Cell, Vol. 92, 587–590, March 6, 1998, Copyright ©1998 by Cell Press

Reversing Time: Origin of Telomerase

Minireview

Toru M. Nakamura and Thomas R. Cech
Howard Hughes Medical Institute
Department of Chemistry and Biochemistry
University of Colorado
Boulder, Colorado 80309-0215

Telomerase, the enzyme responsible in most eukaryotes for replication of the ends of chromosomes, or telomeres, is a ribonucleoprotein (Greider and Blackburn, 1987). Its RNA subunit acts as a template for the synthesis of telomeric DNA, while a protein component catalyzes this process to make up for conventional DNA polymerases' inability to replicate completely the ends of linear DNA. In the past year, the catalytic protein subunit of telomerase was identified in *Euplotes aediculatus* and *Saccharomyces cerevisiae* (Lingner et al., 1997), and subsequently in *Schizosaccharomyces pombe* (Nakamura et al., 1997) and humans (Harrington et al., 1997; Kilian et al., 1997; Meyerson et al., 1997; Nakamura et al., 1997; Nakayama et al., 1998). Because these proteins share clear sequence and functional similarity with previously known reverse transcriptases (RTs) in their active site region, we will refer to them here collectively as Telomerase Reverse Transcriptases (TERTs). [The *S. cerevisiae* and *S. pombe* genes are named *EST2* (Ever shorter telomeres 2; Lendvay et al., 1996) and *trt1*$^+$ (Nakamura et al., 1997) respectively. The GDB (Genome Data Base) has approved the name *hTERT* for the human gene; this gene and protein have previously been called hTRT (Nakamura et al., 1997), hEST2 (Meyerson et al., 1997), TCS1 (Kilian et al., 1997), and TP2 (Harrington et al., 1997)].

RT Motifs in the Telomerase Catalytic Subunit: An Expected Feature?

Finding RT motifs in the catalytic subunit was not unexpected. Telomerase polymerizes DNA using an RNA template and is therefore by definition an RT, although an evolutionary relationship to other RTs was unknown. Now, all seven previously defined RT motifs (Eickbush, 1994) have been identified in TERT genes (Nakamura et al., 1997), and the importance of some of these RT motifs for telomerase activity has been shown by mutagenesis in *S. cerevisiae* (Lingner et al., 1997) and human (Harrington et al., 1997; Weinrich et al., 1997; Nakayama et al., 1998). Therefore, telomerase is not just an RT by mechanistic criteria, but has a related protein structure as well. Additional subunits or proteins associated with telomerase have also been identified (Collins et al., 1995; Lendvay et al., 1996) and may function to enhance or regulate the activity of this catalytic subunit.

Telomerases from evolutionarily diverse organisms all contain an RNA template, first suggesting that the fundamental mechanism of telomeric DNA synthesis is widely conserved. This conclusion is now further supported by the finding that telomerase's TERT subunit is also conserved in evolutionarily diverse organisms (protozoa, fungi, and mammals). In vitro telomerase activities have been detected from many eukaryotes, including some plants, and the presence of telomerase can be inferred in many more species on the basis of characteristic repetitive G-rich telomeric DNA sequences. The near universality of telomerase in eukaryotic species is striking because there are many other solutions, such as protein priming, terminal hairpins, and recombination, that ensure complete replication of linear DNA in viral and bacterial plasmid genomes and in linear mitochondrial genomes of certain eukaryotes.

Why is the use of telomerase so widespread in eukaryotes? The simplest explanation would be that telomere maintenance by telomerase is ancient, coinciding with or even preceding the first eukaryotic cells. But how can we test how ancient telomerase is? Telomerase RNAs, unfortunately, vary widely in size, and their sequence homology is too low to be useful for phylogenetic analysis except among closely related species. In addition, we have no candidate for an RNA ancestor of telomerase RNAs. On the other hand, the identification of the essential RT motifs in TERT genes now makes it possible to place telomerase in the context of other RTs by sequence analysis and therefore obtain some clues about its origin.

Phylogenetic Relationship of TERTs to Other Retroelements

Aside from telomerases, RTs are encoded by a wide variety of parasitic genetic elements (Table 1). These diverse retroelements share little in their genome organizations or their propagation methods. Their only connection is the presence of RT genes. Even their RT mechanisms differ, as seen in their wide-ranging primer specificities that include the 3′ OH of tRNA, a 3′ OH derived from their RNA genomes, an internal 2′ OH of RNA, a 3′ OH at a break in double-stranded DNA, and an OH group from a tyrosine within the RT.

Previous molecular phylogenetic studies, using seven conserved RT motif sequences from more than 80 RTs, have divided these diverse retroelements into two groups when the tree of RTs was rooted with RNA-dependent RNA polymerases (RRPs) (Eickbush, 1994). One group contains LTR retrotransposons, RNA viruses, and DNA viruses, while the other group contains non-LTR retrotransposons and bacterial and organellar elements. Use of RRPs as the root stems from the hypothesis that our current DNA-, RNA-, and protein-based world was preceded by an RNA- and protein-based world. This transition probably required RTs, and the assumption is that RTs evolved from RRPs. Consistent with this assumption, RRPs show greater sequence similarity to RTs than do other polymerases, and the recently solved X-ray structure of the RRP from poliovirus also supports their close relationship (Hansen et al., 1997).

Where does telomerase fit into this RT phylogenetic tree? Comparative sequence analysis of the TERT proteins using the same set of RT motifs places telomerase in the group containing non-LTR retrotransposons when the tree is rooted with RRPs (Nakamura et al., 1997; Eickbush, 1997; see Figures 1A and 1B). This placement of telomerase is intriguing in light of *Drosophila* telomere maintenance by the non-LTR retrotransposons Het-A and TART. Non-LTR retrotransposons and group II introns

Table 1. Distribution of Reverse Transcriptase Encoding Elements

Type of Element	Distribution
Retrovirus	Mammals, birds
Hepadnavirus	Mammals, birds
Caulimovirus	Plants
LTR Retrotransposon	Animals, plants, fungi, protozoa
Non-LTR Retrotransposon	Animals, plants, fungi, protozoa
Telomerase	Animals, plants, fungi, protozoa
Group II intron	Bacteria, fungi and plant mitochondria, chloroplasts, algae plastids
Mitochondrial plasmid	*Neurospora* mitochondria
RTL gene	*Chlamydomonas* mitochondria
Retron (msDNA)	Purple bacteria, other bacteria

transpose by a process called target-primed reverse transcription, in which the RT uses a 3′-OH at the DNA break, created either by an endonuclease or the intron itself, as a primer. Similarly, telomerase uses the 3′-OH at the natural chromosome DNA end as a primer. In addition, certain group II intron and mitochondrial RTs are found in stable complexes with their RNA subunits, similar to telomerase. Therefore, the close relationship of TERTs to the non-LTR RTs can be derived from functional criteria as well as by the RT motif analysis.

Uncertainties in Phylogenetic Analysis of RTs

Although the sequence motifs in TERTs are evolutionarily related to those of other RTs, determination of evolutionary distance or exact topology between telomerase and other retroelements, or even among retroelements, is somewhat problematic. The major assumption in estimating evolutionary distance among homologous proteins is that mutation rates remained relatively constant after these proteins diverged. This assumption may not be true for some RT-encoding elements. Due to the high error rates of their RTs, retroelements accumulate mutations much faster when they are duplicated thorough the cDNA cycle by RT, compared to when they are integrated into chromosomes and replicated by chromosomal DNA polymerase. Telomerase is not a transposable element, and therefore is never expected to be the subject of replication by a low-fidelity RT. However, inclusion of highly transposable elements or viruses that go through a cDNA cycle frequently may obscure the topology of the phylogenetic tree (see also Eickbush, 1994).

Additional difficulties in the analysis stem from the small number of amino acids in the RT motifs (only 178 amino acids can be aligned with reasonable confidence) and the relatively low extent of sequence identity between members of the set. There is already evidence that the telomerase tree suffers from these problems, because the topology of the branches for the human, *S. pombe*, *S. cerevisiae*, and *E. aediculatus* sequences (Nakamura et al., 1997) does not recapitulate the known phylogeny of these organisms. On the other hand, the situation is not beyond hope, because the phylogenetic analysis using seven RT motifs does correctly cluster elements from the same class. The classifications, such as group II introns, telomerases, and LTR retrotransposons, are based on genomic organizations and functional criteria, not on RT motif sequences. Therefore, this agreement gives us some confidence regarding our phylogenetic analysis based on seven RT motifs.

Finally, unlike trees of the phylogeny of organisms (often based on rRNA sequences), the relationships among RTs cannot be double-checked using independent genes, since only RT genes are common to all retroelements. This sole reliance on RT motifs, therefore, also limits the reliability of the relationship between telomerase and other retroelements derived from these studies.

Rooting the RT Phylogenetic Tree: An Evolutionary Guessing Game

Molecular phylogenetic analysis of the RT sequence alone produces only an unrooted tree of RTs (Figure 1A) and gives no information on the order in which various elements arose. Two potential ways to root the tree have been suggested (Eickbush, 1994): Figure 1B roots the tree using RRP with an assumption that it is the ancestor of RTs, while Figure 1C roots the tree with the branch that contains RTs from prokaryotic and organelle-derived elements. These two rootings have different implications for the possible origin of telomerase.

In the rooting of the phylogenetic tree based on RRP sequences (Figure 1B), telomerase branches off earliest among the group containing non-LTR retrotransposons and bacterial/organellar elements. Telomerase precedes

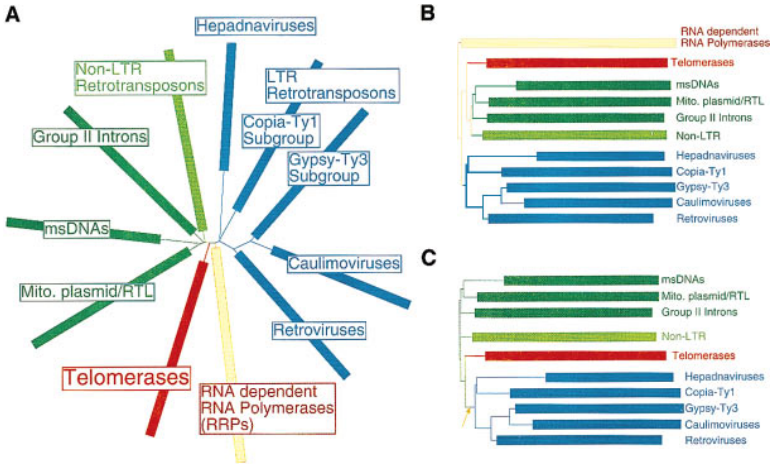

Figure 1. Reverse Transcriptase Trees

An unrooted tree of reverse transcriptases (A), and trees rooted with RRPs (B) and with prokaryotic and organellar retroelements (C). The arrow in (C) indicates where RRPs would fit if they are not removed from the tree. The length of each box corresponds to the most divergent element within that box. Methods are cited in Nakamura et al. (1997).

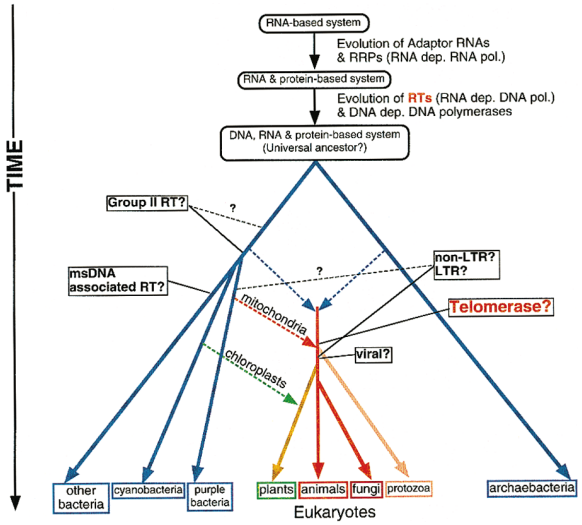

Figure 2. Proposed Origin of RTs, Including Telomerase

The proposed origin of RTs is superimposed on a scheme for the origin of life and the evolution of species (Gupta and Golding, 1996). Question marks denote some of the points of uncertainty or alternative possibilities.

the non-LTR retrotransposons, which implies that the essential cellular TERT gene in early eukaryotes gave rise to parasitic retroelements or that they had a common ancestor. This rooting further suggests that non-LTR retrotransposons are the oldest among mobile retroelements in this branch.

On the other hand, Eickbush (1997) has recently argued for the rooting as in Figure 1C. The implication of this rooting is that among the eukaryotic RTs, non-LTR retrotransposon-encoded RTs are the oldest, and telomerase and LTR-retroelements diverged from this lineage. This interpretation suggests that in early eukaryotes, the important cellular function of telomere maintenance was fulfilled by recruitment of an RT gene from a parasitic mobile element. The main argument in favor of this rooting is that it does not require a transfer of sequences from eukaryotes to prokaryotes (see Table 1 for distribution of RT encoding elements). As mitochondria and chloroplasts are thought to be derived from bacteria (Gupta and Golding, 1996; also see Figure 2), and bacteria are generally thought to be more ancient than eukaryotes, elements residing in bacteria and organelles are assumed to be older. However, we are dealing with mobile elements, which may have a phylogeny distinct from that of their host organisms. For example, a retroelement present in two different species could have been present in their common ancestor or could have been transferred horizontally after these species diverged. There is evidence for relatively recent transfer of msDNA elements in *E. coli* by phage-like particles, and the mobility of group II introns is well documented. The identification of both non-LTR- and LTR-like retroelements in plant mitochondrial genomes (Knoop et al., 1996) also shakes the argument for rooting based on prokaryotic and eukaryotic branches. These elements are thought to have been inserted into plant mitochondrial genomes by transfer of nuclear retrotransposons, and therefore, their distribution cannot be taken as evidence for the presence of non-LTR and LTR retrotransposons in ancient bacterial species.

Either tree is consistent with the TERT gene already being present in ancestral eukaryotes. Progenitor eukaryotic cells are thought to be chimeras formed by fusion of archaebacteria and gram-negative bacteria. This universal ancestor of eukaryotes then acquired mitochondria by symbiosis with purple bacteria, and later plant species acquired chloroplasts by symbiosis with cyanobacteria (Gupta and Golding, 1996; see also Figure 2). The parasitic protozoan *Giardia* has been suggested to resemble closely the hypothetical ancient eukaryote that has not yet acquired organelles, as *Giardia* has neither mitochondria nor chloroplasts. Although the status of *Giardia* as a model for primitive eukaryotes has come under question with the finding that the *Giardia* genome encodes chaperonins, which are thought to be derived from mitochondria (Roger et al., 1998), other genes including the rDNA in *Giardia* have sequences resembling those of bacterial species, and therefore they may still retain characteristics of ancestral eukaryotes. Curiously, *Giardia* has a telomeric sequence of (TAGGG)n, as expected for telomere maintenance by telomerase. Thus, it seems likely that TERT can be found in *Giardia*. Such a finding would support the hypothesis that telomerase is ancient, possibly coinciding with or even preceding the appearance of the first eukaryotic cells.

Telomerase Lost: What Happens in the Absence of Telomerase?

Telomerase is not the only solution to the end-replication problem used in eukaryotes. If telomerase already existed in ancestral eukaryotic cells as suggested by phylogenetic analysis, small subgroups of insect and plant species must have lost telomerase during their evolution (Pardue et al., 1996; Biessmann and Mason, 1997). These species either utilize retrotransposons that specifically transpose to chromosome ends, or they use a recombination mechanism to extend their telomeric DNA.

Insects of the order Diptera, which include *Drosophila*, lack functional telomerase and instead maintain their telomeres with non-LTR retrotransposons called TART and HeT-A. It has been argued that this *Drosophila* system resembles telomerase: a reverse transcriptase uses a specific RNA transcript as the template for telomeric DNA synthesis, except now the telomerase RNA is replaced by either the HeT-A or TART transcript (Pardue et al., 1996; Biessmann and Mason, 1997). As HeT-A does not encode its own RT, it presumably relies on a not-yet-identified master RT gene in the *Drosophila* genome. If this RT turns out to have the specific characteristics of TERT genes, such as the telomerase-specific T motif sequence and unique amino acid signatures within the RT motifs (Nakamura et al., 1997), then it supports the hypothesis that telomerase lost its RNA subunit and replaced it with another transcript. On the other hand, there are many telomere-specific retrotransposons related to TART in insect species where telomerase is functional. For example, the silkworm, *Bombyx mori*, has at least two non-LTR elements called TRAS1 and SART1 that insert within (CCTAA)n telomeric repeats (Takahashi et al., 1997). Like TART, these retrotransposons encode their own RTs. Therefore, TART may represent just another retrotransposon that targets itself into

the heterochromatin at telomeres, much like the yeast Ty5 LTR retrotransposon (Zou et al., 1996).

Recombination may come to the rescue when telomerase has been completely lost. As documented for telomerase-deleted mutant yeast strains, recombination can substitute to maintain telomere length. Perhaps that is what is happening in the onion and the mosquito (Biessmann and Mason, 1997). Also, up to one fourth of immortalized human cell lines maintain long telomeres by a telomerase-independent mechanism that probably involves recombination.

Perspective on the Origin of Telomerase

Although a number of alternative solutions to the chromosome end-replication problem are used in nature, the telomerase solution is the most widespread and perhaps the oldest among eukaryotes. The finding of clear RT motifs in the catalytic subunit of telomerase means we no longer need to qualify it as a "specialized" RT. Indeed, expression of the telomerase RNA and the catalytic subunit (along with whatever components might be provided by a reticulocyte lysate) reconstitutes human telomerase activity in vitro (Weinrich et al., 1997). This suggests that underneath the massive telomerase RNP complex (based on glycerol gradient and sizing column estimates), telomerase may have a simple two-component RNP enzyme at its core, much like simpler RTs encoded by group II introns and non-LTR retrotransposons. Since RTs are thought to have been with us since the transition from RNA- and protein-based systems to the present-day DNA-, RNA-, and protein-based systems (see Figure 2), we now have a satisfying explanation for near universality of telomerase among eukaryotes. While it is still far from clear exactly how telomerase evolved to its present-day form, it is likely to be with us for a long time.

Selected Reading

Biessmann, H., and Mason, J.M. (1997). Chromosoma *106*, 63–69.

Collins, K., Kobayashi, R., and Greider, C.W. (1995). Cell *81*, 677–686.

Eickbush, T.H. (1994). In the Evolutionary Biology of Viruses, S.S. Mores, ed. (New York: Raven Press), pp. 121–157.

Eickbush, T.H. (1997). Science *277*, 911–912.

Greider, C.W., and Blackburn, E.H. (1987). Cell *43*, 405–413.

Gupta, R.S., and Golding, G.B. (1996). Trends Biol. Sci. *21*, 166–171.

Hansen, J.L., Long, A.M., and Schultz, S.C. (1997). Structure *5*, 1109–1122.

Harrington, L., Zhou, W., McPhail, T., Oulton, R., Yeung, D.S.K., Mar, V., Bass, M.B., and Robinson, M.O. (1997). Genes Dev. *11*, 3109–3115.

Kilian, A., Bowtell, D.D.L., Abud, H.E., Hime, G.R., Venter, D.J., Keese, P.K., Duncan, E.L., Reddel, R.R., and Jefferson, R.A. (1997). Human Mol. Genet. *6*, 2011–2019.

Knoop, V., Unseld, M., Marienfeld, J., Brandt, P., Sunkel, S., Ullrich, H., and Brennicke, A. (1996). Genetics *142*, 579–585

Lendvay, T.S., Morris, D.K., Sah, J., Balasubramanian, B., and Lundblad, V. (1996). Genetics *144*, 1399–1412.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). Science *276*, 561–567.

Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddle, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q., et al. (1997). Cell *90*, 785–795.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). Science *277*, 955–959.

Nakayama, J., Tahara, H., Tahara, E., Saito, M., Ito, K., Nakamura, H., Nakanishi, T., Tahara, E., Ide, T., and Ishikawa, F. (1998). Nat. Genet. *18*, 65–68.

Pardue, M.L., Danilevskaya, O.N., Lowenhaupt, K., Slot, F., and Traverse, K.L. (1996). Trends Genet. *12*, 48–52.

Roger, A.J., Svard, S.G., Tovar, J., Clark, C.G., Smith, M.W., Gillin, F.D., and Sogin, M.L. (1998). Proc. Natl. Acad. Sci. USA *95*, 229–234.

Takahashi, H., Okazaki, S., and Fujiwara, H. (1997). Nucleic Acids Res. *25*, 1578–1584.

Weinrich, S.L., Pruzan, R., Ma, L., Ouellette, M., Tesmer, V.M., Holt, S.E., Bodnar, A.G., Lichtsteiner, S., Kim, N.W., Trager, J.B., et al. (1997). Nat. Genet. *17*, 498–502.

Zou, S., Ke, N., Kim, J.M., and Voytas, D.F. (1996). Genes Dev. *10*, 634–645.