The beginning of the end: Links between ancient retroelements and modern telomerases

M. Joan Curcio and Marlene Belfort*

Wadsworth Center, New York State Department of Health, 150 New Scotland Avenue, Albany, NY 12208

Molecular and evolutionary biologists have long debated the origin of telomerases, the enzymes that maintain chromosome ends (1, 2). These ends, called telomeres, are nucleoprotein structures that consist of repetitive DNA sequences bound by specific proteins. Telomeres serve two basic functions: they distinguish chromosome termini from double-strand breaks, and they counteract the erosion of linear chromosomes from their ends. Telomerase, a ribonucleoprotein (RNP) composed of a reverse transcriptase (RT) and an RNA encoded by a different gene (3, 4), synthesizes the telomeric DNA repeats by using the 3'-OH at the end of the chromosome as a primer and telomerase RNA as a template (Fig. 1A). It is assumed that linear eukaryotic chromosomes have coevolved with telomerase. But the question remains, how did their association arise?

A relationship between telomerases and retrotransposon RTs, which are encoded by their template RNA, was surmised from shared amino acid motifs (4, 5) and the observation that telomeres are elongated by retrotransposons in some insects (6). Retroviruses and LTR retrotransposons create a target for insertion of their cDNA, the product of reverse transcription, using an integrase, whereas non-LTR retrotransposons, also termed target-primed (TP) retrotransposons, use an endonuclease. A 3'-OH generated by the nuclease is used to prime cDNA synthesis (Fig. 1A), in contrast to telomerase, which lacks an accompanying endonuclease module. In this issue of PNAS, Gladyshev and Arkhipova (7) report on a phylogenetically diverse family of eukaryotic retrotransposons called *Penelope*-like elements (PLEs), of which they found a nuclease-deficient subset specifically at telomeres. Although the nature of these "terminal PLEs" hints that similar elements could have given rise to early telomerases, it could be argued that this hypothesis is not compelling because the PLEs are a defined class of retrotransposons found only in bdelloid rotifers, basidiomycete fungi, stramenopiles (diatoms), and some plants. Human TP retrotransposons, termed L1s or LINEs, to the rescue! L1s occupy more than one-sixth of the genome (8), which is a huge

Fig. 1. Telomerases and retrotransposons. (A) RTs at telomere and within chromosome (black lines with gray circle). TERT acts on 3'-OH at telomere, whereas the RT of endonuclease-proficient TP-retrotransposons acts on the 3'-OH at a double-strand break. The wavy line indicates RNA. (B) Structure. The canonical nuclease-proficient PLE consists of a single ORF comprising RT and a GIY-YIG endonuclease. Terminal PLEs lack the endonuclease domain, but contain a 5' ORF with similarity to ORF1 of L1 elements, depicted below. The RT of L1 contains an AP-endonuclease domain (AP, apurinic/apyrimidinic). (C) Mechanism. PLE retrotransposition occurs when repeat sequences near the 3' end of PLE RNA hybridize to a telomeric DNA repeat (TGAGGG) at the single-stranded 3' end of the chromosome. With the 3' OH as a primer, RT synthesizes cDNA (blue) from the PLE RNA template. The second (bottom) strand of DNA is presumably synthesized by host-mediated DNA replication. RT dissociates from the PLE RNA prematurely, creating a 5' truncated cDNA, which is capped with telomeric DNA repeats (purple) by telomerase. Retrotransposition of the endonuclease-deficient L1 element begins when the 3' poly(A) tail of L1 RNA associates with the 3' end of a dysfunctional telomere. RT synthesizes cDNA by using the L1 RNA template. (D) Evolution. The schematic shows the ancestral RT giving rise to telomerase or a retrotransposon, which can revert to the primordial-RT-like state.

fraction. Morrish *et al.* (9) recently observed that L1s containing a disabled endonuclease can transpose to dysfunctional telomeres.

PLEs, which were discovered a decade ago, and have now been found in >10 animal phyla (10), encode proteins that are formed by fusion of RTs and GIY-YIG endonucleases (11) (Fig. 1B). The GIY-YIG domain, which was first discovered in T4 phage intron endonuclease I-TevI (12), has been identified in all three kingdoms (13). The *Penelope* RT, which is related to retron, group II

intron, and retrotransposon RTs, is most closely akin to telomerase RTs (5). Strikingly, the terminal PLEs discovered by Gladyshev and Arkhipova (7) lack the nuclease domain. The absence of an endonuclease ORF was determined by

Author contributions: M.J.C. and M.B. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 9352.

*To whom correspondence should be addressed. E-mail: belfort@wadsworth.org.

© 2007 by The National Academy of Sciences of the USA

comparison of clones from genomic fosmid libraries and telomere-enriched chromosomal DNA minilibraries of two bdelloid rotifers, anciently asexual invertebrates, and by informatic searches of public databases. Moreover, the terminal PLEs have a second associated ORF, with surprising resemblance to ORF1 of L1. The ORF1 product, which contains a coiled-coil domain, is an RNA chaperone that forms an RNP required for L1 retrotransposition (14).

The lack of an associated nuclease provides the first clue to the mechanism of chromosome-end localization of the terminal PLEs. Second, telomeric repeats near the 3' end of the RNA of terminal PLEs could hybridize to the telomeric DNA 3' overhang to prime RT (Fig. 1C). Third, the retrotransposons have 5' truncated ends that are telomere-oriented, consistent with reverse transcription beginning at the 3' end of the PLE RNA template and terminating prematurely. Fourth, arrays of terminal PLEs are usually punctuated by telomeric repeats, and the terminal PLE is capped by the addition of telomeric repeats, suggesting that both telomerase and the PLE RT have access to the telomere.

It has been argued for several retroelements, including group II introns and TP retrotransposons, that those lacking endonucleases are ancestral (15, 16). Because these retroelements rely on either DNA replication forks (17) or damaged DNA, including nicks, as target sites (18), acquisition of an endonuclease domain would likely have enhanced the efficiency and range of such elements. Remarkably, the study by Gladyshev and Arkhipova (7) revealed that functional copies of nuclease-deficient PLEs in species from protist, fungal, and plant clades are effectively single-copy genes, unlike nuclease-containing retrotransposons. Thus, nuclease-deficient PLEs cannot be viewed as selfish transposons, as they fail to support the spread of mobile copies of themselves. This characteristic, which is shared by the two genes that encode the telomerase RNA and RT subunits, suggests that the

common ancestor of telomerase genes and TP retrotransposons was a single gene whose RNA and RT activity, like that of nuclease-deficient PLEs, were involved in maintaining the ends of linear chromosomes (Fig. 1D). As predicted by this hypothesis, many of the terminal PLEs occupy a deep-branching phylogenetic position near the point of divergence from telomerase (7). Gladyshev and Arkhipova (7) argue that in early eukaryotic evolution, before the RT-endonuclease association gave rise to selfish, invasive retrotransposons, movement of cDNAs could have been limited to the free DNA ends. Later, the ancestral RT gene could have evolved into two genes, each encoding a subunit of the telomerase RNP. On the other hand, acquisition of an endonuclease ORF by the ancestral gene could have resulted in a mobile retrotransposon.

To complete the picture, the study by Morrish *et al.* (9) suggests that L1s have retained the ancestral ability of TP retrotransposons to recognize chromosome ends. In hamster cell lines, in which the protective nucleoprotein cap at telomeres was dysfunctional, and nonhomologous end joining was disabled, nuclease-deficient L1s retrotransposed adjacent to telomeric repeats. L1 insertions ending in an oligo A/T stretch abutted the reverse complement of a series of telomeric repeats (Fig. 1C), suggesting that reverse transcription of the polyadenylated L1 mRNA template was primed from the 3'-OH at the end of a telomeric repeat. Together, these observations indicate that nuclease-deficient L1s can use the 3'-OH at the end of disabled telomeres to prime reverse transcription.

Two conditions were essential for L1 cDNA to be inserted at telomeres. First, the L1 endonuclease domain had to be inactivated; in its presence, L1 insertions occurred at canonical endonuclease-recognition sites, rather than at telomeres. Thus, addition of L1 sequences to chromosome ends is outcompeted by retrotransposition to 3'-OH sites created by the L1 endonuclease. Second, mammalian cell lines had to be defective for

both telomere capping and nonhomologous end joining, suggesting that the chromosome end must be exposed and stabilized to serve as a target for L1 retrotransposition.

It is likely that other remnants of the ancient relationship between retrotransposons and telomeres will continue to be revealed. For example, the LTR-retrotransposon, Ty1, in *Saccharomyces cerevisiae*, has been implicated in telomere maintenance (19). Ty1 virus-like particles preferentially copackage the RNA of an unrelated telomeric mobile element, Y', and Y':Ty1 cDNA is incorporated into the genome at frequencies high enough to extend telomeres in the absence of telomerase.

When evidence from several directions converges on a point, that point becomes a discovery. Whereas the work of Gladyshev and Arkhipova (7) describes retroelements that are similar to telomerases and transpose to telomeres, the work by Morrish *et al.* (9) shows that artificial disruptions can drive a TP retrotransposon to unprotected chromosome ends. These two articles, which use different approaches, organisms, and experimental systems, combine to provide insight into events at the dawn of eukaryotic evolution, by relating telomeres and nuclease-independent retrotransposition. The findings also raise some intriguing questions. Why is ORF1, a feature of L1s, associated specifically with terminal PLEs and not with PLEs within chromosomes? Might the ORF1 RNA-binding protein, which forms an RNP with L1 and acts as a chaperone, be required by a PLE for terminal chromosomal localization? Regardless of the answers to such questions, the point of convergence of the two studies, the discovery, is that the apparent ancestral ability of RTs to use the free ends of linear chromosomes as substrates for retrotransposition gave birth to present-day telomerases.

Our work on retrotransposons is supported by National Institutes of Health Grants GM39422, GM44844, and GM52072.

- Eickbush TH (1997) *Science* 277:911–912.
- Nakamura TM, Cech TR (1998) *Cell* 587–590.
- Greider CW, Blackburn EH (1987) *Cell* 51:887–898.
- Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR (1997) *Science* 276:561–567.
- Arkhipova IR, Pyatkov KI, Meselson M, Evgen'ev MB (2003) *Nat Genet* 33:123–124.
- Pardue ML, Danilevskaya ON, Traverse KL, Lowenhaupt K (1997) *Genetica* 100:73–84.
- Gladyshev EA, Arkhipova IR (2007) *Proc Natl Acad Sci USA* 104:9352–9357.
- Deininger PL, Moran JV, Batzer MA, Kazazian HH, Jr (2003) *Curr Opin Genet Dev* 13:651–658.
- Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV (2007) *Nature* 446:208–212.
- Evgen'ev MB, Arkhipova IR (2005) *Cytogenet Genome Res* 110:510–521.
- Pyatkov KI, Arkhipova IR, Malkova NV, Finnegan DJ, Evgen'ev MB (2004) *Proc Natl Acad Sci USA* 101:14719–14724.
- Van Roey P, Derbyshire V (2005) in *Homing Endonucleases and Inteins*, eds Belfort M, Derbyshire V, Stoddard BL, Wood DW (Springer, New York), pp 67–83.
- Dunin-Horkawicz S, Feder M, Bujnicki JM (2006) *BMC Genomics* 7:98.
- Martin SL (2006) *J Biomed Biotechnol* 2006:1–6.
- Lambowitz AM, Zimmerly S (2004) *Annu Rev Genet* 38:1–35.
- Toro N, Jimenez-Zurdo JJ, Garcia-Rodriguez FM (2007) *FEMS Microbiol Rev* 31:342–358.
- Ichihyanagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M (2002) *Mol Microbiol* 46:1259–1272.
- Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV (2002) *Nat Genet* 31:159–165.
- Maxwell PH, Coombes C, Kenny AE, Lawler JF, Boeke JD, Curcio MJ (2004) *Mol Cell Biol* 24:9887–9898.