Nucleotide sequence of a mouse full-length F-type L1 element

Nils B.Adey, Mary B.Comer, Marshall H.Edgell and Clyde A.Hutchison III*
Department of Microbiology and Immunology, Curriculum in Genetics, Program in Molecular Biology and Biotechnology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

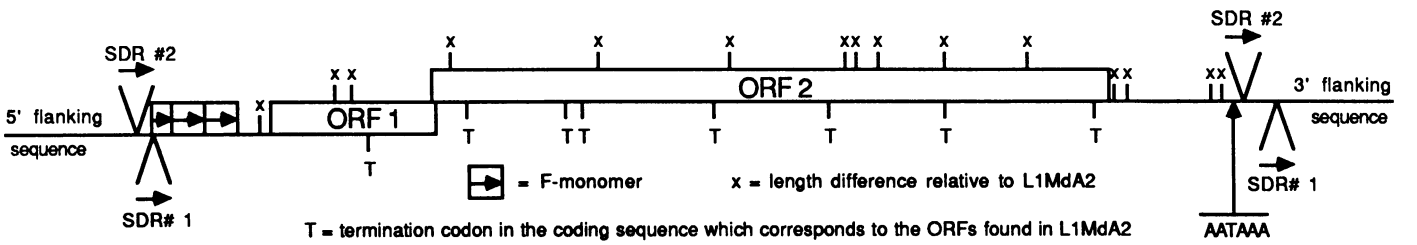
Submitted February 21, 1991 EMBL accession no. X57795

L1 is a repetitive sequence dispersed throughout the mammalian genome and is believed to be a transposable element that replicates via an RNA intermediate (5, 7). For review see Hutchison *et al.* (2). Two types of full-length L1s have been reported to exist in mouse, A-type and F-type, depending on the type of 200 bp sequence (known as a monomer) that is tandemly repeated at the 5′ end (1, 3, 4). Here, we report the first full-length F-type L1, L1MdF15 (F15), to be sequenced in its entirety. F15 was chosen from eight full-length F-type L1s because it was most like the consensus restriction map of the eight. Translation of this sequence indicates F15 contains no long open reading frames (ORFs), in contrast to the sequences of full-length A-type L1s. The first two A-type L1s to be sequenced in their entirety, L1MdA2 (A2) (3) and L1MdA13 (A13) (6), each contained two ORFs of 1.1 and 3.9 kb. The sequence between the monomer array and the polyadenylation signal (AATAAA) were compared in F15 and A2. These sequences are 91% homologous and the sequence differences are evenly distributed throughout. Fifteen length differences exist, and each difference is only one or two bases. When the length differences in F15 are corrected to match A2, eight termination codons still disrupt the regions which correspond to the ORFs in A2. These results indicate F15 is an older L1 that has accumulated many mutations, a result consistent with previous reports that F-type L1s ceased generating progeny a relatively long time ago (4, 6, 9). These results do not reveal any significant differences in the structural organization of A and F-type L1s except for the different monomer sequences.

F15 contains 2.6 F-monomers at the 5′ end. It is unclear which of two pairs of sequences, SDR #1 and SDR #2, are the flanking short direct repeats that define the ends of F15. The SDRs in most L1s lie immediately 5′ of the monomer array and within 100 bp of the polyadenylation signal (8). The 5′ SDR #1 lies immediately 5′ to the monomer array but the 3′ SDR #1 lies 226 bp 3′ of the polyadenylation signal. The 3′ SDR #2 lies 33 bp 3′ to the polyadenylation signal but the 5′ SDR #2 lies 19 bp 5′ of the monomer array.

REFERENCES

1. Fanning,T.G. (1983) *Nucl. Acids Res.* **11**, 5073–5091.
2. Hutchison,C.A., Hardies,S.C., Loeb,D.D., Shehee,W.R. and Edgell,M.H. (1989) In Berg,D.E. and Howe,M.M. (eds) *Mobile DNA*. American Society for Microbiology, Washington DC, vol. 1, pp. 593–617.
3. Loeb,D.D., Padgett,R.W., Hardies,S.C., Shehee,W.R., Comer,M.B., Edgell,M.H. and Hutchison,C.A. (1986) *Mol. Cell. Biol.* **6**, 168–182.
4. Padgett,R.W., Hutchison,C.A. and Edgell,M.H. (1988) *Nucl. Acids Res.* **16**, 739–749.
5. Rogers,J. (1985) *Int. Rev. Cytol.* **93**, 187–279.(?)
6. Shehee,W.R., Chao,S., Loeb,D.D., Comer,M.B., Hutchison,C.A. and Edgell,M.H. (1987) *J. Mol. Biol.* **196**, 757–767.
7. Skowronski,J., Fanning,T.G. and Singer,M.F. (1988) *Mol. Cell. Biol.* **8**, 1385–1397.
8. Voliva,C.F., Martin,S.L., Hutchison,C.A. and Edgell,M.H. (1984) *J. Mol. Biol.* **178**, 795–813.
9. Wincker,P., Jubier-Maurin,V. and Roizes,G. (1987) *Nucl. Acids Res.* **15**, 8593–8606.



* To whom correspondence should be addressed