

# Essential Motifs in the 3' Untranslated Region Required for Retrotransposition and the Precise Start of Reverse Transcription in Non-Long-Terminal-Repeat Retrotransposon SART1

Mizuko Osanai, Hidekazu Takahashi,† Kenji K. Kojima, Mitsuhiro Hamada, and Haruhiko Fujiwara\*

*Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, Japan*

Received 14 October 2003/Returned for modification 14 January 2004/Accepted 2 June 2004

**Non-long-terminal-repeat (non-LTR) retrotransposons amplify their copies by reverse transcribing mRNA from the 3' end, but the initial processes of reverse transcription are still unclear. We have shown that a telomere-specific non-LTR retrotransposon of the silkworm, SART1, requires the 3' untranslated region (3' UTR) for retrotransposition. With an *in vivo* retrotransposition assay, we identified several novel motifs within the 3' UTR involved in precise and efficient reverse transcription. Of 461 nucleotides (nt) of the 3' UTR, the central region, from nt 163 to nt 295, was essential for SART1 retrotransposition. Of five putative stem-loops formed in RNA for the SART1 3' UTR, the second stem-loop (nt 159 to 221) is included in this region. Loss of the 3' region (nt 296 to 461) in the 3' UTR and the poly(A) tract resulted in decreased and inaccurate reverse transcription, which starts mostly from several telomeric repeat-like GGUU sequences just downstream of the second stem-loop. These results suggest that short telomeric repeat-like sequences in the 3' UTR anneal to the bottom strand of (TTAGG)<sub>n</sub> repeats. We also demonstrated that the mRNA for green fluorescent protein (GFP) could be retrotransposed into telomeric repeats when the GFP coding region is fused with the SART1 3' UTR and SART1 open reading frame proteins are supplied *in trans*.**

Non-long-terminal-repeat (non-LTR) retrotransposons are endogenous mobile genetic elements that are widespread among the genome of eukaryotes. Non-LTR retrotransposons multiply their copies through reverse transcription of RNA intermediates with a self-encoding reverse transcriptase (RT). Non-LTR retrotransposons, some of which are called long interspersed nuclear elements (LINEs) in vertebrates, contribute to genome structure and evolution through their replication process (5, 8, 10, 21). In humans, LINEs have accumulated up to 21% of the genome, and they are the major source of insertional mutagenesis. LINEs shape the mammalian genome through exon shuffling, mobilization of short interspersed nuclear elements, and processed pseudogene formation (4, 6, 10, 18, 20). However, we have little knowledge of the molecular basis underlying these genomic events because the retrotransposition mechanisms of non-LTR retrotransposons are insufficiently understood, compared with those of other retroelements, such as LTR retrotransposons and retroviruses.

Non-LTR retrotransposons encode an RT domain and an endonuclease (EN) domain. Pioneering studies of the *Bombyx* non-LTR element R2 showed that the EN domain nicks the bottom strand of target DNA and that the RT domain uses the 3'-hydroxyl end of the nicked DNA as a primer for reverse transcribing non-LTR element RNA (12). This reverse tran-

scription initiation process is termed target-primed reverse transcription (TPRT) and is inherent to all non-LTR retrotransposons. During TPRT, reverse transcription is initiated at the 3' end of non-LTR elements. Most genomic copies of non-LTR retrotransposons are 5' truncated, presumably because of the arrest of reverse transcription. However, all integrated copies have the precise 3'-terminal region and 3' truncation is rarely observed (5, 8). These observations reflect the processes of reverse transcription peculiar to non-LTR retrotransposons and show the necessity of recognizing some structure near the 3' ends for their initial steps. The requirement of the 3' UTR for retrotransposition is reported in R2 (13), and *UnaL2* elements in eels (9). However, there are elements such as human L1s that apparently have no strict sequence requirements in the 3'-end sequences except for the polyadenylation tail (4, 6, 19). Therefore, we do not have the general picture of the initial processes of reverse transcription, especially of how the RT recognizes the 3' UTR RNA of non-LTR elements.

A telomere-specific non-LTR retrotransposon, SART1, that inserts itself between TT and AGG of the silkworm (TTAGG)<sub>n</sub> telomeric repeats (23, 25) is a good model with which to study the TPRT mechanisms. We have developed an *in vivo* assay system in which retrotransposition is easily detected by PCR (24). With this system, we are trying to enumerate all of the functional regions in open reading frame 1 (ORF1) and ORF2 required for SART1 retrotransposition, which should enable us to clarify the TPRT mechanisms from various aspects (17). In a previous report, we showed that SART1 loses its retrotransposition ability when 461 nucleotides (nt) containing the 3' UTR and a poly(A) tract are

\* Corresponding author. Mailing address: Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Bioscience Building 501, Kashiwa, Chiba 277-8562, Japan. Phone: 81-4-7136-3659. Fax: 81-4-7136-3660. E-mail: haruh@k.u-tokyo.ac.jp.

† Present address: Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205.

TABLE 1. Primers used in this study

Primer	Sequence (5'→3')	Generated derivative(s) or use
SART1-S3014-NcoI	AAAAAACCATGGGCAGCAGCCCTTATCATATACTAC	ORF2+3'UTR
SART1-S6222-NotI	AATAATAATTGCGGCCGCGACCGTCGGGCG	SART1 WT (+NotI)-pAcGHLTB
SART1-A6600-BglII	AAAAAAAGATCTGGAAGAAACAGGAAGAAGTTCG	1-379, 73-379
SART1-S6501-NotI	AATAATAATAGCGGCCGCTGAACTCAGCCCAGC	279-461/(A) <sub>20</sub>
SART1-A6682-BglII	AAAAAAAGATCTGGTATCGATGGGGAATCCC	ΔpolyA
SART1-S6294-NotI	AATAATAATTGCGGCCGCGGGCGCTGTGGCTC	73-379, 73-379/(A) <sub>20</sub> , 73-295
SART1-A6516-BglII	AAAAAAAGATCTGGGCTGAGTTCAGCTC	73-295
SART1-A6600+20A-BglII	AAAAAAAGATCT(T) <sub>18</sub> TGGAAGAAACAGGAAGAAGTTCG	1-379/(A) <sub>20</sub> , 73-379/(A) <sub>20</sub>
SART1-A6704-BglII/BamHI	TTTTTTGGATCCAGATCT(T) <sub>19</sub> GGTATCGATGGGGAATC	279-461/(A) <sub>20</sub>
SART1-S6517	CGCGCCTTTTTCAAGGCGTAGTCTCC	Δ163-297, Δ228-297
SART1-A6293	TATACCCTCACCACCACCCTGGACTATCG	Δ73-162, Δ73-227
SART1-S6384	TGTGGGGGGCCTGCGGGG	Δ73-162, Δ73-227
SART1-S6449	GAGCTCGTTGGTTTTGATCGGTAGTCGTTAAG	Δ73-227, Δ163-227
SART1-A6383	CCGGCTCCAGCCTGACGAAC	Δ163-227, Δ163-297
SART1-A6448	CTATCTTCCGGCATAGGGGGAACCTACGATAC	Δ228-297
pAcGHLTB-S2183	CCTATAAATACGGATCTGTATTTCATGTCCC	GST-His <sub>6</sub> probe
pAcGHLTB-A2810	GGCCATGCTATATACTGTGCTGATTTCAAG	GST-His <sub>6</sub> probe
pAcGHLTB-S3032	CGACTCTGCTGAAGAGGAGGAAATTC	Polyhedrin 3' region probe
pAcGHLTB-A3430	CAAGATTGGCAAGTTTTGTGGCGTTGAG	Polyhedrin 3' region probe
hsp_pEGFP1-A2330T-S	GCGGCCGCGACTCTAGATCTTAATCAGCCATAC	hsp-EGFP1-BglII
hsp_pEGFP1-A2330T-A	GTATGGCTGATTAAGATCTAGATCGCGGCCG	hsp-EGFP1-BglII
pEGFP1-S96-EcoRI	AAAAAAGAATTCATGGTGAGCAAGGGCGA	EGFP1/S1-3'UTR-pVL1393
pEGFP1-A901	GGGGGAGGTGTGGG	EGFP1/S1-3'UTR-pVL1393
pEGFP1-S688	GACAACCACTACCTGAGCACC	3' junction PCR
pEGFP1-A576	GTTCTTCTGCTTGTTCGGCCATGATATAG	5' junction PCR

deleted from the element (24). The retrotransposition activity of the 3' UTR-deficient SART1 mutant was not rescued in *trans* by mutants with amino acid substitutions in Zn fingers and EN and RT domains, which lost their activities. However, the activities of the latter mutants were *trans* complemented by the 3' UTR-deficient mutant. These observations show that the 3' UTR of SART1 RNA, but not other RNA regions, is essential for initiation of TPRT.

In this study, to understand how and which region of the 3' UTR functions in SART1 retrotransposition, we have analyzed the effects of deletions in the 3' UTR on transposition of SART1. The retrotransposition assay showed that 132 nt of the central region of the 3' UTR, which features a firm stem-loop structure, are essential for retrotransposition. The remaining part of the 3' UTR and the poly(A) tail were required for precise initiation of reverse transcription and efficient retrotransposition. We also show that possessing the 3' UTR sequences is sufficient for RNA to go through retrotransposition by SART1 proteins.

#### MATERIALS AND METHODS

**Northern hybridization.** Approximately  $3 \times 10^5$  *Spodoptera frugiperda* 9 (Sf9) cells in a 12-well plate were infected with SART1-containing *Autographa californica* nuclear polyhedrosis virus (AcNPV) at a multiplicity of infection of 10 PFU per cell. Total RNA was isolated from Sf9 cells with TRIZOL (GIBCO-BRL) at 48 h postinfection. Aliquots of 3 μg of RNA per lane were electrophoresed at 5 V/cm on 18% formaldehyde–20 mM MOPS (3-morpholinopropanesulfonic acid, pH 7.0)–5 mM sodium acetate–1 mM EDTA–0.9% agarose gels and blotted onto nylon membranes (Biohyne A Membrane; Pall BioSupport) in 10× SSC (1.5 M NaCl, 0.15 M sodium citrate). After prehybridization, the membranes were hybridized with each probe at 42°C overnight in 40% formamide–10× Denhardt's solution (0.2% each bovine serum albumin, Ficoll, and polyvinylpyrrolidone)–5× SSC–250 μg of salmon sperm DNA per ml–50 mM NaPO<sub>4</sub> (pH 7.0)–10% dextran sulfate. The probes were labeled with [ $\alpha$ -<sup>32</sup>P]dCTP with Ex-Taq polymerase (Takara) by PCR. The primer sets used to generate the probes are listed in Table 1.

**Plasmid construction.** To construct plasmid clone SART1 ORF2 + 3'-pAcGHLTB, the SART1 ORF2 3' UTR portion was amplified by PCR from genomic library clone BS103 with primers SART1-S3014-NcoI (Table 1) and SAX3P-NotI (24). PCR was conducted for 30 cycles with *Pfu* Turbo DNA polymerase (Stratagene). The PCR product was subcloned between the NcoI and NotI sites of the pAcGHLTB plasmid (Pharming). Constructs used in the retrotransposition assay to identify the essential region of the 3' UTR were generated in either of two ways. The primers used for plasmid construction are listed in Table 1. Constructs SART1 WT (+NotI)-pAcGHLTB, ΔpolyA, 1-379/(A)<sub>20</sub>, 1-379, 73-379/(A)<sub>20</sub>, 73-379, 73-295, and 279-461-(A)<sub>20</sub> were generated by PCR amplifying the SART1 3' UTR portion of SART1 WT-pAcGHLTB and subcloning it into the NotI and BglII sites of SART1Δ3'-pAcGHLTB (24). To generate constructs SART1 Δ73-295, Δ73-227, Δ73-162, Δ163-227, Δ163-295, and Δ228-295, portions of SART1 WT-pAcGHLTB other than those deleted were amplified by inverse PCR with 5'-phosphorylated primers and then self-ligated.

EGFP1/S1-3'UTR-pVL1393 was generated by the following procedure. The *Drosophila hsp* promoter was subcloned into the HindIII site of pEGFP1 (Clontech). A BglII site was introduced into this plasmid by inverse PCR with primers hsp\_pEGFP1-A2330T-S and hsp\_pEGFP1-A2330T-A. This construct, hsp\_pEGFP1-BglII, contains a NotI site and a BglII site that immediately follows the EGFP1 protein. Next, the SART1 3' UTR portion was PCR amplified with SART1-S6222 NotI and SART1-A6704 BglII/BamHI from SART1 WT-pAcGHLTB and subcloned between the NotI and BglII sites of hsp\_pEGFP1-BglII. In the resulting plasmid, hsp-pEGFP1-SART1-3'UTR, the SART1 3' UTR resides downstream of EGFP1. Third, the EGFP1-SART1 3' UTR portion was amplified with pEGFP1-S96-EcoRI and pEGFP1-A906 from hsp\_pEGFP1-SART1-3'UTR and subcloned into the EcoRI and BglII sites of pVL1393 (Pharming).

**Recombinant AcNPV generation.** Sf9 cells were propagated as monolayer cultures at 27°C in TC-100 medium supplemented with 10% fetal bovine serum (Katakura Co., Nagano, Japan) in the presence of penicillin-streptomycin (GIBCO-BRL). The recombinant baculovirus containing the wild-type, mutant, or chimeric SART1 portion driven by the polyhedrin promoter was produced by cotransfection of the wild-type, mutant, or chimeric SART1-pAcGHLTB/pVL1393 plasmid with BaculoGold DNA (Pharming) into Sf9 cells with the Tfx-20 lipofection reagent (Promega). The medium was collected 4 days later and used for plaque purification and subsequent virus propagation in accordance with the manufacturer's (Pharming) instructions.

**In vivo retrotransposition assay by PCR.** Approximately  $3 \times 10^5$  Sf9 cells were infected in a 12-well plate with SART1-containing AcNPV at a multiplicity of 10

PFU per cell. The genomic DNA was extracted 72 h postinfection as previously described (24). PCR assays were conducted with *Ex-Taq* or *LA-Taq* (Takara) in the presence of TaqStart Antibody (Clontech) with  $\sim 1 \mu\text{g}$  of Sf9 DNA. The reaction mixture was denatured at  $94^\circ\text{C}$  for 3 min, followed by 30, 35, or 40 cycles of  $98^\circ\text{C}$  for 20 s,  $62^\circ\text{C}$  for 30 s, and  $72^\circ\text{C}$  for 1 min for the SART1 3' junction and 40 cycles for the 5' junction. Five microliters of each mixture was subjected to 2% agarose electrophoresis in Tris-acetate-EDTA buffer and visualized by ethidium bromide staining. PCR products were directly cloned into the pGEM-T Easy vector (Promega). The cloned products were sequenced with a BigDye Terminator cycle sequencing kit (Applied Biosystems) on an ABI 310 genetic analyzer and an ABI 3100 genetic analyzer. Sequence analysis was carried out with Vector NTI Suite version 7.1 (Informax).

## RESULTS

### Primary and secondary structures of the SART1 3' UTR.

First, to focus on the 3' UTR sequences essential for SART1 retrotransposition, we compared the 3' UTR sequences of SART1 (GenBank accession number D85594) with those of WISH Bm1 (NV060754), SARTPx1 (AB078931), and the *Bombyx* genomic clone of SART1 (SARTBmGC; AB088394) (Fig. 1A) (11). WISHBm1 is a derivative of SART1 that has lost its sequence specificity. SARTPx1 is a SART1 element from the swallowtail butterfly, *Papilio xuthus*, that also integrates into the same site of  $(\text{TTAGG/CCTAA})_n$  telomeric repeats. SARTPx1 and WISHBm1 constitute a monophyly with SART1. The sequence alignment was conducted by simply matching bases. The 3' UTR sequences between SART1 and SARTBmGC were highly conserved; only 21 of 461 nt in total are substituted, and only 4 nt are deleted in SARTBmGC. Between SART1 and WISHBm1, the only conserved regions are the first  $\sim 60$  nt and the last  $\sim 80$  nt in the 3' UTR. The 3' UTR sequence of SARTPx1 does not show such a remarkable sequence homology with SART1, except for  $\sim 60$  nt of the 3'-terminal region. As far as we compared the primary sequences, only  $\sim 60$  nt at the 3' terminus were highly conserved among the SART1-related elements.

Next, we predicted the RNA secondary structure of the SART1 3' UTR by computer with the mFOLD program (16, 28). The SART1 3' UTR showed five putative firm stem-loop structures (Fig. 1B). Stem-loops 1, 2, and 5 were also conserved in SARTBmGC, where all base substitutions (shown by /N) were compensating changes, and deletions were in paired nucleotides of stems (G-C pairs in stem 2). However, three out of four substitutions in stem-loop 3 altered the secondary structure, suggesting that this region is less important than the others. In WISHBm1 and SARTPx1, the only conserved structure was stem-loop 5, as was predicted from its primary structure.

**Readthrough transcripts of AcNPV-expressed SART1.** To understand the initial process of SART1 reverse transcription, we next attempted to characterize by Northern hybridization the 3' ends of the SART1 transcripts, which were produced in a baculovirus-mediated in vivo assay system (Fig. 2). We extracted total RNAs from Sf9 cells infected with AcNPV includ-

ing various SART1 constructs (SART1 WT, wild-type SART1;  $\Delta 3'$ , SART1 without the 3' UTR; ORF2+3'UTR, SART1 lacking ORF1) (Fig. 2A). We prepared a glutathione *S*-transferase (GST)-His<sub>6</sub> probe (probe a) and a polyhedrin 3' region probe (probe b) that both originated from the baculovirus transfer vector (Fig. 2A). Hybridization with probe a showed smeary bands of ca. 8 kb, which correspond to GST-fused SART1 transcripts in all of the constructs (Fig. 2B, probe a). Probe b hybridized with all of the transcripts from Sf9 cells infected with AcNPV, which express three SART1 constructs, but not with RNAs from Sf9 free of virus infection (Fig. 2B, probe b). The major band that hybridized with probe b was about 8 kb in SART1 WT and in  $\Delta 3'$  but about 6 kb in ORF2+3'UTR, which reflects the deletion of the ORF1 region (2,148 bp long). This indicates that detectable amounts of readthrough transcripts that include the downstream polyhedrin 3' UTR sequences are synthesized in baculovirus-mediated SART1 expression. In SART1 WT and ORF2+3'UTR, the original poly(A) tract ( $A_{20}$ ) contiguous to the 3' UTR was subcloned into the transfer vector. Thus, transcription continues through the poly(A) tract until reaching the polyhedrin 3' regions in these clones.

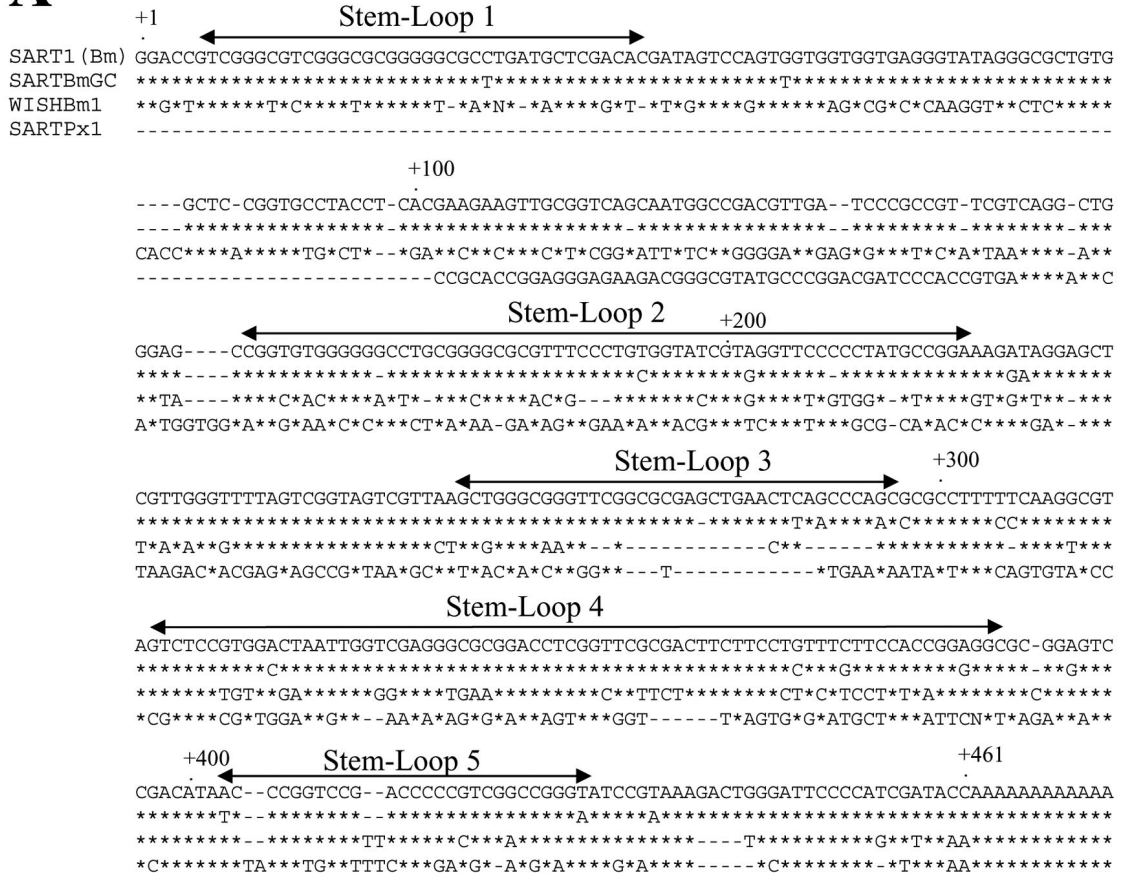
The in vivo retrotransposition assay showed that the 3' junction to the telomeric repeat of retrotransposed SART1 WT is the poly(A) tract adjacent to the 3' UTR sequence (24) (Fig. 3A, WT). These observations suggest that reverse transcription is started from the SART1 poly(A) tract in the readthrough RNA. In this case, SART1 RT must recognize both the 3' UTR and the poly(A) tract of the RNA in initiating reverse transcription. Another possibility is that there are transcripts ending with the SART1 poly(A) tract that could not be effectively distinguished by Northern hybridization and that they serve as the template for reverse transcription.

**A poly(A) tract at the end of the 3' UTR is necessary for efficient and accurate retrotransposition of SART1.** To investigate the domains in the 3' UTR including the poly(A) tract that are required for SART1 retrotransposition, we generated a series of SART1-expressing AcNPVs with partial deletions (Fig. 3A, constructs 1 to 10) and assayed their in vivo retrotransposition abilities with a system we have established (24). The AcNPV-expressed SART1 in Sf9 cells retrotransposes into the telomeric repeats  $(\text{TTAGG})_n$  in a highly sequence-specific manner. We detected the retrotransposition of SART1 by amplifying the 3' junction of SART1 to telomeric repeats by PCR with primers +6096 and  $(\text{CCTAA})_6$  (Fig. 3A) by using genomic DNAs from SART1-expressing AcNPV-infected Sf9 cells as the PCR template.

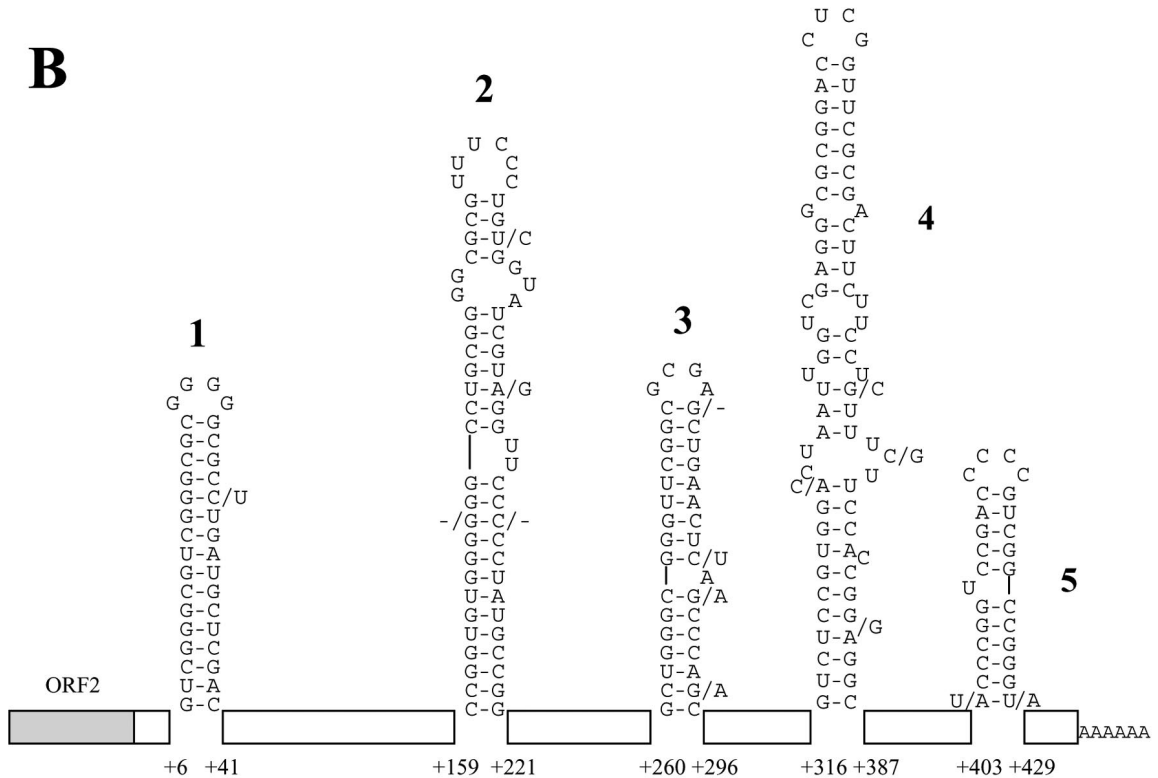
Figure 3B shows the results of the in vivo retrotransposition assay for the 3' UTR mutants listed in Fig. 3A. To clarify the differences in the retrotransposition frequencies (RFs) and abilities of respective SART1 constructs, we performed PCRs with 30, 35, and 40 cycles and compared the band patterns (Fig.

FIG. 1. Primary and secondary structures of the SART1 3' UTR. (A) Alignment of the 3' UTR sequences of SART1 and related non-LTR retrotransposons. The asterisks indicate identical nucleotides. Gaps (-) have been introduced to maximize homology. The SART1 3' UTRs predicted to form stem-loops are indicated by double-headed arrows. (B) Predicted secondary structure of the SART1 3' UTR RNA. The nucleotides are numbered from the 5' terminus of the 3' UTR. The nucleotide substitutions (or deletions) of SARTBmGC in the stem-loops are indicated by /N (or /-). The sequences that do not form stem-loop structures are indicated by boxes.

**A**



**B**



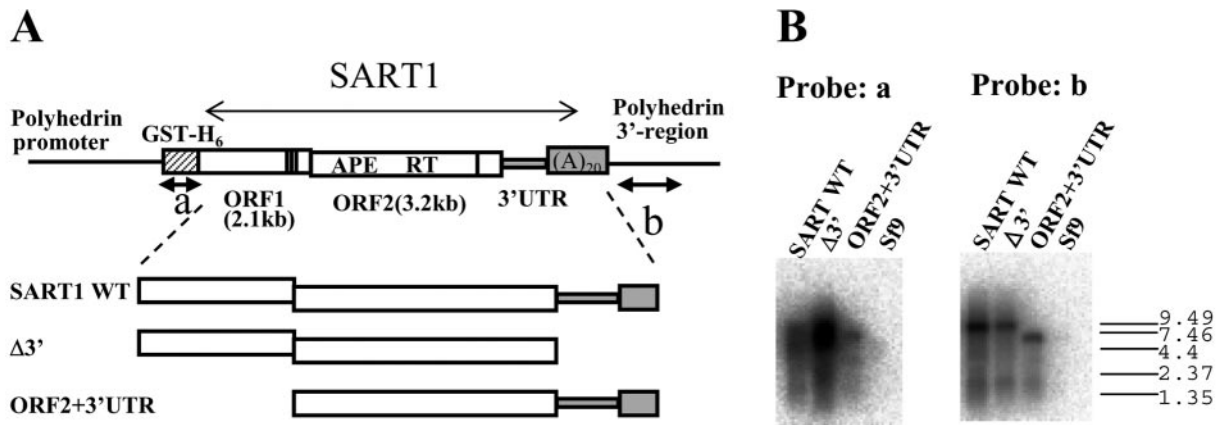


FIG. 2. Analysis of the 3' ends of AcNPV-expressed SART1 transcripts by Northern hybridization. (A) Schematic structure of AcNPV-expressed SART1 constructs used for Northern hybridization assays. Transcription is driven by the polyhedrin promoter (not drawn to scale). The hatched box represents the vector-derived GST-His<sub>6</sub> portion, which is fused in frame with ORF1. RT is the RT domain of ORF2. Vertical lines near the C termini of both ORFs represent cysteine-histidine motifs. The 3' UTR and poly(A) tract are shaded in gray, followed by the polyhedrin 3' region. The black arrows under GST-His<sub>6</sub> and the polyhedrin 3' region indicate the probes (a and b) used for Northern hybridization. APE, apurinic/aprimidinic EN. (B) Aliquots of 3  $\mu$ g of total RNA extracted from Sf9 cells infected with SART1-expressing AcNPVs were hybridized with a GST probe (probe a) and a polyhedrin 3' region probe (probe b). The RNA size marker was electrophoresed simultaneously and is shown on the right (sizes are in kilobases). Lane Sf9 contained RNA samples from uninfected Sf9 cells.

3B). In all cycles, the wild-type SART1 construct, SART1 WT (+NotI), showed an intense 600-bp PCR band, which represents the retrotransposed 3' junction region composed of 570 bp plus telomeric repeats. On the basis of the band intensity, we defined the RF for this construct as +++ (Fig. 3A).

Compared to wild-type SART1,  $\Delta$ polyA (Fig. 3A, construct 3) which is the same as SART1 WT but without a 20-bp poly(A) tract, showed a very weak PCR band of 400 bp at 35 cycles (Fig. 3B, lane 3). Sequence analysis of PCR products indicated that the band represented retrotransposition from the internal sequence of the 3' UTR (Fig. 4A, lane 3). The band density of  $\Delta$ polyA became more intense in 40 cycles; however, it was more dilute than the wild-type band, which seemed to have reached the plateau phase in 30 cycles. These data suggest that RF (shown as +) is less effective and retrotransposition is inaccurate (Fig. 4A) in this construct. Thus, the poly(A) tract at the end of the 3' UTR of SART1 is required not for retrotransposition itself but for accurate and effective reverse transcription of SART1 mRNA.

**Regions in the SART1 3' UTR essential for in vivo retrotransposition.** As with  $\Delta$ polyA, we also found that a number of 3' UTR mutations did not abolish retrotransposition but reduced its efficiency or altered its specificity. Constructs 1-379 and 1-379/(A)<sub>20</sub>, in which the 3'-terminal about 80 bp are deleted, showed weak 400-bp PCR bands (Fig. 3, constructs and lanes 5 and 6), which are about 100 bp smaller than the expected size when the correct reverse transcription occurs. Sequence analyses revealed that these bands also represented inaccurate retrotransposition (Fig. 4A, lanes 5 and 6). Similar results were obtained for constructs 73-379, 73-379/(A)<sub>20</sub>, and 73-295 (Fig. 3B, lanes 7, 8, and 9, respectively), whether or not the constructs include the poly(A) tract, suggesting that at least the most 3'-terminal region (379 to 461) is involved in efficient and accurate retrotransposition but is not required for retrotransposition itself.

A more important finding is that a mutant with a large

deletion in the 5' portion of the 3' UTR [construct 279-461-(A)<sub>20</sub> (Fig. 3A)] showed no band but only smears even in 40 cycles (Fig. 3B, lane 4), as in Sf9 (negative control) and  $\Delta$ 3'UTR (Fig. 3A, construct 2). We cloned and sequenced the PCR products, which were produced from primer-dimers but not from retrotransposed copies (Fig. 4A, lanes 2 and 4). This confirmed that construct 279-461-(A)<sub>20</sub> abolished the retrotransposition ability. Because the 72 nt in the 5'-terminal region of the 3' UTR were not essential in three constructs (73-379, 73-379/(A)<sub>20</sub>, and 73-295), we next generated construct  $\Delta$ 73-295 and found that this mutant did not produce PCR bands (Fig. 3, construct and lane 10) or retrotransposed copies (Fig. 4A, lane 10). These results indicate that region 73-295 is indispensable for SART1 retrotransposition. In construct 73-295, out of the three conserved stem-loops, 1 and 5 are disrupted and only stem-loop 2 presumably remains.

**Internal start of reverse transcription from telomeric repeat-like sequences within the 3' UTR in several mutants.** The in vivo retrotransposition assay demonstrated that mutants without a poly(A) tract or without region 379-461 showed decreased accuracy of retrotransposition. In order to understand how reverse transcription is started in mutants, we cloned and sequenced the PCR products in mutants that showed abnormal-sized bands. We obtained 52 clones from  $\Delta$ polyA, 1-379, 73-379/(A)<sub>20</sub>, 73-379, and 73-295 and have summarized the 3' junction sequences of retrotransposed SART1 mutants in Fig. 4A. Remarkably, we could not detect nontemplated additional nucleotides at the 3' end of the inserted sequence of SART1, although such additional nucleotides were observed in R2Bm (13) and human L1 (3) retrotransposon.

Of the 3' junction sequences of  $\Delta$ polyA (Fig. 4A, lane 3), six out of nine clones initiated reverse transcription from an internal sequence within the 3' UTR. There were three internal initiation sites, +236, +241, and +271 (+1 being the 5' end of the 3' UTR of SART1 [Fig. 4B]), that reside within a short

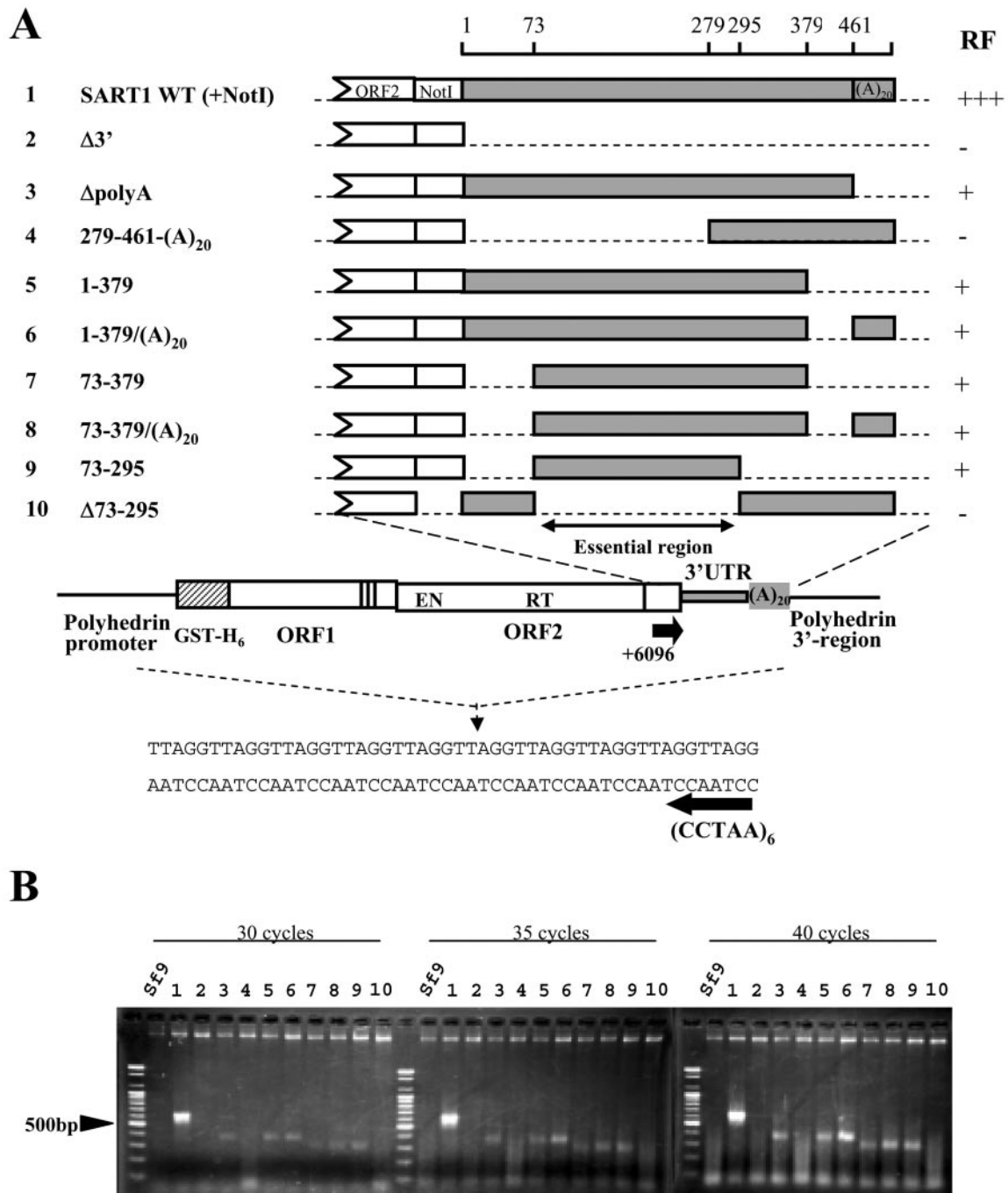


FIG. 3. Regions of the 3' UTR essential for in vivo retrotransposition. Partial deletions were introduced into the SART1 3' UTR and assayed for in vivo retrotransposition. (A) Schematic view of the retrotransposition assay of SART1 mutants containing various deletions in the 3' UTR. All of the constructs possess SART1 sequences prior to the 3' UTR (ORF1 and ORF2) and have partial or complete deletions in the 3' UTR and/or the poly(A) tail. The numbering of the nucleotides is from the 5' end of the 3' UTR. A NotI site was introduced between ORF2 and the 3' UTR to generate these constructs, except for Δ73-295. In vivo retrotransposition of SART1 mutants was detected by PCR. Sf9 cells were infected with SART1-expressing AcNPVs, and the genomic DNAs were extracted 3 days postinfection to be used as PCR templates. The 3' junction between the telomeric repeats and retrotransposed SART1 was detected with primers +6096 and (CCTAA)<sub>6</sub> (indicated by arrows). The RF of wild-type SART1 is indicated by +++. The RF of the SART1 mutant whose PCR bands were light toned at 35 cycles is indicated by +. The RF of the mutant whose PCR bands were not detected is indicated by -. (B) Ethidium bromide-stained gel of the 3' junction PCR assay for in vivo retrotransposition. The PCR was conducted for 30, 35, and 40 cycles to clarify the RF. A molecular size marker was run to the left of each Sf9 lane. The numbers above the lanes correspond to the numbering of the SART1 constructs in panel A. Sf9 stands for the genome of noninfected Sf9 cells.

## A

Lane	construct	SART1/polyhedrin 3'-region→	←telomeric repeats→	52
3	$\Delta$ polyA	---GGAGCTCGTT <sup>+236</sup>	AGG (TTAGG) <sub>6</sub>	1
		---TCGTTGGGTT <sup>+241</sup>	AGG (TTAGG) <sub>n</sub>	4
		---TGGGCGGGTT <sup>+271</sup>	AGG (TTAGG) <sub>2</sub> (TTAG) (TTAGG) <sub>6</sub>	1
		---ATACCAGATC <sup>+466</sup> *1	AGG (TTAGG) <sub>n</sub>	2
		---AAATAGGTT <sup>+3338</sup> *2	AGG (TTGG) (TTAGG) <sub>6</sub>	1
6	1-379	---TCGTTGGGTT <sup>+241</sup>	AGG (TTAGG) <sub>n</sub>	2
		---TAGTCGGTAG <sup>+252</sup>	G (TTAGG) <sub>n</sub>	1
		---TGGGCGGGTT <sup>+271</sup>	AGG (TTAGG) <sub>n</sub>	2
		---TTTTCAAGG <sup>+311</sup>	(TTAGG) <sub>n</sub>	1
7	73-379/(A) <sub>20</sub>	---TCGTTGGGTT <sup>+241</sup>	AGG (TTAGG) <sub>10</sub>	1
		---TAGTCGGTAG <sup>+252</sup>	G (TTAGG) <sub>6</sub>	3
		---TGGGCGGGTT <sup>+271</sup>	AGG (TTAGG) <sub>5-10</sub>	4
		---GTTTCTTCCA <sup>+379</sup> A <sub>18</sub> *3	AGG (TTAGG) <sub>6</sub>	1
8	73-379	---ATCGTAGGTT <sup>+206</sup>	AGG (TTAGG) <sub>4,10</sub>	2
		---GATAGGAGCT <sup>+232</sup>	AGG (TTAGG) <sub>4</sub>	1
		---GGAGCTCGTT <sup>+236</sup>	AGG (TTAGG) <sub>2</sub>	3
		---GCTCGTTGGG <sup>+239</sup> A <sub>23</sub> *4	AGG (TTAGG) <sub>2</sub>	1
		---TCGTTGGGTT <sup>+241</sup>	AGG (TTAGG) <sub>1-10</sub>	9
		---TAGTCGGTAG <sup>+261</sup>	G (TTAGG) <sub>6</sub>	1
		---CTGGGCGGGT <sup>+270</sup>	AGG (TTAGG) <sub>n</sub>	1
		---TGGGCGGGTT <sup>+271</sup>	AGG (TTAGG) <sub>n</sub>	1
9	73-295	---GGAGCTCGTT <sup>+236</sup>	AGG (TTAGG) <sub>3,5</sub>	2
		---CTCGTTGGGTT <sup>+240</sup> A <sub>20</sub> *4	AGG (TTAGG) <sub>6?</sub>	2
		---TCGTTGGGTT <sup>+241</sup>	AGG (TTAGG) <sub>17,10</sub>	2
		---TGGGTTT <sup>+245</sup>	G (TTAGG) <sub>9</sub>	1
		---TGGGCGGGTT <sup>+271</sup>	AGG (TTAGG) <sub>6,11</sub>	2
1	WT*5	---CCATCGATACC <sup>+461</sup> A <sub>n</sub>	GG (TTAGG) <sub>n</sub>	>93%
2	$\Delta$ 3'UTR			0*6
4	279-461-(A) <sub>20</sub>			0*6
10	$\Delta$ 73-295			0*6

## B

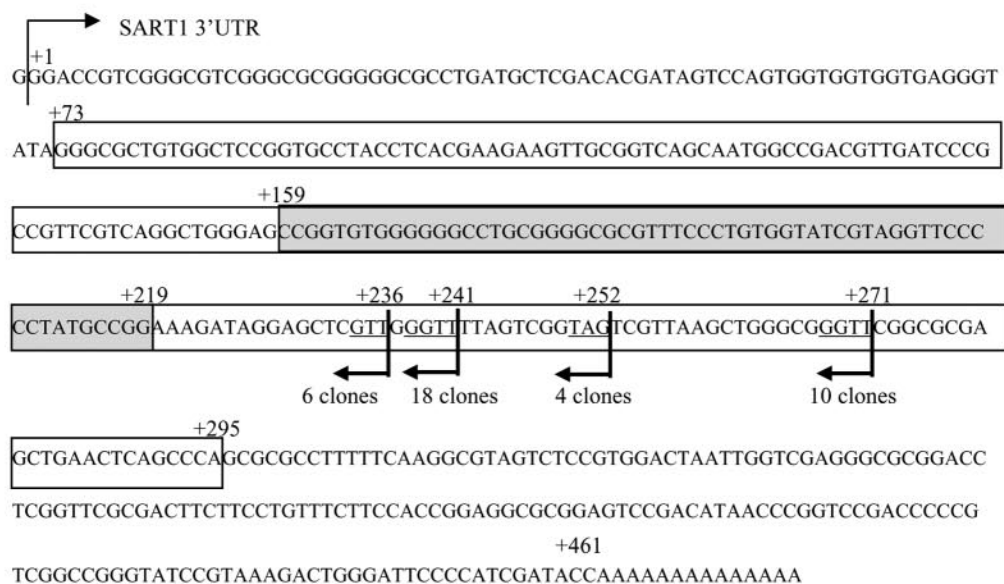


FIG. 4. Reverse transcription initiation sites of SART1 with partial deletions in the 3' UTR or the poly(A) tract. (A) Nucleotide sequences of the 3' junction PCR products in Fig. 3B. The numbering of the PCR lanes in Fig. 3B is shown on the left. The 3' junction sequences, which are the boundaries between the inserted retrotransposon copies (SART1-polyhedrin 3' region) and the host target sites (telomeric repeats), are shown

sequence of only 36 nt. Surprisingly, all of these clones end with GTT or GGTT at the 3' terminus of the retrotransposed copy and are further followed by the telomeric repeats AGG(TTAGG)<sub>n</sub> of the host genome. Of the remaining three, two clones initiated reverse transcription from just 5 nt downstream of the 3' UTR (AGATC is a readthrough sequence and is derived from the vector). The remaining single clone (+3338) initiated reverse transcription from ~250 bp downstream of the 3' UTR in the readthrough RNA derived from the vector sequence. Interestingly, the boundary sequence to telomeric repeats was GGTTT, similar to the case of internal initiation.

The other mutants also initiated reverse transcription, mostly from internal sequences, among which the majority were telomeric repeat-like GGTT (+206, +241, and +271), GTT (+236), and GT (+240 and +270): 4 clones in 1-379, 5 clones in 73-379(A)<sub>20</sub>, 16 clones in 73-379, and 6 clones in 73-295. Other than the GGTT-like sequences above, AG (+245, +252, and +261), AGG (+311), and TAG (+252) were used to start reverse transcription. All of these sequences were part of the TTAGGTTAGG telomeric repeats.

In constructs 73-379 and 73-295, we found three clones that initiated reverse transcription from the poly(A) tail, which was apparently generated by polyadenylation at position +239 or +240. The irregular polyadenylation may be due to severe deletion of 3' UTR sequences and to the lack of (A)<sub>20</sub> in these mutants.

In summary, inaccurate reverse transcription in the above mutants started with GGUU and GUU as templates, which accounted for 71% of the total 3' junction sequences. In these clones, it is interesting that reverse transcription appears to continue telomeric repeats such as GGTTAGGTT (presumed reverse-transcribed portion underlined). Furthermore, other telomeric repeat-like start sites, AG, AGG, and TAG, also ensured the continuity of the telomeric repeats, such as TAGGTTAGG or AGGTTAGG. In all, telomeric repeat-related initiation sites add up to 85%. Most (76%) of the above inaccurate reverse transcription was initiated from only four sites, shown in Fig. 4B.

**The stem-loop structure and the internal reverse transcription initiation sites make up the essential region.** To further characterize the 3' UTR that is essential for retrotransposition, we divided region 73-295 into three sections (73-162, 163-227, and 228-295) and generated mutants with each section deleted (Fig. 5A). Stem-loop 2 comprises region 163-227 except for the first 5 nt. The four major start sites for inaccurate reverse transcription (+236, +241, +252, and +271 [Fig. 4B]) are

included in region 228-295. An in vivo retrotransposition assay showed that a weak PCR band representing the retrotransposition event was only observed in  $\Delta$ 73-162 (Fig. 5B). The decrease in the PCR band size coincides with the size of the deletion in this mutant. However, all of the constructs that lack regions 163-227 and 228-295 did not show retrotransposition, indicating that both the stem-loop 2 region (163-227) and the region including telomeric repeat-like reverse transcription initiation sites (228-295) are necessary for SART1 retrotransposition. Although we do not know the exact function of the telomeric repeat-like sequences from 228-295, we suggest that the GGUU (or UAG) telomeric repeat-like sequences of SART1 RNA interact with the target telomeric DNA (the AACC bottom strand [see Fig. 7 and Discussion]).

**SART1 ORF proteins can retrotranspose GFP mRNA with the SART1 3' UTR into telomeric repeats.** We observed that a large portion of the 3' UTR in SART1 is necessary either for retrotransposition or for precision and efficiency of retrotransposition. Next, to determine whether the SART1 3' UTR is sufficient to confer retrotransposition, we asked whether an enhanced GFP (EGFP) mRNA fused to the SART1 3' UTR could retrotranspose into telomeric repeats.

For this purpose, we made construct EGFP1/S1-3'UTR by connecting the EGFP-encoding gene with SART1 3' UTR sequences and poly(A) (Fig. 6Aa). Because two different SART1 mutants can recover the ability to retrotranspose by *trans* complementation (24), we coinfecting SART1  $\Delta$ 3' (Fig. 6Ab) or the 2D699V mutant (Fig. 6Ac), which has abolished RT activity (24), with EGFP1/S1-3'UTR and assayed if EGFP1/S1-3'UTR was inserted into telomeric repeats. Sets of primers that were designed for amplifying EGFP sequences and telomeric repeats (Fig. 6A) were used for the in vivo retrotransposition assay. A distinct 700-bp band was observed in the 3' junction PCR assay when coinfecting with  $\Delta$ 3' (Fig. 6B, left, lane a + b). To confirm that the retrotransposition occurred accurately, the PCR products were cloned and sequenced. Of the nine clones sequenced, only one initiated reverse transcription from the internal sequences within the SART1 3' UTR but the remaining eight were reverse transcribed precisely from the poly(A) tract at the end of the 3' UTR of EGFP1/S1-3'UTR (data not shown). Furthermore, when primers EGFP1-A576 and (TTAGG)<sub>6</sub> were used to detect the 5' junction of the retrotransposed copy, the PCR band was also detected (Fig. 6B, right, lane a + b). The PCR band size was in good accordance with the full-length insertion of 610 bp plus the telomeric repeat length for the 3' junction

---

in each line. The number of each type is shown on the right. Nucleotide positions are indicated with the 5' end of the 3' UTR defined as +1. Position +461 is the original end of the SART1 3' UTR. The telomeric repeat-like sequences in the inserted copies are underlined. \*1, ATACC is the 3' end of the 3' UTR, and AGATC is the sequence derived from the baculovirus transfer vector. \*2, position +3338 is based on the nucleotide numbering of the pAcGHLT-B baculovirus transfer vector (Pharming). Position +3338 is 317 nt downstream of the SART1 3' end. \*3, A<sub>18</sub> is derived from the construct. \*4, A<sub>n</sub> is due to polyadenylation incidental to transcription. \*5, data from Takahashi and Fujiwara (23). In the WT construct, more than 93% of the clones showed this type of integration. \*6, we cloned and sequenced the PCR products of  $\Delta$ 3'UTR and 279-461-(A)<sub>20</sub> by TA cloning and blue-white selection to eliminate the possibility of retrotransposition. For both constructs, 3 out of 11 white colonies actually had inserts. However, they were all primer-dimers. In  $\Delta$ 73-295, we also obtained five clones that were all primer-dimers. Since we could not see the retrotransposed sequences from these mutants, they were scored as 0. In the WT and the other mutants, the percentages of retrotransposed sequences were 50 to ~90% among the white colonies analyzed (the rest were primer-dimers). (B) Major sites for reverse transcription start in SART1 mutants with partial deletions in the 3' UTR or the poly(A) tract. The boxed region indicates the essential region defined in the retrotransposition assay in Fig. 3. The region that forms stem-loop 2 is shaded. The reverse transcription initiation sites with more than four clones in mutants listed in panel A are indicated by arrows. The sequences with homology to the telomeric repeat are underlined.



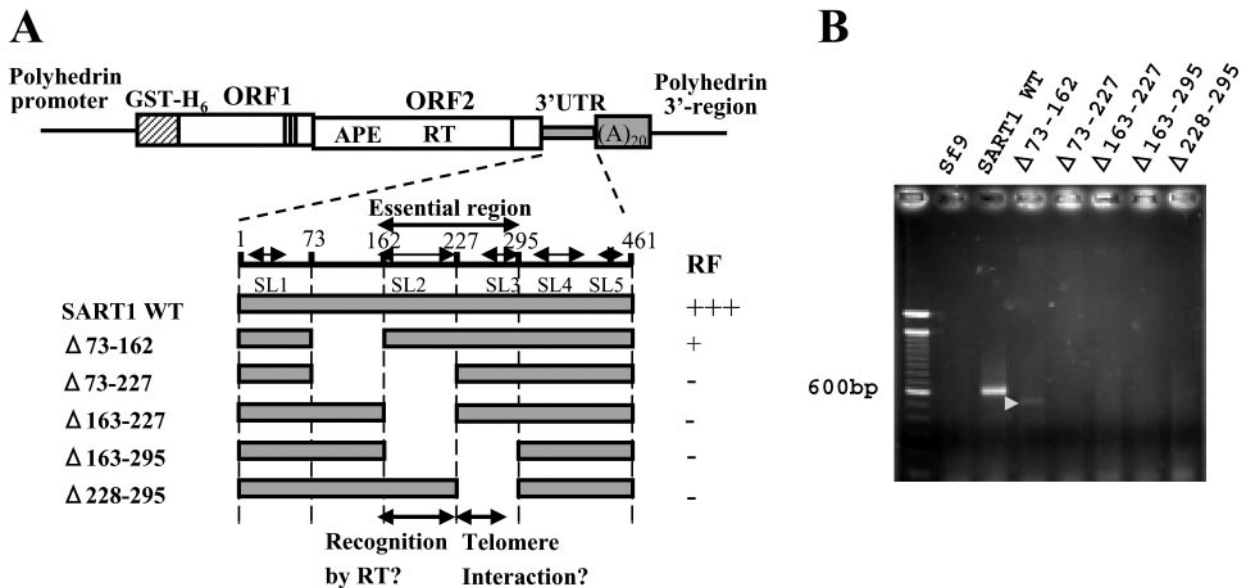


FIG. 5. The central region from nt 163 to 295 of the SART1 3' UTR is essential for in vivo retrotransposition. (A) Schematic of constructs used for the assay examining the effect caused by deleting the stem-loop structure. The positions of five stem-loop structures (SL1 to SL5) are indicated by double-headed arrows. RF was determined on the basis of the results in panel B. The regions essential for in vivo retrotransposition are stem-loop 2, the region including major sites for reverse transcription initiation in several mutants (Fig. 4), and stem-loop 3. APE, apurinic/aprimidinic EN. (B) 3' junction PCR assay of in vivo retrotransposition. Genomic DNAs of Sf9 cells infected with AcNPV containing the SART1 elements shown in panel A were PCR amplified with the primers in Fig. 3B. A faint band with the expected size, which indicates in vivo retrotransposition, appeared only for  $\Delta 73-162$  (shown by the white arrowhead).

and with 543 bp plus the telomeric repeat length for the 5' junction.

In contrast, EGFP1/S1-3'UTR did not show retrotransposition when infected alone (Fig. 6B, lane a) or with the 2D699V mutant (Fig. 6B, lanes a+c), indicating that retrotransposition of EGFP1/S1-3'UTR is mediated by the RT activity of the SART1 ORF proteins provided in *trans* from  $\Delta 3'$ . This suggests that the 3' UTR sequence of SART1 is sufficient for RNA recognition of the SART1 RT unit and that any genes having the SART1 3' UTR can be retrotransposed effectively into the telomeric repeats when *trans* complemented by 3' UTR-deficient SART1.

## DISCUSSION

**Functional structures in the 3' UTR of SART1.** Previous studies of the R2 and UnaL2 elements have suggested that the RT unit in non-LTR retrotransposons recognizes a specific RNA secondary structure within the 3' UTR in the initial step of TPRT (9, 15), although the protein-RNA interaction has not been characterized.

In this study, we found five putative stem-loops in the 3' UTR of SART1. Comparative studies showed that stem-loop 5 (+403 to +429) near the 3' end of the 3' UTR is the most conserved feature of the SART1-related elements (Fig. 1). Deletion of the 3'-terminal 3' UTR (+380 to +461) resulted in a decrease in retrotransposition efficiency and loss of precision in reverse transcription initiation. Stem-loop 5 may function in initiating reverse transcription from the poly(A) tail, because mutants rarely initiate reverse transcription from the poly(A) tract when region 380-461 is deleted. However, in vivo retrotransposition assays revealed that stem-loops 1, 4, and 5 are

not involved in essential steps of retrotransposition. Stem-loop 2 was conserved only among the two SART1 genomic clones of the silkworm, but when deleted, the retrotransposition ability was abolished, suggesting that this stem-loop is an essential structure recognized by the RT unit. Because stem-loop 2 is not conserved in SARTPx1 and WISHBm1, stem-loops formed at different positions may be recognized for these forms.

From the retrotransposition assay, we found that deletion of poly(A) from SART1 WT causes aberrant and inefficient reverse transcription initiation (Fig. 3 and 4,  $\Delta$ polyA). This indicates that the poly(A) tract is not critical for retrotransposition but is important in initiating reverse transcription. This is different from human L1, which is believed to recognize just the poly(A) tract itself (19). *Drosophila* I factor has TAA repeats at the 3' end instead of the poly(A) tail. The deletion of TAA repeats also affects the reverse transcription initiation site and the efficiency of reverse transcription in I factor. The functional role of the poly(A) tract in SART1 may be similar to TAA repeats of I factor (1) rather than the poly(A) tail of L1.

**Reverse transcription from telomeric repeat-like sequences.** Deletion of the poly(A) tract and the latter part of the 3' UTR decreased retrotransposition efficiency and altered the reverse transcription initiation sequences. The 3' junction of wild-type SART1 on the genome is composed of 3' UTR-(A)<sub>n</sub>AGGTTAGGTTAGG, although most of the 3' junction sequences of the above mutants were GGTTAGGTTAGG or TAGTTAGG (the presumed SART1-derived sequence is underlined). In these mutants, reverse transcription was initiated mainly at the internal GGUU sequences concentrated from +236 to +271 within the 3' UTR, not from the poly(A) tract. There are many GGUU or UAG sequences in SART1 mRNA, but only four sites in a very restricted region of the 3' UTR are

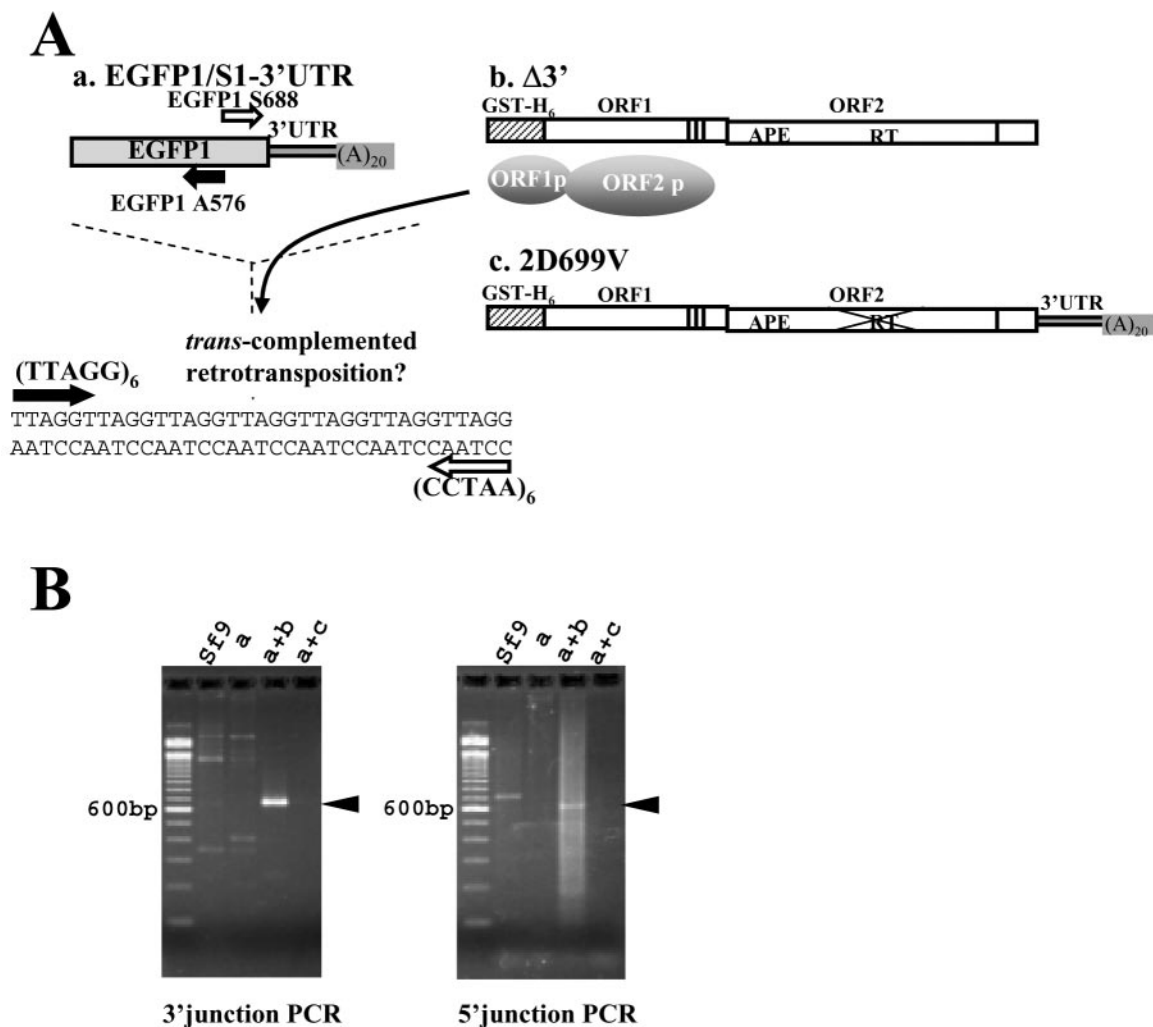


FIG. 6. The EGFP gene connected with the SART1 3' UTR is retrotransposed into the telomeric repeats. (A) Schematic of *trans* complementation between EGFP-S1/3'UTR (a) and Δ3'UTR (b). EGFP1/S1-3'UTR is an AcNPV-expressed clone that includes the EGFP gene directly followed by the complete SART1 3' UTR. Δ3'UTR does not have retrotransposition activity because it lacks the 3' UTR but can provide the ORF1-ORF2 protein complex. 2D699V is a mutant that has lost its RT activity (24). The 3' junction of *in vivo* retrotransposed copies was detected by PCR with primers EGFP1 S688 and (CCTAA)<sub>6</sub> (shown by white arrows), and the 5' junction was detected by PCR with primers EGFP1 A576 and (TTAGG)<sub>6</sub> (shown by black arrows). APE, apurinic/apyrimidinic EN. (B) Ethidium bromide-stained gels of the junction PCR assay for *in vivo* retrotransposition assay. Left, 3' junction PCR; right, 5' junction PCR; a, EGFP1/S1-3'UTR infection of Sf9 cells; a+b, coinfection with EGFP1/S1-3'UTR and Δ3'UTR; a+c, coinfection with EGFP1/S1-3'UTR and 2D699V; Sf9, cells without AcNPV infection. The black arrowhead indicates the successful *in vivo* retrotransposition band with the expected size. Molecular sizes are shown on the left.

selected for initiating reverse transcription. Why does reverse transcription start exactly from the telomeric repeat-like sequences in region 236-271? One possible explanation is that GGUU telomeric repeat-like sequences might interact with the target telomeric (CCTAA) strand.

We propose a hypothetical model in Fig. 7, which shows the specific interaction between the SART1 RNA and the target telomeric DNA during initiation of the TPRT reaction. Although the biochemical features of the SART1 EN domain have not been analyzed, SART1 EN is presumed to nick between CCT and AA of the (CCTAA) bottom strand specifically, on the basis of junction sequence analyses of SART1 in the *Bombyx* genome (25), *in vivo* retrotransposition assays (24), and *trans* complementation experiments (T. Anzai and H. Fujiwara, unpublished data). After the bottom strand of the

target telomeric DNA is nicked by the SART1 EN, RNA of SART1 mutants is presumably bound to the RT domain by stem-loop 2 and bound to the 3'-CCAA-5' sequence of the future primer strand DNA by the 5'-GGUU-3' sequences (Fig. 7A). When the mutant SART1 RNA anneals to the telomeric DNA CCAA sequence, a thymine residue (T) at the 3' end of the nicked bottom strand will be removed because the T does not pair with the RNA strand. Recently, human apurinic/apyrimidinic EN 1, with a structure similar to that of the EN domains of many LINES including SART1 (14), was reported to have 3'-to-5' exonuclease activity in mismatched DNA pairs (2). Thus, the EN domain of SART1 might remove the superfluous mismatched nucleotides on the bottom target DNA, which results in the start of reverse transcription from the nucleotide next to the CCAA site. If this hypothesis is correct,

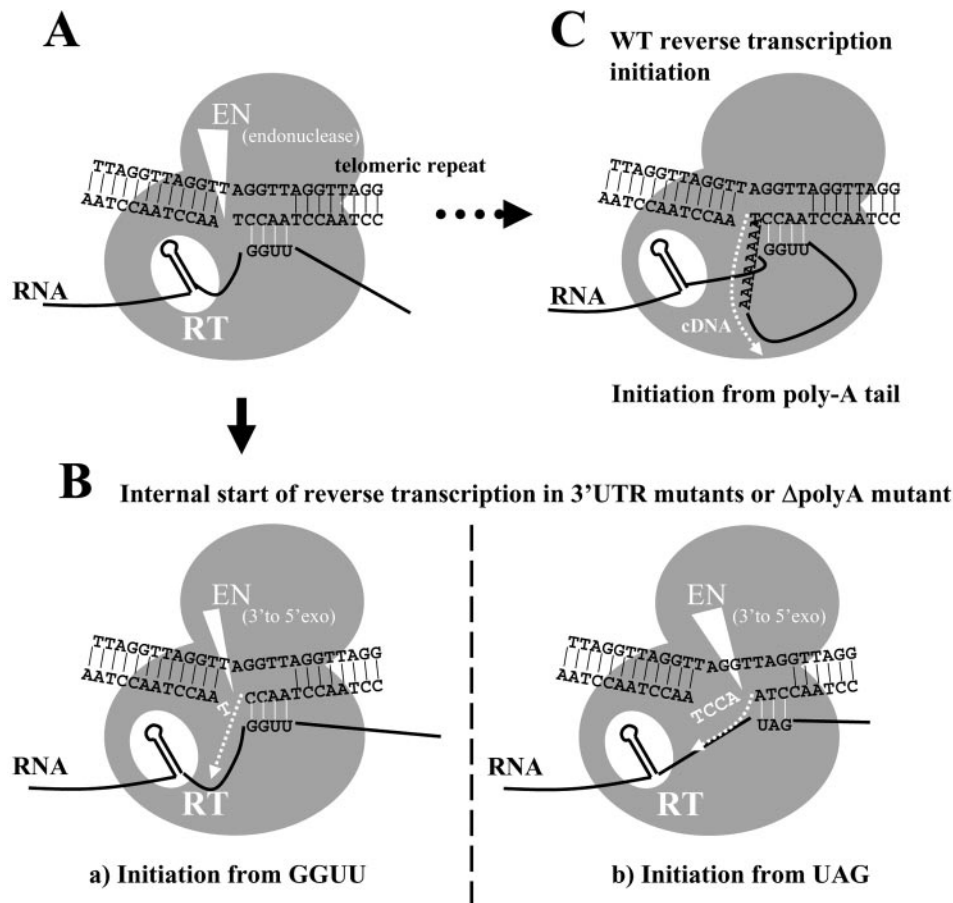


FIG. 7. Hypothetical models for initial processes of SART1 reverse transcription. (A) First, the bottom strand of telomeric repeats, which is the target DNA of SART1, is presumably nicked at the specific site between A and T of the AACCT repeats by the specific EN activity of EN encoded by SART1 ORF2. The present study suggests the interaction between mRNA of SART1 3' UTR mutants and the AACCT bottom strand (see text). GGUU of the SART1 RNA is bound to the future primer strand of the telomeric DNA. SART1 protein units are shown as gray circles. The putative interaction is indicated by short white lines. (B) Model of reverse transcription in the 3' UTR mutants. cDNA synthesis is shown by the dotted white arrow. (a) Reverse transcription from GGUU sites. The thymine residue in the nicked end of the bottom strand is presumably removed by the 3'-to-5' exonuclease activity of the EN domain (shown by a large white triangle). cDNA synthesis starts from the 5'-AACCT-3' sequence of the bottom strand as the primer, with the SART1 mRNA sequence next to the GGUU as the template. This model can explain the results shown in Fig. 4. (b) Reverse transcription from UAG sites. As in scheme a, the 5'-ACCT-3' sequence is presumably removed by the presumed exonuclease activity of EN (shown by a large white triangle). The cDNA synthesis starts from the 5'-CCTA-3' sequence of the bottom strand as the primer, with the SART1 mRNA sequence next to the UAG as the template. (C) Hypothetical reverse transcription model of wild-type SART1. There is a possibility for RNA-DNA annealing in wild-type SART1 as in mutants. Reverse transcription starts from the primary nicking site made by EN with the poly(A) tail as the template.

reverse transcription must start from the next nucleotide 5' to the GGUU sequence as an RNA template with the CCAA sequence as a DNA primer (Fig. 7Ba). However, we do not know exactly whether the junction GGTT sequence in the retrotransposed copies of mutants originates from the host genome or newly synthesized by reverse transcription.

Other telomeric repeat-like sequences, AGG, AG, and TAG, in the SART1 3' UTR were also selected as initiation sites for reverse transcription in the above mutants (Fig. 4A). Similar to the GGUU initiation mechanisms, the selection of these sites seems to be dependent on the annealing of 3' UTR RNA to the telomeric DNA (Fig. 7Bb). The SART1 EN originally nicks the bottom strand between A and T, but its 3'-to-5' exonuclease activity further eliminates the CTCA sequence because it is not annealed to the 3' UTR RNA.

Recently, we found similar phenomena in an *in vivo* retro-

transposition assay of the R1 element (Anzai and Fujiwara, unpublished data), which is a 28S ribosomal DNA-specific retrotransposon (27). Wild-type R1 frequently used TG or TGT in the 3' UTR of RNA as the initiation sites for reverse transcription. The TG or TGT nucleotides correspond to the 5' end of the target DNA of R1 (TGTCCTATCTACT). These observations suggest that the interaction between the RNA template and the target DNA may occur and facilitate the initial step of TPRT in wild-type SART1 (Fig. 7C), although we do not have direct evidence of this. The fact that four of the initiation sites of SART1 RNA mentioned above are located within the essential region between positions +228 and +295 supports this hypothesis.

**Recognition of 3' UTR by RT of non-LTR retrotransposons.** The necessity of the 3' UTR for retrotransposition has been reported in other non-LTR retrotransposons. For instance, the

common 3' UTR structure of UnaL2 and UnaSINE1 enables retrotransposition of both elements (9). SART1 is similar to UnaL2 in the sense that they both have stem-loop structures in the essential region. However, SART1 requires the full-length 3' UTR sequence for efficient and precise insertion, so the entire 3' UTR may play a role in determining RNA structure, as was suggested for R2 (15).

Human L1 is able to retrotranspose under conditions in which most of the 3' UTR is deleted, and it is likely that RT recognizes the poly(A) tail to initiate reverse transcription (19). According to such relaxed recognition of template RNA, human L1 possibly recognizes huge numbers of cellular mRNAs with a poly(A) tail. Both L1-mediated processed pseudogene formation and 3' transduction have been demonstrated in the human genomes and in experimental results (6, 8, 18, 26). Human L1 prefers acting on self-mRNA (*cis* preference) rather than other mRNA (*trans* complementation) (26). This feature of L1 should minimize the frequency of genome rearrangement caused by its relaxed recognition of L1 mRNA.

In contrast, SART1 recognizes the 3' UTR sequences strictly and works in *trans* effectively on other SART1 mRNAs. This is supported by the observation that the complete length of the SART1 3' UTR fused with the gene for EGFP is more effectively retrotransposed than SART1 mutants with deletions in the 3' UTR when SART1 ORF proteins are provided in *cis*. Because SART1 has strict recognition of its 3' UTR, its action in *trans* on other mRNAs is only restricted for SART1 species and does not cause processed pseudogene formation. These specialized features of SART1—strict targeting, strict 3' UTR recognition, and *in-trans* action—may be correlated with the control of telomere length in the silkworm, which has lost or attenuated its telomerase activity (7, 22).

The above-described transposition system has several advantages for gene delivery. A long mRNA gene to be delivered can be manipulated easily because only the 460-bp SART1 3' UTR is put on the 3' end of the desired mRNA. Furthermore, the retrotransposed copies are stable in the genome because the enzymatic unit SART1 ORF proteins are provided separately in *trans*.

#### ACKNOWLEDGMENTS

This work was supported by a grant from the Ministry of Education, Science and Culture of Japan (15370003) and by a Grant-in-Aid from the Research for the Future Program of the Japan Society for the Promotion Science.

#### REFERENCES

- Chambeyron, S., A. Bucheton, and I. Busseau. 2002. Tandem UAA repeats at the 3'-end of the transcript are essential for the precise initiation of reverse transcription of the I factor in *Drosophila melanogaster*. *J. Biol. Chem.* **277**:17877–17882.
- Chou, K. M., and Y. C. Cheng. 2002. An exonucleolytic activity of human apurinic/aprimidinic endonuclease on 3' mispaired DNA. *Nature* **415**:655–659.
- Cost G. J., Q. Feng, A. Jacquier, and J. D. Boeke. 2002. Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.* **21**:5899–5910.
- Dewannieux, M., C. Esnault, and T. Heidmann. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**:41–48.
- Eickbush, T. H. 1992. Transposing without ends: the non-LTR retrotransposable elements. *New Biol.* **4**:430–440.
- Esnault, C., J. Maestre, and T. Heidmann. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**:363–367.
- Fujiwara, H., Y. Nakazato, S. Okazaki, and O. Ninaki. 2000. Stability and telomere structure of chromosomal fragments in two different mosaic strains of the silkworm, *Bombyx mori*. *Zool. Sci.* **17**:743–750.
- International Human Genome Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
- Kajikawa, M., and N. Okada. 2002. LINES mobilize SINEs in the eel through a shared 3' sequence. *Cell* **111**:433–444.
- Kazazian, H. H., Jr. 2000. Genetics. L1 retrotransposons shape the mammalian genome. *Science* **289**:1152–1153.
- Kojima, K. K., and H. Fujiwara. 2003. Evolution of target specificity in R1 clade non-LTR retrotransposons. *Mol. Biol. Evol.* **20**:351–361.
- Luan, D. D., M. H. Korman, J. L. Jakubczak, and T. H. Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**:595–605.
- Luan, D. D., and T. H. Eickbush. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.* **15**:3882–3891.
- Malik, H. S., W. D. Burke, and T. H. Eickbush. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**:793–805.
- Mathews, D. H., A. R. Banerjee, D. D. Luan, T. H. Eickbush, and D. H. Turner. 1997. Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA* **3**:1–16.
- Mathews, D. H., J. Sabina, M. Zuker, and D. H. Turner. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**:911–940.
- Matsumoto, T., H. Takahashi, and H. Fujiwara. 2004. Targeted nuclear import of open reading frame 1 protein is required for *in vivo* retrotransposition of a telomere-specific non-long terminal repeat retrotransposon, SART1. *Mol. Cell. Biol.* **24**:105–122.
- Moran, J. V., R. J. DeBerardinis, and H. H. Kazazian, Jr. 1999. Exon shuffling by L1 retrotransposition. *Science* **283**:1530–1534.
- Moran, J. V., S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, and H. H. Kazazian, Jr. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**:917–927.
- Okada, N., M. Hamada, I. Ogiwara, and K. Ohshima. 1997. SINEs and LINES share common 3' sequences: a review. *Gene* **205**:229–243.
- Ostertag, E. M., and H. H. Kazazian, Jr. 2001. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* **11**:2059–2065.
- Sasaki, T., and H. Fujiwara. 2000. Detection and distribution patterns of telomerase activity in insects. *Eur. J. Biochem.* **267**:3025–3031.
- Takahashi, H., and H. Fujiwara. 1999. Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyx mori*. *Nucleic Acids Res.* **27**:2015–2021.
- Takahashi, H., and H. Fujiwara. 2002. Transplantation of target site specificity by swapping the endonuclease domains of two LINES. *EMBO J.* **21**:408–417.
- Takahashi, H., S. Okazaki, and H. Fujiwara. 1997. A new family of site-specific retrotransposons, SART1, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. *Nucleic Acids Res.* **25**:1578–1584.
- Wei, W., N. Gilbert, S. L. Ooi, J. F. Lawler, E. M. Ostertag, H. H. Kazazian, J. D. Boeke, and J. V. Moran. 2001. Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell. Biol.* **21**:1429–1439.
- Xiong, Y., and T. H. Eickbush. 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. *Mol. Cell. Biol.* **8**:114–123.
- Zuker, M., D. H. Mathews, and D. H. Turner. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide, p. 11–43. *In* J. Barciszewski and B. F. C. Clark (ed.), *RNA biochemistry and biotechnology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.