Solution structure of an RNA stem–loop derived from the 3' conserved region of eel LINE UnaL2

SEIKI BABA,[1] MASAKI KAJIKAWA,[2] NORIHIRO OKADA,[2] and GOTA KAWAI[1]

[1]Department of Industrial Chemistry, Faculty of Engineering, Chiba Institute of Technology, Narashino, Chiba 275-0016, Japan
[2]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Midori-ku, Yokohama, Kanagawa 226-8501, Japan

ABSTRACT

The eel long interspersed element (LINE) UnaL2 and its partner short interspersed element (SINE) share a conserved 3' tail containing a stem–loop that is critical for their retrotransposition. Presumably, the first step of retrotransposition is the recognition of their 3' tails by UnaL2-encoded reverse transcriptase. The solution structure of a 17-nucleotide RNA derived from the 3' tail of UnaL2 was determined by NMR. The GGAUA loop forms a specific structure in which the uridine is exposed to solvent with the third and fifth adenosines stacked. A sharp turn in the phosphodiester backbone occurs between the second guanosine and third adenosine. When the uridine is mutated (but not deleted), all mutants form the loop structure, indicating that the loop structure requires an exposed fourth residue. The retrotransposition assay in HeLa cells revealed that retrotransposition requires the second guanosine, although any nucleoside functions at the fourth position, suggesting that UnaL2 reverse transcriptase specifically recognizes the 5' side of the GGANA loop.

Keywords: LINE; NMR; retrotransposition; solution structure; RNA

INTRODUCTION

Long interspersed elements (LINEs) and short interspersed elements (SINEs) are transposable elements and mobilize through an RNA intermediate. These elements are first transcribed into RNA, and the RNA is then reverse transcribed into complementary DNA that is subsequently integrated into a new location within the host genome. This "copy-and-paste" mechanism is called retrotransposition, and the number of LINEs and SINEs has been expanding by this process.

LINEs of ~4–7 kilobase pairs (kbp) are abundant in many eukaryotic genomes, contributing to genome structure and evolution (Eickbush 1992; Boeke and Devine 1998; Kazazian 2000; International Human Genome Sequencing Consortium 2001; Ostertag and Kazazian 2001). For example, ~20% of the haploid human genome is composed of LINEs (International Human Genome Sequencing Consortium 2001). LINEs usually encode two open reading frames, ORF1 and ORF2, each of which is required for LINE ret-

rotransposition (Feng et al. 1996; Moran et al. 1996). ORF1 encodes a nucleic acid binding protein that is thought to bind LINE RNA to form a retrotransposition intermediate (Hohjoh and Singer 1997). ORF2 encodes an endonuclease and a reverse transcriptase (RT). SINEs of ~100–500 bp are another abundant component of many eukaryotic genomes (Weiner et al. 1986; Britten et al. 1988; Okada 1991; Schmid and Maraia 1992; Deininger and Batzer 1993). In fact, SINEs constitute ~13% of the human genome (International Human Genome Sequencing Consortium 2001). Although SINEs do not encode any proteins, they are also thought to mobilize via retrotransposition, with each SINE recruiting the enzymatic machinery for its own retrotransposition. Hence, SINEs are nonautonomous transposable elements.

Eickbush's group demonstrated that the silkworm LINE R2Bm protein specifically recognizes the 3' untranslated region (UTR) of its own RNA and reverse transcribes the R2Bm RNA (Luan and Eickbush 1995; Mathews et al. 1997). Okada's group used sequence comparisons to show that many LINEs and SINEs share the same 3' tail sequence, proposing that LINE-encoded RTs not only recognize their own 3' tail during retrotransposition, but also specifically recognize the shared 3' tail of their partner SINEs in *trans* (Ohshima et al. 1996; Okada and Hamada 1997). In contrast, the RT of human L1 does not specifically recognize

the 3′ UTR and it is thought that the 3′ poly(A) tail is only required for the reverse transcription of L1 RNA (Moran et al. 1996). These observations led to the proposal that LINEs can be divided in two groups, a stringent type and a relaxed type (Okada and Hamada 1997). The former type requires a specific sequence at the 3′ end for retrotransposition, whereas the latter type does not require such a sequence but probably requires a poly(A) tail.

Recently, Okada's group isolated a LINE/SINE pair, denoted UnaL2 and UnaSINE1, respectively, from the eel genome (Kajikawa and Okada 2002). UnaL2 and UnaSINE1 share a common 3′ tail of ~60 bp, and it was demonstrated that the conserved 3′ tail of UnaL2 is required for UnaL2 retrotransposition and that the retrotranspositional machinery of UnaL2 can recognize the 3′ conserved region in *trans* (Kajikawa and Okada 2002). Furthermore, the 3′ conserved region of UnaSINE1 also can be recognized by the UnaL2 enzymatic machinery in *trans*, resulting in efficient mobilization (Kajikawa and Okada 2002). These data show that SINEs having a 3′ tail similar to that of LINEs are mobilized by the proteins encoded by the partner LINEs.

The conserved 3′ tail of UnaL2 is ~60 bp and comprises two regions (Fig. 1), one of which is a stem–loop (a GGAUA loop) that probably forms when this region is transcribed. The other region is a pentanucleotide repeat [TGTAA]$_n$ (usually $n = 3$) that is positioned at the extreme 3′ end. A retrotransposition assay for UnaL2 in HeLa cells demonstrated that both the stem–loop and the pentanucleotide repeat are required for UnaL2 retrotransposition (Kajikawa and Okada 2002). Although both regions are required, the functional significance of these two regions appears to differ. Mutational analyses revealed that more than one repetition of the [TGTAA] repeat is necessary, because
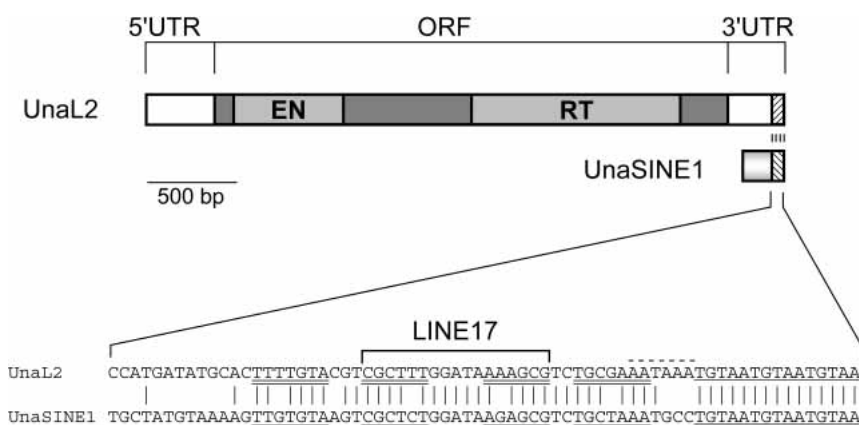

FIGURE 2. Secondary structure of the 3′ tail of UnaL2 and the 17-mer RNAs used in this study. (*A*) Secondary structure of the UnaL2 3′ tail that is putatively recognized by the UnaL2 reverse transcriptase. (*B*) The conserved RNA step-loop structure of the LINE17 3′ tail. (*C–F*) Structures of U10C, U10A, U10A, and U10del, respectively.

one TGTAA is not functional. It was suggested that a slippage reaction occurs during the initiation of reverse transcription of UnaL2 RNA and that the TGTAA repeats are required for this slippage reaction (Kajikawa and Okada 2002). On the other hand, it appears that the RNA within the stem–loop (Fig. 2A) constitutes the recognition site for UnaL2-encoded protein, because the conserved 3′ tail of UnaL2 is required only for efficient *trans* retrotransposition and the TGTAA repeats do not seem to be the recognition site for retrotransposition.

In the present study, we determined the solution structure of an RNA stem–loop containing 17 residues (LINE17, Fig. 2B) including the sequence GGAUA corresponding to the upper region of the UnaL2 3′ stem–loop. Using the retrotransposition assay in HeLa cells, we determined the retrotranspositional frequencies of mutant UnaL2s containing a point mutation within the loop. From these data, we discuss the requirement of the loop structure for retrotransposition of UnaL2 and related LINEs.

RESULTS

Analysis of NMR spectra of LINE17

The imino proton resonances (Fig. 3A) of LINE17 were assigned on the basis of the NOESY spectrum observed in H$_2$O. The minor resonances around 12.3 ppm were observed only for samples prepared by the transcription reaction and not by the chemical synthesis, indicating that these signals were due to the heterogeneity of the 5′ terminal. The formation of the LINE17 stem–loop was also confirmed by comparing mobility via native polyacrylamide gel electro-

FIGURE 1. Schematic representation of full-length UnaL2 and UnaSINE1 from the eel and an alignment of their conserved 3′ tail regions. The single ORF comprises the shaded boxes, and the putative endonuclease (EN) and reverse transcriptase (RT) domains are indicated. The 5′ and 3′ untranslated regions (UTRs) are also indicated. The sequence alignment of the conserved 3′ tail is shown. The upper and lower stem regions are underlined by double lines and the region corresponding to LINE17 is indicated. The 3′ terminal repeats in UnaL2 and UnaSINE1 are underlined by a single line. The putative poly(A) signal in the UnaL2 sequence is indicated by a dotted line.
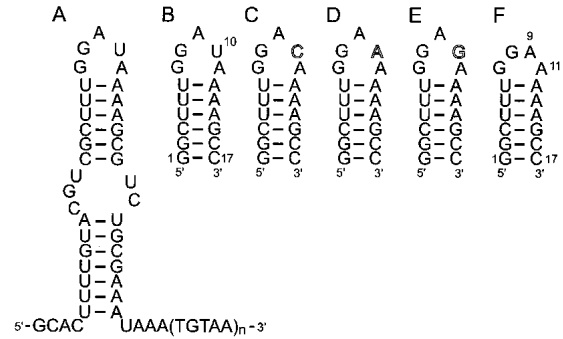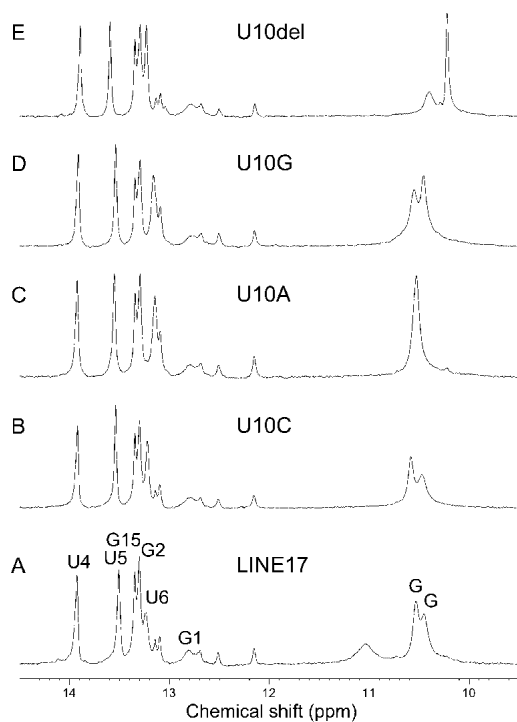
FIGURE 3. Imino proton spectra of LINE17 and its mutants. (*A*) LINE17. (*B*) U10C. (*C*) U10A. (*D*) U10G. (*E*) U10del. Resonance assignments of LINE17 are shown.

phoresis (PAGE) with other known RNA molecules (data not shown). The resonances of nonexchangeable protons were assigned by the conventional method (Allain and Varani 1995). Analysis of the TOCSY spectrum indicated that most of the residues were adapted to C3′-*endo* conformation except for G8, A9, and U10 in the GGAUA loop, which may have contained a mixture of C2′-*endo* and C3′-*endo* conformations. It should be noted that the resonances due to the GGAUA loop of the 36-mer RNA including the entire 3′ conserved stem–loop were identical to those of LINE17 (data not shown). A total of 204 NOE and 80 dihedral angle restraints were obtained and structures were calculated using restrained molecular dynamics calculations with a simulated annealing protocol (Nilges et al. 1988). The structure is well defined, having a heavy atom r.m.s. deviation of 1.50 Å for 20 converged structures (Fig. 4). R.m.s. deviations of bonds (Å) and angle (°) from the idealized geometry were 0.0083 ± 0.0001 and 2.31 ± 0.04. The structural statistics are summarized in Table 1.

Structure of the GGAUA loop

The local structure of the GGAUA loop is well defined in the calculated structure of LINE17 (heavy atom r.m.s. deviation was 1.93 Å for G7-A11, Table 1) as shown in Figure 4, although the conformation for the glycosidic bond (χ angle) of U10 is not defined. The uracil base moiety of U10

was exposed to the solvent and strong intraresidue NOEs for both H6-H1′ and H6-H2′ of U10 were observed, indicating rotation around the glycosidic bond. Note that the heavy atom r.m.s. deviation for G7-A11, except the U10 base, was 1.12 Å. A schematic diagram of the LINE17 loop structure is shown in Figure 5A. One of the structural characteristics is the stacking of G8 onto G7. A sharp turn in the phosphodiester backbone occurs between G8 and A9, and this is consistent with the NOE connectivities for H1′/H4′(G8)-H8(A9) and H1′(G8)-H2′/H3′(A9). The three adenosine residues, A9, A11, and A12, are continuously stacked, and this is consistent with the NOE connectivities for H8/H1′(A9)-H8(A11), H2(A9)-H1′(A11), and H1′/H2(A11)-H1′(A12). The H1′ resonance of A12 is shifted slightly down field by the ring current of the adenosine base in A11. In the calculated structures, G8, A9, and U10 adopt C2′-*endo*, C2′-*endo*/C3′-*endo*, and C3′-*endo* conformations, respectively. The calculated structures revealed that G7 and A11 form a sheared-type G:A base pair, which is similar to the G:A base pair in the GGAA loop (Fig. 5B; Schmitz et al. 1999) and the GAAAA loop (Fig. 5C; Schärpf et al. 2000). The distances between NH2(G7)-N7(A11) and N3(G7)-NH6(A11) are 2.14 and 2.05 Å, respectively, suggesting the formation of two hydrogen bonds. However, the structure of the wild-type GGAUA loop in LINE17 differed
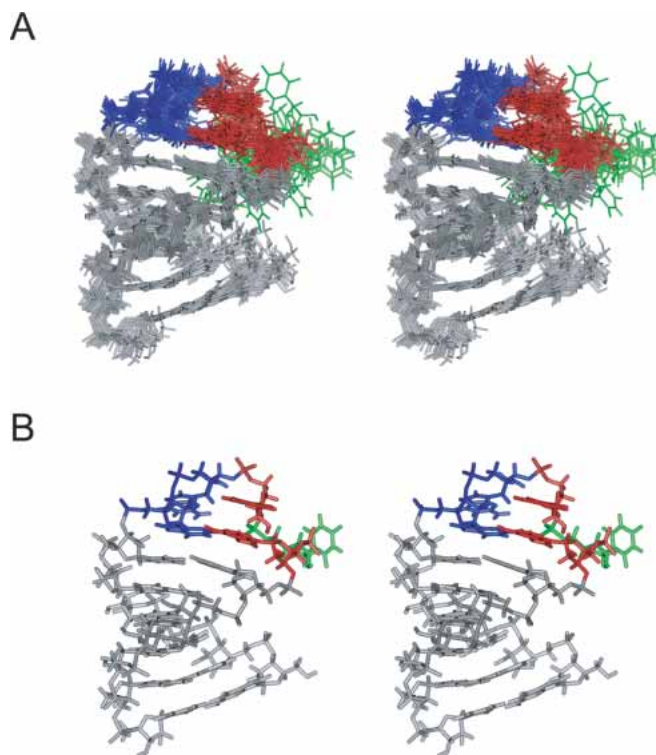

FIGURE 4. Stereo view of the solution structures of LINE17. (*A*) The superposition of the final 20 structures. (*B*) The minimized average structure of LINE17. G, A, and U residues in the GGAUA loop of LINE17 are colored in blue, red, and green, respectively.

from the GNRA and GAAAA structures (Fig. 5), especially at the second guanosine (G8 in LINE17). For the GNRA loop, the H1′ resonance of the 3′ adjacent residue is shifted to around 4.0 ppm because of the ring current effect of the adenine base of the fourth residue of the loop (Jucker et al. 1996; Schmitz et al. 1996). In the case of LINE17, the H1′ resonance of A12 is also shifted to 4.85 ppm, whereas H1′ resonances of other adenosine residues resonate around 5.7 ppm. The position of G:A base pair relative to the loop closing base pair for LINE17 is slightly different from those for the GNRA and GAAAA structures and this may cause the difference in the shift of the H1′ resonances. One of the characteristic NOEs determining the location of G8 [H1′(G8)-H3′(A9)] is shown in Figure 6A,B.

Structures of the LINE17 mutants

One of the remarkable structural characteristics of LINE17 is the exposed U10 residue. Thus, U10 was mutated or deleted to elucidate the effect of the residue on the structure and function of the LINE RNA (Fig. 2C–F). The imino proton spectra and chemical shift differences of H6/H8 and H1′ for the mutant RNAs are shown in Figure 3B–E and Figure 7, respectively. The imino proton spectrum of the U10C mutant is nearly identical to that of LINE17, and the H6/H8 and H1′ chemical shifts also are similar to LINE17 except for the mutated residue, indicating that the U10C and LINE17 structures are essentially identical. For the U10A and U10G mutants, some imino proton resonances are slightly shifted, but most of the chemical shifts of H6/H8 and H1′ of the mutants are nearly the same as those of LINE17. In the TOCSY spectra, it should be noted that the
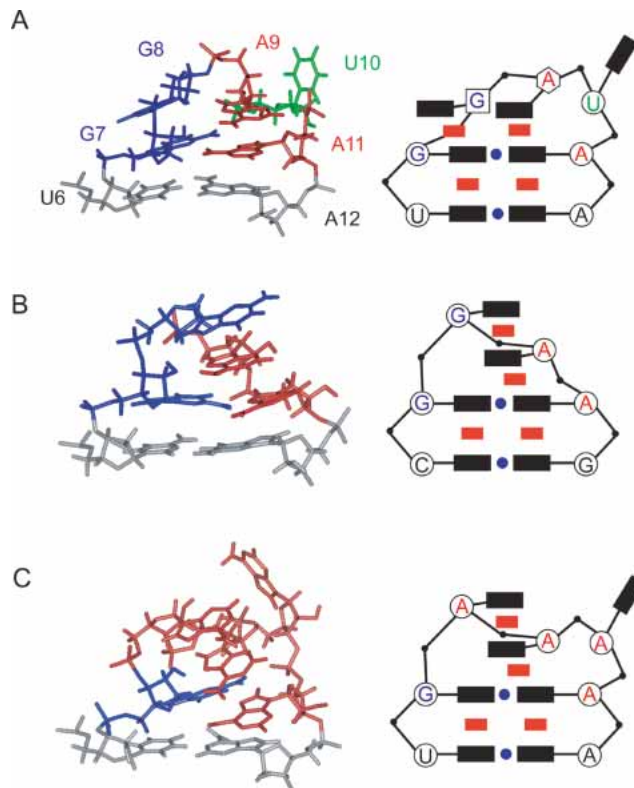

FIGURE 5. Comparison of the GGAUA, GGAA, and GAAAA loop structures and schematic representations. The following symbols are used in the schematic representation: black rectangle, base; red rectangle, stacking interaction; blue circle, hydrogen bonding interaction; open circle, C3′-*endo* ribose; open square, C2′-*endo* ribose; open hexagon, intermediate between C3′-*endo* and C2′-*endo*. (*A*) The GGAUA loop structure in LINE17 determined in this study. (*B*) The GGAA loop structure in *Escherichia coli* SRP RNA (Schmitz et al. 1999). (*C*) The GAAAA loop structure in *boxB* RNA in complex with the 36-mer N-terminal peptide of the N protein (N36) from bacteriophage λ (Schärpf et al. 2000).

cross peaks due to H1′ and H2′ of G8, A9, and N10, which were originally observed for LINE17, were also observed for U10A and U10G, as well as for U10C. More important, the characteristic NOE for H1′(G8)-H3′(A9) was observed for each of U10C, U10A, and U10G (Fig. 6C–E). For U17C, the NOE signal is overlapped with other NOE (H1′(G7)-H3′(G7)), but the relative volume of the overlapped peak corresponds to the sum of two NOE volumes for other spectra (Fig. 6B–E). Thus, we conclude that these three mutants share the same loop structure as LINE17, in which the base moiety of N10 is exposed to the solvent and a sharp turn occurs between G8 and A9.

The deletion of U10 results in the loop sequence of GGAA, which is one of the well-characterized GNRA loops (Fig. 5B; Jucker et al. 1996). In fact, the NMR spectra for U10del were consistent with those of the GGAA loop-structure determined by Schmitz et al. (1999) in that G8 is stacked on A9 instead of G7. Thus, N10 is required for the formation of the characteristic structure of the GGANA loop of the LINE RNA.

TABLE 1. NMR restraints and structural statistics

Number of restraints	
Intraresidue distance restraints	108
Interresidue distance restraints	96
Dihedral restraints	80
Hydrogen bonding distance restraints	14
R.m.s. deviations from the experimental restraints[a]	
Distance (Å)	0.0105 ± 0.0009
Dihedral (°)	0.20 ± 0.08
R.m.s. deviations from the idealized geometry	
Bonds (Å)	0.0083 ± 0.0001
Angle (°)	2.31 ± 0.04
Impropers (°)	1.6 ± 0.2
Heavy-atoms r.m.s. deviations (Å)[b]	
All	1.50
Loop (G7-A11)	1.93
Loop without U10 (G7-A9, A11)	1.12

[a]The converged structures did not contain an experimental distance violation of >0.1 Å or a dihedral violation >3°. The converged structures did not contain any hydrogen bonding distance restraints.
[b]Averaged r.m.s. deviations between the average structure and the 20 converged structures were calculated.
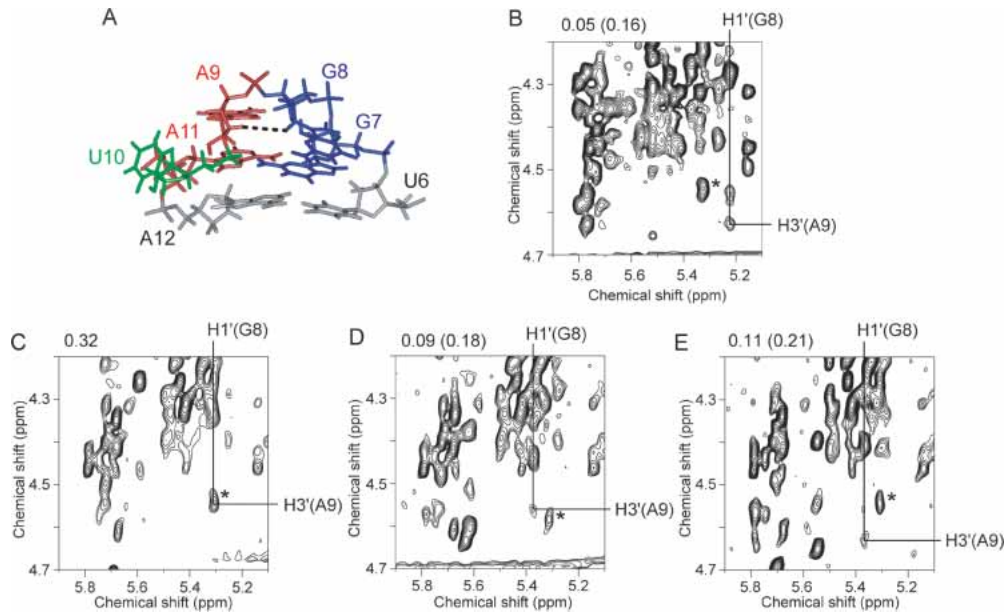
FIGURE 6. The characteristic NOE (H1′(G8)-H3′(A9)) in the GGARA loop structures of LINE17 and mutants thereof. (*A*) The black dotted line indicates a NOE between H1′(G8) and H3′(A9) in the GGAUA loop of LINE17. (*B–E*) NOESY spectra measured in D$_2$O at 20°C with a mixing time of 200 msec. Resonance of the G8 H3′ proton is labeled as a horizontal line and that of the A9 H1′ proton is labeled as a vertical line. NOE cross peaks between H3′(G8) and H1′(A9) were observed. (*B*) LINE17. (*C*) U10C. (*D*) U10A. (*E*) U10G. The NOE signals due to H1′(G7)-H3′(G7) are indicated by asterisks. Note that the signal-to-noise ratio for LINE17 is higher than others because of higher sample concentration. The NOE volumes for H1′(G8)-H3′(A9) and H1′(G7)-H3′(G7) in parentheses relative to the average for H5–H6 of U4, U5, U6, and C16 were shown on the *top* of spectra. For U10C, the value for the overlapped peak is shown.

Structural requirements of the GGANA loop for recognition by RT

To examine whether the characteristic loop structure of the 3′ tail is requisite for retrotransposition, we constructed a series of loop mutants of the UnaL2 expression plasmid and determined their retrotranspositional frequencies. These UnaL2 expression plasmids contained a reporter cassette (designated mneol; Freeman et al. 1994; Moran et al. 1996) in the 3′ UTR for the detection of retrotransposition (Kajikawa and Okada 2002). The retrotransposition assay results are summarized in Figure 8. Mutation of the second guanosine (G8 in LINE17) in the loop abolished or substantially reduced the retrotransposition activity. Mutation of the third adenosine (A9) of the loop also decreased the activity, although the A9G mutant retained considerable activity. In contrast, mutation of the solvent-exposed fourth uridine (U10) did not severely reduce the retrotransposition frequency. Deletion of either the second, third, or fourth residues (i.e., tetraloop variants) essentially abrogated activity. We conclude that the guanosine residue in the second position of the GGAUA loop is critical for retrotransposition, and that the adenosine residue in the third position also is important. Purines at the third position of the loop yielded higher retrotransposition frequencies, most likely owing to the stabilizing effect of stacking between the third and fifth purine bases. In contrast, mutations at the fourth position do not substantially affect function, although this

residue must not be deleted. These results indicate that the specific loop structure determined in this study is required for the retrotransposition of UnaL2.

DISCUSSION

On the basis of the results, we speculate that the second guanosine residue is directly recognized by UnaL2 RT and that this contact is required for the specific recognition of the UnaL2 3′ tail by its own RT. The fourth residue of the loop is not specifically recognized by the RT but is required for the formation of the specific structure of the loop. The third purine residue might also be required for the formation of the structure. Thus, our data suggest that the sequence GGRNA within the loop yields the structure that is recognized by UnaL2 RT. The first guanosine and the fifth adenosine in GGRNA form a base pair. This base pair is in the same conformation as a sheared-type base pair of the first guanosine and the fourth adenosine of the GNRA loop (Jucker et al. 1996). Also, Sakamoto et al. (2002) found that the GGAG loop forms a structure similar to that of the GNRA loop, and therefore the loop family can be extended to GNRR. Thus, the required sequence for the RT recognition could be GGRNR.

The structures of the GAAGA loop of the 19-mer *boxB* RNA stem–loop complexed with the 22-mer N-terminal peptide of the N protein (Legault et al. 1998) and the GAAAA loop of the 15-mer *boxB* RNA stem–loop com-
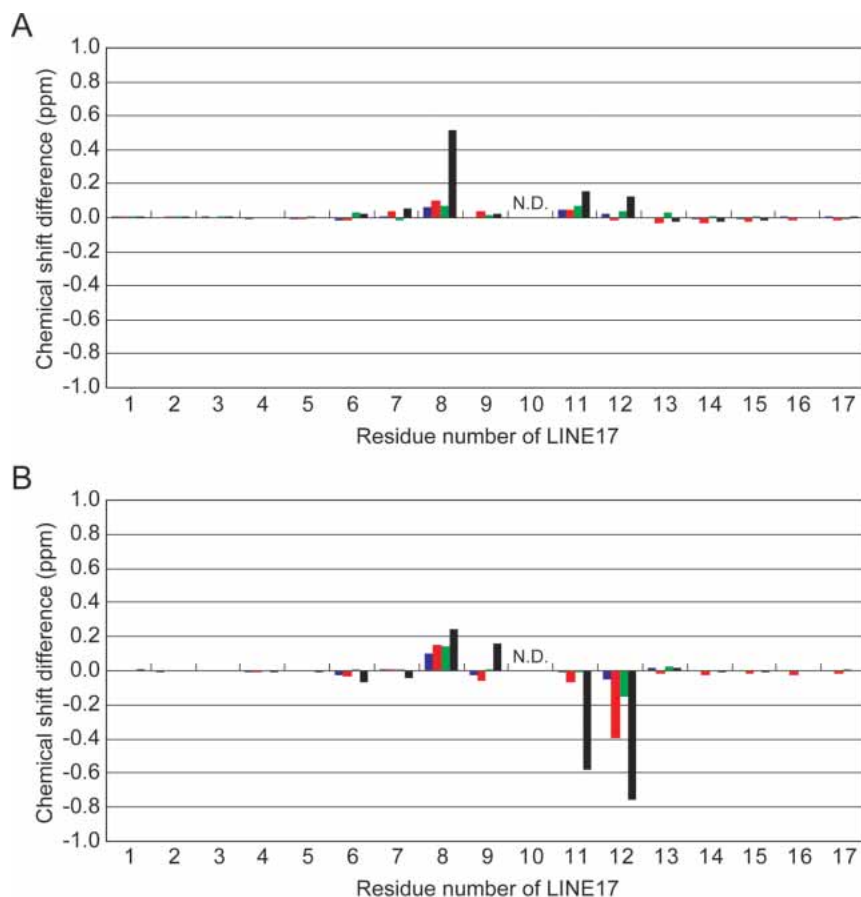
FIGURE 7. Chemical shift differences between LINE17 and the LINE17 mutants: blue, U10C; red, U10A; green, U10G; black, U10del. (*A*) Chemical shift differences of H8/H6 protons between LINE17 and each of the LINE17 mutants. (*B*) Chemical shift differences of the H1′ proton between LINE17 and each of the LINE17 mutants. The tenth residue is not shown because it is the mutated position in the LINE17 mutants.

plexed with the 36-mer N-terminal peptide of the N protein (Schärpf et al. 2000) have been determined. The two structures are similar to each other. This binding is involved in transcriptional antitermination of phage λ. The 36-mer peptide binds tightly to the major groove of the *boxB* stem–loop. The conformation of a GAAA tetraloop other than the fourth A in the GAAAA pentaloop forms a GNRA-like tetraloop in the complex. The formation of the GAAAA pentaloop involves a sheared-type base pair involving the first guanosine and the fifth adenosine, as well as stacking of the second, third, and fifth adenine bases and extrusion of one nucleotide (fourth adenosine) from the loop (Fig. 5C; Schärpf et al. 2000). In the structure of the GAAAA loop, the second adenosine stacks with the third adenosine to yield a loop structure similar to that of the GNRA loop (Fig. 5B,C). For the GGANA loop determined in this study, the second guanosine stacks with the first guanine and not to the third adenosine. Structure of the GAAAA loop of UnaL2 complexed with UnaL2 RT is the important problem to be solved to elucidate the mechanism of binding between the UnaL2 3′ RNA and UnaL2 RT.

To date, several LINEs (including UnaL2) belonging to the L2 clade have been reported. The 3′ tails of these LINEs are predicted to form a stem–loop structure similar to that within UnaL2 (data not shown). The LINE RTs are probably able to distinguish their own 3′ tails from those of others. The 3′ stem structures seem to be well conserved, whereas the loop sequences vary and apparently are not conserved. The LINE/SINE stem regions may represent the common binding site for LINE RTs, whereas the loops may provide the specificity for binding. In the case of UnaL2, the second guanosine in the loop may be involved in binding specificity. Further determination of the 3′ tail structures of various LINEs belonging to the L2 clade will enable structural comparisons from which we may formulate a general mechanism for the specificity of RT binding to the 3′ tail of its own LINE RNA.

MATERIALS AND METHODS

RNA synthesis, purification, and preparation

For structural determination, nonlabeled LINE17 was synthesized chemically by the phosphoramidite method with an automatic DNA/RNA synthesizer, Expedite model 8909 (PerSeptive Biosystems, Inc.). The protection groups were removed with ammonia and tetra-n-butylammonium fluoride. For structural comparison, the mutant RNAs as well as LINE17 were synthesized enzymatically by in vitro transcription with AmpliScribe T7 transcription kits (Epicentre Technologies Co.). Purification for each RNA sample was performed with PAGE using 30-cm × 40-cm glass plates (Nihon Eido Co. Ltd.) under denaturing conditions, and the RNAs were recovered from gel slices and salt was removed by ultrafiltration (Centricon YM3, Amicon Inc.). RNA samples were annealed by heating for 5 min at 90°C followed by snap-cooling on ice. To confirm the formation of the stem–loop structure, we subjected RNAs to native PAGE. For NMR measurements, RNA samples were dissolved in 10 mM sodium phosphate buffer (pH 6.0) containing 50 mM NaCl. The final concentration of chemically synthesized LINE17 was 1.0 mM. The concentrations for LINE17 (transcript), U10C, U10A, U10G, and U10del were 1.0, 0.4, 0.3, 0.5, and 0.5 mM, respectively.

For stable isotopic labeling by in vitro transcription, we used a larger RNA molecule with 36 residues including the entire 3′ conserved stem–loop (Fig. 3A) to increase the efficiency of the incorporation of ^{13}C- and ^{15}N-labeled NTPs (Nippon Sanso). The concentration of guanosine- and cytidine-specific ^{13}C- and ^{15}N-labeled RNA ([G, C-^{13}C/^{15}N] RNA) was 0.3 mM.
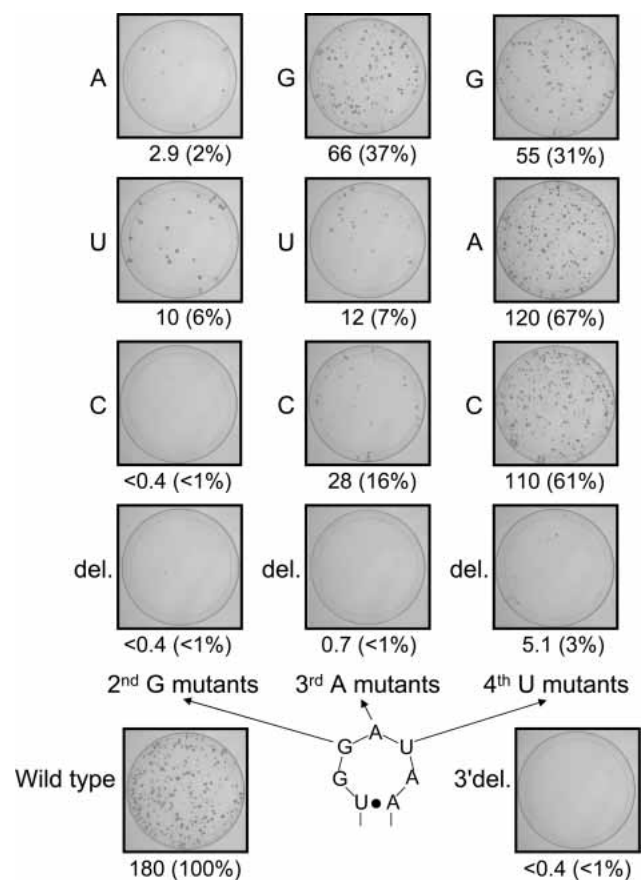
FIGURE 8. Retrotransposition frequency (RF) in mutants of the stem–loop within the conserved 3′ tail region. RFs are calculated as described in Materials and Methods. RFs and relative value of RFs for mutants compared with the wild type are shown. Images show each 100-mm plate with G-418[R] colonies selected from ~2–4 × 10[6] Hyg[R] cells. 3′ del indicates the UnaL2 mutant in which the 3′ conserved region is deleted.

NMR measurements

NMR spectra were measured using Bruker DRX-500 and DRX-600 spectrometers. Spectra were recorded at probe temperatures of 10°C–30°C and NMR data at 25°C were used for structure calculations. The imino proton resonances of G and U residues within RNAs in H_2O were distinguished by the HSQC-selected and HSQC-filtered spectra measured with [G, C-^{13}C/^{15}N] RNA. Exchangeable proton resonances were assigned by NOESY in H_2O with a mixing time of 150 msec using the jump-and-return scheme and gradient pulses for water suppression. Well-established procedures were used for resonance assignments for nonexchangeable protons (Allain and Varani 1995). H2 protons of adenosine were assigned using natural-abundance HSQC experiments (Legault et al. 1994). NOE distance restraints from nonexchangeable protons were obtained using NOESY (mixing times of 100, 200, and 400 msec) in D_2O (Jeener et al. 1979). Dihedral restraints were obtained from TOCSY (mixing time of 50 msec) and DQF-COSY, as described below. For TOCSY experiments, the modified composite pulse was used to eliminate the ROESY effect with a delay time equal to the 90-degree pulse (Griesinger et al. 1988). Five restraints (>5 Å) were added to the distance restraints described earlier on

the basis of the absence of NOE cross peaks. Absence of cross peaks between H1′ and H2′ in TOCSY and DQF-COSY experiments was interpreted as the residue being in the C3′-endo conformation (Altona and Sundaralingam 1973).

Structure calculation

A set of 100 structures was calculated using a simulated annealing protocol with the InsightII/Discover package (Accelrys) and the amber force field was used. A total of 204 NOE distance restraints including the five restraints for the absence of NOE cross peaks, 80 dihedral restraints, 14 hydrogen bonding restraints, and 68 chiral restraints were used. NOE intensities between exchangeable protons were interpreted as distances of 2.1–5.0 Å. NOE intensities from nonexchangeable protons were interpreted as distances with a margin of −1.5 to +1.5 Å for the 100-msec NOESY and −1.0 to +2.0 Å for the 200-msec NOESY. The five restraints for the absence of NOE cross peaks were defined as distances of 5.0–99.0 Å. The hydrogen bonding restraints were defined as distances of 1.8–2.2 Å, except for U6-A12 as distances of 1.8–5.0 Å, because the imino proton resonances of U6 are broader than other imino proton resonances due to the stem region. The information of anticonformation (G1-G7, A11-C17), the C3′-endo conformation (G2-G7, A11-C16), and RNA-A conformation for backbone in the stem region (G2-U5, A13-C16) were added for the dihedral torsion restraints. The force constants were 100 kcal mole^{-1} Å^{-2} for distance restraints and 100 kcal mole^{-1} rad^{-2} for dihedral restraints. The starting coordinates were randomized and the randomized structures were heated to 2000K over 5 psec, and the temperature maintained at 2000K for another 5 psec. After all restraints were increased to full value over 20 psec, they were decreased to 10% of full value over 5 psec. The van der Waals radius was increased from 10% to 82.5% over 20 psec. All restraints were increased to full value over 10 psec again. The nonbond scale was increased to full value over 20 psec. The temperature was kept at 2000K for another 5 psec. The temperature was then scaled to 300K over 10 psec. After the structure was heated from 300 to 1000K over 5 psec, the van der Waals radius was increased from 82.5% to full value, after which it was decreased from 82.5%. An additional 5 psec of dynamics was performed at 1000K, and the temperature was scaled to 300K over 10 psec. A final minimization step was performed, which included a Lennard-Jones potential and electrostatic terms with a dielectric constant of 7 according to Puglisi and Puglisi (1998). The 20 final structures that had the lowest total energy were chosen.

Retrotransposition assay

The plasmids containing UnaL2 loop mutants were constructed as described previously (Kajikawa and Okada 2002). Complete experimental procedures for construction of the plasmids are available on request.

The retrotranspositional frequencies of UnaL2 loop mutants were calculated as described previously (Kajikawa and Okada 2002). HeLa cells (2 × 10^5 cells/well) were seeded in six-well plates and transfected with 1 µg of a plasmid DNA with 3 µL of Fu-GENE6 reagent (Roche Diagnostics Co.) according to the manufacturer's instructions. After transfection, hygromycin resistance (Hyg[R]) cells were selected with 200 µg/mL hygromycin. By com-

paring the data with cell survival results from negative controls, we estimated that at least 95% of transfected cells became hygromycin resistant. The Hyg^R cells were then trypsinized and seeded to a density of ~2×10^5 to 4×10^6 cells per 100-mm plate and grown in medium containing 400 µg/mL of G-418. After G-418 selection, plates were fixed with 100% ethanol and stained with 2% Giemsa's solution. G-418^R colonies were counted and retrotranspositional frequencies were calculated as the number of G-418^R colonies per 10^6 plated Hyg^R cells.

Coordinates

The structure has been deposited with the Protein Data Bank (accession code 1WKS).

REFERENCES

Allain, F.H. and Varani, G. 1995. Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **250:** 333–353.

Altona, C. and Sundaralingam, M. 1973. Conformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants. *J. Am. Chem. Soc.* **95:** 2333–2344.

Boeke, J.D. and Devine, S.E. 1998. Yeast retrotransposons: Finding a nice quiet neighborhood. *Cell* **93:** 1087–1089.

Britten, R.J., Baron, W.F., Stout, D.B., and Davidson, E.H. 1988. Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci.* **85:** 4770–4774.

Deininger, P.L. and Batzer, M.A. 1993. Evolution of retroposons. *Evolutionary Biol.* **27:** 157–196.

Eickbush, T.H. 1992. Transposing without ends: The non-LTR retrotransposable elements. *New Biol.* **4:** 430–440.

Feng, Q., Moran, J.V., Kazazian Jr., H.H., and Boeke, J.D. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87:** 905–916.

Freeman, J.D., Goodchild, N.L., and Mager, D.L. 1994. A modified indicator gene for selection of retrotransposition events in mammalian cells. *Biotechniques* **17:** 46–52.

Griesinger, C., Otting, G., Wuthrich, K., and Ernst, R.R. 1988. Clean TOCSY for ^{1}H spin system identification in macromolecules. *J. Am. Chem. Soc.* **110:** 7870–7872.

Hohjoh, H. and Singer, M.F. 1997. Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon. *EMBO J.* **16:** 6034–6043.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. 1979. Investigation of exchange processes by two-dimensional NMR spectroscopy. *J. Phys. Chem.* **71:** 4546–4553.

Jucker, F.M., Heus, H.A., Yip, P.F., Moors, E.H.M., and Pardi, A. 1996. A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **264:** 968–980.

Kajikawa, M. and Okada, N. 2002. LINEs mobilize SINEs in the Eel through a shared 3′ sequence. *Cell* **111:** 433–444.

Kazazian Jr., H.H. 2000. L1 retrotransposons shape the mammalian genome. *Science* **289:** 1152–1153.

Legault, P., Farmer II, B.T., Mueller, L., and Pardi, A. 1994. Through-bond correlation of adenine protons in a 13C-labeled ribozyme. *J. Am. Chem. Soc.* **116:** 2203–2204.

Legault, P., Li, J., Mogridge, J., Kay, L.E., and Greenblatt, J. 1998. NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: Recognition of a GNRA fold by an arginine-rich motif. *Cell* **93:** 289–299.

Luan, D.D. and Eickbush, T.H. 1995. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol. Cell. Biol.* **15:** 3882–3891.

Mathews, D.H., Banerjee, A.R., Luan, D.D., Eickbush, T.H., and Turner, D.H. 1997. Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA* **3:** 1–16.

Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian Jr., H.H. 1996. High frequency retrotransposition in cultured mammalian cells. *Cell* **87:** 917–927.

Nilges, M., Clore, G.M., and Gronenborn, A.M. 1988. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding. *FEBS Lett.* **239:** 129–136.

Ohshima, K., Hamada, M., Terai, Y., and Okada, N. 1996. The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. *Mol. Cell. Biol.* **16:** 3756–3764.

Okada, N. 1991. SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends Ecol. Evol.* **6:** 358–361.

Okada, N. and Hamada, M. 1997. The 3′ ends of tRNA-derived SINEs originated from the 3′ ends of LINEs: A new example from the bovine genome. *J. Mol. Evol.* **44:** S52–S56.

Ostertag, E.M. and Kazazian Jr., H.H. 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35:** 501–538.

Puglisi, E.V. and Puglisi, J.D. 1998. HIV-1 A-rich RNA loop mimics the tRNA anticodon structure. *Nat. Struct. Biol.* **5:** 1033–1036.

Sakamoto, T., Morita, S., Tabata, K., Nakamura, K., and Kawai, G. 2002. Solution structure of a SRP19 binding domain in human SRP RNA. *J. Biochem.* **132:** 177–182.

Schärpf, M., Sticht, H., Schweimer, K., Boehm, M., Hoffmann, S., and Rosch, P. 2000. Antitermination in bacteriophage λ. The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem.* **267:** 2397–2408.

Schmid, C. and Maraia, R. 1992. Transcriptional regulation and transpositional selection of active SINE sequences. *Curr. Opin. Genet. Dev.* **2:** 874–882.

Schmitz, U., Freymann, D.M., James, T.L., Keenan, R.J., Vinayak, R., and Walter, P. 1996. NMR studies of the most conserved RNA domain of the mammalian signal recognition particle (SRP). *RNA* **2:** 1213–1227.

Schmitz, U., Behrens, S., Freymann, D.M., Keenan, R.J., Lukavsky, P., Walter, P., and James, T.L. 1999. Structure of the phylogenetically most conserved domain of SRP RNA. *RNA* **5:** 1419–1429.

Weiner, A.M., Deininger, P.L., and Efstratiadis, A. 1986. Nonviral retroposons: Genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55:** 631–661.