Sequence-Specific Recognition and Cleavage of Telomeric Repeat (TTAGG)$_n$ by Endonuclease of Non-Long Terminal Repeat Retrotransposon TRAS1

TOMOHIRO ANZAI, HIDEKAZU TAKAHASHI, AND HARUHIKO FUJIWARA*

Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

Received 13 March 2000/Returned for modification 4 May 2000/Accepted 13 October 2000

The telomere of the silkworm Bombyx mori consists of (TTAGG/CCTAA)$_n$ repeats and harbors a large number of telomeric repeat-specific non-long terminal repeat retrotransposons, such as TRAS1 and SART1. To understand how these retrotransposons recognize and integrate into the telomeric repeat in a sequence-specific manner, we expressed the apurinic-apyrimidinic endonuclease-like endonuclease domain of TRAS1 (TRAS1 EN), which is supposed to digest the target DNA, and characterized its enzymatic properties. Purified TRAS1 EN could generate specific nicks on both strands of the telomeric repeat sequence between T and A of the (TTAGG)$_n$ strand (bottom strand) and between C and T of the (CCTAA)$_n$ strand (top strand). These sites are consistent with insertion sites expected from the genomic structure of boundary regions of TRAS1. Time course studies of nicking activities on both strands revealed that the cleavages on the bottom strand preceded those on the top strand, supporting the target-primed reverse transcription model. TRAS1 EN could cleave the telomeric repeats specifically even if it was flanked by longer tracts of nontelomeric sequence, indicating that the target site specificity of the TRAS1 element was mainly determined by its EN domain. Based on mutation analyses, TRAS1 EN recognizes less than 10 bp around the initial cleavage site (upstream 7 bp and downstream 3 bp), and the GTTAG sequence especially is essential for the cleavage reaction on the bottom strand (5'...TTAGGT T↓AG...3'). TRAS1 EN, the first identified endonuclease digesting telomeric repeats, may be used as a genetic tool to shorten the telomere in insects and some other organisms.

* Corresponding author. Mailing address: Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan. Phone and Fax: 81-3-5841-4447. E-mail: haruh@k.u-tokyo.ac.jp.

The extreme ends of the silkworm chromosome are composed of the simple telomeric repeat (TTAGG)$_n$ (26). When studying the telomeric structure of the silkworm, we found that more than 2,000 copies of non-long terminal repeat (non-LTR) retrotransposons have accumulated within the telomeric repeat (25). These Bombyx mori retrotransposons are sequence-specific elements that have been classified by differences in insertion sites and in amino acid sequences into two distinct, large families called TRAS (telomeric repeat-associated sequence) and SART (32). TRAS1 is the best-characterized family among the TRAS group (25). Most of the 250 copies of TRAS1 cluster at the telomeres of dozens of chromosomes and are highly conserved in structure without 5'-end truncation. The 5’ end of a 7.85-kb stretch of TRAS1 is precisely adjacent to CC of the (CCTAA)$_n$ telomeric strand ([CCTAA],-CC-[5’ untranslated region (UTR)]). TRAS1 is transcribed in many tissues, and its transcription starts just after the 5’ end under the control of internal promoter in the 5’ UTR (31). These facts suggest that TRAS1 is an active retrotransposon. All TRAS1s analyzed so far terminate with oligo(A). We cannot define the exact junction site at the 3’ end because this oligo(A) sequence connects directly to CC of the (CCTAA) strand (Fig. 1A). Based on structural analysis of junction regions of TRAS1, it is also hard to identify target site duplication (TSD) of the telomere sequence, which is generated in the process of retrotransposition, because the target of TRAS1, the telomeric repeat (TTAGG/CCTAA) itself, is repetitious. To identify the exact junction site and TSD structure for TRAS1, the target specificity of TRAS1 must be studied by another functional assay.

Non-LTR retrotransposons are a very abundant family and are found in most eukaryotes (11). Some of them are inserted within specific sites on the chromosome. Tx1 of Xenopus laevis inserts within another transposable element (13). CRE1, SLACS, and CZAR are found in the spliced leader exons of trypanosomes (1, 12, 34). R1 and R2 are located at specific sites of 28S rDNA in most insects (7, 10, 16). In two mosquito species, RT1 and RT2 are inserted at the same position about 630 bp downstream of the R1 insertion site (2, 27). Phylogenetic analysis showed that TRAS1 and SART1 are most similar to the insect 28S rDNA-specific retrotransposons R1 and RT1, respectively (21, 32). Presumably, the target specificity of these elements might have changed from rDNA to telomeric repeat or vice versa during insect evolution.

Recent reports have shown that two different types of endonuclease encoded in non-LTR retrotransposons function to define and cleave target site DNA. R2 and some other elements such as CRE1 and CZAR have a single open reading frame (ORF) that encodes an endonuclease domain near its C-terminal end. This type of endonuclease is similar in some residual motifs to various prokaryotic restriction endonucleases (38). The R2 protein first makes a specific nick on the noncoding bottom strand. Reverse transcription of the RNA template is then primed by the exposed 3'-hydroxyl group, a process termed target-primed reverse transcription (TPRT),...
Before cleavage of the coding top strand (19). The second type of endonuclease is an apurinic-apyrimidinic endonuclease (APE)-like protein encoded in the N terminus of ORF2 of many non-LTR retrotransposon groups, such as L1, R1, TxL, and TRAS/SART elements, which have two ORFs. The APE-like domain of the human non-sequence-specific retroposon L1 preferentially cleaves the sequence similar to in vivo target sequences (6, 9). It has also been reported that R1 and TxL endonuclease specifically cleaves 28S rDNA and transposable element (Tx1D) sequence at the precise target site in vitro (5, 8), respectively.

To understand how TRAS1 recognizes and cuts the 5-bp repetitive telomeric repeats (TTAGG/CCTAA), we have expressed and purified the endonuclease (EN) domain of TRAS1 in Escherichia coli and characterized TRAS1 EN activities. Digestion patterns of end-labeled (TTAGG), or (CCTAA), by TRAS1 EN were clear 5-bp ladder signals, demonstrating that both telomeric strands were cleaved at specific sites. Various mutations on the substrate DNA also revealed that some sequences in the telomeric repeat are involved in the target-specific digestion by TRAS1 EN. This is the first evidence of a telomere-specific endonuclease. This endonuclease, which specifically digests and shortens the telomeric repeat, may thus be used as a genetic tool to study telomere function.

**MATERIALS AND METHODS**

Cloning of TRAS1 EN in the expression vector. The TRAS1 EN domain was amplified by PCR with Pfu Turbo DNA polymerase (Stratagene) using primers s788 (‘5’-AAAAAAAACATATGCACCGGACGACTGGA-3’) and a5532 (‘5’-AAAAAAACATATGCTATTTTGGGACGTCCAATATGGAT ACCATACC-3’) from the XblI clone as a template (25). Each primer was designed to contain Ndel and Xhol sites, respectively (underlined in the primers above). The PCR product was digested with Ndel and Xhol and cloned into the Ndel and Xhol sites of the pET16b expression vector (Novagen), and the insert sequence of a candidate clone was confirmed. The cloned plasmid, named pHs1Tien, includes the first 741 bp, corresponding to the first 247 amino acids of ORF2, as shown in Fig. 1B, and the sequence encoding an MGHHHHHHHHHHHHHHHHHHHHHHHHH tag on the N terminus of TRAS1 EN. This clone was designed to stop translation at the bases of TTA, which is a complement to the stop codon TAA (bold letters in the a5532 primer above). A point mutation in the putative active site of EN (His→Ala at 258) was generated by a Quick Change site-directed mutagenesis kit (Stratagene). The primers used for constructing this mutant clone, named pHs1Tien (H258A), were s747 (‘5’-CGATGAAAGACCATGCTGCTCCGAGTCCAAAAGGATGATG G-3’) and a784 (‘5’-CCATACCGTGGCAGCTCATTCGACTTCGTCGTA CG-3’).

Expression and purification of TRAS1 EN. The pHs1Tien and pHs1Tien (H258A) DNAs were transformed into E. coli strain BL21(DE3)/pLysS (Nova- gen). A 50-ml culture of transformant was grown at 37°C until the optical density at 600 nm (OD600) reached 0.6. Isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM, and incubation was continued for 8 h at 25°C. An accumulating protein was run on sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and detected as a band with a molecular mass of 30.3 kDa. We confirmed that this band corresponds to the target protein by Western blot analysis with α-(His)6 antibody (Boehringer Mannheim). Cells were pelleted by centrifugation and stored at −80°C. Purification processes followed the protocol of the His- Bind kit from Novagen (catalog number 70156). Freeze-thawed cell pellets were suspended in 6 ml of binding buffer (5 mM imidazol, 500 mM NaCl, 20 mM Tris-HCl [pH 7.9]), and Triton X-100 was added to a final concentration of 0.1% and incubated for 10 min at 0°C. After the cell extracts were centrifuged at 39,000 × g for 20 min at 4°C, the supernatants containing soluble proteins were filtered through a 0.45-μm membrane (Millipore) twice and applied to a Quick 900 Cartridge (Novagen). The cartridge was washed with 20 ml of binding buffer and subsequently with 10 ml of washing buffer (50 mM imidazol, 500 mM NaCl, 20 mM Tris-HCl [pH 7.9]). Most of the target proteins were eluted with 2 ml of elution buffer (300 mM imidazol, 500 mM NaCl, 20 mM Tris-HCl [pH 7.9]). The eluted proteins were ultracentrifuged in storage buffer (500 mM NaCl, 50 mM Tris-HCl [pH 7.9], 20% glycerol, 10 mM 2-mercaptoethanol) with an Ultrafree-MC centrifugal filter unit (Millipore). The concentrated proteins (about 2 ml) were stored at −80°C in 10-μl aliquots. The concentration of the eluted proteins was determined to be approximately 0.5 mg/ml by comparing the intensity of the band on a Coomassie blue-stained SDS-PAGE gel with that of known amounts of bovine serum albumin (BSA) protein.

**Assays for endonucleolytic activities.** To generate the substrates for TRAS1 EN, oligonucleotides (Nishinbo) were end labeled with T4 polynucleotide kinase (Toyobo) and [γ-32P]ATP (ICN). In the assay for site specificity shown in Fig. 5, 60 bp of nontelomeric sequence was randomly selected from pBR322

FIG. 1. Telomeric sequence-specific retrotransposons of the silkworm, B. mori. (A) Chromosome structure of B. mori. The shaded triangle indicates the telomeric repeat sequence 5’TETAGG/5’-CCTAA and the positions of the ORFs of TRAS1 are depicted by open boxes, and the positions of the promoter, EN domain, and RT domain are indicated. (B) Structure of TRAS1. Arrows show the directions of transcription by these elements. The 5’ and 3’ junction regions between TRAS1 and the telomeric repeat are shown below. The precise 3’ end of TRAS1 has not been determined. Dotted boxes indicate possible junction sites.
sequence (accession number J01749, positions 81 to 140) (see Fig. 5C). The 5' end-labeled oligonucleotide was annealed with the nonlabeled complementary strand by incubation at 95°C for 2 min, followed by gradual cooling to room temperature. The samples were separated on 28% polyacrylamide gels, and the location of a precisely annealed oligonucleotide was detected by autoradiography. The band was cut out from the gel, incubated in elution buffer (500 mM ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, and 0.1% SDS) at 37°C for 4 h, and ethanol precipitated. The purified double-stranded oligonucleotides were dissolved in 0.5 ml of TE buffer (1 mM EDTA, 10 mM Tris-HCl [pH 7.6]) and stored at −30°C.

To determine the optimal conditions for assay of TRAS1 EN, we quantified the nicking activities for the (TTAGG)5, substrate by measuring the radioactivity of a specifically cleaved product at a 17-base position, (TTAGG)5-TT (see Results). Radiolabeled DNA substrate (1 ng) was cleaved most efficiently (approximately 13% of substrate DNA cleaved) with 0.2 μg of TRAS1 EN DNA (data not shown). Optimum cleavage of substrates was observed at 25°C, and the reaction at 37°C showed a great decrease in activity (approximately 7% of the activity at 25°C). The reaction buffer for TRAS1 EN determined to show the highest endonucleolytic activity for oligonucleotide substrates contained 50 mM PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid)]-HCl (pH 6.0), 10 mM NaCl, 2 mM MgCl2, and BSA (100 μg/ml). Substrate DNA was treated with purified TRAS1 EN protein in 10 μl of reaction buffer for 60 min at 25°C. The reaction was stopped by addition of EDTA to a final concentration of 50 mM. The reaction mixture was denatured in a loading buffer containing 75% formamide for 5 min at 95°C, immediately chilled on ice, and run on a 28% polyacrylamide denaturing sequencing gel. Quantitation of the reaction products was carried out with BAS 2500 imaging analyzer system (Fujifilm).

Various-sized oligonucleotides for telomeric repeat were end labeled with T4 polynucleotide kinase and [γ-32P]ATP and used as precise molecular size markers: dG10 (5'-TTAGGTTAGG-3'), dG12 (5'-TTAGGTTAGGTT-3'), dG15 (5'-TTAGGTTAGGTTAGG-3'), dG17 (5'-TTAGGTTAGGTTAGGTT-3'), dC10 (5'-CCTAACCTAACC-3'), dC12 (5'-CCTAACCTAACCCT-3'), dC15 (5'-CCTAACCTAACCCT-3'), and dC17 (5'-CCTAACCTAACCCTT-3'). The sequences of molecular size markers b42, b37, b32, t42, t37, and t32 are shown in Fig. 5.

**RESULTS**

**TRAS1 EN has a specific nicking activity for telomeric repeats.** Sequence-specific retrotransposons generate a nick in the target DNA using their encoded endonucleases (5, 6, 8, 37). TRAS1 EN encodes an AP endonuclease-like domain, located upstream of the reverse transcriptase (RT) domain (Fig. 1B). Based on amino acid sequence alignment with other non-LTR retrotransposons, we have defined the EN domain of TRAS1 without excluding any of the proposed catalytic motifs (9) (Fig. 1B). TRAS1 EN was expressed in E. coli with an N-terminal (His)10 tag and purified by nickel chelation chromatography. A single band of the predicted size, 30.3 kDa, was observed in SDS-PAGE (Fig. 2A).

To investigate the target site specificity of TRAS1 EN, we introduced a sensitive assay, using double-stranded oligonucleotide substrate: 5'-TTAGG/TTAGG-3'. The 5' end of (TTAGG)5 (G-strand) or (CCTAA)5 (C-strand) was radiolabeled and annealed with the nonlabeled complementary oligonucleotide of the same length. In all experiments, only one strand of the double-stranded substrate was end labeled. The optimal conditions for endonucleolytic activity of the TRAS1 EN protein were determined for pH, salt concentration, and temperature (see Materials and Methods). When 0.2 μg of TRAS1 EN was added to the reaction mixture with end-labeled (TTAGG)5, ladder bands representing cleavage sites were clearly observed at intervals of 5 bp (Fig. 2B). Similar ladder patterns were also detected for various lengths of substrates (n = 6, 9, and 12; data not shown). These results demonstrate that TRAS1 EN cleaves the telomeric repeats at specific sites in vitro.

We have also expressed a mutated protein, H258A, which has a single-amino-acid substitution (His→Ala) (Fig. 1B) to test whether these nicking activities indeed result from TRAS1 EN itself. The histidine residue mutated in H258A is known to be essential for the catalysis of the L1 EN domain (9), exonuclease III (24), and DNase I (30). The H258A endonuclease could not make any nicks on the telomeric repeats (Fig. 2B), suggesting that other endonucleolytic activities from E. coli did not contaminate the purified protein. Therefore, we conclude that TRAS1 EN itself is responsible for the sequence-specific nicking activity on the TTAGG/CCTAA repeats.

Telomeres are known to have a longer G-strand than the complementary C-strand, creating the single-stranded 3' overhanging structure at the end of the chromosome (15, 36). To exclude the possibility that TRAS1 EN associates with the telomeric G-strand overhang, we tested a single-stranded telomeric repeat as a substrate (Fig. 2C). We did not observe any ladder patterns for the 5'-end-labeled (TTAGG)5, indicating that TRAS1 EN has no nicking activity on the single-stranded telomeric repeats. To confirm that the missing activity on the single strand is not due to buffer conditions, we conducted another experiment as follows. After the TRAS1 EN reaction on the single-stranded substrate was complete, its complementary strand was readded and annealed. Under these conditions, we retracted the cleavage reaction and observed 5-bp ladder bands (data not shown). This demonstrates that the absence of activity in the single-stranded reactions could be reconstituted by adding the complementary strand to the reaction. These results shown above strongly suggest that the target of TRAS1 EN in vivo is the telomeric region forming double-stranded DNA.
nucleotides. Reactions with TRAS1 EN were conducted as in Fig. 2 (CCTAA)₆ (top strand) were annealed with the complementary oligo-
dG₁₇, (TTAGG)₃TT] that were end labeled with ³²P at the 5-
position of (TTAGG)
DNA products with those of the (TTAGG)
n₂; dG₁₂, (TTAGG)₂TT; dG₁₅, (TTAGG)₃; and
and run on PAGE alongside four DNA size markers [dG₁₀,
strand (G-strand) was cleaved by TRAS1 EN (as shown above)
CCTAA)
products were specifically cleaved between C and T of (CCTAA)₆
(Fig. 3B). In the experiments shown in Fig. 3, twofold the 
amount of TRAS1 EN (0.4 µg) was added. An excess amount of TRAS1 EN seemed to produce nonspecific bands, which were helpful for identifying 5-bp intervals of the main bands.

Previous studies showed that the 5’ end of TRAS starts
between C and T of the CCTAA top strand (Fig. 1A) (25, 31). The specific cleavage between C and T of the top CCTAA strand in vitro shown here is consistent with the 5’ junction site of TRAS1 in the silkworm genome. We could not, however, define the exact TRAS1-(CCTAA) junction site at the 3’ end, since the oligo(A)-oligo(T) sequence connects directly to the telomeric (CCTAA/TTAGG) sequence, proposing three possible 3’ junction structures: (TTAGG)ₙ-T-(T)ₙ₋₁(3’UTR), (TTAGG)ₙ-T-(T)ₙ₋₂(3’UTR), and (TTAGG)ₙ-(T)ₙ₋₁(3’UTR) (Fig. 1A). The result shown above indicates that TRAS1 cleaves its target sequence in the T-A junction on the (TTAGG)ₙ bottom strand, suggesting that the 3’ junction structure between TRAS1 and the (TTAGG)ₙ repeats in the genome is (TTAGG)ₙ-(T)ₙ₋₂(3’UTR).

TRAS1 EN cleaves the G-strand before the C-strand. To
determine whether TRAS1 EN has a strand preference for
nicking, we examined the time course of TRAS1 EN endonu-
cleolytic activities for the telomeric C-strand and G-strand.
The double-stranded oligonucleotides, including only one end-
labeled strand, were incubated with 0.2 µg of TRAS1 EN 
protein for 60 min. The changing patterns of 5-bp ladders 
during a 60-min reaction are shown in Fig. 4A and B. Cleaved
products from the C-strand gradually increased over the course 
of 60 min. The cleavage reaction on the G-strand seemed to
reach a maximum within the first 30 min, indicating that cleav-
age on the G-strand precedes that on the C-strand. To dem-
ostrate this more clearly, the amounts of nicked substrates 
were quantified with the BAS2500 imaging analyzer and plotted 
against the reaction time (Fig. 4C). The G-strand was preferentially cleaved before the C-strand. Judging from the
insertion structure of the TRAS1 element, the bottom and top 
strands correspond to the G- and C-strands, respectively.

Interpretation of the TPRT model (19) suggests that cleavage on the bottom strand generates an exposed 3’-hydroxyl that serves as a primer for reverse transcription of the RNA template prior to the second-strand cleavage. Thus, cleavage of the G-strand before the C-strand is consistent with the proposed TPRT model.

Target-specific cleavage of a long DNA substrate by TRAS1 
EN. All genomic copies of TRAS1 identified so far exist in the
specific site of the telomeric repeat (25). To investigate
whether this site specificity of TRAS1 for the telomere is defined by the EN domain itself, we assayed the nicking activities of TRAS1 EN for a long DNA substrate including both telomeric and nontelomeric sequences. The substrate contained three repeats of telomeric sequence (15 bp) flanked by 30 bp of nontelomeric sequences at both ends, which were randomly selected from pBR322 (Fig. 5C). Based on comparison of sizes of bands with several 32P-labeled oligonucleotides (b32, b37, b42, t32, t37, and t42), major bands were thought to be produced by specific cleavage within the telomeric sequence but not nontelomeric sequences. The TTAGG bottom strand was mainly cleaved between T and A of the telomeric unit at positions 37 and 42 from the 5' end (shown as solid arrowheads) (Fig. 5A). The CCTAA top strand was cleaved between C and T at position 32 (open arrowhead in Fig. 5B). These cleaved products gradually increased during the 60-min reaction. It was also shown that the top-strand cleavage reaction occurred more slowly, compatible with the TPRT model. The observation that other probable target sites in the (TTAGG) unit (b32, t37, and t42) were not cleaved may reflect the loss of essential sequences recognized by TRAS1 EN around them (see below).

Several minor bands were observed in both bottom- and top-strand reactions. The nontelomeric sequence adjacent to the 3' end of the telomeric unit (5'-[TTAGG]-[CCCT]-3') is similar to TTA of the (TTAGG) unit and may cause the minor cleavages downstream of b42 in the bottom-strand reaction. The nontelomeric sequences, because these bands were also observed before the reaction with TRAS1 EN (time zero). When we used another long substrate including (TTAGG)$_5$ surrounded by 125 bp of nontelomeric sequences at both ends, TRAS1 EN cleaved DNA only within the telomeric repeats at the same cleavage patterns shown above (data not shown). Thus, we conclude that the EN domain of TRAS1 mainly determines the target specificity of TRAS1 elements for the telomere sequence.

Sequence involved in target site specificity by TRAS1 EN. To investigate how TRAS1 EN protein recognizes a target site, we made several mutated substrates for the bottom strand and compared TRAS1 EN nicking activities for each substrate on PAGE. We first altered one nucleotide of the TTAGG unit to a cytosine (C) residue and made five different substrates (CTAGG)$_5$, (TCAGG)$_5$, (TTCGG)$_5$, (TTACG)$_5$, and (TTAGC)$_5$ (mutated nucleotides are underlined) (Fig. 6). When the first two thymines of TTAGG were mutated, the nicking reaction of TRAS1 EN itself was not disturbed, although nonspecific bands could be observed. This suggests that the first and second T's of the TTAGG unit are important for cleavage site definition by TRAS1 EN. When the third A and the fourth and fifth G's of the TTAGG unit were changed, the nicking activities of TRAS1 EN were greatly reduced, indicating that the third A and the fourth and fifth G's of the TTAGG unit are essential for the endonucleolytic activity of TRAS1 EN.

The results in Fig. 6 indicate that all residues in the TTAGG unit are somehow involved in site-specific cleavage of the bottom strand by TRAS1 EN. In this assay, however, we cannot specify the nucleotide essential for TRAS1 EN activity in successive TTAGG sequences, since a nucleotide of TTAGG was altered in all repeating units. Hence, we made a series of mutated substrates for the (TTAGG)$_5$ strand with a single base substitution and tried to figure out how TRAS1 EN recognizes the bases around the cleavage site. These mutated substrates contained substitutions to cytosine in the various positions flanking the cleavage site, from −7 to +8 (where 0 is cleavage site) (Fig. 7). We did not observe significant changes in the nicking activities for the substrates with mutations in positions +8 to +4 (Fig. 7) and in sites upstream of −8 (data not shown). When the 5' flanking bases to the cleavage site in −7 to −3 were changed, the nicking activity of TRAS1 EN was
reduced to about 40 to 80% of that for the nonmutated substrate (Cont., Fig. 7). Among them, the 3-mutated substrate had a severe effect on TRAS1 EN activity (38% of control). The nicking activity was also greatly inhibited by substitutions in the 3'-flanking sequence (to TTTGATC), especially at +1 [(TTAGG)₂TTCGG(TTAGG)₃] and +2 [(TTAGG)₂TTACG(TTAGG)₂] (20 and 25%, respectively). The results in Fig. 6 showing the great reduction in nicking activity for three substrates (TTCGG)ₙ, (TTACG)ₙ, and (TTAGC)ₙ support the above observation. In the case of (TTACG)ₙ, both the +3 and −3 positions from the cleavage sites were altered, which should have stronger blocking effects on nicking activity than a single substitution in either +3 or −3 in Fig. 7.

In contrast, the TRAS1 EN activity increased for the substrates mutated at −1 and −2 in appearance. Both substrates were, however, cleaved in many nonspecific sites upstream of the cleavage site (Fig. 7A), suggesting that the target site specificity of TRAS1 EN is relaxed by the −2 and −1 mutations. Combined with the aberrant cleavage patterns of (CT ↓ AGG)ₙ and (TC ↓ AGG)ₙ, these results suggest that the TT sequence of TT ↓ AGG is involved in determining the precise cleavage site by TRAS1 EN.

In summary, mutations −7 to +3 to the cleavage site affect the cleavage reaction, especially the −3 to +2 bases (5'...TTAGGTT ↓ AGGTT...3'; underlined), and are essential for sequence-specific recognition and digestion on the bottom (TTAGG)ₙ strand by TRAS1 EN.

**DISCUSSION**

TRAS1 EN recognizes and cleaves a specific DNA structure. TRAS1 comprises two ORFs and encodes an AP endonuclease (APE)-like endonuclease at the N terminus of ORF2 (25). TRAS1 EN cleaves the bottom TTAGG strand between T and A at the boundary of the pyrimidine tract, TT, and the following purine tract, AGG. In the human L1 element, a similar target structure is observed in the first major nicking sites of the bottom strand (6, 9). The L1 endonuclease preferentially targets 5'(dTₙ₋ₐ) tracts and cleaves its T-A junction, illustrating that L1 and TRAS1 may have similar features for target site recognition in the bottom strand. Dinucleotide sequences
such as TA, CA, and TG are known as specific DNA sites where kinks may occur under bending constraints (22). DNA kinks are defined as abrupt deflection of the double helical structure, leading to unstacking of two neighboring base pairs. It has been suggested that the EcoRV restriction enzyme (35) and the endonuclease of non-sequence-specific retrotransposons from plant and mammal cleave kinked DNA sites (17, 33). TRAS1 may also cleave the DNA kink at the T-A junction. Consistent with this, TRAS1 EN also cleaved a kinked C-A junction of the −1 mutant, 5′-(TTAGG)2-TCAGG-(TTAGG)2-3′, at a rate nearly threefold that of T-A cleavage in (TTAGG)2 (Fig. 7). On the top strand, in contrast, TRAS1 EN makes a specific nick between C and T of the CCTAA sequence. The C-T junction, which cannot form a kinked structure, may not be cleaved by TRAS1 EN without recognizing these DNA structures.

The results shown in Fig. 3 suggest that two units of telomeric repeats, the 7 bp upstream and 3 bp downstream from the T-A junction (5′-TTAGGTT 3′) are the minimum structures to ensure the endonucleolytic activity of TRAS1 EN. Consistent with this, single-base substitutions to cytosine upstream of −8 (data not shown) and downstream of +4 had little effect on the nicking activity of TRAS1 EN (Fig. 7). TRAS1 EN activity was greatly influenced by mutations in the −2 to +3 bases just adjacent to the cleavage site (5′-TTAGGTT 3′-AGGGT . . . -3′; underlined). This suggests that the G-T-A-G pentamer may contain the first recognition region by TRAS1 EN on the bottom strand of the telomeric sequence. Such asymmetric recognition of the bases around the target site was also proposed for the endonuclease of R2 (37) and Tx1L (5).

Site-specific cleavage on the top strand by TRAS1 EN. During retrotransposition by non-LTR-type retroelements, first-strand cleavage to prime reverse transcription of the RNA template was suggested to be followed by second-strand cleavage on the top strand. For site-specific integration, retrotransposons have to make a precise nick on the top strand. However, it is still ambiguous whether the APE-like (EN) encoded by site-specific retrotransposons is responsible for the second-strand cleavage reaction. Although R1 EN was shown to cleave its target 28S rDNA sequence in the precise site on the bottom strand, top-strand cleavage seemed to be less specific, with a few nonspecific products (8). Similarly, specific cleavage site on the top strand by Tx1L EN to ensure target site duplication has not yet been observed clearly (5). In this study, we first showed site-specific cleavage of the second strand by a site-specific non-LTR retrotransposon, TRAS1 (Fig. 5). The cleavage site between C and T of the (CCTAA)n, top strand is consistent with insertion sites expected from the genomic structure of 5′ boundary regions of TRAS1 ([CCTAA]n-C-C[5′UTR of TRAS1]) (Fig. 1). Specific cleavage at the top strand by TRAS1 EN indicates that APE-like endonuclease could carry out the second-strand cleavage even in the absence of its own RNA template or reverse transcription reaction. This shows a striking contrast to a site-specific endonuclease encoded in the C-terminal region of the R2 element, which requires an RNA transcript for cleavage of the second DNA strand (20, 39).

How does TRAS1 EN accomplish a relatively high specificity for the top strand? R2Bm is able to cleave the top strand in a very specific manner. R2Bm protein is thought to accomplish these specific cleavages on both strands by remaining attached to the target DNA during the TPRT reaction (39). In contrast, it was suggested that Tx1L EN protein is released after cleaving the target DNA and that the full-length Tx1L sequences may be needed for the specific cleavage reaction on the top strand (5). Similar to Tx1L EN, TRAS1 EN seems to be capable of enzymatic turnover (data not shown), indicating that TRAS1 EN may carry out the second-strand cleavage without
attachment to the target DNA. It is of interest that Tx1L EN and TRAS1 EN show different specificities for the top strands, though they may have similar enzymatic properties.

In host genomes, site-specific retrotransposons with APE-like endonuclease are usually flanked by TSD sequences at both ends. It is postulated that TSD is generated by cleavage patterns of APE-like endonuclease which cut the top strand at a site downstream of bottom-strand cleavage (4, 14). Although the TRAS1 element also encodes an APE-like endonuclease, its TSD sequence is unclear, because the target sequence of TRAS1 (i.e., the telomeric repeat) itself is repetitious. Figure 5B shows that the second cleavage of the top strand seemed to take place 6 bp upstream of the bottom-strand cleavage. This type of cleavage pattern by TRAS1 EN may generate 6-bp deletions at both ends instead of TSD, as suggested by former reports (4, 14).

**Specific integration of TRAS1 into the telomere region.** If the TRAS1 element can integrate into its target site by recognizing only 10 bp (5'-TAGGGTTAGG-3'), as suggested above, TRAS1 should retrotranspose into many and various genomic locations that have the recognition sequence. The locations of TRAS1 were, however, restricted to the telomere region (25), indicating that integration of TRAS1 seems to be more specific to the telomeric repeat sequence in vivo.

Two models could explain the higher integration specificity of TRAS1 to telomeres in vivo. The first is that some other regions in TRAS1 might be required for recognizing the longer arrays of telomeric repeats or telomeres. In fact, TRAS1 EN structurally lacks the cysteine-histidine motifs which are located in the C terminus of ORF1 and ORF2 and supposed to be involved in DNA binding in R2 (38). Recently, we found a novel and putative myb-like DNA-binding structure between the EN and RT domains of ORF2 for the TRAS1 element (Y. Kubo and H. Fujiwara, submitted for publication). There is a possibility that some cysteine-histidine motifs or the myb-like domain may be involved in the observed higher specificity for telomeres.

The second model postulates that a TRAS1 integration complex interacts with a chromatin component specific for telomeres before the retrotransposition and then integrates by recognizing relatively short arrays of telomeric repeats. In the *Saccharomyces cerevisiae* LTR-type retrotransposon Ty5, which preferentially integrates into domains of silent chromatin at HM loci and telomeres, it has been proposed that target site specificity results from the recognition of specific chromatin components such as Sir3p and Sir4p (41).

As shown in Fig. 5, TRAS1 EN could make a specific nick on both strands of oligonucleotides even if it contained additional nontelomeric sequences. The EN domain of TRAS1 can cleave the target sequence in vitro. Therefore, we conclude that telomere-specific chromatin components or telomere-binding proteins may interact with TRAS1 protein first and then the EN domain of TRAS1 mainly defines the precise sequence for integration.

**Application of TRAS1 EN for the telomere-nicking enzyme in vivo.** The data shown here suggest that TRAS1 has the potential to integrate into telomere sequences. These telomere-associated non-LTR retrotransposons may be involved in telomere maintenance, like He-T-A and TART in *Drosophila* telomeres (3, 18). On the other hand, this study raises the possibility that this telomere-specific endonuclease can be used as a negative regulator of telomeres. To our knowledge, this is the first evidence of a telomere-specific endonuclease. Recently, it was reported that human topoisomerase II cleaves tandem repeats of telomeric DNA, (TTAGGG), only in the presence of the topoisomerase II poison etoposide (40). This nicking activity, however, was not specific for the telomere sequence, since topoisomerase II itself is known to cleave a wide range of consensus sequences (29).

We suggest that TRAS1 EN cleaves TTAGG repeats interacting with the 5 bases of the GGTTA sequence, as shown in Fig. 7, indicating that this telomere-nicking enzyme may be applied to the study of other organisms which have a GGTTA sequence in their telomeric repeats, such as (TTAGG), of insects (26, 28) and (TTAGGG), of many species, including humans (23). To test this possibility, we tried to examine whether TRAS1 EN could cleave the human-type telomeric repeats (TTAGGG/CCCTAA). The results demonstrated that TRAS1 EN also nicked the TTAGGG bottom strand between T and A and that the cleaved products represented banding patterns with 6-bp but not 5-bp ladder signals (data not shown). Although further experiments will be necessary to clarify the specific digestion of other GGTTA-including telomeric repeats than (TTAGG), by TRAS1 EN, this enzyme activity may be used in insects and other organisms to study telomere function in vitro and in vivo.

**ACKNOWLEDGMENTS**

This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan.

**REFERENCES**

1. Aksoy, S., S. Williams, S. Chang, and F. F. Richards. 1990. SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs. Nucleic Acids Res. 18:785–792.
2. Besansky, N. J., S. M. Paskewitz, D. M. Hamm, and F. H. Collins. 1992. Distinct families of site-specific retrotransposons occupy identical positions in the rDNA genes of *Anopheles gambiae*. Mol. Cell. Biol. 12:5102–5110.
3. Biessmann, H., L. E. Champion, M. O'Hair, K. Ikenaga, B. Kasravi, and J. M. Mason. 1992. Frequent transpositions of *Drosophila melanogaster* HeT-A transposable elements to receding chromosome ends. EMBO J. 11:4469–4469.
4. Burke, W. D., F. Muller, and T. H. Eickbush. 1995. R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. Nucleic Acids Res. 23:4628–4634.
5. Christiansen, S., G. Pont-Kington, and D. Carroll. 2000. Target specificity of the endonuclease from the *Xenopus laevis* non-long terminal repeat retrotransposon Tx1L. Mol. Cell. Biol. 20:1219–1226.
6. Cost, G. J., and J. D. Boeke. 1998. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry 37:18081–18093.
7. Eickbush, T. H., and B. Robins. 1985. *Rhomboid* mRNAs 2S ribosomal genes contain insertion elements similar to the Type I and II elements of *Drosophila melanogaster*. EMBO J. 4:2281–2285.
8. Feng, Q., G. Schumann, and J. D. Boeke. 1998. Retrotransposon R1Ibm endonuclease cleaves the target sequence. Proc. Natl. Acad. Sci. USA 95:2083–2088.
9. Feng, Q., J. V. Moran, H. H. Kazazian, Jr., and J. D. Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87:905–916.
10. Fujiwara, H., T. Ogura, N. Takada, N. Miyajima, H. Ishikawa, and H. Maskawa. 1984. Introns and their flanking sequences of *Rhomboid* mRNAs. Nucleic Acids Res. 12:6861–6869.
11. Gabriel, A., and J. D. Boeke. 1993. Retrotransposon reverse transcription, p. 275–328. In A. M. Skalka and S. F. Goff (ed.), Reverse transcriptase. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
12. Gabriel, A., T. J. Yen, D. C. Schwartz, C. L. Smith, J. D. Boeke, B. Sollner-Webb, and D. W. Cleveland. 1990. A rapidly rearranging retrotransposon
within the minixenon gene locus of *Crithidia fasciculata*. Mol. Cell. Biol. 16:615–624.

13. Garrett, J. E., D. S. Knutson, and D. Carroll. 1989. Composite transposable elements in the *Xenopus laevis* genome. Mol. Cell. Biol. 9:3038–3037.

14. George, J. A., W. D. Burke, and T. H. Eickbush. 1996. Analysis of the 5′ junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. Genetics 142:853–863.

15. Henderson, E. 1995. Telomere structure, p. 11–34. In E. H. Blackburn, and C. W. Greider (ed.), Telomeres. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

16. Jakubczak, J. L., W. D. Burke, and T. H. Eickbush. 1991. Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc. Natl. Acad. Sci. USA 88:3295–3299.

17. Jurka, J., P. Klonowski, and E. N. Trifonov. 1998. Mammalian retrotransposons integrate at kinkable DNA sites. J. Biomol. Struct. Dyn. 15:717–721.

18. Levis, R. W., R. Ganesan, K. Houtchens, L. A. Tolar, and F. M. Sheen. 1989. Conservation of the human telomere sequence (TTAGGG) among vertebrates. Proc. Natl. Acad. Sci. USA 86:7049–7053.

19. Mol, C. D., C. F. Kuo, M. M. Thayer, R. P. Cunningham, and J. A. Tainer. 1995. Structure and function of the multifunctional DNA-repair enzyme exonuclease III. Nature 332:2801–2809.

20. Okazaki, S., H. Ishikawa, and H. Fujiwara. 1995. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyxs mori*. Mol. Cell. Biol. 15:4545–4552.

21. Okazaki, S., K. Tsuchida, H. Markawa, H. Ishikawa, and H. Fujiwara. 1993. Identification of a pentanucleotide telomeric sequence, (TTAGGG) repeats, in the silkworm *Bombyxs mori* and in other insects. Mol. Cell. Biol. 13:1424–1432.

22. Paszewitz, S. M., and F. H. Collins. 1989. Site-specific ribosomal DNA insertion elements in *Anopheles gambiae* and *A. arabiensis*: nucleotide sequence of gene-element boundaries. Nucleic Acids Res. 17:8125–8133.

23. Sahara, K., F. Marec, and W. Traut. 1999. TTAGG telomeric repeats in chromosomes of some insects and other arthropods. Chromosome Res. 7:449–460.

24. Spitzner, J. R., J. K. Chung, T. D. Gootz, P. R. McGuirk, and M. T. Muller. 1995. Analysis of eukaryotic topoisomerase II cleavage sites in the presence of the quinolone CP-115,953 reveals drug-dependent and -independent recognition elements. Mol. Pharmacol. 48:238–249.

25. Takahashi, H., and H. Fujisawa. 1999. Transcription analysis of the telomeric repeat-specific retrotransposons TRAS1 and SART1 of the silkworm *Bombyxs mori*. Nucleic Acids Res. 27:2015–2021.

26. Takahashi, H., S. Okazaki, and H. Fujiwara. 1997. A new family of site-specific retrotransposons, SART1, is inserted into telomeric repeats of the silkworm, *Bombyxs mori*. Nucleic Acids Res. 25:1578–1584.

27. Tatout, C., L. Lavie, and J. M. Deragon. 1998. Similar target site selection occurs in integration of plant and mammalian retroelements. J. Mol. Evol. 47:463–470.

28. Villanueva, M. S., S. P. Williams, C. B. Beard, F. F. Richards, and S. Aksoy. 1991. A new family of site-specific retrotransposons is present in the spliced leader RNA genes of *Trypanosoma cruzi*. Mol. Cell. Biol. 11:6139–6148.

29. Winkler, F. K., D. W. Banner, C. Oefner, D. Tsernoglou, R. S. Brown, S. P. Heathman, R. K. Bryan, P. D. Martin, K. Petraitos, and K. S. Wilson. 1993. The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. EMBO J. 12:1781–1795.

30. Wright, W. E., V. M. Tesmer, K. E. Huffman, S. D. Levine, and J. W. Shay. 1997. Normal human chromosomes have long G-rich telomeric overhangs at one end. Genes Dev. 11:2801–2809.

31. Xiong, Y. E., and T. H. Eickbush. 1988. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell 55:235–246.

32. Yang, J., H. S. Malik, and T. H. Eickbush. 1999. Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. Proc. Natl. Acad. Sci. USA 96:7847–7852.

33. Yoon, H. J., I. Y. Choi, M. R. Kang, S. S. Kim, M. T. Muller, J. R. Spitzner, and J. R. Spitzner. 1998. DNA topoisomerase II cleavage of telomeres in vitro and in vivo. Biochim. Biophys. Acta 1395:110–120.

34. Zhu, Y., S. Zou, D. A. Wright, and D. F. Voytas. 1999. Tagging chromatin with retrotransposons: target specificity of the *Saccharomyces* Ty5 retrotransposon changes with the chromosomal localization of Sir3p and Sir4p. Genes Dev. 13:2738–2749.