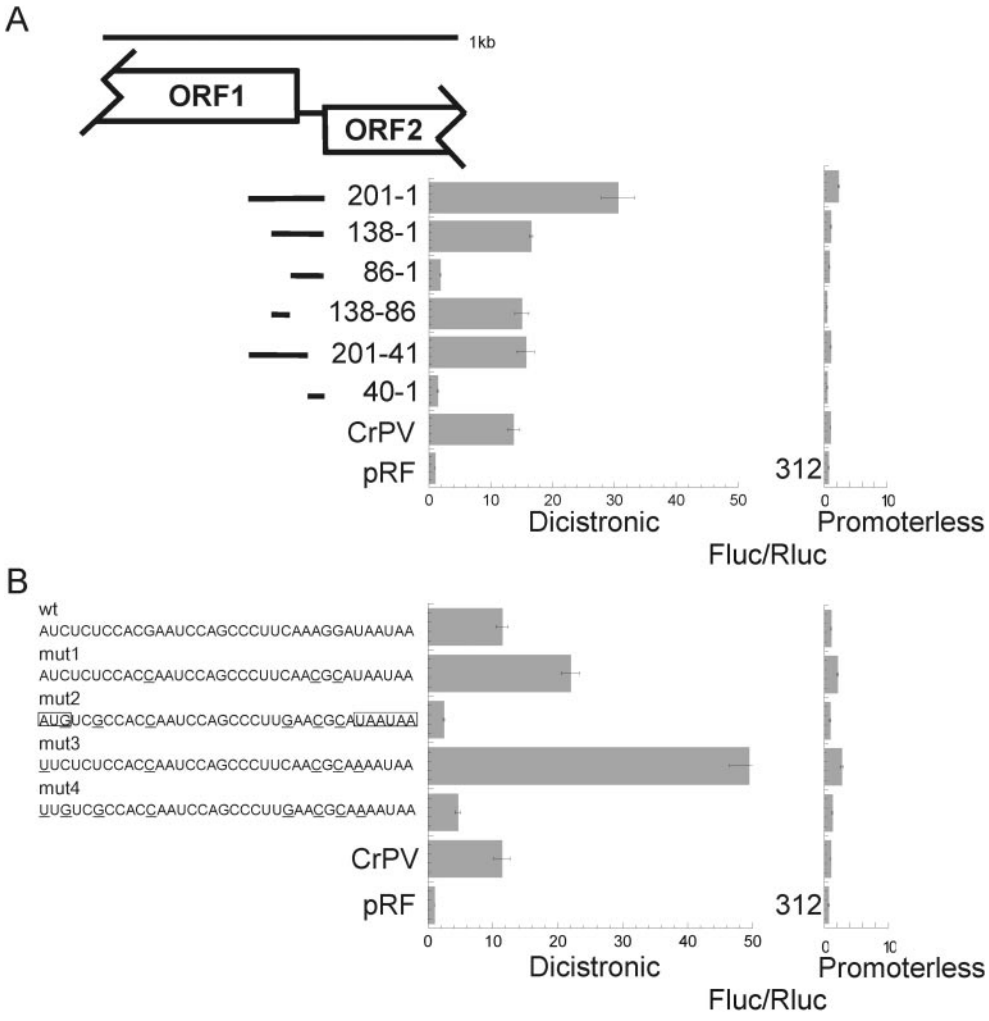

Figure 7. Mapping the ORF2 IRES. (**A**) Truncations and fragments of the ORF2 IRES tested in the dicistronic reporter assay with (dicistronic, middle) and without (promoterless, right) a promoter. Relative luciferase activities are plotted as in Figure 6. (**B**) The indicated point mutations in the putative stem–loop tested as in (A), point mutations are underlined, the new AUG in mut 2 and the resulting in frame stop codons (UAAUAA) are boxed.

the inability of mut2 to facilitate translation is not due to the inadvertent uORF, but rather due to one or more of the 3 nt substitutions that distinguish it from mut1. This is likewise the case for mut3 and mut4. Additionally, in both pairs of mutants Fluc translation is facilitated by the nt substitutions that disrupt base pairing, and severely impaired when base pairing is restored (compare mut1 to mut2, and mut3 to mut4 in Figure 7B, middle panel). The free energies of the predicted stem–loops are virtually identical among the original wt L1 sequence of 138-1, mut2 and mut4 (−3.4, −3.7 and −3.7 kcal/mol, respectively), yet Fluc expression in wt is equivalent to the known IRES in CrPV, whereas mut2 and mut4 are more like that of the negative control, pRF (Figure 7B, middle panel). The finding that restored base pairing inactivates the IRES suggests that specific sequences in 138-86, and not simply the melting of the stem–loop structure (35), are important for the function of the ORF2 IRES in mouse L1.

ORF2 translation and L1 retrotransposition

Naturally expressed mouse ORF2 protein has only been detected indirectly by the requirement for its activities during retrotransposition (18). This requirement was determined using the autononomous retrotransposition assay, which detects L1 retrotransposition events by expression of an antisense-intron marked reporter gene (36). We used the autonomous retrotransposition assay to examine the importance of the IGR and the putative stem–loop for ORF2 translation, as well as to test whether a ribosomal frameshift in the overlapping region of ORF1 and ORF2 in mouse L1 is responsible for ORF2 translation (37). As shown in Table 2, deletion of the IGR caused a modest decrease in retrotransposition frequency, similar to the decrease observed with the same deletion in the dicistronic reporter gene assay, confirming that the IGR is not crucial for translation of ORF2. It was also possible to introduce the three single nucleotide substitutions of mut1 into the retrotransposition vector, since these are silent in ORF1p. Retrotransposition of T_{FC}mut1 was indistinguishable from wt, also consistent with the results of the dicistronic assays. Mut2-4 were not tested because they altered the sequence of ORF1, which is essential for retrotransposition (9,36,38). To test whether a frameshift translates ORF2 during L1 retrotransposition, a UAA stop codon was introduced just 5′ of the first AUG in ORF2. This mutant retrotransposed

Table 2. Retrotransposition frequency of mouse L1 with mutations to test ORF2 translation

Construct	% Green fluorescent protein (GFP)
T_{FC}	4.83 ± 0.12
IRES control	4.70 ± 0.17
ΔIGR	2.50 ± 0.26
UAA-AUG	5.80 ± 0.26
AUG-UAA	2.00 ± 0.10
T_{FC}mut1	5.00 ± 0.10

IRES control contains a silent substitution in ORF2 that introduces a HpaI site for cloning into T_{FC} element (9), all other constructs are derived from it as indicated by the name: ΔIGR has the IGR between ORF1 and ORF2 removed, juxtaposing the UAA stop of ORF1 with the AUG start of ORF2. UAA-AUG has a stop codon inserted immediately 5′, and AUG-UAA immediately 3′, of the ORF2 AUG. T_{FC}mut1 contains the 3 nt substitutions of mut1 in Figure 7.

at least as well as wt, eliminating the possibility that ORF2 is translated by a ribosomal frameshift between ORF1 and ORF2. Intriguingly, introduction of a UAA just 3′ of the first AUG in ORF2 also failed to abolish retrotransposition. Because the next downstream AUG in ORF2 is 3′ of critical residues in the endonuclease domain (11), ORF2 translation must not be strictly AUG dependent.

DISCUSSION

Dicistronic mRNAs are generally considered rare in eukaryotes, although perhaps this view should be reconsidered given that the ∼3000 active (39) dicistronic copies of L1 in the mouse comprise roughly 1% of its genome. Because insertion of a new, full-length copy of L1 (e.g. retrotransposition) requires the products of both ORFs (36) to be supplied *in cis* (14,15), L1 RNA is necessarily dicistronic when its two ORFs are translated. This feature, taken together with the enormous potential for heterogeneity in both length and base composition of the region upstream of ORF1 in mouse L1, suggests that neither of these two proteins could be efficiently translated by the standard cap-dependent recognition system followed by ribosome scanning to the first AUG. An alternative mechanism for translational initiation is to recruit ribosomes internally, using an IRES. IRESes are found in a wide variety of viral and cellular RNAs, and are functionally defined by their ability to promote independent translation of the second cistron in a dicistronic RNA (40). By this criterion, the data presented here demonstrate that the sequences upstream of both ORF1 and ORF2 in mouse L1 RNA are IRESes.

Results of dicistronic reporter assays using 5′ and 3′ truncated sequences, as well as four internal 100 nt fragments, localized the ORF1 IRES to the 300 nt 5′ of the first AUG. This conclusion is based upon three observations: (i) in the 5′ truncation series there was a progressive reduction, but never complete loss of activity, even when only 44 nt of the 5′-UTR remained; (ii) complete loss of IRES function was observed when the 3′ 200 nt were removed in the 3′ truncation series, suggesting that the 3′ 200 nt are critical; (iii) the three 3′ 100 nt fragments, but not the 5′ 92 nt, had IRES activity. Interestingly, when two active, adjacent 100 nt fragments were joined to form 200 nt fragments, IRES activity was significantly decreased compared with the 100 nt fragments alone (Figure 6C and D). Similar modular activities have been reported for other cellular IRESes, including TrkB (41), c-myc (42), N-myc (43) and L-myc (44). These L1 results raise the possibility that RNA structure is critical for the function of the ORF1 IRES, which is common for viral IRESes (45–48) as well as some cellular IRESes (35,44,49).

The ORF2 IRES was localized to 53 nt in the 3′ end of ORF1. Surprisingly, the IGR was not required for either IRES function or L1 retrotransposition, since greater than half of these activities remained when the IGR was deleted. Although the highest IRES activity was observed with the longest construct (201-1), any deletion that retained the 53 nt region of ORF1, including the 53 nt alone, maintained robust IRES activity. In contrast, any construct with these 53 nt deleted lost IRES activity. The loss of activity between 201-1 and 138-1 may simply be due to elimination of the small amount of cryptic promoter activity in 201-1 (Figure 7B). Alternatively, because removing either the 5′ 63 nt or the IGR reduces IRES activity equally, an interaction between these two regions may enhance IRES activity. Just three point mutations, those that distinguish mut4 from mut3 and mut2 from mut1, in the 53 nt IRES destroyed activity, demonstrating that IRES function is abolished by one or more of these changes. These mutations may disrupt a direct interaction with the 40S subunit (50), or a factor that recruits the 40S subunit, e.g. an IRES *trans*-acting factor [ITAF (20,21)].

These experimental results also provide new information to address the question of how ORF2 is translated. Because ORF1 and ORF2 in mouse L1 overlap by 14 nt, it is theoretically possible that ORF2 could be translated by a +1 frameshift of the ribosome during elongation on ORF1 (37). Furthermore, ribosomal frameshifting is a widely-used mechanism for translation of the enzymatic components of other retrotransposons and retroviruses (51), which is exactly the role of ORF2 in L1 retrotransposition. To test the possibility that a frameshift is used to translate mouse ORF2, a stop codon was introduced just 5′ of the first AUG in ORF2. Retrotransposition of L1 was unaffected, if not improved (Table 2), effectively ruling out a role for frameshifting in the translation of ORF2 in mouse L1. This finding is consistent with experiments that eliminated frameshifting for human (52) and rat L1 (53). Given the juxtaposition of ORF2 relative to ORF1, reinitiation following ORF1 translation is another possibility. Ribosomal reinitation was first revealed in studies of the yeast GCN4 gene and has more recently been documented for the ATF4 gene in mammals (23). Translational coupling is also formally a possibility and has been shown to be responsible for translation of the ORF2 protein in the non-LTR-retrotransposon SART1 (54). However, translation of the downstream ORF by either of these mechanisms is dependent upon translation of the uORF. This is not the case for the L1 ORF2 IRES based upon our *in vitro* translation data, where translation of the downstream ORF occurs independently of translation of the uORF (Figure 5). These findings argue against ORF2 translation by frameshifting, ribosome reinitation or translational coupling in mouse L1.

How are translation of ORF1 and ORF2 regulated during retrotransposition? Although the stoichiometry of ORF1p, ORF2p and L1 RNA in the L1 RNP is unknown, ORF1p binds sequence non-specifically to RNA with high affinity (7). The protein binds as a homotrimer (55), with each trimer occupying 50 nt (56). Large RNPs containing ORF1p and

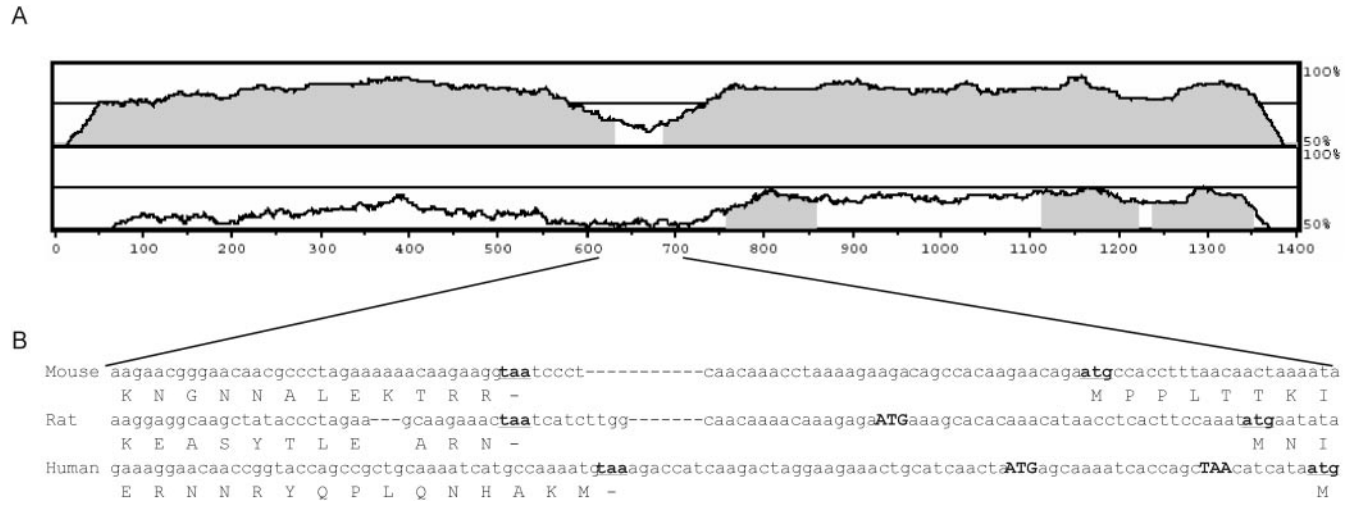

Figure 8. Rapid evolution in the IGR of L1. (**A**) VISTA plot (http://genome.lbl.gov/vista/mvista/submit.shtml) to visualize sequence similarity between mouse L1 and rat (top) or human (bottom) from the conserved region of ORF1 [amino acid 150 (55)] through the conserved endonuclease domains of ORF2 (10) including the IGR. Percent identity (between 50–100%) is plotted for 100 nt windows across the length compared, gray fills are regions with at least 60% identity for 100 nt. (**B**) Enlargement of the IGR of mouse L1 aligned with the homologous rat and human sequences. Sequence begins in the poorly conserved C-terminal region of ORF1 and extends through the first AUG of human ORF2. The ORF1 stop and ORF2 start codons are bold and underlined in all three sequences; the AUGs in the IGR of rat and human are bold and capitalized, as are their in frame stops which would lead to translation of 18 or 6 amino acids peptides, respectively. Both of these uORFs in the IGR require a +1 frameshift to match the frame of the ORF2 AUG.

L1 RNA have been observed in mouse and human cells (4,57,58) and are known to be a necessary intermediate in L1 retrotransposition (38). Assuming that the final RNP complex contains full-length L1 RNA (7521 nt) coated with ORF1p, 450 translations of ORF1 are needed. On the other hand, ORF2p contains endonuclease and reverse transcriptase activities, and is likely needed in just one or two copies for the TPRT reaction (59,60). Because ORF2p overexpression is highly toxic to cells (61,62), it is likely critical that ORF2 translation is tightly controlled, such that far less is made compared with ORF1p. Yet in our dicistronic reporter assays, there is little difference between the efficiencies of the sequences upstream of the ORF1 or ORF2 AUG in promoting Fluc expression. In order to explain the paradox that both IRESes appear similarly efficient in our assays, even though far fewer translations of ORF2 than of ORF1 are required for L1 retrotransposition, we propose a model whereby ORF1 translation inactivates the ORF2 IRES. In this model, the first round of translation is initiated efficiently on both ORF1 and ORF2. However, once ORF1 is translated, the ORF2 IRES is inactivated, thereby stopping all further translation of ORF2. This could be achieved by ORF1p binding to L1 RNA immediately as it is released from the ribosome, which is in the near vicinity of the ORF2 IRES. If the ORF2 IRES uses a land and scan mechanism as reported for c-myc, L-myc and Apaf1 (35,44,49), perhaps ORF1 protein bound to the IGR simply blocks 43S ribosome scanning. Self-regulation in IRES-mediated translational initiation has been proposed for *Idefix*, an LTR-retrotransposon in *Drosophila melanogaster* (63) and HCV (64,65), although neither of these are dicistronic mRNAs. Alternatively, it may be that the ORF2 IRES is only recognized when bound by an ITAF (20), perhaps added as a consequence of the nuclear experience. Once a ribosome elongating through ORF1 causes this factor to dissociate, the ORF2 IRES is silenced. Additional experiments are needed to test the predictions of these models.

Significantly, neither of the two regions in mouse L1 identified in this study as IRESes are well-conserved among mammalian L1s. The 5′ end of L1, including the promoter, 5′-UTR and N-terminal third of ORF1, appears to have an independent evolutionary origin in humans, rodents and rabbits, rather than being related by descent [reviewed in (2)]. Thus, the mouse L1 sequence containing the ORF1 IRES detected here is not shared among mammalian L1s. Another significant difference between the 5′-UTRs of the mouse and human elements is that human L1 typically contains at least one AUG upstream of ORF1 (52), whereas mouse L1 does not. In contrast, the ORF2 IRES is in a region that is homologous among mammalian L1s (37), but within that region the sequences surrounding the ORF2 AUG, particularly extending to the 5′, are strikingly divergent (Figure 8). The sequences encompassing the last eight amino acids of ORF1 and the first seven amino acids of ORF2 as well as the IGR in mouse and human L1 are difficult to align. There are also several differences between mouse and human elements in sequences that are significant for translation: the mouse IGR is 40 nt, has no AUGs and has just one termination codon (UAA) in the third reading frame that does not encode either ORF1 or ORF2; the human IGR is 63 nt (keeping ORF1 and ORF2 in the same frame) and contains an AUG as well as five stop codons with at least one in each of the three reading frames. Intriguingly, one of these stop codons overlaps the ORF2 AUG, creating the sequence UAAUG, which is identical to the sequence associated with translational coupling for ORF2 expression in SART1 (54). Furthermore, a strong stem–loop structure, similar to one required for the translation of SART1 ORF2, can be found downstream of the ORF2 AUG in human L1 by Mfold analysis. Neither the overlapping stop-start codon, nor the structural element is evident in mouse L1. These sequence differences likely indicate, or even necessitate, the use of different strategies for initiation of translation of ORF2 in mouse compared with human L1. It is intriguing to speculate

that these regions could be targets for regulation (and perhaps suppression) of L1 retrotransposition by altering the translation of ORF1p and ORF2p. If this is the case, the high sequence divergence observed between transposition-competent mouse and human L1s may indicate a shift in the strategy used by L1 in these two species to recruit ribosomes for translation of both of the L1-encoded essential retrotransposition proteins. Thus the rapid evolution of this region may reflect positive selection to either attract enhancers or lose inhibitors that regulate translation of ORF2 in L1 mRNA, rather than a lack of functional constraints.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr P. Sarnow (Stanford University) for pRF and the CrPV IRES, A. Knox for preliminary experiments, and the members of the Martin lab and the UCHSC translation club for helpful comments on the data and the manuscript. This work was supported by NIH GM40367 to S.L.M. and NIH NS42631 to L.A.K. Funding to pay the Open Access publication charges for this article was provided by the NIH GM40367.

Conflict of interest statement. None declared.

REFERENCES

1. Consortium,M.G.S. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
2. Furano,A.V. (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. *Prog. Nucleic Acid Res. Mol. Biol.*, **64**, 255–294.
3. Moran,J.V. and Gilbert,N. (2002) In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington,D.C., pp. 836–869.
4. Hohjoh,H. and Singer,M.F. (1996) Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.*, **15**, 630–639.
5. Hohjoh,H. and Singer,M.F. (1997) Sequence specific single-strand RNA-binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.*, **16**, 6034–6043.
6. Kolosha,V.O. and Martin,S.L. (1997) *In vitro* properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl Acad. Sci. USA*, **94**, 10155–10160.
7. Kolosha,V.O. and Martin,S.L. (2003) High affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J. Biol. Chem.*, **278**, 8112–8117.
8. Martin,S.L. and Bushman,F.D. (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.*, **21**, 467–475.
9. Martin,S.L., Cruceanu,M., Branciforte,D., Li,P.W.-I., Kwok,S.C., Hodges,R.S. and Williams,M.C. (2005) LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.*, **348**, 549–561.
10. Feng,Q., Moran,J.V., Kazazian,H.H.J. and Boeke,J.D. (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*, **87**, 905–916.
11. Mathias,S.L., Scott,A.F., Kazazian,H.H.,Jr, Boeke,J.D. and Gabriel,A. (1991) Reverse transcriptase encoded by a human transposable element. *Science*, **254**, 1808–1810.
12. Cost,G.J., Feng,Q., Jacquier,A. and Boeke,J.D. (2002) Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.*, **21**, 5899–5910.
13. Luan,D.D. and Eickbush,T.H. (1995) RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.*, **15**, 3882–3891.
14. Esnault,C., Maestre,J. and Heidmann,T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.*, **24**, 363–367.
15. Wei,W., Gilbert,N., Ooi,S.L., Lawler,J.F., Ostertag,E.M., Kazazian,H.H., Boeke,J.D. and Moran,J.V. (2001) Human L1 retrotransposition: *cis* preference versus trans complementation. *Mol. Cell. Biol.*, **21**, 1429–1439.
16. Kozak,M. (1989) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229–241.
17. DeBerardinis,R.J. and Kazazian,H.H.,Jr (1999) Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics*, **56**, 317–323.
18. Naas,T.P., DeBerardinis,R.J., Moran,J.V., Ostertag,E.M., Kingsmore,S.F., Seldin,M.F., Hayashizaki,Y., Martin,S.L. and Kazazian,H.H.,Jr (1998) An actively-retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.*, **17**, 590–597.
19. Kingsmore,S.F., Giros,B., Suh,D., Bieniarz,M., Caron,M.G. and Seldin,M.F. (1994) Glycine receptor beta-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. *Nature Genet.*, **7**, 136–142.
20. Hellen,C.U. and Sarnow,P. (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev.*, **15**, 1593–1612.
21. Stoneley,M. and Willis,A.E. (2004) Cellular internal ribosome entry segments: structures, *trans*-acting factors and regulation of gene expression. *Oncogene*, **23**, 3200–3207.
22. Hutchison,C.A.,III, Hardies,S.C., Loeb,D.D., Shehee,W.R. and Edgell,M.H. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, D.C., pp. 593–617.
23. Gebauer,F. and Hentze,M.W. (2004) Molecular mechanisms of translational control. *Nature Rev. Mol. Cell Biol.*, **5**, 827–835.
24. Yueh,A. and Schneider,R.J. (2000) Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. *Genes Dev.*, **14**, 414–421.
25. Johannes,G., Carter,M.S., Eisen,M.B., Brown,P.O. and Sarnow,P. (1999) Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. *Proc. Natl Acad. Sci. USA*, **96**, 13118–13123.
26. Barettino,D., Pombo,P.M., Espliguero,G. and Rodriguez-Pena,A. (1999) The mouse neurotrophin receptor trkB gene is transcribed from two different promoters. *Biochim. Biophys Acta.*, **1446**, 24–34.
27. Han,B. and Zhang,J.T. (2002) Regulation of gene expression by internal ribosome entry sites or cryptic promoters: the eIF4G story. *Mol. Cell. Biol.*, **22**, 7372–7384.
28. Wilson,J.E., Powell,M.J., Hoover,S.E. and Sarnow,P. (2000) Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol. Cell. Biol.*, **20**, 4990–4999.
29. Ostertag,E.M., Prack,E.T.L., DeBerardinis,R.J., Moran,J.B. and Kazazian,H.H.,Jr (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.*, **28**, 1418–1423.
30. Stoneley,M., Subkhankulova,T., Le Quesne,J.P., Coldwell,M.J., Jopling,C.L., Belsham,G.J. and Willis,A.E. (2000) Analysis of the c-myc IRES; a potential role for cell-type specific *trans*-acting factors and the nuclear compartment. *Nucleic Acids Res.*, **28**, 687–694.
31. Wang,Z., Weaver,M. and Magnuson,N.S. (2005) Cryptic promoter activity in the DNA sequence corresponding to the pim-1 5′-UTR. *Nucleic Acids Res.*, **33**, 2248–2258.
32. Shiroki,K., Ohsawa,C., Sugi,N., Wakiyama,M., Miura,K., Watanabe,M., Suzuki,Y. and Sugano,S. (2002) Internal ribosome entry site-mediated translation of Smad5 *in vivo*: requirement for a nuclear event. *Nucleic Acids Res.*, **30**, 2851–2861.
33. Van Eden,M.E., Byrd,M.P., Sherrill,K.W. and Lloyd,R.E. (2004) Translation of cellular inhibitor of apoptosis protein 1 (c-IAP1) mRNA is IRES mediated and regulated during cell stress. *RNA*, **10**, 469–481.
34. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 1–10.
35. Mitchell,S.A., Spriggs,K.A., Coldwell,M.J., Jackson,R.J. and Willis,A.E. (2003) The Apaf-1 internal ribosome entry segment attains the correct structural conformation for function via interactions with PTB and unr. *Mol. Cell*, **11**, 757–771.

36. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H.,Jr (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
37. Loeb,D.D., Padgett,R.W., Hardies,S.C., Shehee,W.R., Comer,M.B., Edgell,M.H. and Hutchison,C.A.,III (1986) The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. *Mol. Cell. Biol.*, **6**, 168–182.
38. Kulpa,D.A. and Moran,J.V. (2005) Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Gen.*, **14**, 3237–3248.
39. Goodier,J.L., Ostertag,E.M., Du,K. and Kazazian,H.H.J. (2001) A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.*, **11**, 1677–1685.
40. Pisarev,A.V., Shirokikh,N.E. and Hellen,C.U. (2005) Translation initiation by factor-independent binding of eukaryotic ribosomes to internal ribosomal entry sites. *C. R. Biol.*, **328**, 589–605.
41. Dobson,T., Minic,A., Nielsen,K., Amiott,E. and Krushel,L. (2005) Internal initiation of translation of the TrkB mRNA is mediated by multiple regions within the 5′ leader. *Nucleic Acids Res.*, **33**, 2929–2941.
42. Stoneley,M., Paulin,F.E., Le Quesne,J.P., Chappell,S.A. and Willis,A.E. (1998) C-Myc 5′ untranslated region contains an internal ribosome entry segment. *Oncogene*, **16**, 423–428.
43. Jopling,C.L. and Willis,A.E. (2001) N-myc translation is initiated via an internal ribosome entry segment that displays enhanced activity in neuronal cells. *Oncogene*, **20**, 2664–2670.
44. Jopling,C.L., Spriggs,K.A., Mitchell,S.A., Stoneley,M. and Willis,A.E. (2004) L-Myc protein synthesis is initiated by internal ribosome entry. *RNA*, **10**, 287–298.
45. Jan,E. and Sarnow,P. (2002) Factorless ribosome assumbly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.*, **324**, 889–902.
46. Kanamori,Y. and Nakashima,N. (2001) A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. *RNA*, **7**, 266–274.
47. Evstafieva,A.G., Ugarova,T.Y., Chernov,B.K. and Shatsky,I.N. (1991) A complex RNA sequence determines the internal initiation of encephalomyocarditis virus RNA translation. *Nucleic Acids Res.*, **19**, 665–671.
48. Brown,E.A., Zhang,H., Ping,L.H. and Lemon,S.M. (1992) Secondary structure of the 5′ nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res.*, **20**, 5041–5045.
49. Le Quesne,J.P., Stoneley,M., Fraser,G.A. and Willis,A.E. (2001) Derivation of a structural model for the c-myc IRES. *J. Mol. Biol.*, **310**, 111–126.
50. Hu,M.C., Tranque,P., Edelman,G.M. and Mauro,V.P. (1999) rRNA-complementarity in the 5′ untranslated region of mRNA specifying the Gtx homeodomain protein: evidence that base pairing to 18S rRNA affects translational efficiency. *Proc. Natl Acad. Sci. USA*, **96**, 1339–1344.
51. Farabaugh,P.J. (1996) Programmed translational frameshifting. *Annu Rev Genet.*, **30**, 507–528.
52. McMillan,J.P. and Singer,M.F. (1993) Translation of human LINE-1 element, L1Hs. *Proc. Natl Acad. Sci. USA*, **90**, 11533–11537.
53. Ilves,H., Kahre,O. and Speek,M. (1992) Translation of the rat LINE bicistronic RNAs *in vitro* involves ribosomal reinitiation instead of frameshifting. *Mol. Cell. Biol.*, **12**, 4242–4248.
54. Kojima,K.K., Matsumoto,T. and Fujiwara,H. (2005) Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Mol. Cell. Biol.*, **25**, 7675–7686.
55. Martin,S.L., Branciforte,D., Keller,D. and Bain,D.L. (2003) Novel trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl Acad. Sci. USA*, **100**, 13815–13820.
56. Basame,S., Li,P.W.-I., Howard,G., Branciforte,D., Keller,D. and Martin,S.L. (2006) Spatial assembly and RNA binding stoichiometry of a protein essential for LINE-1 retrotransposition. *J. Mol. Biol.*, in press.
57. Martin,S.L. and Branciforte,D. (1993) Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol. Cell. Biol.*, **13**, 5383–5392.
58. Martin,S.L. (1991) Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.*, **11**, 4804–4807.
59. Christensen,S.M. and Eickbush,T.H. (2005) R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol. Cell. Biol.*, **25**, 6617–6628.
60. Martin,S.L., Li,P.W.-I., Furano,A.V. and Boissinot,S. (2005) The structures of mouse and human L1 elements reflect their insertion mechanism. *Cytogenet. Genome Res.*, **110**, 223–228.
61. Goodier,J.L., Ostertag,E.M., Engleka,K.A., Seleme,M.C. and Kazazian,H.H.,Jr (2004) A potential role for the nucleolus in L1 retrotransposition. *Hum. Mol. Genet.*, **13**, 1041–1048.
62. Clements,A.P. and Singer,M.F. (1998) The human LINE-1 reverse transcriptase: effect of deletions outside the common reverse transcriptase domain. *Nucleic Acids Res.*, **26**, 3528–3535.
63. Meignin,C., Bailly,J.L., Arnaud,F., Dastugue,B. and Vaury,C. (2003) The 5′ untranslated region and Gag product of Idefix, a long terminal repeat-retrotransposon from *Drosophila melanogaster*, act together to initiate a switch between translated and untranslated states of the genomic mRNA. *Mol. Cell. Biol.*, **23**, 8246–8254.
64. Zhang,J., Yamada,O., Yoshida,H., Iwai,T. and Araki,H. (2002) Autogenous translational inhibition of core protein: implication for switch from translation to RNA replication in hepatitis C virus. *Virology*, **293**, 141–150.
65. Shimoike,T., Mimori,S., Tani,H., Matsuura,Y. and Miyamura,T. (1999) Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *J. Virol.*, **73**, 9718–9725.