**TOWARDS UNDERSTANDING OF REGULATION OF NON-LTR RETROTRANSPOSONS: RTE-1 ELEMENT OF THE NEMATODE CAENORHABDITIS ELEGANS.**

BEREZIKOV E.1+, BLINOV A.1, BERGTROM G.2

1Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, 10 Lavrentjeva Ave., Novosibirsk 630090, Russia;

2University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA;

+Corresponding author e-mail: eberez@bionet.nsc.ru

Keywords: NON-LTR, retrotransposons, nematode caenorhabditis elegans, long-terminal repeat

Retrotransposons are transposable elements whose mobility involves reverse transcription of an RNA intermediate. They are of two types: long-terminal repeat (LTR) retrotransposons and non-LTR retrotransposons (also called Long Interspersed Nuclear Elements, or LINEs). The mechanism of transposition and regulation of LTR-retrotransposons seems clear based on their structural similarity to retroviruses whose mechanisms of integration are well understood. In contrast, the molecular mechanisms responsible for retrotransposition of LINEs remain to be fully elucidated. The essential step in the retrotransposition of a LINE is its initial transcription. Several LINEs were shown to have a 5' region internal promoter which can initiate transcription at the first nucleotide of the element. Common multiple promoter modules were shown to regulate the transcription of I, F and Doc elements in *Drosophila melanogaster*1. Further understanding of the regulation of transcription of non-LTR retrotransposons will benefit from a comparison of regulatory systems in different organisms. To this end we initiated investigations of non-LTR retrotransposons in the model organism *Caenorhabditis elegans*. There are three principal sources of data that contribute to an understanding of mechanisms of transposition: (1) assays of suspected promoter activities of sequence motifs found in 5’ region of mobile elements; (2) assays of enzymatic activities of proteins encoded by elements; (3) locus, sequence and structural organization of the elements themselves, providing information about context and conserved sequence motifs with possible roles in regulation and mobility. While first two sources of information are of an experimental nature, analysis of the structural organization of elements dispersed in a genome is a intensive task in bioinformatics.

*C. elegans* is the first multicellular organism whose genome is scheduled to be fully sequenced by end of 1998. This provides an unprecedented opportunity for an exhaustive structural characterization of families of transposable elements that invaded this genome. The essential steps in the discovery and characterization of retrotransposons in a genome are: (1) the identification of sequences encoding reverse transcriptase (RTase) - the principal and most conserved protein of retrotransposons; (2) the determination of borders of the element and an analysis of flanking sequences to distinguish between LTR and non-LTR retrotransposons; (3) a search for target site duplications (TSDs), a hallmark of the integration of the element into genome; (4) a search for open reading frames (ORFs) and an analysis of the potential coding capacity of the element; (5) a search for characteristic for putative non-LTR elements, an analysis of 5’-regions for potential promoter activity and 3’-regions for poly(A) tracts.

We analyzed 78 Mb of *C. elegans* DNA following the plan outlined above. RTase containing sequences were discovered utilizing Hidden Markov models (HMMs) and software (HMMer) provided by Dr. Sean Eddy, who previously used HMMs to analyze 4 Mb of *C. elegans* sequences. Our search confirmed presence of 3 principal groups of RTase-containing sequences. However only 1 group seems to be a family of true retrotransposons, represented by several dozen nearly identical elements in the genome. This family includes the previously described Rte-1 non-LTR retrotransposon2. We focused our efforts on a structural characterization of this family of potentially active elements to which the question of the regulation of non-LTR retrotransposons may be addressed.

Rte-1 is 3298 bp long with a single ORF encoding an RTase similar in structure to those of the R4 and Dong elements of *Bombyx mori*, though less similar to those of the L1 elements of mammals and I element of *Drosophila***.** There are 59 Rte-1 elements in 78 Mb of DNA: 9 are full-length, 26 are truncated at different positions from 5’-end, 6 are truncated from 3’-end,14 are truncated from both ends, and 4 have intact ends but extensive internal rearrangements. Only 16 of the 59 Rte-1 elements retain TSDs, which range from 8 bp to an unusually long 1400 bp. Included are 5 full-length elements, 9 elements truncated at their 5' ends, 1 truncated at its 3' end, and 1 truncated at both ends. Truncated elements could arise from genomic rearrangements (deletions) at any time after transposition, but the existence of truncated elements still flanked by TSDs are more likely due to damage of RNA intermediates during reverse transcription or to faulty cDNA integration reactions. Among 5 elements with TSDs and intact 5’-ends, 2 seem to have additional nucleotides between one of the TSDs and the conventional border of the element. This could be explained by the existence of alternative 5’-end borders of the element. More likely however, both TSDs were originally the same, but shortly after transposition, one suffered deletion of bases at the border of the element. The absence of TSDs from 4 otherwise full-length elements suggests that genomic rearrangements occurred at their integration at new sites coincident with or soon after transposition. Among elements with intact 3’-ends, sequence comparisons revealed that length polymorphisms are due to the generation of repeated sequences at this end of the elements. The presence of many truncated elements without TSDs, the otherwise complete elements lacking parts of a TSD, and the repeated 3'-end structure of some elements suggest that that Rte-1 elements often integrate into hot spots of genomic rearrangement. These genomic rearrangements are yet another appealing problem in bioinformatics where analysis may provide additional information about retrotransposition mechanisms at the genomic level.

To date there have been no reports of mobile retroposons or retrotransposon-induced mutations in *C. elegans,*leading several authors to conclude that none of the *C. elegans* non-LTR elements is currently active*.*Our structural survey suggests for the first time that the Rte-1 family may still be active in *C. elegans* since several almost identical copies of the element are found in the genome. This fact is at least evidence of recent transposition events in Rte-1 family. A comparison of the 5’-UTR of Rte-1 with known non-LTR promoter sequences revealed that Rte-1 shares similarity with the de2 domain, which is a conserved region in the promoters of *Drosophila melanogaster*I, F and Doc elements1, further supporting the possibility that Rte-1 is an active retroposon. A direct test of this hypothesis would be to show that the element is transcribed by amplifying the element from RNA extracts by RT-PCR. If transcripts can be detected and characterized as Rte-1-derived by DNA sequencing, the *C. elegans* Rte-1 retrotransposon would become a model for the investigation of the regulation of non-LTR retrotransposons. If all Rte-1 family members prove to be quiescent, comparative structural studies should nevertheless suggest mechanisms that have inactivated the element.

**References**

1. Minchiotti, C. Contursi, P.P. Di Nocera, "Multiple downstream promoter modules regulate the transcription of the *Drosophila melanogaster*I, Doc and F elements" J. Mol. Biol. **267**, 37-46 (1997).
2. Youngman, H. van Luenen, R. Plasterk, "Rte-1, a retrotransposon-like element in *Caenorhabditis elegans*" FEBS Letters **380,** 1-7 (1996).