Alu RNA Secondary Structure Consists of Two Independent 7 SL RNA-like Folding Units*

(Received for publication, November 15, 1990)

Daniel Sinnett, Chantal Richer,
Jean-Marc Deragon, and Damian Labuda
From the Service de Génétique Médicale, Hôpital Sainte-
Justine, Département de Pédiatrie, Université de
Montréal, 3175 Chemin de la Côte Ste-Catherine,
Montreal, Québec H3T 1CS, Canada

The amplification of genomic Alu elements by retro-
position, i.e., by reintegration of reverse-transcribed
RNA, suggests that Alu RNA plays an important role
in this process. We report enzymatic studies of the
secondary structure of Alu RNA transcribed in vitro
from two recently retroposed Alu elements. These
experiments show that the dimeric organization of an Alu
sequence is reflected in its RNA folding. Alu subunits
fold independently, conserving secondary structure
motifs of their progenitor 7 SL RNA molecule. Energy
minimization analysis indicates that this folding pat-
tern is also characteristic of different Alu and Alu-like
sequences and has been conserved since primate diver-
gence. By analogy to 7 SL RNA, the Alu RNA folding
may be important for specific interactions with pro-
tins. This could indicate a physiological function for
Alu transcripts. However, this can be also seen as a
structural adaptation leading to efficient retroposition
of these sequence elements.

Alu repeats are ubiquitous sequence elements that occupy
about 6% of the human genome (1, 2). Numerous reports (3-
5) of Alu involvement in mutations causing genetic disorders
rekindled interest in the mechanism of their dispersion and
its significance for evolution of the genome. As the prolifera-
tion of Alu elements originates in Alu RNA, its structure
might be responsible for the specificity of this process that
gave rise to hundreds of thousands of Alu copies, largely
outnumbering different families of pseudogenes also gener-
ated by retroposition (reviewed in Refs. 6 and 7). The dimeric
Alu element is composed of two homologous subunits, Alu-
left and Alu-right, held together by an A-rich linker and
terminated by an oligo-A tail. It has been shown (8) that each
of the subunits originated from 5' and 3' terminal segments
of 7 SL RNA in which these segments form a common folding
domain referred to as the "Alu fragment" (9–10). Since
typical Alu element retains about 80% sequence similarity
with homologous regions of 7 SL RNA, one may ask (i)
whether Alu RNA did conserve the folding pattern of its

* This work was supported by a grant from the Cancer Research
Society, Inc. The costs of publication of this article were defrayed
in part by the payment of page charges. This article must therefore be
hereby marked "advertisement" in accordance with 18 U.S.C. Section
1734 solely to indicate this fact.

† To whom correspondence should be addressed: Génétique Méd-
cale, Hôpital Sainte-Justine, 3175 Chemin de la Côte Ste-Catherine,
Montréal, Québec H3T 1CS, Canada.

progenitor 7 SL RNA molecule and (ii) whether the left and
right subunits fold independently or, due to regions of ex-
clusive complementarity, interact to form a common secondary
structure. Because no abundant source of the specific in vitro
Alu transcripts is known, it was difficult to address these
questions experimentally.

To circumvent this problem, we analyzed the structure of
Alu RNAs transcribed in vitro from DNA templates of two
members of the youngest Alu family (class IV (11) or Sb (12)).
Since these Alu elements, HAPF-Alu from the human α-
fetoprotein locus (13) and GBG-Alu from the gorilla β-globin
gene cluster (14), retroposed recently (i.e., after the divergence
of human and great ape lineages), they are expected to closely
represent the Alu sequence that is active in retroposition (11,12,15–18).

MATERIALS AND METHODS

DNA templates for an in vitro Alu RNA transcription by T7 RNA
polymerase (19) were prepared by polymerase chain reaction (cf., Ref.
20 for amplification conditions) from the genomic clone comprising
1.15 kilobase of intergenic sequence between β- and β-globin genes of
gorilla (GBG-Alu template) (11) and from the plasmid pHAPP.5
containing a 5.5-kilobase insert from human α-fetoprotein gene (13)
(HAPF-Alu template). The gorilla clone, originally from Trabuchet
et al. (14), was provided by R. J. Britten (California Institute of
Technology), whereas the human clone (13) was a gift from A.
Dugasceky (University of California, Riverside). The 5' polymerase
chain reaction primer, TAATACGACCTATAGGGGCCC-
GTTGGCTCA included the T7 RNA polymerase promoter (under-
lined sequence), whereas the 3' primer, TTTTTGGAGACGGAGT-
CTCGCTC, limited the oligo-A tail of the Alu sequence to 6 A residues
(both primers were synthesized using Gene Assembler, Pharmacia
LKB Biotechnology Inc.). RNA transcripts (19) were labeled with
32Pphosphate at either the 5' or 3' end as described (21) and purified
by polyacrylamide gel electrophoresis. By DNA sequencing, we found
two positions different from the published GBG-Alu sequence (14);
there is no deletion of C in position 158, and position 218 is G instead
of C.

To probe Alu RNA secondary structure, we used S1 nuclease, U2
(A-residue specific) and T1 (G-residue specific) RNases that cleave
in nonpaired, and V nucleotide that cleaves in double-stranded RNA
regions (all enzymes were from Pharmacia). The reactions were
carried out as described (21) using four 10-fold enzyme dilutions,
typically containing 20,000–40,000 cpm of the end-labeled gel-purified
Alu RNA in a final volume of 5 μl. The results were analyzed on 12%
polyacrylamide sequencing gel (if not stated otherwise). Minimal
energy RNA folding was analyzed using version 4.0 (for IBM PC) of
the FOLD program (22), kindly provided by M. Zuker (National
Research Council of Canada, Ottawa).

RESULTS

The templates for in vitro synthesis of both Alu RNAs by
T7 RNA polymerase (19) were prepared from the correspond-
ging genomic clones by polymerase chain reaction, placing
the RNA polymerase promoter such that transcription started
with the first G nucleotide of the Alu sequence. The resulting
Alu RNA were labeled with radioactive phosphate either at
the 5' or at the 3' end and subjected to enzymatic digestions
using single-strand-specific or double-strand-specific nu-
clease (Fig. 1). The structure of Alu RNA fragment that

8675
Evolutionary Conservation of Alu RNA Secondary Folding

Fig. 1. Enzymatic digestions of 5' end-labeled HAFP-Alu RNA (a) and GBG-Alu RNA and its R fragment labeled at the 3' ends (b). S₁, V₁, U₂, and T₁, corresponding digestions carried out in native conditions with serial 10-fold enzyme dilutions 1-4; T₁S, sequencing reaction using T₁ nuclease; L, partial formamide hydrolysis of full-length RNA in b or its truncated fragment in a; n and R, undigested full-length RNA and the R fragment, respectively. RNA fragments resulting from S₁, V₁, cleavage (indicated on the left side) that terminate in 3'-OH do not migrate exactly as those ending with 3'-phosphate (indicated on the right). The numbers refer to nucleotide positions in the Alu general consensus sequence (18). Note that doubling of bands in the 3'-labeled RNA is due to the heterogeneity in the number of A residues at the 3' end of the in vitro Alu transcript.

digestion pattern being virtually identical with that in the corresponding region of the full-length RNA (Fig. 1b, data not shown) pointed to an independent folding of the Alu subunits. In this case, a partial digestion of Alu RNA with S₁ nuclease, shown to cut within the linker (Fig. 1), would lead to a separation of subunits during polyacrylamide gel electrophoresis in nondenaturing conditions.

As shown in Fig. 2a, the incubation of the 3'-labeled full-length Alu RNA with increasing amounts of S₁ nuclease leads to a single major band that corresponds to the 3'-labeled Alu right subunit migrating with the same electrophoretic mobility as the R fragment. In a control experiment, the aliquots of the reactions analyzed in Fig. 2a were subjected to electrophoresis under denaturing conditions. The results shown in Fig. 2c demonstrate that S₁ nuclease also cuts susceptible sites of Alu RNA in other than the linker region (cf. also Figs. 1 and 3). However, these fragments, being kept together by base pairing within the secondary structure, are not revealed by the nondenaturing electrophoresis shown in Fig. 2a. Associated, they migrate as an intact Alu RNA or its subunits, unless the S₁-incubated samples are heat-denatured before loading the native gel (Fig. 2b). In conclusion, there is no evidence for secondary interactions between Alu-left and Alu-right that are strong enough to keep the structure together once the covalent bonds within the linker have been cleaved. This indicates that Alu RNA subunits fold independently, which is consistent with the more detailed analysis of the enzymatic digestion data presented below.

Fig. 3, a and b, summarizes the results of enzymatic digestions obtained in a series of experiments such as illustrated in Fig. 1. In the case of the HAFP-Alu RNA (Fig. 3a), the cleavage pattern is consistent with a model where each subunit folds like the Alu fragment of 7 SL RNA shown in Fig. 3c (representing a predicted minimal energy structure of 7 SL RNA that is virtually identical with that derived from enzymatic digestions and evolutionary comparisons (9)). Both Alu-left and Alu-right RNA conserve hairpins I and II as well as the domain III. Nearly the same structure is found for GBG-Alu RNA, except that hairpins I and II in the right subunit form a mixed helix I/II (Fig. 3b), thus departing from the 7 SL RNA scheme. This seems to involve a U → G transversion at position 174 in GBG-Alu RNA that favors the helix I/II by allowing a wobble pair between G-174 and U-144. This substitution is absent in the HAFP sequence that conforms at this position with consensus sequences of human Alu families (cf. Ref 18 for the weighted consensus sequences of human Alu families used in this work, according to the original classification (12)). Although there is no evidence for a mixed helix I/II in the right subunit of HAFP-Alu RNA, a weak T₁ RNase susceptibility of G at positions 176-178 (cf. Fig. 3a) might indicate the lability of hairpins I and II in the right subunit and a possibility of conformational flexibility in this region.
be common among different primate Alu and Alu-like RNAs

age of 3'-labeled HAFP-Alu RNA by S1 nuclease was analyzed on

same reactions were analyzed in heat-denatured prior to native electrophoresis. The numbers at the sides indicate the fragment sizes.

amounts of enzyme (from 10⁻⁴ to 37°C) samples 0 and n had no enzyme added with and without incubation, respectively; R denotes the R fragment. Aliquots of the same reactions were analyzed in a-c, except that samples in b were heat-denatured prior to native electrophoresis. The numbers at the sides indicate the fragment sizes.

The 7 SL-like folding pattern shown in Fig. 3a appears to be common among different primate Alu and Alu-like RNAs that were analyzed by Zuker’s minimal energy approach (22) using as RNAs the consensus sequences of the corresponding genomic elements (data not shown). This is the case for Galago AluI RNA (23) and the right 7 SL-like subunit of Galago AluII (24). This turns out to be the case for RNAs of different human Alu families when energy minimization is carried out separately for the right and the left subunit. Although the minimal energy structure predicted using the whole sequence involves subunit interaction through regions of complementarity corresponding to hairpins I in the Alu-left and Alu-right (cf. Ref 18), its energy minimum is only slightly lower (2.5% in the Alu Sb and 3.3% in the Alu J family) than that of the separate arrangement of subunits. The pattern of nucleotide substitutions among consensus sequences of human Alu families analyzed earlier (18) is in agreement with the model in Fig. 3a (most of these substitutions are either compensatory or conservative in character as far as the secondary structure model is concerned). Interestingly, the “consensus” RNAs of the Alu-like rodent B1 elements also fold according to the 7 SL RNA scheme (31) and this appears to be the case for BC200 RNA from the brain of cynomolgus monkey (25), similar to the left subunit of the human Alu family J.

DISCUSSION

As shown by recent analyses of genomic Alu elements, different Alu families were generated via retroposition from a limited number of conserved master sequences (11, 12, 15-18). Experimental evidence presented in this paper provides arguments for the notion that evolutionary conservation of Alu master sequences is due to selection for invariance in folding characteristics from 7 SL RNA (18). This appears to be also true for Alu-like retroposons as judged from their folding pattern predicted using Zuker’s approach. Dimeric arrangement of 7 SL-like folding units is presumably associated with a certain functional advantage for Alu RNA and/or its proliferation. Examples of monomeric (rodent B1 elements, BC200 RNA) or heterodimeric arrangements (Galago Alu II) of different 7 SL-like molecules corroborate well with our observation that Alu subunits fold independently and suggest that the dimeric structure represents an advanced form in Alu evolution.

Although sequences of Alu RNAs used in our experiments differ in some positions from the exact consensus of their family, these differences occur in sequence positions that do not affect the overall 7 SL-like folding pattern of these molecules. Enzymatic digestions confirm the minimal energy/folding of HAFP-Alu RNA and, except for the hairpin I/II, that of GBG-Alu RNA, both of which are consistent with the folding of Alu RNA predicted separately for right and left subunits derived from the four human Alu family consensus sequences. Among nine sequence positions, disregarding a different number of A residues in the linker by which HAFP-Alu RNA differs from the consensus sequence of the Alu family Sb, only one (C⁻²→G) affects base pairing in the model in Fig. 3a. A similar situation is seen with the GBG sequence. The question may be asked whether differences between these sequences do reflect different evolution of the parental templates or are simply due to substitutions introduced after their insertion in the genome (26). Compared to the consensus, four out of the nine substitutions in HAFP-Alu and three out of the six in GBG-Alu are due to transitions in CpG dinucleotides, as expected from the number of CpG dinucleotides in the consensus and their higher substitution rate (18, 26). Assuming a substitution rate of about 0.15%/nucleotide/million years (27), one would expect an accumulation of 0.42 substitution/Alu sequence/million years since its retroposition, or about twice as much when the effect of CpG is taken into account. For HAFP-Alu and GBG-Alu that could have been inserted 1–5 million years ago (13, 14), the number of differences with respect to the consensus appears, therefore, to be higher, from the estimated figure of 0.8–4 substitutions/sequence. This indicates that their parental templates could have already evolved from the precise consensus and that sequences of their RNAs may represent even closer Alu master sequences active in retroposition.

In signal recognition particle, 7 SL RNA interacts with protein components to form a specific ribonucleoprotein complex (10, 28). This suggests that 7 SL-like folding in Alu RNA maintains specific interactions with proteins (27). The resulting ribonucleoprotein could mediate retroposition of Alu RNA and thus explain specificity of this process that has led to proliferation of Alu repeats to an extent far exceeding other sequences that only occasionally generate retropseudogenes. The requirement for appropriate folding could also explain why Alu retroposition only occurs from the conserved master sequences (11, 12, 14-18). The relatively high number of 7 SL-RNA pseudogenes (29) indicates that 7 SL RNA, itself prone to retroposition, was a good starting point for an efficient retroposition element to evolve, presumably in the context of a more complex retroposition system (6, 7, 30). Further experiments are required to establish the nature of the putative Alu ribonucleoprotein and its relation to retroposition and to the host cell. An in vitro transcription system of Alu RNAs, such as used in the present study, could provide a useful tool for these investigations.

Acknowledgments—We are grateful to George Striker for critical reading of the manuscript, to Roy J. Britten and Achilles Dugnicyz for sending us clones containing the Alu sequences, and to Michael Zuker for providing the latest PC version of the FOLD program.
Fig. 3. Secondary structure model of Alu RNA based on digestions results (a and b) compared with that of the Alu fragment of 7 SL RNA (c). Cleavage sites are identified by different symbols for enzymes used as indicated. In a, a model based on the HAFP-Alu RNA sequence compiles the data of different digestion experiments for both RNAs, except for the fragment corresponding to hairpins I and II in the GBG-right subunit (L and II), where digestion results point to a different folding shown in b, and for the fragment between positions 250 and 260, where the cleavage pattern slightly differs in GBG-Ah, presumably due to deletion of G-213. Numbering of sequence positions is according to general Alu consensus (18); i.e., deletions at positions 65 and 66, characteristic of Alu family Sb, are counted, whereas three A residues following position 133 are not. In a, sequence positions that differ between HAFP and GBG sequences are indicated in lower case. In c are shown only 7 SL RNA sequence segments that are homologous to Alu subunits; lower case letters indicate fragments that are absent in the Alu-left subunit. Subscripts L and r are used to distinguish homologous domains I, II, and III in the left and right subunits, respectively.

REFERENCES

1. Rinehart F. P., Ritch, T. G., Deininger, P. L., and Schmid, C. W. (1981) Biochemistry 20, 3003-3010
2. Hwu, H. R., Roberts, J. W., Davidson, E. H., and Britten, R. J. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 3875-3879
3. Lehman, M. A., Goldstein, J. L., Russell, D. W., and Brown, M. S. (1987) Cell 48, 827-835
4. Merowitz, R., and Hogikyan, N. D. (1987) J. Biol. Chem. 262, 15396-15399
5. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
6. Rogers, J. H. (1985) Int. Rev. Cytol. 93, 187-279
7. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
8. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
9. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
10. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
11. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
12. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
13. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
14. Rouyer, F., Simmler, M.-C., Page, D. C., and Weissenbach, J. (1987) Cell 51, 417-425
15. Willard, C., Nguyen, H. T., and Schmid, C. W. (1987) J. Mol. Biol. 206, 180-186
16. Deininger, P. L., and Slagel, V. K. (1988) Mol. Cell. Biol. 8, 4566-4569
17. Quentin, Y. (1989) J. Mol. Biol. 208, 299-305
18. Lahaida, D., and Striker, G. (1988) Nucleic Acids Res. 17, 2477-2491
19. Milligan, J. F., Groebe, D. R., Witherell G. W., and Uhlenbeck, O. C. (1987) Nucleic Acids Res 15, 8783-8789
20. Sinnett, D., Deragon, J.-M., Simard, L. R., and Labuda, D. (1990) Genomics 7, 331-334
21. D'Alessio, J. M. (1982) Gel Electrophoresis of Nucleic Acids (Rickwood, D., and Hames, B. D., ed) pp. 173-198, JRL Press, Oxford
22. Zuker, M. (1989) Methods Enzymol. 180, 262-288
23. Daniels, G. R., and Deininger, P. L. (1983) Nucleic Acids Res. 11, 7599-7610
24. Daniels, G. R., and Deininger, P. L. (1983) Nucleic Acids Res. 11, 7599-7610
25. Watson, J. B., and Sutcliffe, J. G. (1987) Mol. Cell. Biol. 7, 3324-3327
26. Britten, R. J., Stout, D. B., and Davidson, E. H. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 3718-3722
27. Britten, R. J. (1987) Science 231, 1393-1398
28. Ullu, E., and Weiner, A. M. (1984) EMBO J. 3, 2325-2332
29. Ullu, E., and Weiner, A. M. (1984) EMBO J. 3, 2325-2332
30. Ullu, E., and Weiner, A. M. (1984) EMBO J. 3, 2325-2332
31. Deragon, J.-M., Sinnett, D., and Labuda, D. (1990) EMBO J. 9, 3363-3369
32. Deragon, J.-M., Sinnett, D., and Labuda, D. (1990) EMBO J. 9, 3363-3369