

Molecular evolution

On the origin of the *Alu* family of repeated sequences

from Andrew Leigh Brown

7SL RNA IS an example of a eukaryotic RNA species that was identified long before a cellular role could be ascribed to it. Only recently was it established¹ that it forms an essential part of the signal recognition particle — the cellular apparatus that aids the transport of secreted proteins across the endoplasmic reticulum. Subsequently nucleotide sequence analysis revealed that parts of the 7SL RNA sequence seem to derive from the *Alu* element², which is repeated about 300,000 times in the human genome³. While *Alu* is usually found as a head-to-tail dimer of a sequence about 150 base pairs long, in the 7SL sequence a single *Alu* monomer seems to have suffered an insertion of 155 base pairs. This was not in itself a dramatic discovery because *Alu* elements are dispersed so widely throughout the human genome that they could easily be a target for random DNA insertion. However, the latest developments concerning *Alu* and 7SL RNA^{4,5}, some of which are published on page 171 of this issue, have taken a quite unexpected turn.

Ullu and Tschudi⁴ have sequenced 7SL RNA (as a DNA clone) from *Xenopus* and *Drosophila*. Both are clearly related to the human counterpart throughout their length — that is the sequence homology extends to the *Alu*-like part of the human molecule. The presence of an *Alu*-like sequence in *Drosophila*, also found by Gundelfinger *et al.*⁵, means that *Alu* has had a much longer evolutionary history than was previously thought. More interesting yet, in *Drosophila* it seems only to exist as part of the 7SL DNA, of which there are two copies. On this basis, Ullu and Tschudi suggest that *Alu* sequences represent defective 7SL RNA molecules that have been reverse-transcribed into DNA and inserted into the genome. An analogous origin has been suggested for alpha-globin pseudogenes in the mouse⁶, and the multiple pseudogenes for small nuclear RNAs in man⁷. Pseudogenes are generally thought not to play an important role in the cell. Perhaps those who have argued that *Alu*, by its very abundance, must have an important function will recognize that this argument has now lost some of its weight.

Although it seems likely that the 7SL sequence with its single *Alu* evolved in the absence of repeated *Alu* elements, an alternative interpretation of Ullu and Tschudi's data presents itself, at first sight. In the 7SL-specific region of the sequences the match between the human and *Drosophila* sequences is 67 per cent, and the correct alignment is obvious. In the *Alu*-related re-

gions the degree of match is less and the two groups opt for rather different alignments, especially at the 3' end. If, in fact, these regions of the sequence are completely unrelated, then the earlier hypotheses could still be tenable. However, using the method of Fitch and Smith⁸, I find that the alignment of the *Drosophila* sequence with the 3' *Alu*-related part of the human 7SL sequence suggested by Ullu and Tschudi has a probability of less than 0.01 of occurring by chance. Thus the flanking regions of the *Drosophila* 7SL RNA are homologous to *Alu* in man and the *Alu* sequence first existed as a part of this molecule.

As expected, for *Xenopus laevis* the overall match with the human sequence is much higher⁵ than either is with *Drosophila*, although it drops off in 3' *Alu*-related region. Although their previous results indicate only weak homology to *Alu* in *Xenopus* DNA⁹, Ullu and Tschudi describe the *Xenopus* genome as being 'rich' in sequences related to the *Alu*-like parts of the *Xenopus* 7SL molecule. So, in *Xenopus* as well as man, there is a close relationship between the 7SL RNA and highly repetitive elements, although the *Xenopus* and human sequences have diverged during evolution. How did this relationship arise? The fact that 7SL has a longer evolutionary history, being found in the absence of any *Alu*-related repeats in *Drosophila*, strongly suggests it was the antecedent. Moreover, the 7SL RNA is arranged in the signal recognition particle in such a way that three RNA fragments similar to the 5' *Alu*-related, 7SL-specific and 3' *Alu*-related sequence domains are generated by limited enzymatic digestion of the intact particle¹⁰. Put that together with the circumstance in which the 7SL RNA was first discovered — in a retroviral particle¹¹ and therefore close to a source of reverse transcription activity — and the generation of DNA copies of *Alu* is easy to imagine.

Amplification

Once *Alu* had appeared, what happened next? It has been known for some time that *Alu* is a remarkably stable component of the genome. Having arisen at a specific location, it does not seem to disappear quickly. For example, in the gamma-globin region, two *Alu* elements are present in exactly the same place in both man and chimpanzee¹². This is in striking contrast to the vagility shown by the middle repetitive DNA elements in *Drosophila*, and until now there has been no real explanation for the difference. However, if *Alu* originated

as a 7SL pseudogene it may not possess or respond to any mechanism for excision or transposition. Thus, unless it kills the chromosome in which it inserts, it might just have to stay there. If it retains polymerase III promoter activity (not all copies do retain it), it could even generate more transcripts, each of which might be another possible template for reverse transcription and reinserion. One would therefore expect copies to accumulate in the genome.

Two questions are left. First, did the events which gave rise to the *Alu*-related repetitive elements occur once in each of the *Xenopus* and human lineages independently? If so, it might explain the sequence divergence between repeats of the two species; within the species the repeat sequence very closely resembles the 7SL sequence. However, we should remember that some concerted evolution has probably taken place. It is possible to envisage repeats already existing in the primitive amphibian DNA and gradually, after the two lineages separated, changing *en bloc*. Any subsequent generation of pseudogenes would be of the species-specific 7SL sequence; a process of concerted evolution which kept the 7SL and *Alu*-related repeats reasonably similar within each species is not unlikely.

The second question is, why has none of this occurred in *Drosophila*? Here there are no easy answers. The situation is similar to the case of the small nuclear RNAs, where many pseudogenes are found in man⁷, but not in *Drosophila*¹³. Processed pseudogenes of protein coding regions also seem to be absent in the fly. There is certainly no absence of reverse transcriptase activity which might explain this — generation of reverse transcripts may be a major mechanism for replication of *copia*¹⁴. Nevertheless, reverse transcripts do not accumulate in the genome of *Drosophila* in the way they have in vertebrates. When we know why this is so, we shall be well on the way to understanding why the genomes of insects are so very much smaller than those of vertebrates despite containing much the same amount of information. □

1. Walter, P. & Blobel, G. *Nature* **299**, 691 (1982).
2. Ullu, E., Murphy, S. & Melli, M. *Cell* **29**, 195 (1982).
3. Jelinek, W.R. & Schmid, C.W. *A. Rev. Biochem.* **51**, 813 (1982).
4. Ullu, E. & Tschudi, C. *Nature* **312**, 171 (1984).
5. Gundelfinger, E.D. *et al.* *EMBO J.* **3**, 2325 (1984).
6. Nishioka, Y., Leder, A. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2806 (1980).
7. Denison, R. A., van Arsdell, S. W., Bernstein, L. B. & Weiner, A. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 810 (1981).
8. Fitch, W. F. & Smith, T. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1382 (1983).
9. Ullu, E., Esposito, V. & Melli, M. *J. molec. Biol.* **161**, 195 (1982).
10. Gundelfinger, E. D., Krause, E., Melli, M. & Doberstein, B. *Nucleic Acids Res.* **11**, 7363 (1983).
11. Bishop, J. M. *et al.* *Virology* **42**, 927 (1970).
12. Maeda, N., Bliska, J. B. & Smithies, O. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5012 (1983).
13. Saluz, H. P. *et al.* *Nucleic Acids Res.* **11**, 77 (1983).
14. Flavell, A. *Nature* **310**, 514 (1984).

Andrew Leigh Brown is in the Department of Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK.