elles when crude axoplasmic extracts are applied to polar microtubule arrays. Natural organelles exhibit smooth continuous movement in only one direction, whereas polystyrene beads have a jerky motion, reversing direction frequently. The beads, which have a negative surface charge, might be expected nonspecifically to bind many proteins from the crude extract. Organelle movement in this assay is consistent with the presence on the surface of force-generating machinery for both directions of movement. This raises the possibility that the surface properties of endogenous organelles specify which direction they will travel by determining which mechanochemical system they can bind, and suggests a simple mechanism by which the cell can direct traffic. □

1. Brady, S.T., Lasek, R.J. & Allen, R.D. *Cell Motil.* **5**, 81 (1985).
2. Vale, R.D. *et al. Cell* **43**, 623 (1985).
3. Hollenbeck, P. *Nature News and Views* **317**, 17 (1985).
4. Vale, R.D., Reese, T.S. & Sheetz, M.P. *Cell* **41**, 39 (1985).
5. Mitchison, T. & Kirschner, M. *Nature* **312**, 232 (1984).
6. Miller, R.H. & Lasek, R.J. *J. Cell Biol.* **101**, 2181 (1985).

Peter Hollenbeck is in the MRC Cell Biophysics Unit, 26–29 Drury Lane, London WC2B 5RL, UK.

Repeated sequences

The origin of retroposons

from John Rogers

THE chromosomal DNA of mammals is prolifically infested with dispersed repeated nucleotide sequences of various types, scattered through the non-coding regions of the genome. One of the most abundant and most mysterious types is the long interspersed nucleotide element or LINE class (see figure). Two recent advances have produced an extraordinary new view of these repeated sequences.

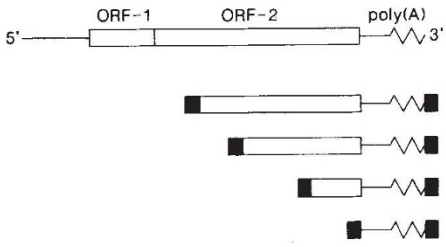

Diagram of the arrangement of a full-length LINE-1 sequence (top line) and of 5′-truncated dispersed repeats (lower lines) flanked by short direct repeats (black squares). ORF, open reading frame.

Understanding of these sequences has been hindered by their complexity, but it is clear that they all belong to a single family (LINE-1), both in rodents and in primates, which has a consensus structure about 6 kilobases long, although most copies are randomly truncated 3′ fragments[1–4]. Most copies have a poly(A) (or A-rich) tail at the 3′ end and are flanked by short repeats. This suggests that LINEs are inserted reverse transcripts (retroposons[5]), similar to processed pseudogenes and many other dispersed repeats. Their origin and operation remain to be elucidated, but the new developments allow these problems to be formulated with greater precision.

First, a full-length LINE-1 transcript has been discovered in a human teratocarcinoma cell line[6]. Heterogeneous LINE-1 transcripts are common in nuclear RNA, but this is the first report of an abundant, full-length, 'plus'-strand, cytoplasmic LINE-1 RNA. Such a transcript is required by the reverse transcription theory of LINEs. Its presence in a teratocarcinoma cell line, which may represent the earliest cleavage stage of the embryo, and its down-regulation when these cells differentiate, are of special interest with respect to retroposon origins. A family of short retroposons in mice, called B2, is also abundantly transcribed in the early embryo[7]. And it is noteworthy that mammals, which begin transcribing the zygotic genome at the two-cell stage, contain numerous retroposon copies of many kinds of RNA, whereas other animals, which do not begin transcription until later, have few retroposons[5]. Therefore, the cleavage-stage embryo could be where new retroposons are made.

The second major advance is the compilation of full-length LINE-1 DNA sequences, which show that LINEs are derived from parental copies with long conserved open reading frames. Maxine Singer and colleagues have compiled a collection of random primate specimens into a 'pastiche' complete sequence, which has several open reading frames that could be joined by a few point mutations[4]. Others have sequenced complete single specimens from human[8] and mouse[9]; the human specimen has many interruptions in the reading frames but the mouse has two perfect open reading frames spanning 5 kilobases. This seems to be the arrangement from which other LINE-1 specimens have degenerated. Although most of them are defective in the open reading frames because of 5′-truncation, frameshifts and/or stop codons, comparisons of many specimens show clearly that the parental open reading frames have been evolving under selction for their coding potential since the mammalian radiation[9–11].

So what is the function of the open reading frames? And why have these genes produced a thousand-fold more processed pseudogenes than have other genes? Two alternative theories can now be formulated. One possibility is that LINEs are 'selfish' genes in the manner of retroviruses or transposons, propagating via transcription and reverse transcription, and that the open reading frames encode proteins involved in this process. Indeed, in their size and arrangement, the two open reading frames (joined by a short overlap) resemble the *gag–pol* region of retroviruses[9]. Loeb *et al.*[9] point out two other features of the full-length mouse LINE-1 sequence which enable them to construct a novel theory along these lines. First, there is a possible homology (on the borderline of significance) between the larger open reading frame and the reverse transcriptases of retroviruses. Second, there is an array of tandem repeats at the 5′ end, 4⅔ in one specimen and 1⅔ in another, which Loeb *et al.* suggest may carry promoters. Stochastic expansions of this array could solve the major problem in propagating a gene through multiple rounds of transcription and reverse transcription, which is to regenerate the promoter at the 5′ end.

However, these 5′ repeats may be a recent acquisition in mouse LINE-1, as they are absent from primate LINE-1. In contrast, the coding sequences have been conserved in all mammalian orders, evolving in parallel with the whole organism[12]. This is consistent with the alternative possibility, that there is a conserved parental LINE gene with a function for the organism. The vast number of retroposon copies could be explained if this gene normally undergoes reverse transcription as a physiological process, of which the dispersed copies are occasional by-products.

Both theories propose that LINE-1 RNA is the normal substrate for reverse transcription to form retroposons. If so, all the other retroposons in mammals may be by-products of the same machinery, which occasionally acts on the wrong substrate RNA. Fortunately, the complete DNA sequences not only permit these theories to be formulated, but also open up the possibility of testing them. □

1. Voliva, C.F. *et al. Nucleic Acids Res.* **11**, 8847 (1983).
2. Grimaldi, G., Skowronski, J. & Singer, M.F. *EMBO J.* **3**, 1753 (1984).
3. Rogers, J.H. *Nature News and Views* **306**, 113 (1983).
4. Singer, M.F. & Skowronski, J. *Trends biochem. Sci.* **10**, 119 (1985).
5. Rogers, J.H. *Int. Rev. Cytol.* **93**, 187 (1985).
6. Skowronski, J. & Singer, M.F. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6050 (1985).
7. Vasseur, M., Condamine, H. & Duprey, P. *EMBO J.* **4**, 1749 (1985).
8. Hattori, M., Hidaka, S. & Sakaki, Y. *Nucleic Acids Res.* **13**, 7813 (1985).
9. Loeb, D.D. *et al. Molec. cell. Biol.* **6**, 168 (1986).
10. Martin, S.L. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 2308 (1984).
11. Hardies, S.C. *et al. Molec. Biol. Ecol.* (in the press).
12. Burton, F.H. *et al. J. molec. Biol.* **187**, 291 (1986).

John Rogers is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.