Alu: a parasite's parasite?

Carl W Schmid

The technical problem of detecting the retrotransposition—amplification and genomic dispersion of sequences through an RNA intermediate—of Pol III–directed transcripts has been solved. This provides a long-sought experimental system to learn how Alu repeats amplified to a high copy number in human DNA.

About fifty percent of human DNA consists of currently active retrotransposable elements, most notably Alu and LINE1 elements, and highly decomposed fossils of similar elements, some of which predate mammalian radiation[1]. There are over one million Alu family members, which share a 282-nucleotide consensus sequence, account for 10% of human DNA alone and are ubiquitously dispersed throughout the genome but highly overrepresented in gene-rich regions of human DNA. At a coarse level of resolution, the organization of our genome resembles islands of genes surrounded by an ocean of Alu elements[1]. LINE1 repeats account for a comparable mass fraction of our genome[1].

In this issue, Dewannieux *et al.*[2] show that LINE repeats drive the amplification of Alu elements. The broad dispersion of retrotransposed elements has shaped most eukaryotic genomes. Alu repeats, which first appeared in early primate phylogeny, are still actively expanding in human populations. Their insertion alters the structure and expression of genes, providing another source of genetic variation including that associated with inherited diseases. Given these consequences, how and why Alus amplified to such a high copy number has been a mystery since they were first reported in *Nature* in 1980 (ref. 3).

Line1 function

Only a few members of the LINE1 family of highly repetitive retrotransposable sequences are capable of autonomous amplification. Most LINE elements share consensus 3′ sequences but are truncated at various positions in their 5′ sequences. Full-length LINE1 is bicistronic: the product of ORF1 is an RNA-binding protein (ORF1p), and ORF2 encodes a protein (ORF2p) with endonuclease and reverse transcriptase activities (**Fig. 1**). The downstream location of ORF2 ensures that translational initiation of these catalytic activities is downregulated with respect to ORF1p expression. Both ORF1p and ORF2p are required for LINE1 retrotransposition, but surprisingly, ORF2p acts efficiently only on the active LINE element that encoded its expression (in *cis*; **Fig. 1**). A retrotranspositionally successful LINE must sequester its limiting ORF2p, preventing the amplification of either defective LINE elements or entirely unrelated sequences, such as Alu (**Fig. 1**).

Our understanding of LINE1 retrotransposition primarily results from an ingenious experimental system that uses a backward selectable marker gene that is activated only by the retrotransposition of its LINE RNA intermediate. Post-transcriptional processing of this Pol II–directed transcript removes an intron that otherwise inactivates the backward marker gene[2]. But because Alu elements are transcribed by Pol III, this system is useless for examining their retrotransposition. Consequently, compared with large conceptual advances in our understanding of LINE elements, our knowledge of Alu retrotransposition has, until now, languished. By substituting a self-splicing intron and making other targeted changes, Dewannieux *et al.* have now adapted this assay to the efficient detection of Alu retrotransposition[2].

A close and complex relationship

Pol III transcripts generally do not encode proteins, and circumstantial evidence has long suggested that the retrotransposition of Alu elements, and analogous sequences in other eukaryotes, requires *trans*-acting factors, which are provided by companion LINE elements[4,5]. By showing high levels of *de novo* Alu retrotransposition in cells transiently overexpressing Alu RNA and LINE1 proteins, Dewannieux *et al.*[2] provide the first direct evidence for this model and elucidate the

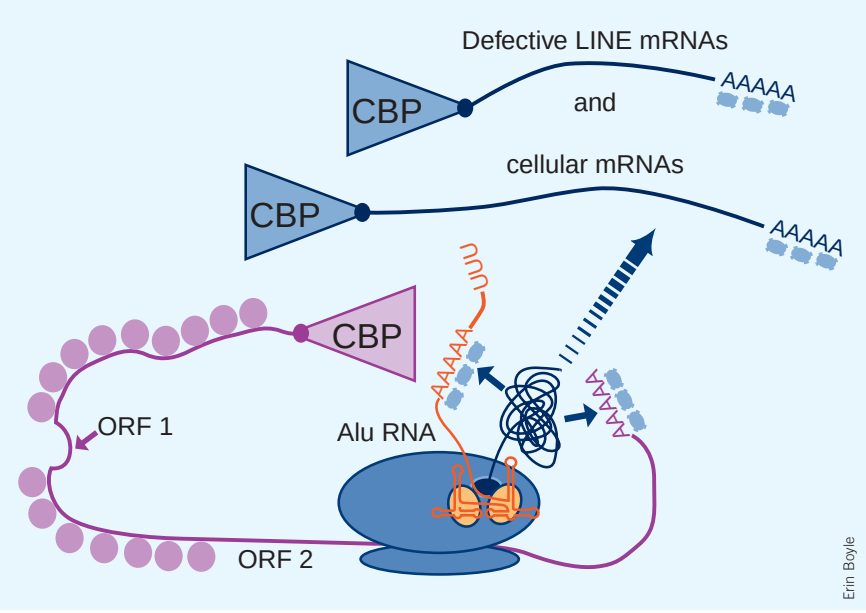

Figure 1 Co-compartmentalization of Alu RNA and LINE1 mRNA with *cis*-acting ORF2p. Bicistronic LINE1 mRNA encodes an RNA binding protein, ORF1p (circles), and catalytic functions, ORF2p (tangled line). A-rich regions, presumably in association with poly-A binding protein, are required for retrotransposition. ORF2p acts more efficiently on colocalized (*cis*) Alu and LINE1 RNAs than on cellular and defective LINE1 mRNAs. SRP 9/14 proteins may direct Alu RNA to the ribosome, and the interaction of cap-binding protein (CBP; triangle) with poly-A binding protein is another possible link between LINE1 and Alu RNAs with ORF2p activity.

The author is in the Molecular and Cellular Biology Section, University of California-Davis, Davis, California 95616-8535, USA.
e-mail: cwschmid@ucdavis.edu

required components for Alu amplification.

As expected, catalytic activities encoded by ORF2 are essential for efficient Alu retrotransposition but exogenous ORF1p is dispensable. One caveat in interpreting this result is that endogenous ORF1p may suffice for Alu retrotransposition[2], a possibility that will be resolved in future experiments. If Alu does not require ORF1p, a large number of truncated LINE elements, lacking ORF1 and incapable of autonomous retrotransposition, might provide sufficient ORF2 activity for Alu amplification. Accordingly, the requirements of Alu for retrotransposition could be far more permissive than those of LINE1, partially explaining the relative success of Alu elements. But even if Alu can scavenge ORF2p from defective LINE elements, how does Alu outcompete other potential RNA targets for ORF2p, which is both limiting and *cis*-acting (**Fig. 1**)?

Alu elements, like analogous elements in other eukaryotes, typically have an A-rich 3′ tail. Based on earlier results for LINE1 retrotransposition, Boeke proposed that this region is essential for Alu mobility[6] and, indeed, deletion of the A-rich tail virtually abolishes Alu retrotransposition[2]. Nearly all mRNAs have 3′ poly-A tails, implying that other *cis*-acting elements must deliver Alu elements to the source of ORF2p. Alu elements are ancestrally derived from the SRP RNA gene and Alu RNA binds SRP proteins 9 and14 (ref. 7). The translational role of SRP and its ribosomal location provide a plausible mechanism for co-compartmentalizing Alu RNA with nascent *cis*-acting ORF-2p (**Fig. 1**). Unfortunately, the SRP 9/14–binding site of Alu happens to overlap its internal promoter, which complicates an immediate test of this attractive model[2].

The biological function of LINE elements is unknown, and, as evidenced by their persistence in eukaryotic evolution, LINE1 may just be a genomic parasite, or selfish DNA that evolved to ensure its own perpetuation. We shall soon learn the biochemistry of Alu retrotransposition and, in particular, how it efficiently exploits the limited but essential LINE1-encoded factors. The function of Alu elements is also unknown. If Alu elements are also functionless, one genomic parasite has very successfully parasitized another genomic parasite and collectively, these sequences, as well as their now-fossilized predecessors, have populated, indeed generated, most of our genome. A competing possibility is that Alu elements serve one or more functions that require and drive their ubiquitous dispersion throughout the genome. The observation that Alu elements and similar elements in other animals behave like classic cell-stress genes suggests intriguing but unproven possibilities[8]. Our complete understanding of eukaryotic genome evolution will ultimately require determining both how and why Alu elements have amplified to such high copy numbers. The impressive progress that is now being made toward answering the first question may ultimately address the second as well.

1. Lander, E.S. *et al. Nature* **409**, 860–921 (2001).
2. Dewannieux, M., Esnault, C. & Heidmann, T. *Nat. Genet.* **35**, 41–48 (2003).
3. Rubin, C.M. *et al. Nature* **284**, 372–374 (1980).
4. Kajikawa, M. & Okada, N. *Cell* **111**, 433-444 (2002).
5. Jurka, J. *Proc. Natl. Acad. Sci. USA* **94**, 1872–1877 (1997).
6. Boeke, J.D. *Nat. Genet.* **16**, 6–7 (1997).
7. Sarrowa, J., Chang, D.Y. & Maraia, R.J. *Mol. Cell Biol.* **17**, 1144–1151 (1997).
8. Li, T-H. *et al. Gene* **239**, 367–372 (1999).