Journal of Virology

Structural and Functional Analysis of the RNA Transport Element, a Member of an Extensive Family Present in the Mouse Genome

Sergey Smulevitch, Daniel Michalowski, Andrei S. Zolotukhin, Ralf Schneider, Jenifer Bear, Patricia Roth, George N. Pavlakis and Barbara K. Felber

J. Virol. **2005, 79(4):2356. DOI: 10.1128/JVI.79.4.2356-2365.2005.**

JOURNAL OF VIROLOGY, Feb. 2005, p. 2356–2365 Vol. 79, No. 4
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.4.2356–2365.2005

Structural and Functional Analysis of the RNA Transport Element, a Member of an Extensive Family Present in the Mouse Genome

Sergey Smulevitch,[1]† Daniel Michalowski,[1]† Andrei S. Zolotukhin,[1] Ralf Schneider,[2] Jenifer Bear,[1] Patricia Roth,[3] George N. Pavlakis,[3] and Barbara K. Felber[1]*

Human Retrovirus Pathogenesis Section[1] and Human Retrovirus Section,[3] National Cancer Institute—Frederick, Frederick, Maryland, and Institut für Experimentelle Genetik/AG BIODV, GSF-Forschungszentrum für Umwelt und Gesundheit GmbH, Oberschleissheim, Germany[2]

Received 16 July 2004/Accepted 21 September 2004

We previously identified an RNA transport element (RTE), present in a subclass of rodent intracisternal A particle retroelements (F. Nappi, R. Schneider, A. Zolotukhin, S. Smulevitch, D. Michalowski, J. Bear, B. Felber, and G. Pavlakis, J. Virol. 75:4558–4569, 2001), that is able to replace Rev-responsive element regulation in human immunodeficiency virus type 1. RTE-directed mRNA export is mediated by a still-unknown cellular factor(s), is independent of the CRM1 nuclear export receptor, and is conserved among vertebrates. Here we show that this RTE folds into an extended RNA secondary structure and thus does not resemble any known RTEs. Computer searches revealed the presence of 105 identical elements and more than 3,000 related elements which share at least 70% sequence identity with the RTE and which are found on all mouse chromosomes. These related elements are predicted to fold into RTE-like structures. Comparison of the sequences and structures revealed that the RTE and related elements can be divided into four groups. Mutagenesis of the RTE revealed that the minimal element contains four internal stem-loops, which are indispensable for function in mammalian cells. In contrast, only part of the element is essential to mediate RNA transport in microinjected *Xenopus laevis* oocyte nuclei. Importantly, the minimal RTE able to promote RNA transport has key structural features which are preserved in all the RTE-related elements, further supporting their functional importance. Therefore, RTE function depends on a complex secondary structure that is important for the interaction with the cellular export factor(s).

mRNA export is a critical step in gene expression. The detailed analysis of the fundamental cellular processes that guide the complex assembly of mRNA and proteins (mRNP) from the nucleus to the cytoplasm is important for understanding the regulation of gene expression. The use of retroviral systems, pioneered by research on human immunodeficiency virus type 1 (HIV-1), has led to major discoveries in the field of mRNA metabolism and transport. Research over the past decade has revealed that many retroviruses depend on an efficient mechanism for nucleocytoplasmic export of their full-length RNA in its unspliced form, since this transcript encodes the Gag-Pol polyprotein and in addition serves as genomic RNA to be packaged into progeny virions in the cytoplasm. In HIV-1 and other lentiviruses, this process depends on the Rev protein. HIV-1 Rev is essential for production of structural proteins and infectious virions. Rev promotes export and expression of *gag-pol* and *env* mRNAs by binding to the *cis*-acting RNA recognition signal, the Rev-responsive element (RRE) (12, 15, 21). A similar mechanism was found in human T-cell leukemia virus type 1 (HTLV-1), where the transport of the unspliced transcript depends on the interaction of the viral Rex protein with the *cis*-acting Rex-responsive element RXRE (1), and in human endogenous retroviruses, which use the viral RNA export system termed cORF-RcRE or K-Rev-K-RRE for export (2, 19, 20, 35). The Rev proteins of all lentiviruses, the Rex proteins of the HTLV family of retroviruses, the Rev-like proteins of human endogenous retroviruses, and many cellular proteins share a leucine-rich nuclear export signal which is recognized by the cellular export receptor CRM1, thereby linking the mRNP cargo to the nuclear pore complex (for reviews, see references 6, 13, and 17).

In contrast, simple retroviruses, such as simian type D retrovirus (4, 9, 10, 33), and the avian sarcoma viruses, like Rons sarcoma virus (23–26), depend on cellular *trans*-acting factors for the transport of their unspliced RNA. Type D virus expression depends on the viral *cis*-acting constitutive transport element (CTE) (4, 9, 10, 32, 33, 38). Structure-function analysis of the CTE revealed an extended stem-loop structure with two essential symmetrical internal loop regions (9, 33), which are conserved among all type D retroviral CTEs (9, 32, 33). In addition, we had also identified a CTE-related element in an intracisternal A particle retroelement (IAP) (32) which is located within an intron of the murine osteocalcin-related gene. This element, CTE_{IAP}, preserves only the overall RNA structure and the sequence of the internal loop regions of the type D retroviral CTEs. The sequences of the double-stranded regions of CTE_{IAP} are divergent, but the changes are of compensatory nature, demonstrating that the stem structure and not the sequence of this region is important for function. The conserved internal loops contain the direct binding sites for the cellular mRNA export factor NXF1 (previously named TAP) (14). NXF1 not only promotes the nucleocytoplasmic export of the CTE-containing RNA but, importantly, is also a key factor in the export of cellular mRNAs (3, 8, 14), a function which is

* Corresponding author. Mailing address: Human Retrovirus Pathogenesis Section, Vaccine Branch, Bldg. 535, Rm. 110, NCI—Frederick, Frederick, MD 21702. Phone: (301) 846-5159. Fax: (301) 846-7152. E-mail: felber@ncifcrf.gov.

† S.S. and D.M. contributed equally to the present study.

conserved among eukaryotes (16, 29, 34). We previously identified an evolutionary precursor, the TAP-binding element, of type D retroviral CTEs and CTE_{IAP}, which consists of a minisatellite sequence encoding the essential motifs found in the internal loops of the CTE. This sequence also folds into a structure that is similar to that of type D retroviral CTEs and CTE_{IAP}, presenting the essential sequence motifs within a CTE-like context, and the TAP-binding element is functionally analogous to the CTE (36).

The discovery of the RNA transport element (RTE) identified another potent RNA export element which uses a cellular factor(s) for export. The RTE was identified in complementation experiments using a *rev*- and RRE-defective HIV-1 molecular clone. This screen was designed to identify RNA export elements in the mammalian genome by selecting for rescue of virus replication. Viruses rescued by this method contained the RTE, which belongs to a mouse IAP. Using the virus rescue assay, it was found that the RTE is as potent as the CTE in complementing a Rev- and RRE-mutated HIV (22).

Interestingly, on the basis of the presence of the RTEs (22) or CTE-related element, CTE_{IAP} (32), the murine IAPs can be grouped into two subclasses. The RTE functions in many cell types of different species, indicating that its export factor(s) is widely expressed and evolutionarily conserved. Microinjection experiments have shown that the RTE also functions in *Xenopus laevis* oocytes (22). Here we present a detailed structure and function analysis of the originally identified RTE (22) by using assays of both microinjected *Xenopus* oocyte nuclei and transfected human cells. We further identified an extensive family of RTEs and RTE-related elements in the mouse genome. Comparison of the sequences and the predicted structures of these RTE-related elements further supported the experimentally identified RTE structure and the role of its key features in RNA export.

MATERIALS AND METHODS

RNA structure probing. RNA end labeling and purification after in vitro transcription were previously described (37). Unlabeled tRNA (final concentration, 8 μM) was added to the RNA solution. The in vitro-transcribed RTE RNA was denatured at 55°C and allowed to renature at 25°C for 20 min. The buffer contained 10 mM Tris-Cl (pH 7.2), 10 mM $MgCl_2$, and 40 mM NaCl. Reactions were carried out at 25°C for 10 min. The following were used at the concentrations indicated: Pb^{2+} (lead acetate; Sigma), 1.25 mM; S1 nuclease (Pharmacia), 1,500 U/ml in the presence of 1 mM $ZnCl_2$; T_1 RNase (Pharmacia), 62.5 U/ml; mung bean nuclease I (Boehringer Mannheim), 1,000 U/ml in the presence of 1 mM $ZnCl_2$; P_1 nuclease (Pharmacia), 3 μg/ml; and T_2 RNase (Gibco BRL), 25 U/ml. The reactions were stopped by the addition of an equal volume of loading buffer (8 M urea, 20 mM EDTA, dyes). The products of enzymatic and metal ion cleavage were analyzed by electrophoresis on a denaturing 8% polyacrylamide gel followed by autoradiography. The RNA products were subjected to these processes along with the RTE products of formamide RNA hydrolysis and limited RNase T_1 digestion. The formamide hydrolysis ladder was generated by incubating RNA with 5 volumes of formamide containing 0.5 mM $MgCl_2$ at 100°C for 11 min. The partial T_1 nuclease digestion of the denatured RNA (to obtain cleavage after every G) was performed in the presence of a solution containing 25 U of enzyme/ml, 10 mM sodium citrate (pH 4.5), 0.5 mM EDTA, and 3.5 M urea at 55°C for 15 min.

RNA export from *Xenopus* oocyte nuclei. RTE mutants were generated by PCR, sequenced, and cloned into the unique SacII site in plasmid pBSAd1 (18) linearized with EcoRV (22). The CTE-containing adenovirus precursor RNA, U1ΔSm RNA, and U6Δss RNA were described previously (18, 27). Capped, radiolabeled RNA was prepared from 1 μg of the linearized plasmids by using a MaxiScript kit (Ambion) as described previously (22). One injection contained approximately 2×10^4 cpm of pre-mRNA. Cold RNAs were obtained using an mMessageMachine kit (Ambion). Nuclear and cytoplasmic fractions were obtained by manual enucleation. RNA was extracted from a pool of five oocytes after proteinase K digestion, and equivalents of one-half oocyte were loaded onto 10% polyacrylamide gels containing 7 M urea.

HIV-1 Gag expression in mammalian cells. The RTE mutants were cloned into the unique SacII site of the pNLgag plasmid (30) or the BamHI-KpnI sites of pB37gag located between *gag* and the 3′ long terminal repeat. pNLgag contains the major splice donor of HIV-1 located 5′ of *gag* and a cryptic splice acceptor site located 3′ of *gag* and the RTE, whereas pB37gag lacks the major HIV-1 splice donor and does not undergo splicing (28, 33). One microgram of plasmid was used to transfect HeLa-derived HLtat cells, which constitutively produce HIV-1 tat (11), with FuGene reagent (Roche). In addition to pNLgag, each transfection mixture contained 0.5 μg of the green fluorescent protein expression vector pFRED25 (31) as an internal control. Transfection efficiency was determined by fluorimetry of green fluorescent proteins. Expression of $p24^{gag}$ was determined with an HIV-1 p24 antigen enzyme-linked immunosorbent assay kit from Zeptometrix.

Nucleotide sequence accession numbers. The GenBank accession numbers for the RTE-related elements described in Figure 2 are AY876078 (group B) for the element found in AL663101, AY876079 (group C) for the element found in AL671215, and AY876077 (group D) for the element found in AC003061.

RESULTS

RTE folds into an extended RNA secondary structure. We previously identified the functional 226-nucleotide (nt) RTE RNA element referred to as RTE 391–616 (22). This fragment, referred to as RTE herein (the nucleotide numbering has been changed to 1 to 226), was used for further analysis. We applied the Zuker algorithm (mfold version 3.1) for structure predictions. Figure 1A shows the predicted RTE secondary structure containing four stem-loops (SL I to SL IV). The same secondary structure was predicted using a larger RTE (positions −10 to +237) and a smaller RTE (positions 66 to 206). The majority of the predicted structures shared the strong extended stem-loop SL IV with a ΔG of −84 kcal/mol.

To verify the predicted RTE structure, we used enzymatic and chemical methods of structure probing. The RTE RNA was tested in vitro by using lead ions and a set of specific enzymes (Fig. 1B). Pb^{2+} ions in general cleave the flexible single-stranded RNA regions (5). Nuclease P_1, mung bean nuclease I, and nuclease S1 cleave single-stranded RNA in a relatively base-independent manner; T_1 preferentially cleaves GpN bonds in single-stranded regions; and T_2 cleaves internucleotide bonds with preference for unpaired adenosine residues. Figure 1A shows a summary of these results, indicating strong and very strong cleavages in predicted single-stranded regions. Overall, the predicted structure is in very good agreement with the experimental data. One exception is the stem of SL II, where the cleavages indicate that this stem is more relaxed than indicated in the computer-predicted structure. As predicted from the structure, the distal (nt 123 to 145 and nt 151 to 181) and proximal (nt 104 to 109 and nt 194 to 199) parts of the stem of SL IV do not show enzymatic cleavage in these regions. Together with the weak reactivity induced by Pb^{2+} ions, this finding confirms the strongly paired and stacked character of this region. The distribution of Pb^{2+} cleavages in the adjacent internal loops suggests flexibility in specific areas of this region. In particular, the strong lead-induced cleavage at nt 103, together with the relaxed structure of the single-stranded sequence UUGCAU (nt 71 to 77) on the opposite side, suggests flexibility in the interhelical angle of the three-way junction and possible changes in the orientation of the stems in this region of the RTE molecule.

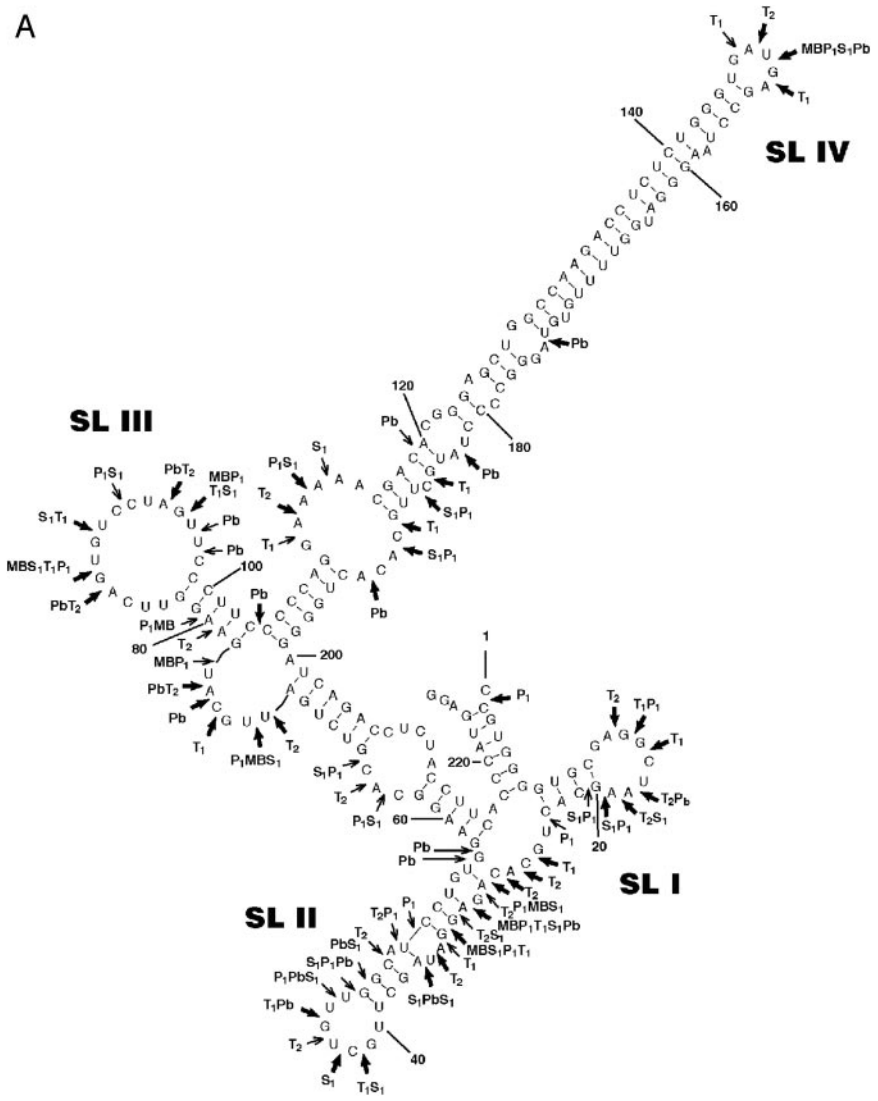


FIG. 1. Secondary structure of RTE. (A) The secondary structure of RTE is modeled using the mfold program and modified by experimental data. Arrows show the locations of specific sites of digestion induced by enzymatic probing and lead ion-induced cleavages. Thin and thick arrows indicate strong and very strong cuts, respectively. (B) Enzymatic and chemical probing of the 5′-end-labeled RTE molecule. Lanes: C, control (RTE RNA incubated without treatment); MB (mung bean nuclease I), P_1, S1, T_2, and T_1, digestions using the indicated enzymes; Pb, cleavages induced by lead ions; LE, RNase T_1 hydrolysis ladder; LF, formamide hydrolysis ladder. The positions of selected guanine residues and SL I to SL IV are indicated on the right.

RTE belongs to a family of related elements present in the mouse genome. We previously identified the presence of additional RTE-related elements in the rodent genomes (22). For a more complete analysis of these elements, the sequence of RTE (nt 1 to 226) was used as a query in a database search which was based on a rigorous Smith-Waterman algorithm (SSearch, implemented in the Genetics Computer Group package with the default parameters) by using the nonredundant rodent database, as available in 2001. The raw SSearch output was used with Pileup (Genetics Computer Group package) to generate multiple sequence alignments, which were manually refined to represent only full-length matches. This process resulted in a library of 326 elements sharing at least 70% identity with RTE. Upon visual examination of the alignment, the elements were subdivided into four groups, named A (which contains the previously identified RTE) through D. Figure 2A shows the comparison of the consensus sequences of RTE in group A (83 sequences) and the RTE-related elements in groups B (27 sequences), C (72 sequences), and D (144 sequences). Figure 2B shows the alignment of each group's representative, namely, RTE (A) and the RTE-like elements identified under GenBank accession numbers AL663101 (B), AL671215 (C), and AC003061 (D). Compared to RTE, there are nucleotide insertions in the region separating loops I and II in groups C and D and within a loop in SL II of groups B, C, and D. Interestingly, the sequences of loops in SL I and SL III are conserved, whereas the loop in SL IV in groups B and D has a 7-nt insertion (5′-AAGCCTG-3′). Figure 2C shows the predicted structures of the RTE-related elements identified in groups B, C, and D in comparison to the RTE structure. These

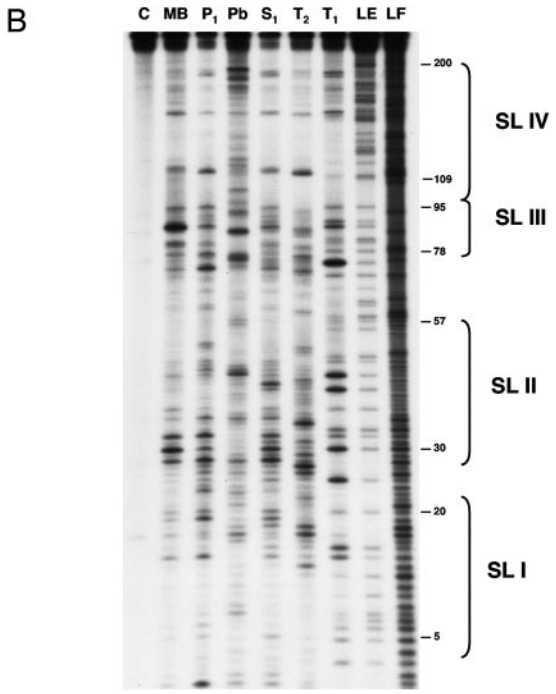


FIG. 1—*Continued*.

data reveal that all the RTE-related elements are highly similar to the previously identified RTE (22) and share the key structural features of SL I to SL IV. Of note, all elements share the long stem of SL IV. This structural conservation is maintained despite the differences in sequence (Fig. 2A and B) because the sequence changes are found to be mostly compensatory to maintain the predicted stem structures. This finding is reminiscent of the mouse CTE_{IAP} element, which has the same overall secondary structure as the type D retroviral CTEs (32) but shares only the sequences of the two internal loops while the sequences of the stem structures are different. There are also marked differences among the predicted structures of the RTE-related elements and the experimentally defined structure of RTE. The elements in groups A and C share a small loop in SL IV, while the elements in groups B and D have an insertion of 7 nt, resulting in a bigger loop. On the other hand, the elements of groups A and D have a big loop in SL III, while this loop is smaller in the elements of groups C and B. Most of the changes are concentrated in the SL I and SL II regions of the elements, affecting the predicted structures in these regions more severely. However, functional analysis (see below) showed that these regions are not essential for nuclear export of the RTE RNA, which may explain the tolerance of bigger variation in this part of the RTE. On the other hand, most of the changes in the remaining regions are of a compensatory nature and are predicted not to disturb the overall secondary structure.

Due to the subsequent availability of the complete mouse genome sequence, we performed new searches using SSearch with RTE as the query sequence, which identified more than 3,000 RTE-related elements, of which 105 elements are identical to the 226-nt RTE. The elements are distributed to all the chromosomes, and no clusters of RTEs were found. For comparison, the previously identified functional CTE-related element (CTE_{IAP}) (32) is found in 275 highly related copies in the mouse genome (of which 1 element has 100% identity to the primary sequence, 272 elements have 82 to 97%, and 2 elements have 66 to 67%). Therefore, these searches show a high abundance of RTE and RTE-related elements in the genome, suggesting an association of these elements with more active IAPs. Whereas some of the RTEs are found to be located within IAPs like the originally identified RTE (22), the vast majority of IAP-related sequences in rodents represents molecular fossils of once intact ancestral retroelements, and it is therefore likely that the majority of the identified RTEs and RTE-related elements may not be found associated with IAP sequences.

Minimal RTE. To identify the minimal functional RTE, we generated a series of 5′- and 3′-end deletion mutants from RTE (nt 1 to 226), such as M31 (nt 28 to 226), M2 (nt 57 to 217), M12 (nt 65 to 206), M13 (nt 104 to 199), and M14 (nt 116 to 189), as shown in Fig. 3A. The mutant RTEs were inserted into the intron of an adenovirus pre-mRNA and were tested for their ability to export intron lariats from microinjected *Xenopus* oocyte nuclei. As shown in Fig. 3B and previously (22), the intron lariat containing the intact RTE is exported to the cytoplasm. For a negative control, we tested the export of the empty adenovirus lariat, which remains in the nuclear compartment and shows only about 10 to 13% export. U1ΔSm RNA and U6Δss RNA were coinjected to serve as quality controls and demonstrated the functionality of the nuclear export machinery and the intactness of the nuclei, respectively. Mutant RTEs M2 and M12 mediated export of the intron lariat, while mutants M13 and M14 were inactive. Quantitation of the export efficiency showed that although M12 is active, its function is impaired (20 to 24% export) compared to that of RTE (45 to 54% export) or M2 (30 to 40% export).

M2 and M31 were also tested for their ability to promote expression of HIV-1 *gag* in mammalian cells. For a reporter plasmid, we used pNLgag, in which the p55 *gag* gene is located within an intron and Gag can be produced only from the unspliced RNA (12, 15), or pB37, expressing the p37 *gag* gene (consisting of p17 and p24) from a plasmid lacking splice sites (28). In the absence of an RNA export element, the *gag* mRNA is confined to the nucleus (in pB37 and pNLgag) and subjected to splicing (in pNLgag). Gag production was determined upon transfection of HeLa cells by an HIV-1 Gag antigen capture assay. The mutant RTEs M2 (Fig. 4) and M31 (data not shown) were severely impaired and unable to mediate *gag* expression of pNLgag, while the intact RTE led to an ~10-fold-higher activation. We also tested M2 in a simpler reporter plasmid, pB37, which lacks functional splice sites, and similar data were obtained (data not shown). These data indicated that the bottom portion of the RTE structure spanning SL I and SL II (Fig. 1A) is essential for function in mammalian cells. Since M2 was able to transport the intron lariat from the *Xenopus* oocyte nucleus but failed to mediate Gag production in the mammalian assay, we conclude that mutant M2 still retains the key interaction site(s) for the RNA export factor(s). As shown in Fig. 2, the region spanning SL I and SL II is most variable among the different groups in both sequence and structure. Our data indicate that this region may also play a role in the overall structure of the RTE. We cannot exclude the

A

```
A(RTE)   ccgtGgGgTGcgaGgCTAAGcACTGCAc...AgaGgaTagcttG.....ctgttGgCatcctgtGgaAGgcAyGTCtgaTTgCAtGAaGGTTcAGTGtcCtAGttcCC
B        skkaGrGtTGtaaGrCTAAGyACTGCAc...AgaGrtTagtctGgaagctgttwGaCartctctGrgAGgyAtGTCtgaTTgCAtGArGGTTgAGTGycCtAGkgtCC
C        gagaGaGtTGcacGgCTAAGcACTGCAgtagAagGgcTctgcgGcatawaatgaGcCtattctaGggAGacAtGTCatcTTtCAaGAaGGTTgAGTGtyCaAGtgtCC
D        gagaGaGtTGcamGgCTAAGcACTGCAatggAaaGgcTctgcgGcatata.tgaGcCtattctaGggAGacAtGTCatcTTtCAtGAaGGTTcAGTGtcCtAGttcCC
```

loop I loop II loop III

```
A(RTE)   TTCcCCCagGaaaAAAACgaCACGGGAgCtGgcCAaGacctCTCTGGGTgA.......TgaGCCTaaGgGatggTttTGTgtAgGGCcCCTatgct.TgCACACTGGGGA
B        TTCcCCCmrGaaaAAAACggCACGGGAgCaGgtCAgGgttrCTCTGGGTaAaaayctgTgrGCCTraGaGtcaaTccTGTacAtGGCcCCTawywy.TgCACACTGGGGA
C        TTCtCCCcaG.gaAAAACraCACGGGAcCaGacCAgGacccCTCTGGGTgA.......TgaGCCTggGaGgaggTtaTGTgtAcGGCtCCTattaccTgCACACTGGGGA
D        TTCtCCCcaG.gcAAAACgaCACGGGAgCaGgtCAgGgttgCTCTGGGTaAaagcctgTraGCCTaaGaGctaaTccTGTacAtGGCtCCT.ttaccTaCACACTGGGGA
```

loop IV

```
A(RTE)   TcaGACCTCTAcCTtcACcCatgagg
B        TcaGACCTCTAyCTyyACcCayggag
C        TttGACCTCTAtCTccACtCtcatta
D        TttGACCTCTAtCTccACtCtcatta
```

B

```
     RTE CCGTGGGGTGCGAGGCTAAGCACTGCAC...AGAGGATAGCTTG.....CTGTTGGCATCCTGTGGAAGGCACGTCTGATTGCATGAAGGTTCAGTGTCCTAGTTCCCT
al663101 TGTAGGGTTGTACGGCTAAGCACTGCAC...AGAGGTTAGTCTGGAAGCTGTTTGACAATCTCTGGGAGGTATGTCTGATTACATGAGGGTTGAGTGTCCTAGTGTCCT
al671215 GAGAGAGTTGCACGGCTAAGCACTGCAGTAGAAGGGCTCTGGGCCATAT.ATGAGCCTATTCTAGGGAGACATGTCATCTTTCAAGAAGGTTGAGTGTCCAAGTGTCCT
ac003061 GAGAGAGTTGCAAGGCTAAGCACTGCAATGGAAAGGCTCTGGGCCATAT.ATGAGCCTATTCTAGGGAGACATGTCATCTTTCATGAAGGTTCAGTGTCCTAGTTCCCT
```

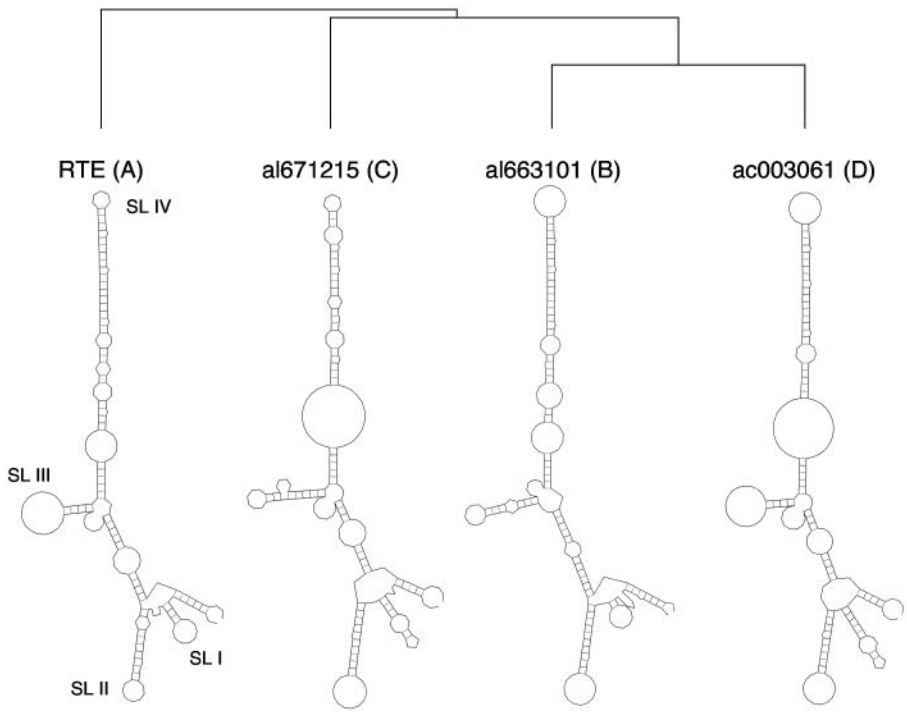
loop I loop II loop III

```
     RTE TCCCCCAGG.AAAAACGACACGGGAGCTGGCCAAGACCTCTCTGGGTGA.......TGAGCCTAAGGGATGGTTTTGTGTAGGGCCCCTATGCTTGCACACTGGGGATCA
al663101 TCCCCCAGGAAAAAACGGCACGGGAGCAGGTCAGGGTTGCTCTGGGTAAAAACCTGTGAGCCTAAGAGTCAATCCTGTACATGGCCCCTATTTCTGCACACTGGGGATCA
al671215 TCTCCCCAGGTAAAACGACACGGGACCAGACCAGGACCCCTCTGGGTGA.......TAAACCTGGAAGGAGGTTATGTGTACGGCTCCTTTACCTGCACACTGGGGATTT
ac003061 TC.CCCCAGGCAAAACGACACGGGAGCAGGTCAGGGTTGCTCTGGGTAAAAGCCTGTAAGCCTAAGAGCTAATCCTGTACATGGCTCCTTTACCTACACACTGGGGATTT
```

loop IV

```
     RTE GACCTCTACCTTCACCCATGAGG
al663101 GACCTCTATCTTCACCCATGGAG
al671215 GACCTCTATCTCCACTCTCATTA
ac003061 GACCTCTATCTCCACTCTCATTA
```

C



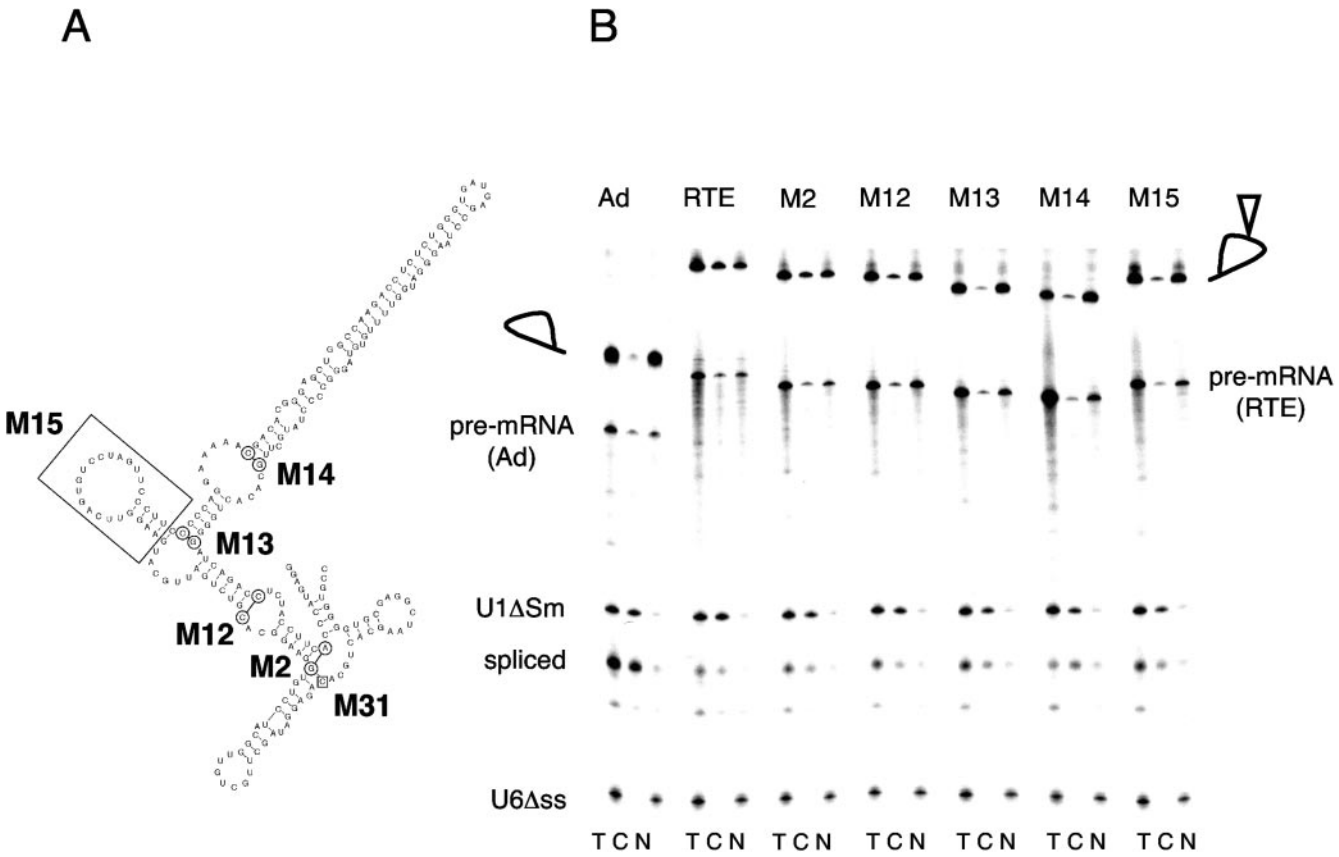


FIG. 3. Analysis of RTE deletion mutants in oocytes. (A) Schematic representation of RTE deletion mutants. Circled nucleotides indicate the beginnings and ends of the corresponding RTE mutants. The boxed nucleotide indicates the start of the 5′-end deletion in RTE M31. M15 lacks SL III, as indicated by the boxed area. (B) RNA export activity of RTE mutants in microinjected *X. laevis* oocyte nuclei. Radiolabeled adenovirus pre-mRNAs containing the indicated mutant RTE were injected into *Xenopus* oocyte nuclei. Total (T), cytoplasmic (C), and nuclear (N) fractions were prepared after a 3-h incubation. The injected RNA mixture also contained U1ΔSm RNA, which is exported into the cytoplasm (an indicator of RNA export), and U6Δss RNA, which remains in the nucleus (an indicator of nuclear integrity). The positions of the pre-mRNAs and intron lariats are indicated. The presence of the RTE is indicated by an open triangle in the lariat loop.

possibility that this region also contributes to the posttransport steps by promoting translation of the mRNA, which is the readout of the transfection assay. Alternatively, the different nature of the test RNAs containing the RTE mutant M2 or the test systems may explain the discrepancy between the two assay systems. For these reasons, the complete 226-nt RTE was chosen for further mutational analysis.

The loop sequence in SL IV is not essential for function. From the sequence comparisons shown in Fig. 2, it became apparent that another difference between RTE and its related elements is the sequence of the end loop in SL IV. To address the role of this loop, we generated mutant M20, in which the original RTE loop (5′-UGAUGAG-3′) was replaced with the 5′-UAAAAGCCUGUGAG-3′ loop as present in groups B and D (Fig. 5A). Figure 5B shows that the intron lariats containing either the wild-type RTE or M20 are efficiently exported to the cytoplasm of injected *Xenopus* oocytes. The presence of M20 in pNLgag was further shown to increase Gag production in a manner similar to that of the RTE (Fig. 4). Together, these experiments demonstrated that the sequence of the loop in SL IV of RTE is not essential for function.

The structure of the extended stem but not the sequence is essential. We generated a series of mutants designed to change the nucleotide sequence but not the predicted structure, focusing on the role of the stem of SL IV. M7 had an internal deletion of the distal portion of the stem (Fig. 6A) and was unable to mediate intron lariat export (Fig. 6B). This finding suggests that this portion of the stem either has a structural

FIG. 2. Family of RTE-related elements in the mouse genome. (A) Alignment of the consensus sequences of RTE (group A) and of the RTE-related groups B, C, and D. The 226-nt RTE was used to identify related elements in the mouse genome. The shaded areas indicate the loops defined for RTE as shown in Fig. 1A. The letters k, m, r, s, w, and y are the degenerate DNA alphabet. (B) Alignment of RTE and a representative member of each group of RTE-related elements, found under GenBank accession numbers AL663101 (group B), AL671215 (group C), and AC003061 (group D). The shaded areas indicate the loops defined for RTE as shown in Fig. 1A. (C) Phylogenetic tree and comparison of the RTE structure (from Fig. 1A) and the predicted secondary structures for the RTE-related elements shown in panel B.

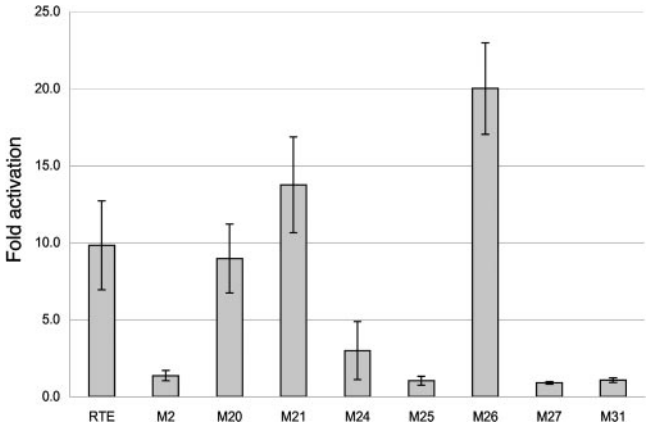


FIG. 4. Analysis of RTE mutants by *gag* expression in human cells. The HIV-1 *gag* expression vector pNLgag, containing either no insert, the RTE, or the RTE mutant indicated, was transfected into HeLa-derived HLtat cells. Two days later, the cells were harvested and the level of Gag production was determined using an antigen capture assay. The increase in Gag production in the absence (pNLgag) or presence of RTE and RTE mutants was determined. The mean *n*-fold induction ± standard deviations of the results for the tested RTE and mutant RTEs are shown. Two independent clones for each construct containing RTEs were tested in duplicate.

role or contains a crucial binding site for the export factor. To distinguish between these possibilities, we generated a series of mutants designed to maintain the secondary structure but with the sequence changed as shown in Fig. 6C for mutants M21, M22, and M24. Interestingly, M21, M22, and to a lesser extent M24 (about 50% activity) promote export of the intron lariat from *X. laevis* oocyte nuclei (Fig. 6D). We further examined the ability of M21 and M24 to mediate export and expression of the unspliced *gag* mRNA when they were inserted into the pNLgag vector and transfected into mammalian cells. Figure 4 shows that M21 is as active as the intact RTE, whereas the function of M24 is also impaired in this assay. Together, these data indicate that the distal portion of the extended stem structure plays primarily a structural role.

The sequence and structure of the middle region are essential for RTE function. The next series of mutations focused on the middle region of RTE, including SL III, the proximal portion of the stem of SL IV, and the internal stem structure. Removal of the complete SL III (mutant M15) (Fig. 3A) or part of the loop of SL III (mutant M27) (Fig. 7A) resulted in inactivation of RTE (Fig. 3B and 7B). Sequence changes in the left-hand portion of the internal loop of SL IV (nt 110 to 115) from GAAAAA to CCCCCC (M25) (Fig. 7A) also inactivated RTE (Fig. 7B). Interestingly, changes in the 3′ portion of the internal loop (nt 190 to 193) from CACA to GCGG (M26) had no effect. Testing these mutants upon transfection in the mammalian system (Fig. 4) revealed that the function of M26 was enhanced (twofold) compared to that of wild-type RTE, while M25 was inactive as expected.

We further found that extending the bottom portion (C-G stem) of SL IV by another 10 bp (M30) or by introducing changes from $G_{74}C_{75}A_{76}$ to UUU (M28) also inactivated RTE (Fig. 7A and B). We cannot exclude the possibility that some of these mutations, as in M30 and M28, alter the flexibility of the structure and/or separate potential factor interaction site(s). On the other hand, changes in the adjacent internal stem (M29), designed to preserve the structure but changing the primary sequence, had no effect on the function. Together, the bottom portion of the stem of SL IV, especially the left-

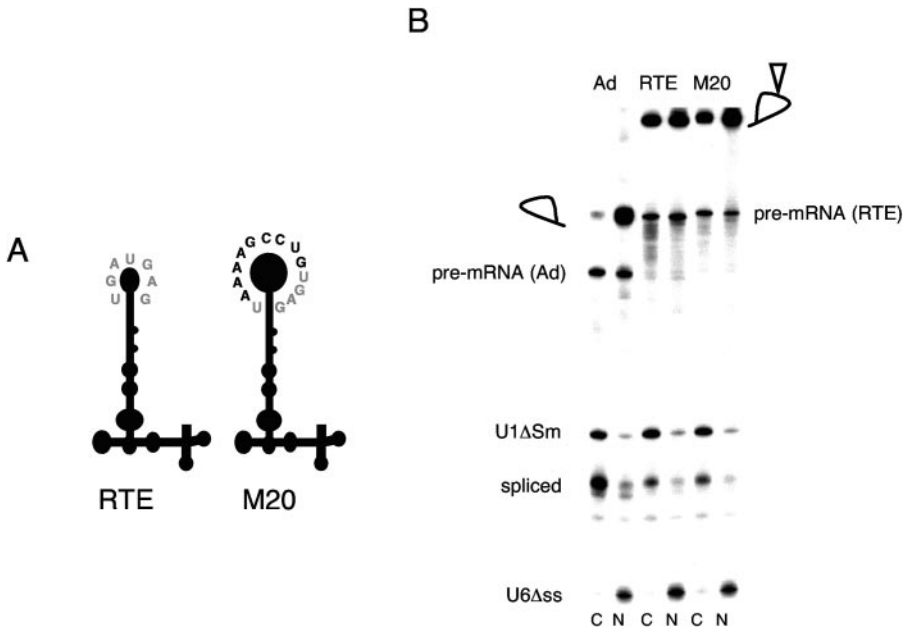


FIG. 5. The two variants of the SL IV loop are functional. (A) Cartoon of RTE and M20. Mutant M20 was engineered to have a larger SL IV as present in RTE-related elements in groups B and D. The nucleotide sequence of the RTE loop is shown in grey; the extra nucleotides in the longer variant loop are shown in black. (B) M20 promotes RNA export from the nucleus of *X. laevis* oocyte nuclei. Adenovirus pre-mRNAs containing either no insert (Ad), RTE, or M20 were generated. Upon injection of *X. laevis* oocyte nuclei with the radiolabeled precursor RNA, cytoplasmic (C) and nuclear (N) fractions were analyzed as described in the Fig. 3 legend.

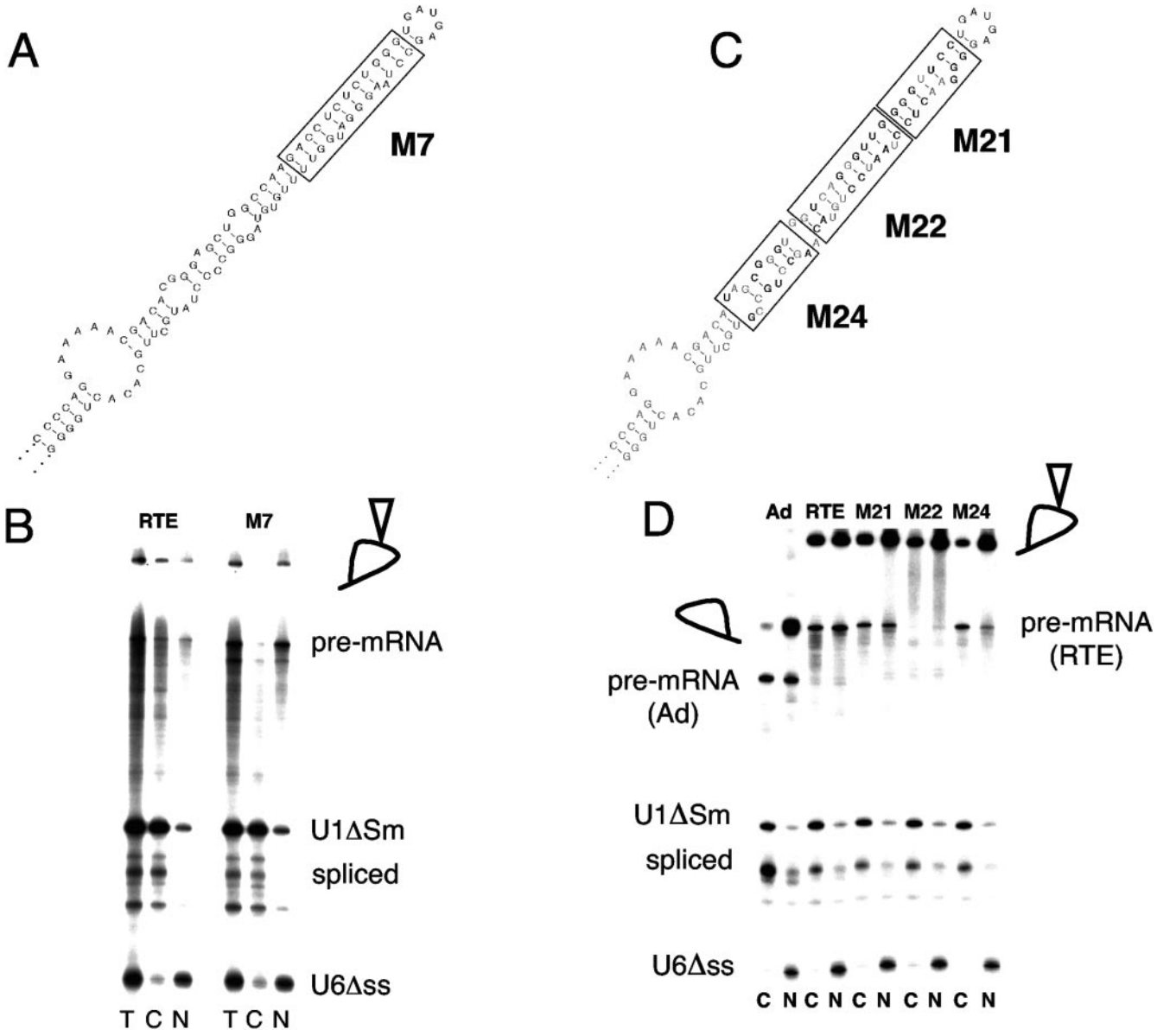


FIG. 6. Analysis of the stem structure of SL IV. (A) Mutant M7 contains a deletion of the distal stem (boxed). (B) RNA export of M7 from microinjected *X. laevis* oocyte nuclei. Adenovirus pre-mRNAs containing RTE or M7 were injected into nuclei, and the total (T), nuclear (N), and cytoplasmic (C) fractions were analyzed as described in the Fig. 3 legend. (C) Compensatory changes in SL IV were designed to change the primary sequence but to preserve the base-paired structure. Changed regions (boxed) for M21, M22, and M24 are shown. Mutated nucleotides are shown in bold. (D) RNA export of M21, M22, and M24 from microinjected *X. laevis* oocyte nuclei. Adenovirus pre-mRNAs lacking (Ad) or containing RTE or RTE mutants were injected into nuclei, and nuclear and cytoplasmic fractions were analyzed as described in the Fig. 3 legend.

hand portion of the internal loop, and SL III together with the stem of SL IV play an essential role in RTE function (Fig. 1A). In some of the examined areas, the primary sequence in unpaired regions appears to be important, whereas the sequence of double-stranded regions can be changed as long as the secondary structure is preserved.

DISCUSSION

To understand the mechanism of RTE function, we performed a detailed structure-function analysis. Here we demonstrate that RTE folds into a complex, extended RNA stem-loop structure, which is distinct from other elements such as RRE of HIV-1, RXRE of HTLV-1, CTE of the type D retroviruses, and RcRE/K of human endogenous retroviruses. Mutational analysis revealed that most parts of RTE are sensitive to changes resulting in inactivation of the element.

Comparison of the RTE secondary structure to the computer-predicted structures of the RTE-related elements showed major differences in SL I and SL II and the loop of SL IV, whereas the stems of SL III and SL IV are highly conserved. We noted that several of the RTE-related elements have compensatory nucleotide changes preserving the structure of dou-

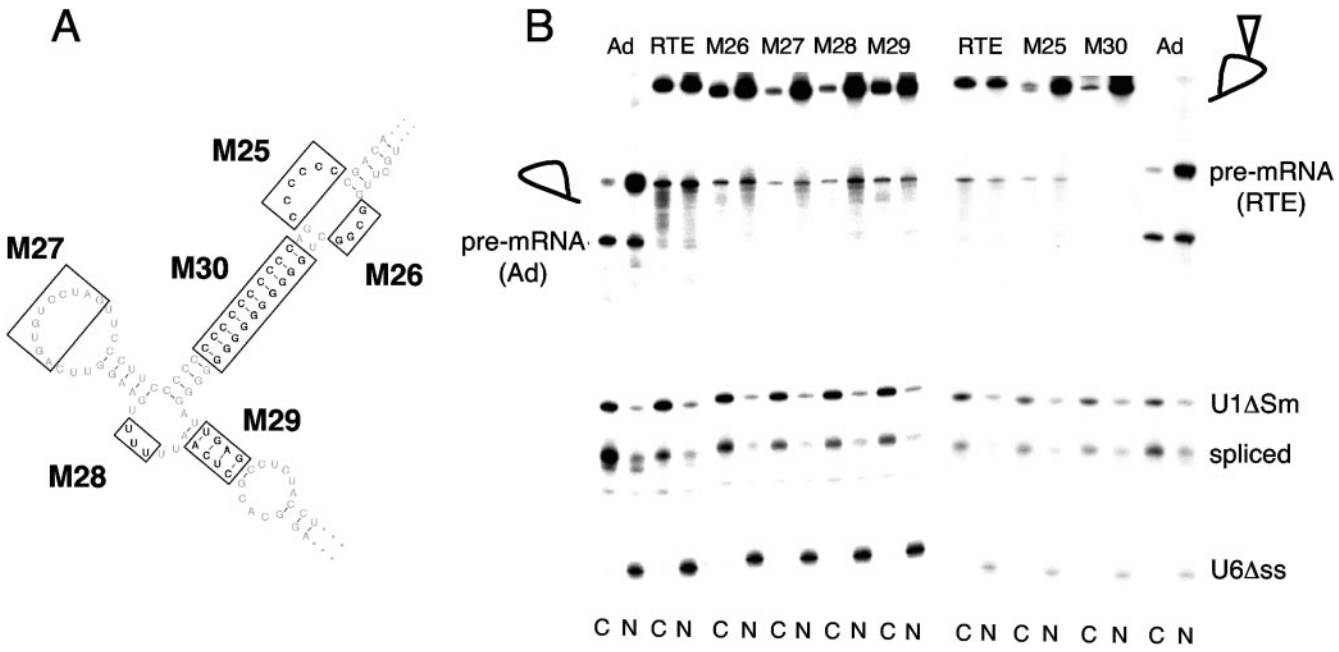


FIG. 7. Mutants containing changes in the middle region of RTE. (A) Boxed areas indicate the mutated regions, and bold letters indicate the nucleotide changes in the indicated RTE mutants. M30 contains an insertion of 10 G-C base pairs. (B) RNA export of the indicated RTE mutants from *X. laevis* oocyte nuclei. Adenovirus pre-mRNAs lacking (Ad) or containing RTE or RTE mutants were injected into nuclei, and the nuclear (N) and cytoplasmic (C) fractions were analyzed as described in the Fig. 3 legend.

ble-stranded stem regions. Although we did not individually test the RTE-related elements experimentally, we used these comparisons to design several of our RTE mutants. We addressed some of these changes experimentally and found that the regions of SL I and SL II are dispensable for export from *X. laevis* oocyte nuclei. In addition, both variants of the loops of SL IV are functional (wild-type RTE and mutant M20). Interestingly, all the RTEs share the long stem in SL IV, and experimental data showed that its role is of functional importance. Deletions in this region inactivate the element, but changes in its primary structure while preserving the base pairing are tolerated (compare M7 to M21, M22, and M24). This finding indicates that the structure of SL IV is important for function. Both the stem and the loop of SL III show a high degree of secondary structure conservation. SL III is essential for RTE function (M15 and M27). Notably, some mutants disrupting the potential tertiary interactions of SL III with SL IV (extending the lower stem of SL IV [M30] and changing the sequence between SL III and SL IV [M28]) were inactive. Together, these data indicate not only that SL III and SL IV are necessary but that they may be involved in higher-order interactions which are required for function. Therefore, the putative cellular binding factor(s) may interact with more than a single region, or, alternatively, more than one factor interacting with distinct regions is necessary to mediate RTE function. The interaction(s) appears to require proper spatial arrangement within RTE. Interestingly, we found one mutant (M26) able to further improve RNA export function. Notably, the nucleotide changes introduced into M26 involve four unpaired nucleotides which are conserved among all the elements.

The studies of retroviruses and retroelements have revealed the existence of potent RNA export elements which serve as important tools to understand retrovirus expression as well as mechanisms mediating cellular mRNA transport. These studies led further to the discoveries of the major pathways mediating mRNA and protein trafficking through the nucleus. Two distinct export pathways have been revealed: CRM1, utilized by the retroviral Rev, Rex, cORF, and an array of nuclear export signal-containing cellular proteins; and NXF1, utilized to export CTE-containing mRNA as well as cellular mRNAs. Similar to CTE, RTE utilizes a conserved cellular mRNA export pathway and its function is independent of CRM1 (22). Although previous work showed that the NXF1 protein itself is not a high-affinity binder to RTE (22), competition experiments have indicated NXF1 involvement (S. Smulevitch and B. K. Felber, unpublished data). One proposed model is that RTE export involves a factor(s) leading to NXF1 recruitment and thus linking RTE to a major export pathway.

Although the respective export factors interact with few nucleotides within RRE, RXRE, CTE, and RcRE/K, these elements contain several regions that are essential for their secondary structure, and the protein interaction site is part of a complex structure. Our findings for RTE are reminiscent of observations of the other studied retroviral *cis*-acting RNA elements, implying the existence of a complex RNA structure that is essential for function, which places the protein-binding motif into a context essential for the specific interaction. Since retroviruses and retroelements have evolved to take advantage of cellular pathways, the use of these systems has provided important tools for the identification of the cellular factors that mediate key steps in controlling eukaryotic gene expression. Recently, retrotransposition-competent IAPs have been identified in the mouse genome (7). Our analysis indicates that

these retroelements contain intact RTEs belonging to the D group (our unpublished data). It will be of interest to study the contribution of these RTEs to IAP expression. The identification of active IAPs will allow us to further study the role of these elements in retrotransposition.

ACKNOWLEDGMENTS

We are grateful to P. Sood, A. Witten, A. Waddelow, A. Arwindekar, and C. Tabernero for their assistance; E. Izaurralde for advice; and T. Jones for editorial assistance.

REFERENCES

1. **Ahmed, Y. F., S. M. Hanly, M. H. Malim, B. R. Cullen, and W. C. Greene.** 1990. Structure-function analyses of the HTLV-I Rex and HIV-1 Rev RNA response elements: insights into the mechanism of Rex and Rev function. Genes Dev. **4:**1014–1022.
2. **Bogerd, H. P., H. L. Wiegand, J. Yang, and B. R. Cullen.** 2000. Mutational definition of functional domains within the Rev homolog encoded by human endogenous retrovirus K. J. Virol. **74:**9353–9361.
3. **Braun, I., E. Rohrbach, C. Schmitt, and E. Izaurralde.** 1999. TAP binds to the constitutive transport element through a novel RNA-binding motif that is sufficient to promote CTE-dependent RNA export from the nucleus. EMBO J. **18:**1953–1965.
4. **Bray, M., S. Prasad, J. W. Dubay, E. Hunter, K.-T. Jeang, D. Rekosh, and M.-L. Hammarskjold.** 1994. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. Proc. Natl. Acad. Sci. USA **91:**1256–1260.
5. **Ciesiolka, J., D. Michalowski, J. Wrzesinski, J. Krajewski, and W. J. Krzyzosiak.** 1998. Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. J. Mol. Biol. **275:**211–220.
6. **Conti, E., and E. Izaurralde.** 2001. Nucleocytoplasmic transport enters the atomic age. Curr. Opin. Cell Biol. **13:**310–319.
7. **Dewannieux, M., A. Dupressoir, F. Harper, G. Pierron, and T. Heidmann.** 2004. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat. Genet. **36:**534–539.
8. **Erkmann, J. A., and U. Kutay.** 2004. Nuclear export of mRNA: from the site of transcription to the cytoplasm. Exp. Cell Res. **296:**12–20.
9. **Ernst, R. K., M. Bray, D. Rekosh, and M.-L. Hammarskjold.** 1997. Secondary structure and mutational analysis of the Mason-Pfizer monkey virus RNA constitutive transport element. RNA **3:**210–222.
10. **Ernst, R. K., M. Bray, D. Rekosh, and M.-L. Hammarskjold.** 1997. A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. Mol. Cell. Biol. **17:**135–144.
11. **Felber, B. K., C. M. Drysdale, and G. N. Pavlakis.** 1990. Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. J. Virol. **64:**3734–3741.
12. **Felber, B. K., M. Hadzopoulou-Cladaras, C. Cladaras, T. Copeland, and G. N. Pavlakis.** 1989. rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. USA **86:**1495–1499.
13. **Gorlich, D., and U. Kutay.** 1999. Transport between the cell nucleus and the cytoplasm. Annu. Rev. Cell Dev. Biol. **15:**607–660.
14. **Grüter, P., C. Tabernero, C. von Kobbe, C. Schmitt, C. Saavedra, A. Bachi, M. Wilm, B. K. Felber, and E. Izaurralde.** 1998. TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Mol. Cell **1:**649–659.
15. **Hadzopoulou-Cladaras, M., B. K. Felber, C. Cladaras, A. Athanassopoulos, A. Tse, and G. N. Pavlakis.** 1989. The *rev* (*trs*/*art*) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a *cis*-acting sequence in the *env* region. J. Virol. **63:**1265–1274.
16. **Herold, A., L. Teixeira, and E. Izaurralde.** 2003. Genome-wide analysis of nuclear mRNA export pathways in Drosophila. EMBO J. **22:**2472–2483.
17. **Izaurralde, E., and S. Adam.** 1998. Transport of macromolecules between the nucleus and the cytoplasm. RNA **4:**351–364.
18. **Jarmolowski, A., W. C. Boelens, E. Izaurralde, and I. W. Mattaj.** 1994. Nuclear export of different classes of RNA is mediated by specific factors. J. Cell Biol. **124:**627–635.
19. **Löwer, R., R. R. Tönjes, C. Korbmacher, R. Kurth, and J. Löwer.** 1995. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. J. Virol. **69:**141–149.
20. **Magin, C., R. Löwer, and J. Löwer.** 1999. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. J. Virol. **73:**9496–9507.
21. **Malim, M. H., J. Hauber, S.-Y. Le, J. V. Maizel, and B. R. Cullen.** 1989. The HIV-1 *rev trans*-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature **338:**254–257.
22. **Nappi, F., R. Schneider, A. Zolotukhin, S. Smulevitch, D. Michalowski, J. Bear, B. Felber, and G. Pavlakis.** 2001. Identification of a novel posttranscriptional regulatory element using a *rev*- and RRE-mutated human immunodeficiency virus type 1 DNA proviral clone as a molecular trap. J. Virol. **75:**4558–4569.
23. **Nasioulas, G., S. H. Hughes, B. K. Felber, and J. M. Whitcomb.** 1995. Production of avian leukosis virus particles in mammalian cells can be mediated by the interaction of the human immunodeficiency virus protein Rev and the Rev-responsive element. Proc. Natl. Acad. Sci. USA **92:**11940–11944.
24. **Ogert, R. A., and K. L. Beemon.** 1998. Mutational analysis of the Rous sarcoma virus DR posttranscriptional control element. J. Virol. **72:**3407–3411.
25. **Ogert, R. A., L. H. Lee, and K. L. Beemon.** 1996. Avian retroviral RNA element promotes unspliced RNA accumulation in the cytoplasm. J. Virol. **70:**3834–3843.
26. **Paca, R. E., R. A. Ogert, C. S. Hibbert, E. Izaurralde, and K. L. Beemon.** 2000. Rous sarcoma virus DR posttranscriptional elements use a novel RNA export pathway. J. Virol. **74:**9507–9514.
27. **Saavedra, C., B. Felber, and E. Izaurralde.** 1997. The simian retrovirus-1 constitutive transport element, unlike the HIV-1 RRE, uses factors required for cellular mRNA export. Curr. Biol. **7:**619–628.
28. **Schneider, R., M. Campbell, G. Nasioulas, B. K. Felber, and G. N. Pavlakis.** 1997. Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation. J. Virol. **71:**4892–4903.
29. **Segref, A., K. Sharma, V. Doye, A. Hellwig, J. Huber, R. Lührmann, and E. Hurt.** 1997. Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)$^+$ RNA and nuclear pores. EMBO J. **16:**3256–3271.
30. **Solomin, L., B. K. Felber, and G. N. Pavlakis.** 1990. Different sites of interaction for Rev, Tev, and Rex proteins within the Rev responsive element of human immunodeficiency virus type 1. J. Virol. **64:**6010–6017.
31. **Stauber, R. H., K. Horie, P. Carney, E. A. Hudson, N. I. Tarasova, G. A. Gaitanaris, and G. N. Pavlakis.** 1998. Development and applications of enhanced green fluorescent protein mutants. BioTechniques **24:**462–466
32. **Tabernero, C., A. S. Zolotukhin, J. Bear, R. Schneider, G. Karsenty, and B. K. Felber.** 1997. Identification of an RNA sequence within an intracisternal-A particle element able to replace Rev-mediated posttranscriptional regulation of human immunodeficiency virus type 1. J. Virol. **71:**95–101.
33. **Tabernero, C., A. S. Zolotukhin, A. Valentin, G. N. Pavlakis, and B. K. Felber.** 1996. The posttranscriptional control element of the simian retrovirus type 1 forms an extensive RNA secondary structure necessary for its function. J. Virol. **70:**5998–6011.
34. **Tan, W., A. S. Zolotukhin, J. Bear, D. J. Patenaude, and B. K. Felber.** 2000. The mRNA export in *Caenorhabditis elegans* is mediated by Ce-NXF-1, an ortholog of human TAP and *Saccharomyces cerevisiae* Mex67p. RNA **6:**1762–1772.
35. **Yang, J., H. Bogerd, S. Y. Le, and B. R. Cullen.** 2000. The human endogenous retrovirus K Rev response element coincides with a predicted RNA folding region. RNA **6:**1551–1564.
36. **Zolotukhin, A. S., D. Michalowski, S. Smulevitch, and B. K. Felber.** 2001. Retroviral constitutive transport element evolved from cellular TAP(NXF1)-binding sequences. J. Virol. **75:**5567–5575.
37. **Zolotukhin, A. S., W. Tan, J. Bear, S. Smulevitch, and B. K. Felber.** 2002. U2AF participates in the binding of TAP (NXF1) to mRNA. J. Biol. Chem. **277:**3935–3942.
38. **Zolotukhin, A. S., A. Valentin, G. N. Pavlakis, and B. K. Felber.** 1994. Continuous propagation of RRE(−) and Rev(−)RRE(−) human immunodeficiency virus type 1 molecular clones containing a *cis*-acting element of simian retrovirus type 1 in human peripheral blood lymphocytes. J. Virol. **68:**7944–7952.