Nucleic Acids Research, 1999, Vol. 27, No. 9 **2015–2021**

Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyx mori*

Hidekazu Takahashi and Haruhiko Fujiwara*

Department of Biological Sciences, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Received January 7, 1999; Revised and Accepted March 4, 1999

ABSTRACT

The telomeres of the silkworm *Bombyx mori* consist of $(TTAGG)_n$ repeats and harbor a large number of sequence-specific non-LTR retrotransposons such as TRAS1 and SART1. In order to ascertain if TRAS1 and SART1 are transcribed *in vivo* and if there is a novel transcription mechanism peculiar to the sequence-specific retrotransposons, we studied their transcription. We detected transcripts of TRAS1 and SART1 by northern hybridization in many tissues and the BmN4 cell line of the silkworm. 5′-Rapid amplification of cDNA ends analysis showed that transcription of both elements was initiated precisely from their own 5′-ends and that most of their genomic copies contained these initiation sites. TRAS1 contained an internal promoter and positively regulating elements in the +1/+581 nucleotides in its 2432 bp 5′-untranslated region (UTR). We could not, however, detect any promoter activity in the SART1 5′-UTR. This difference may be related to the fact that only TRAS1 contained an initiator-like element at its 5′-end. Placing 1–52 units of the telomeric repeat $(TTAGG)_n$ upstream of TRAS1 reduced transcription 5-fold. The evidence suggests that most of the TRAS1 genomic copies within the telomeric repeats are weakly transcribed *in vivo*.

INTRODUCTION

Most eukaryotic telomeres consist of direct repeats of oligonucleotides (telomeric repeats) which are elongated by a reverse transcriptase named telomerase (1). Some insects, however, lack the telomeric repeats and offer interesting models for the study of the function of telomeres. These insects have alternative mechanisms to maintain their telomeres. The telomeres of *Drosophila melanogaster* are elongated by addition, approximately once every 100 generations per one chromosomal terminus, of the retrotransposable elements HeT-A and TART (2,3). The telomeres of the dipteran *Chironomus papallidivittatus* and *Anopheles gambiae* are likely to be maintained by recombinational events (4,5). These telomerase-independent telomere maintenance mechanisms may represent general back-up for telomerase in other eukaryotes, including yeast, mouse and human (6–10).

The telomere of the silkworm *Bombyx mori* has a structure intermediate between canonical repeat-type and *Drosophila*-type telomeres (Fig. 1A). In the *Bombyx* telomeres, hundreds of copies of several families of non-LTR retrotransposons (LINE-like elements) are accumulated within the telomeric repeats, $(TTAGG)_n$, although these retrotransposons are unlikely to exist within 6–8 kb from the extreme ends (11,12). We have been characterizing these telomeric repeat-specific retrotransposons in order to study the evolution of retrotransposon-type telomeres. These *Bombyx* retrotransposons are sequence-specific, non-LTR elements that have been classified, by differences in sequence specificity and in amino acid sequences, into two distinct, large families called telomeric repeat-associated sequence (TRAS) and SART (13; Fig. 1B). Two members of these large families, TRAS1 and SART1, are similar in their amino acid sequences to the insect 28S rDNA-specific retrotransposons R1 (14) and RT1 (15), respectively. However, TRAS1 and SART1 are not closely related to the *Drosophila* HeT-A or TART elements (13). TRAS1 and SART1 comprise two open reading frames (ORFs). Their ORFs 1 are reminiscent of the retroviral Gag ORF, in that they contain three cysteine–histidine motifs near the C-terminus. Their ORFs 2 contain an endonuclease domain that probably determines the sequence specificity upon insertion, as in R1 (16). The ORFs 2 also contain a reverse transcriptase-like domain that probably conducts the target-primed reverse transcription (17) and a cysteine–histidine motif. TRAS1 and SART1 together make up ~1.5% of the total *Bombyx* genomic DNA. In addition to TRAS1 and SART1, at least five more families have been identified so far (unpublished data), suggesting that altogether the families of telomeric repeat-specific retrotransposons make up nearly 10% of the genomic DNA.

Transcription is an important regulatory step for non-LTR retrotransposons because their RNA serves not only as the messenger for translation but also as the template for reverse transcription. Although TRAS1 and SART1 encode fully conserved ORFs, it is unknown if they are actually transcribed *in vivo* and if a few master copies or most of the genomic copies are transcribed. It is also unknown whether the sequence-specific, non-LTR elements have a unique transcription mechanism. Many non-sequence-specific, non-LTR elements such as *Drosophila* Jockey (18), I (19), F (20) and Doc (21) and human LINE-1 (22,23) have been shown to contain promoter activities in the

*To whom correspondence should be addressed. Tel: +81 3 3812 2111; Fax: +81 3 3816 1965; Email: haruh@biol.s.u-tokyo.ac.jp

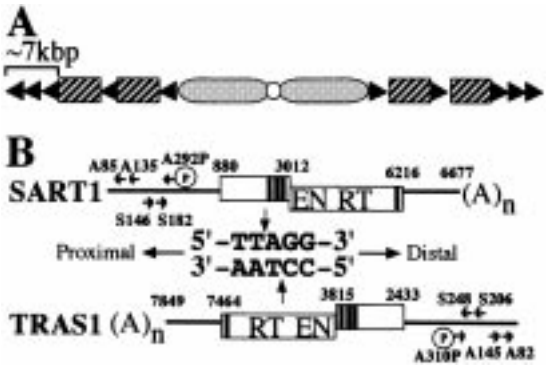

Figure 1. The telomeric repeat-specific retrotransposons in *B.mori*. (**A**) Schematic structure of the *Bombyx* chromosome. Black triangles represent the telomeric repeats, $(TTAGG)_n$. Hatched boxes indicate the telomeric repeat-specific retrotransposons. The extreme ends (~6–8 kb) consist only of the telomeric repeats, while a large number of sequence-specific retrotransposons are mixed with the subtelomeric repeats. 50–100 copies of the retrotransposons may exist at one terminus. (**B**) Sequence specificity of the TRAS/SART families. The position at which the members of the TRAS or SART family insert into the telomeric repeats are shown with vertical arrows, based on the 5′-RACE analysis in Figure 3. The ORFs and UTRs are depicted as open boxes and horizontal lines, respectively, above which nucleotide positions are indicated. Both large families comprise an endonuclease domain (EN), a reverse transcriptase domain (RT), zinc finger-like motifs (vertical lines near the C-termini of both ORFs) and a poly(A) tail. The TRAS families exist with their poly(A) tails facing toward the termini and the SART families are in the opposite orientation.

5′-untranslated regions (UTRs). It seems rational for those non-LTR elements that transpose to many different genomic locations to have promoters downstream of the transcription initiation sites in order to retain their own promoter sequences after transcription and reverse transcription. TRAS1 and SART1, however, insert only at specific positions within the telomeric repeats and thus it is unclear if they have downstream promoters, as do other elements. There is also a possibility that their promoters are dependent on the existence of the telomeric repeats. It is interesting in this regard that transcription of the *Drosophila* HeT-A elements that exist as tandem repeats at telomeres is promoted by the neighboring copies (24).

In this report, we focus on the transcription of TRAS1 and SART1, showing that TRAS1 and SART1 are transcribed in many tissues and in the BmN4 cell line. We demonstrate that the two elements are transcribed from their own 5′-ends of the retrotransposon units without the telomeric repeats at the 5′-ends. Using transient reporter assays, we identified promoter activity in the upstream region of the 5′-UTR of TRAS1 in the BmN4 cell line. We discuss whether the elements are transcribed *in vivo* from most of the copies within the telomeric repeats or from a few master copies.

MATERIALS AND METHODS

Nucleotide sequence accession number

Nucleotide positions indicated in the text are referred to based on D38414 (TRAS1) and D85594 (SART1). In this paper, however, we redefine the transcription initiation sites as the +1 positions which were determined by 5′-rapid amplification of cDNA ends (5′-RACE).

Northern hybridization

Poly(A)$^+$ RNAs were isolated from testes, ovaries, fat bodies, malpighian tubules and posterior silk glands of fifth instar larvae and from BmN4 cultured cells of *B.mori* with Micro-Fast Track™ (Invitrogen). Aliquots of 5 µg of poly(A)$^+$ 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 (Hybond-N; Amersham) in 10× SSC (1.5 M NaCl, 0.15 M sodium citrate). After prehybridization, the membranes were hybridized with ~10^6 c.p.m. of each probe at 42°C overnight in 40% formamide, 10× Denhart's solution (0.2% each of BSA, Ficoll and polyvinylpyrrolidone), 5× SSC, 250 µg/ml salmon sperm DNA, 50 mM $NaPO_4$ (pH 7.0), 10% dextran sulfate. The double-strand DNA probes shown in Figure 2 were amplified by PCR with LA *Taq* polymerase (Takara) and labeled with [α-^{32}P]dCTP by random priming using the BcaBEST DNA labeling kit (Takara). The primer sets used for PCR were as follows (Table 1): TRAS1, S2427 and A7867; SART1, S869 and A6799. A *Bombyx* EF-1α probe (310 bp) was generated by PCR using testis cDNA with a pair of primers, BmEF-1α S840 and A1150. The probes were purified with Microspin S-200 HR columns (Pharmacia Biotech) and heat denatured prior to hybridization.

Table 1. List of primers

Name	Sequence (5' to 3')
TRAS1 S2427	GCTCTAGATTAACGATGTACGCAAATATTAGAACCCCA
TRAS1 A7867	GCTCTAGATTTTTTTTTTTTTTTTTTTTAAGTCACTCT
SART1 S869	GCTCTAGAAAAAATGTCCAGTTATAAAGAAGAATTACCCCAGGA--GGGTACCCC
SART1 A6799	GCTCTAGATTTTTTTTTTTTTTTTTTTGGTATCGATG
EF-1α S840	GGCAGAGTTGAAACTGGTGTG
EF-1α A1150	ATGTGGGCAGTGTGGCAATC
TRAS1 A307P	CACCTGATTGTAGCT[a]
SART1 A287P	AATGGGAACGAGGAG[a]
TRAS1 A142	ACACACGGATAGTATTCGCGAACG
TRAS1 S203	CCTGCAAATTAGATATTCCAGACGC
TRAS1 A79	ATAACCCGTTCGTTCGATTT
TRAS1 S245	AGCGACAACTTACGTTTCTT
SART1 A130	TGTGACTGCTCCCAGAAGCCCGTTAT
SART1 S141	TATCGCTCCTTCTTTCTCGTCGTGTCA
SART1 A80	TGGAAGTCCAGCAAAACTCC
SART1 S177	AGATTGAGCGACGCGGA

[a]The 5′-end is phosphorylated for later ligation.

5′-RACE analysis

5′-RACE was conducted through self-ligation of synthesized cDNA (25) using the 5′-Full Race Core Set (Takara). An aliquot of 20 ng of BmN4 poly(A)$^+$ RNA was reverse transcribed with 5 U of AMV reverse transcriptase at 42°C for 60 min and then at 50°C for 30 min using 200 pmol of TRAS1 A307P primer or SART1 A287P primer, whose 5′-ends were phosphorylated for later ligations. The cDNA was treated with 30 U of RNase H at 32°C and ethanol precipitated. The phosphorylated 5′-end and the 3′-end of the single-strand cDNA were self-ligated with 40 U of T4 RNA ligase. The ligated circular cDNA was amplified in two steps by nested PCR with LA *Taq* polymerase (Takara). The primer pairs for the nested PCR were as follows (Fig. 1B and Table 1): TRAS1 first PCR, A142 and S203; TRAS1 second PCR, A79 and S245; SART1 first PCR, A130 and S141; SART1 second PCR, A80 and S177. The PCR products were cloned into the pGEM T Easy Vector (Promega) and sequenced with an automatic DNA sequencer SQ5500 (Hitachi).

Inverse PCR

Inverse PCR was carried out as described by Triglia *et al.* (26). Genomic DNA was extracted from the silk glands of fifth instar larvae as described previously (13). Aliquots of 15 ng of the genomic DNA were digested with a four-base cutter, *Msp*I or *Alu*I, which cut restriction sites at the +323 or +544 positions in TRAS1 and SART1, respectively. The digested DNA was circularized through self-ligation using the DNA Ligation Kit V.2 (Takara). The DNA was amplified in two steps by PCR with Ex *Taq* polymerase (Takara). The primer sets for the inverse PCR were the same as those used for the 5′-RACE PCR: TRAS1 first PCR, A142 and S203; TRAS1 second PCR, A79 and S245; SART1 first PCR, A130 and S141; SART1 second PCR, A80 and S177. The PCR products were cloned into the pGEM T Easy vectors (Promega) and sequenced as described above.

Promoter assay

The *in vivo* promoter assay was conducted by transient transfection of each experimental DNA cloned in the pGL3-Enhancer vector containing a firefly luciferase gene as reporter (the Dual-Luciferase™ Reporter system; Promega) into the BmN4 cell line. To normalize expression of the firefly luciferase, the pRL vector containing the *Drosophila* HSP70 promoter, which provides constitutive expression of *Renilla* luciferase in BmN4 cells, was co-transfected with each experimental DNA in the firefly luciferase vector. The *Drosophila* HSP70 promoter was a gift from Dr H. Maekawa and was subcloned into the *Hin*dIII site of the pRL-null vector (HSP70-pRL vector). In order to construct experimental firefly luciferase vectors, each DNA was amplified by PCR with LA *Taq* polymerase (Takara), with a pair of primers designed to make sequences for a *Nhe*I and a *Hin*dIII site at each end of the PCR product. The PCR products digested with *Nhe*I and *Hin*dIII were cloned into the pGL3-Enhancer vector. Each construct was sequenced and turned out to be free from PCR errors. The experimental DNAs were transfected into the BmN4 cell line by a liposome-mediated method using the Tfx™ 20 reagent (Promega). To ~1×10^5 BmN4 cells adherent to a well in a 96-well plate was added 40 µl of serum-free TC100 medium (Gibco-BRL) which contains 0.75 µg of each pGL3 construct, 0.01 µg of the HSP70-pRL vector and 2.3 µl of Tfx™ 20. Four hours later, 200 µl of TC100 medium with 10% fetal bovine serum was added. The luciferase assay was conducted 72 h later according to the manufacturer's instructions using a Luminescence Reader BLR-201 (ALOKA).

RESULTS

TRAS1 and SART1 were transcribed as full-length units in various tissues and in the BmN4 cell line

To determine if TRAS1 and SART1 are transcribed *in vivo*, we extracted 5 µg of poly(A)$^+$ RNAs from several tissues of fifth instar larvae and from BmN4 cells and carried out northern hybridization. We hybridized a single filter with a TRAS1 probe (Fig. 2A), reprobed it with a SART1 probe (Fig. 2B) and then with a *Bombyx* EF1-α probe as a positive control (Fig. 2C). The probes were randomly labeled, double-strand DNAs derived from both ORFs and the 3′-UTR of TRAS1 and SART1 (Fig. 2D). Figure 2A and B and longer exposure of the same autograms showed that both TRAS1 and SART1 were transcribed in all the tissues we examined. It is noteworthy that only single bands were seen for TRAS1 and SART1 (Fig. 2A and B), in contrast to most non-LTR retrotransposons, which comprise various sizes of transcripts. An ~2 kb band detected with the TRAS1 probe was probably a non-specifically hybridized rRNA resulting from incomplete purification of the synthesized probe. That band was not seen reproducibly (data not shown). When differences in the signal intensity of EF-1α (Fig. 2C) among tissues were taken into account, we found that similar levels of TRAS1 or SART1 were transcribed among tissues apart from in the BmN4 cells. The lengths of the TRAS1 and SART1 transcripts were ~8 and 7 kb, respectively, which correspond to the full-length units of the retrotransposons themselves. Detection of the full-length RNAs supports the idea that TRAS1 and SART1 can retrotranspose *in vivo*, because full-length RNAs can serve as the template for target-primed reverse transcription. Here, we have demonstrated the first example of sequence-specific retrotransposons that are actively transcribed *in vivo*.

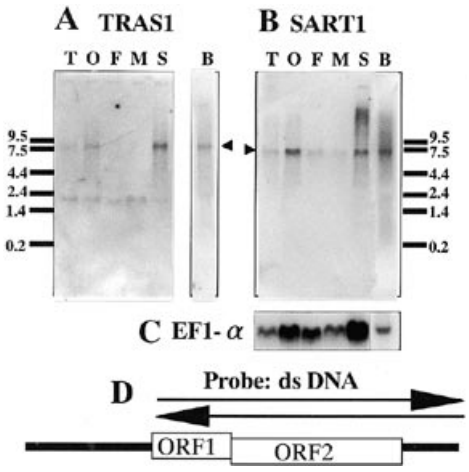

Figure 2. Northern hybridization. Aliquots of 5 µg of poly(A)$^+$ RNA from testis (T), ovary (O), fat body (F), malpighian tubule (M) and posterior silk gland (S) of fifth larva and from the BmN4 cultured cells (B) were hybridized with a TRAS1 probe (**A**), a SART1 probe (**B**) and an EF1-α probe (**C**). (**D**) The regions for the double-strand DNA probes used for TRAS1 and SART1 are shown by arrows (both ORFs and the 3′-UTR). The band around 2 kb in (A) is probably rRNA and was never seen reproducibly. Note that hybridization of RNAs from the BmN4 cells with the TRAS1 probe was conducted in a separate experiment and thus the expression level cannot be compared to those in other tissues.

Most genomic copies of TRAS1 and SART1 contained the transcription initiation sequences

Many non-sequence-specific, non-LTR elements are transcribed from their 5′-ends by downstream promoters (18–23). It is unclear, however, if this is the case in sequence-specific non-LTR elements. In the R1 element, in fact, a low level of read-through transcript from 28S rDNA was found (27). To investigate the transcription initiation mechanism of the sequence-specific elements, we amplified the 5′-ends of TRAS1 and SART1 transcripts (85 nt) from BmN4 cells by 5′-RACE. Although we used RNAs from BmN4 cells as a template, the same result may be expected from other tissues because there seemed to be no apparent RNA size differences among tissues and BmN4 cells (Fig. 2). Sequencing of 8 and 15 clones for TRAS1 and SART1,

A TRAS1

λ B1	(CCTAA)nCC CGAGTTCCCCCTCAGCTCT	
Inverse PCR	(CCTAA)nCC -GAGTTCCCCCTCAGCTCT	2/10
	(CCTAA)nCC --AGTTCCCCCTCAGCTCT	7/10
	(CCTAA)nCC --CGTTCCCCCTCAGCTCT	1/10
5'-RACE	GAGTTCCCCCTCAGCTCT	7/8
	AGTTCCCCCTCAGCTCT	1/8

B SART1

BS103	(TTAGG)nTT CCCGG-CCC--GGGACCTG	
Inverse PCR	(TTAGG)nT- -----CCCC--GGGACCTG	1/18
	(TTAGG)nTT ------CCCGAGGGACCTG	4/18
	(TTAGG)nTT ------CCC--GGGACCTG	13/18
5'-RACE	CCC--GGGACCTG	15/15

Figure 3. Sequence comparison of the 5′-ends of the transcripts of TRAS1 and SART1 with the counterparts in their genomic copies using 5′-RACE (25) and inverse PCR (26). The position of the 5′-phosphorylated primers used for reverse transcription in the 5′-RACE and the primers for the nested PCRs in the 5′-RACE and the inverse PCR are shown in Figure 1B. All the 5′-RACE clones started at the same position (with one base uncertainty in the case of TRAS1) and contained almost the same sequences through 85 nt. Most of their genomic copies (9/10 and 13/18 for TRAS1 and SART1, respectively) which were amplified by inverse PCR contained the transcription initiation sites.

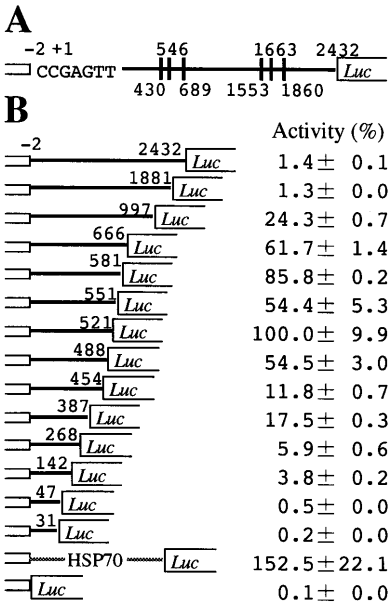

Figure 4. Transient luciferase activity of 3′-deletion derivatives of the TRAS1 5′-UTR in BmN4 cells. (**A**) Structure of the –2/2432 construct which contains all the 5′-UTR of TRAS1. The linker sequence and the firefly luciferase gene (*Luc*) are depicted as a gray box and a white box, respectively. This construct contains cytosine dinucleotides (CC) upstream of the transcription initiation sequence, GAGTT. The positions of ATG triplets that reduce translation of luciferase are indicated by vertical lines. (**B**) Luciferase activity of 3′-deletion derivatives of the TRAS1 5′-UTR. The activity for each experimental pGL3-Enhancer vector was normalized to those of the HSP70-containing pRL vector. The activity for –2/521 (maximal) is designated as 100% and values of the mean ± standard deviation for three independent experiments per construct are shown. HSP70 denotes the *Drosophila* HSP70 promoter that was used as a positive control. Activity for the pGL3-Enhancer vector alone is also shown below as a negative control.

respectively, showed that the 5′-ends of their RNAs started at their first nucleotide positions and did not contain any other sequences, such as telomeric repeats (Fig. 3). The sequences of the 5′-end 85 nt of the TRAS1/SART1 RNAs were identical to the counterparts of the genomic clones that we previously characterized, suggesting that the copies we previously isolated from the genomic library are functional.

Because the 5′-ends of non-LTR retrotransposons are usually variable, probably due to incomplete reverse transcription and aberrant integration processes (28), only a few copies of TRAS1 and SART1 may contain their transcription start sequences. To examine what proportions of their genomic copies retain the transcription initiation sequence, we amplified DNA sequences around their 5′-ends by inverse PCR (26). All the 5′-end clones of TRAS1 and SART1 obtained by inverse PCR were adjacent to the telomeric repeats and retained their same sequence specificities as shown in the genomic clones (Fig. 3). Seven of 10 TRAS1 clones and 13 of 18 SART1 clones contained the transcription start sites which were defined by 5′-RACE. These inverse PCR products may not represent all types of the TRAS1/SART1 copies because we could not detect some of the 5′-end sequences that were found in the previously isolated full-length elements (Fig. 3). These results, however, suggest that most of the genomic copies of TRAS1 and SART1 that exist in the telomeric repeats can be transcribed *in vivo*.

TRAS1 possesses a promoter in its 5′-UTR

The discovery that TRAS1 and SART1 contain their own transcription initiation sites prompted us to examine whether their promoters lie downstream of the initiation sites, as do many non-sequence-specific, non-LTR elements. We subcloned their 5′-UTRs into the pGL3-Enhancer vector, which contains the firefly luciferase gene as a reporter gene (Promega), and transiently transfected them into BmN4 cells. BmN4 cells are biologically relevant with regard to transcription of TRAS1 and SART1 as we proved by northern hybridization (Fig. 2) and 5′-RACE analysis (Fig. 3).

The luciferase assays showed that TRAS1 contained its own promoter activity in its 5′-UTR. The luciferase activities obtained from a series of 3′-deleted constructs of the 5′-UTR are shown in Figure 4. The –2/521 construct demonstrated the largest activity, which was designated as 100%. This activity was about two-thirds as strong as that of the *Drosophila* HSP70 promoter. The –2/454 and –2/551 constructs showed lower activity than the –2/387 and –2/521 constructs, respectively, probably because translation from the methionine codons at positions 430 and 546 decreased translation of luciferase. Longer constructs, such as –2/997, –2/1881 and –2/2432, showed decreased activities, probably because of translation from the methionine codons at positions 689, 1553, 1663 and 1860 or because of the difficulty of ribosomes scanning for long distances from the 5′-terminus of mRNAs. Taking these factors into account, we concluded that the sequences responsible for transcription of TRAS1 resided in the +1 to +581 interval. Within this portion, two regions were particularly important for the activity. The nucleotides +47/+142 and +454/+521 increased transcription approximately 7- and 8-fold, respectively.

We tried to detect promoter activity in the SART1 5′-UTR but the constructs, which contained all or parts of the SART1 5′-UTR, did not show greater activity than the pGL3-Enhancer vector alone (data not shown). We then examined the possibility that transcription of SART1 is promoted by the 3′-UTR of a neighboring SART1 copy, as is the case with the *Drosophila* HeT-A element (24), considering that at least half of all SART1 elements exist as tandem arrays interrupted by short (<100 bp)

I	CAGTACCACTTCAACCTCCGAAGAGATAAGTCGTGCCTCT
Jockey	CATTCGCATGGGAGATGAGCAATCGAGTGGACGTGTTCAC
F	GATTTCAATTCGATCGCCGACGTGTGAAGACGTTTTTATC
Doc	CACTCGTGGATTCGCATTCGAGATTCGCGGACGTGTTTCT
TRAS1	GAGTTCCCCCTCAGCTCTCGTGGCGGTCGGATCGTTTTGC
SART1	CCCGGGACCTGGGCGGGCCCCCCGGCGCGCACTCAGCGTG

Figure 5. Sequence comparison of the first 40 bp of non-LTR retrotransposons in insects for which transcription initiation sites have been determined. The initiator at the 5′-end and the downstream CGT(G/T) motif are underlined. TRAS1 contains an initiator-like sequence while SART1 does not.

telomeric repeats (13). The constructs which contain the 3′-UTR upstream of the telomeric repeats and the 5′-UTR of SART1 did not, however, show significant activity (data not shown).

Influence of the first 40 bp conserved in most insect retrotransposons on transcription of TRAS1

Most non-LTR retrotransposons that have been well characterized in insects have downstream promoters in their 5′-UTR. These elements have in common a TATA-less RNA polymerase II promoter, which have also been found in some LTR retrotransposons in insects (29), many *Drosophila* homeotic genes such as *Antennapedia* (30) and *Engrailed* (31) and the mammalian terminal deoxynucleotidyltransferase (TdT) gene (32), and so on. These retrotransposons share two conserved sites in the 5′-terminal regions (29). The initiator [the (C/G)A(C/G/T)T motif] is located around the transcription initiation site and the CGT(G/T) motif is found at position +31 or +32 (Fig. 5). We compared the first 40 bp of TRAS1 and SART1 with those of some non-LTR elements in insects for which transcription initiation sites have been determined. Although transcription of TRAS1 was not as strongly dependent on the first 40 bp as were the *Drosophila* retrotransposons (Fig. 4), TRAS1 similarly contained the two conserved sites in this region. These two sites were also conserved in the corresponding regions of another member of the TRAS family (TRAS3), those regions retaining the same sequence specificity as TRAS1 (data not presented), suggesting involvement of the two sites in transcription of the TRAS family. On the other hand, SART1 did not contain an initiator-like sequence and only had the CGT(G/T) motif at position +36, which was 4 bp downstream of the position of other elements. These results suggest that an initiator is involved in transcription of TRAS1 in some way, while SART1 has a somewhat different transcription initiation mechanism.

To examine the importance of an initiator on transcription of TRAS1, we created a series of 5′-deleted TRAS1 5′-UTR constructs and performed luciferase assays (Fig. 6). To our surprise, removing the first 16 bp did not significantly inhibit transcription of TRAS1. Deletion of the first 43 bp, the counterparts of which have been shown to be crucial in transcription of many *Drosophila* retrotransposons (18–21), reduced transcription by only 40%. This result indicates that the first 40 bp of TRAS1 are important but not indispensable for the activity. Deletion of 126 bp decreased the activity by 100-fold. Combined with the observation that nucleotides +1/+142 showed 7-fold stronger activity than +1/+47 (Fig. 4), the results from these deletions indicate that nucleotides +47/126 are essential for activity.

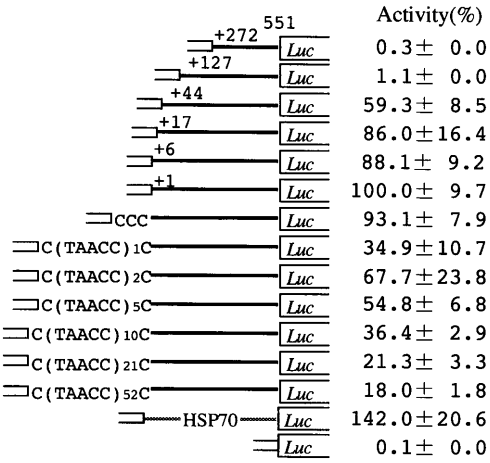

Figure 6. Luciferase activity of 5′-variation derivatives of the TRAS1 5′-UTR in BmN4 cells. The activity for +1/551 (maximal) is taken as 100% and values of the mean ± standard deviation for three independent experiments per construct are shown. The telomeric repeats are shown as $(CCTAA)_n$.

The neighboring telomeric repeats inhibit transcription of TRAS1

TRAS1 contained an internal promoter in its 5′-UTR as described above. There is a possibility, however, that the neighboring telomeric repeats also have some influence on transcription of TRAS1. To examine this possibility, we constructed a series of plasmids that contained various short lengths of the telomeric repeats upstream of the first 551 bp of the TRAS1 element and performed luciferase assays. As shown in Figure 6, placing the telomeric repeat upstream of TRAS1 decreased the promoter activity. Constructs with longer telomeric repeats appeared to repress transcription more strongly. For example, insertion of 52 telomeric repeats reduced transcription by 82% while insertion of five repeats decreased it by 45%. These results suggest that transcription of TRAS1 is down-regulated *in vivo* according to the length of the neighboring telomeric repeats.

DISCUSSION

Previous structural analyses, which showed the existence of undisrupted ORFs, a few truncated copies and sequence conservation among genomic copies, suggested that most copies of TRAS1 and SART1 in the genome have the ability to transpose via RNA (12,13). In this paper, we provide additional evidence for this idea by showing that TRAS1 and SART1 are actually expressed in many tissues and a cultured cell line as approximately full-length RNAs, which can serve not only as messengers for translation but as templates for target-primed reverse transcription. Their expression in testis and ovary tissues and in BmN4 cells suggests that they have been surviving by retrotransposition through vertical transmission and/or that they play some important role in telomere maintenance. Although most non-LTR retrotransposons show various lengths of transcripts due to 5′-truncation and read-through transcription from neighboring genes, only single sizes of the TRAS1/SART1 RNAs were found. This is probably because only a few of their genomic copies are truncated and because the TRAS1 and SART1 copies are not transcribed

from adjoining genes, being situated exclusively in the telomeric repeats. TRAS1 and SART1 are the first examples of sequence-specific retrotransposons that are actively transcribed *in vivo*. Although sequence-specific retrotransposons (R1 and R2) may be transcribed as read-through RNAs from neighboring rRNA genes (27,33), TRAS1 and SART1 were expressed as unit length RNAs. This may reflect differences in genomic localization between the rRNA genes and non-transcribed telomeric repeats.

Major questions regarding transcription of TRAS1 and SART1 are concerned with whether the copies within the telomeric repeats are transcribed and whether they possess internal promoters. The data presented here suggest that most TRAS1 elements within the telomeric repeats are weakly transcribed *in vivo* by internal promoters. First, the TRAS1 RNA that was weakly detected on the northern blot corresponds in size (~8 kb) to the previously cloned TRAS1 element (TRAS1-λB1) that was adjacent to the telomeric repeats (12). Second, 5′-RACE analysis showed that the 5′-end of the TRAS1 RNA started precisely from the 5′-end of TRAS1-λB1 and that the nucleotide sequence of the RNA was completely identical over 85 nt to that of a genomic TRAS1 element. Third, we have never found, by inverse PCR, a TRAS1 element that was inserted outside the telomeric repeats (Fig. 3). Finally, the 5′-UTR of TRAS1 contained a promoter activity for its own transcription (Fig. 4). Most of the 600 TRAS1 copies may be transcribed weakly *in vivo* if they are, like SART1, adjacent to less than 20 units of the telomeric repeats, transcription being reduced in such instances by 20–80% (13; Fig. 6). It is uncertain, however, if the amount of transcription we detected in this promoter assay could account for that we discovered by northern hybridization. On the other hand, it is still not known from which copies the SART1 RNA is transcribed. The results of northern hybridization, 5′-RACE and inverse PCR showed that the SART1 transcript had approximately the same length, the same 5′-end nucleotide sequence and the identical 5′-end position as the most common copy within the telomeric repeats, as was the case with TRAS1. We were unable, however, to detect any promoter activity in the 5′-UTR of SART1 and its 5′-end lacked an initiator motif, which was found in most insect TATA-less RNA polymerase II promoters, including TRAS1. The 3′-UTR of SART1 also did not promote transcription of the tandem neighboring SART1 copy, as mentioned above. A possible explanation may be that there are some SART1 master copies that are transcribed by an external promoter, although we were not able to detect such a genomic copy by inverse PCR. Since SART1 lacks an initiator-like sequence, the SART1 RNA we detected might also have been processed from the primary transcript, which contains a promoter and an initiator-like sequence. Another possibility is that some part of the SART1 element possesses a promoter, but we were unable to detect it in the experimental system presented here.

Internal promoters have already been identified for some non-sequence-specific, non-LTR retrotransposons such as Jockey (18), I (19), F (20) and Doc (21) from *D.melanogaster*, the human LINE-1 element (22,23) and DRE (34), which integrates upstream of *Dictyostelium discoideum* tRNA genes with position specificity but not sequence specificity. This study showed that a sequence-specific, non-LTR element TRAS1 also comprised an internal promoter, as do many non-sequence-specific elements. Transcription of TRAS1 occurred independently of the telomeric repeats, which rather inhibited transcription. Non-LTR retrotransposons may be autonomous elements with regard to transcription and be independent of surrounding environments. This result does not necessarily mean, however, that all sequence-specific elements are transcribed by internal promoters. For example, R1 and R2 may be transcribed as one unit with rRNA (27,33).

TRAS1 possesses an initiator-like sequence at its 5′-end as do the retrotransposons I, Jockey, F and Doc. In *Drosophila* retrotransposons I, F and Doc, the first 40 bp segment contains all the promoter activity in cultured *Drosophila* S2 cells (19–21). Further, the sequences downstream of the first 40 bp regulate the tissue specificity in the case of I (35) and Fex (36). On the other hand, eliminating the first 43 bp of TRAS1 did not completely eliminate the promoter activity (Fig. 6) and the nucleotides +47/+142 and +454/+521 were crucial for activity. This difference for the promoter regions may reflect the fact that TRAS1 has a much longer 5′-UTR (2432 bp) than do the *Drosophila* retrotransposons (<300 bp). These results suggest that TRAS1 utilizes a quite different mechanism for transcription. Conservation of the initiator sequence in TRAS1 and another member of the TRAS family, however, suggests involvement of this sequence for the purpose of their transcription by RNA polymerase II. An initiator may be important in TRAS1 for accurate initiation rather than for promoter activity itself.

Drosophila depends for maintenance of its telomeres upon retrotransposons (2,3), which have an intimate evolutionary association with telomerase (37). The fact that TRAS1 and SART1 are actively transcribed retrotransposons suggests the importance of retrotransposition of the TRAS/SART families on the evolution of insect telomeres. One can imagine that some insects lost the ability to produce telomerase at some point during their evolution, but maintained telomere length by insertion of TRAS/SART-like elements into the telomeric repeats. Various mechanisms such as addition of retrotransposons onto the very ends of telomeres and recombinational events may have replaced telomerase activity. Since so far we have not succeeded in detecting telomerase activity (Y.Sasaki and H.Fujiwara, unpublished data), telomerase might have been lost and an alternative mechanism might have arisen in *B.mori*, although some other insects may have telomerase. Functional analysis of the TRAS/SART families may provide insight into the mechanism of telomere maintenance in the silkworm.

ACKNOWLEDGEMENTS

We are grateful to Dr H. Maekawa for providing a plasmid including a *Drosophila* HSP70 promoter and to Dr T. Shimada for helping to use the luminometer. This work was supported by grants from the Ministry of Education, Science and Culture of Japan (MESCJ) and Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists of MESCJ.

REFERENCES

1 Greider,C.W. (1995) In Blackburn,E.H. and Greider,C.W. (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 35–68.
2 Biessmann,H., Mason,J.M., Ferry,K., d'Hulst,M., Valgeirsdottir,K., Traverse,K.L. and Pardue,M.L. (1990) *Cell*, **61**, 663–673.
3 Levis,R.W., Ganesan,R., Houtchens,K., Tolar,L.A. and Sheen,F.M. (1993) *Cell*, **75**, 1083–1093.
4 Cohn,M. and Edstrom,J.E. (1992) *J. Mol. Evol.*, **35**, 114–122.
5 Roth,C.W., Kobeski,F., Walter,M.F. and Biessmann,H. (1997) *Mol. Cell. Biol.*, **17**, 5176–5183.
6 Lundblad,V. and Blackburn,E.H. (1993) *Cell*, **73**, 347–360.

7 Blasco,M.A., Lee,H.W., Hande,M.P., Samper,E., Lansdorp,P.M., DePinho,R.A. and Greider,C.W. (1997) *Cell*, **91**, 25–34.
8 Higashiyama,T., Noutoshi,Y., Fujie,M. and Yamada,T. (1997) *EMBO J.*, **16**. 3715–3723.
9 Biessmann,H. and Mason,J.M. (1997) *Chromosoma*, **106**, 63–69.
10 Zakian,V.A. (1997) *Cell*, **91**, 1–3.
11 Okazaki,S., Tsuchida,K., Maekawa,H., Ishikawa,H. and Fujiwara,H. (1993) *Mol. Cell. Biol.*, **13**, 1424–1432.
12 Okazaki,S., Ishikawa,H. and Fujiwara,H. (1995) *Mol. Cell. Biol.*, **15**, 4545–4552.
13 Takahashi,H., Okazaki,S. and Fujiwara,H. (1997) *Nucleic Acids Res.*, **25**, 1578–1584.
14 Xiong,Y. and Eickbush,T.H. (1988) *Mol. Cell. Biol.*, **8**, 114–123.
15 Besansky,N.J., Paskewitz.S.M., Hamm,D.M. and Collins,F.H. (1992) *Mol. Cell. Biol.*, **12**, 5102–5110.
16 Feng,Q., Schumann,G. and Boeke,J.D. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 2083–2088.
17 Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (1993) *Cell*, **72**, 595–605.
18 Mizrokhi,L.J., Georgieva,S.G. and Ilyin,Y.V. (1988) *Cell*, **54**, 685–691.
19 McLean,C., Bucheton,A. and Finnegan,D.J. (1993) *Mol. Cell. Biol.*, **13**, 1042–1050.
20 Minchiotti,G. and Di Nocera,P.P. (1991) *Mol. Cell. Biol.*, **11**, 5171–5180.
21 Contursi,C., Minchiotti,G. and Di Nocera,P.P. (1995) *J. Biol. Chem.*, **270**, 26570–26576.
22 Swergold,G.D. (1990) *Mol. Cell. Biol.*, **10**, 6718–6729.
23 Minakami,R., Kurose,K., Etoh,K., Furuhata,Y., Hattori,M. and Sakaki,Y. (1992) *Nucleic Acids Res.*, **20**, 3139–3145.
24 Danilevskaya,O.N., Arkhipova,I.R., Traverse,K.L. and Pardue,M.L. (1997) *Cell*, **88**, 647–655.
25 Maruyama,I.N., Rakow,T.L. and Maruyama,H.I. (1995) *Nucleic Acids Res.*, **23**, 3796–3797.
26 Triglia,T., Peterson,M.G. and Kemp,D.J. (1988) *Nucleic Acids Res.*, **16**, 8186.
27 Long,E.O. and Dawid,I.B. (1979) *Cell*, **18**, 1185–1196.
28 George,J.A., Burke,W.D. and Eickbush,T.H. (1996) *Genetics*, **142**, 853–863.
29 Arkhipova,I.R. and Ilyin,Y.V. (1991) *EMBO J.*, **10**, 1169–1177.
30 Perkins,K.K. and Dailey,G.M. and Tjian,R. (1988) *Genes Dev.*, **2**, 1615–1626.
31 Soeller,W.C., Poole,S.J. and Kornberg,T. (1988) *Genes Dev.*, **2**, 68–81.
32 Smale,S.T. and Baltimore,D. (1989) *Cell*, **7**,103–113.
33 Luan,D.D. and Eickbush,T.H. (1996) *Mol. Cell. Biol.*, **16**, 4726–4734.
34 Schumann,G., Zundorf,I., Hofmann,J., Marschalek,R. and Dingermann,T. (1994) *Mol. Cell. Biol.*, **14**, 3074–3084.
35 Udomkit,A., Forbes,S., McLean,C., Arkhipova,I. and Finnegan,D.J. (1996) *EMBO J.*, **15**, 3174–3181.
36 Kerber,B., Fellert,S., Taubert,H. and Hoch,M. (1996) *Mol. Cell. Biol.*, **16**, 2998–3007.
37 Nakamura,T.M. and Cech,T.R. (1998) *Cell*, **92**, 587–590.