Telomere-End Processing: the Terminal Nucleotides of Human Chromosomes

Agnel J. Steir, Weihang Chai, Jerry W. Shay, and Woodring E. Wright
1Department of Cell Biology
5323 Harry Hines Boulevard
University of Texas Southwestern Medical Center
Dallas, Texas 75390

Summary

Mammalian telomeres end in single-stranded, G-rich 3’ overhangs resulting from both the “end-replication problem” (the inability of DNA polymerase to replicate the very end of the telomeres) and postreplication processing. Telomeric G-rich overhangs are precisely defined in ciliates; the length and the terminal nucleotides are fixed. Human telomeres have very long overhangs that are heterogeneous in size (35–600 nt), indicating that their processing must differ in some respects from model organisms. We developed telomere-end ligation protocols that allowed us to identify the terminal nucleotides of both the C-rich and the G-rich telomere strands. Up to ~80% of the C-rich strands terminate in CCAATC-5’, suggesting that after replication a nuclease with high specificity or constrained action acts on the C strand. In contrast, the G-terminal nucleotide was less precise than Tetrahymena and Euplotes but still had a bias that changed as a function of telomerase expression.

Introduction

Telomeres are the ends of linear chromosomes, and their “end-capping” function helps maintain the integrity of the genome by preventing end-end fusions and degradation. Mammalian telomeric DNA contains TTAGGG repeats bound by specialized proteins (Smogorzewska and de Lange, 2004). Human telomeres end in a single-stranded, G-rich 3’ overhang (35–600 nt in length) (Makarov et al., 1997; Stewart et al., 2003; Wright et al., 1997). Ultrastructural studies have shown that this overhang invades the preceding double-stranded region of the telomere to form the t loop, a lariat-like structure stabilized by base pairing and protein-protein interactions (Griffith et al., 1999). The t loop is proposed to protect the end from being perceived as a double-stranded break (de Lange, 2002; Stansel et al., 2001). Telomeric overhangs result from both the end-replication problem and postreplication processing events. Upon the completion of lagging-strand synthesis, removal of the final primer should result in a short overhang (around 9–12 nt long) if it were placed at the extreme end of the template telomere. Alternatively, a longer and more variable overhang would result if the last priming event did not take place at the very end. Leading-strand synthesis, which is continuous, should produce a blunt end if it proceeds to the very end of the telomere or might leave a 5’ overhang if the replication machinery falls off prematurely (Cimino-Reale et al., 2001).

Because overhangs are present on both chromosome ends, a nuclease must act upon the 5’ end of at least the leading strand to generate overhangs. It is not known if leading and lagging daughters are subject to the same regulatory processing.

Most of the evidence for overhang processing comes from yeast and ciliates that possess short overhangs (10–21 nt) of relatively uniform length. Euplotes overhang generation gives precise terminal nucleotides at both ends, GGGTTTG-3’ at the G-rich strand and AAAACCC-5’ at the C-rich strand, and the length of the overhang is always 14 nt (to facilitate alignment, G-rich sequences are presented in the 5’ to 3’ orientation, C-rich sequences in the reversed 3’ to 5’ orientation, and the terminal nucleotides are shown bolded and underlined). Extensive studies characterizing mechanisms of overhang processing have been done in Tetrahymena (Fan and Price, 1997; Jacob et al., 2001, 2003). Instead of ending in GGGTTG-3’, as would be expected if the terminus is generated by dissociation of telomerase during the translocation step, most Tetrahymena telomeres end in TGGGGT-3’. Tetrahymena C-rich strands end with CAACC-5’ or CCAAC-5’. This suggests that overhang generation is mediated by two separate processing steps: one cleaves the G strand and the other resects the C strand, and both steps are distinctively terminated at a specific base.

As a first step in characterizing overhang-processing events in human cells, we established ligation-mediated methods to identify the last nucleotide of both the C-rich and the G-rich strands. The human pattern differs from that seen in Tetrahymena and Euplotes, suggesting a divergence in the underlying processing mechanisms. The terminal nucleotide of the C strand was uniform: ~80% of human C strands begin with the sequence CCAATC-5’. The G strand terminal nucleotide was much more variable, with a bias that changed with telomerase expression. This definition of the terminal nucleotide should facilitate the identification of the factors involved in telomeric end-processing events.

Results

C Strand Processing of Human Telomeres Is Very Tightly Regulated

We used two independent assays to identify the terminal nucleotide of the C-rich telomeric strand. In the primer ligation assay, we tagged the C strand with a known oligonucleotide. Six permutations of the telomeric sequence AATCCC were synthesized at the 3’ end of a cassette containing two PCR primers. The primers were separated by a variable spacer so that each permutation could be identified by its spacer size (Figure 1A). Only the oligonucleotide(s) annealing immediately adjacent to the 5’ terminus could be ligated to the C strand. After ligation of an equal mix of these
Figure 1. Strategy for Primer Ligation Assay to Determine the Identity of the C Strand Terminal Nucleotide

(A) Six oligonucleotides are annealed to the overhang at the chromosome end. Each contains a permutation of the telomeric sequence (AATCCC). Only the linker that anneals in correct register and adjacent to the last base of the C strand terminal nucleotide can be ligated. The six oligonucleotides contain four AATCCC repeats followed by a tail containing two known sequences that are used for PCR amplification. Each permutation of the sequence AATCCC is paired with a different length spacer between the two PCR primer sequences. The first three nucleotides of the labels represent the terminal nucleotides of the G-rich strand with the last base in bold and underlined, whereas the boxed sequence represents the sequence of the ligated nucleotide.

(B) Two artificially generated telomeric overhangs possessing either one specific terminal nucleotide (top) or a mix of six nucleotides representing all permutations of the telomeric repeat (AATCCC) (bottom) showed the predicted patterns of band distribution and intensities on an acrylamide gel. The background appearing in the later cycles is due to the failure to remove all unligated primers.

(C) The primer ligation assay was applied to genomic DNA from BJ cells. Semiquantitative PCR utilizing $^{32}$P-labeled primers and an increasing number of cycles was performed. (Ca) is a negative control (ligase). (Cb) shows major preference for the oligonucleotide possessing a
six oligonucleotides, free oligonucleotides were removed, and the telomere bound oligonucleotides were PCR amplified. The amplification products were resolved on acrylamide gels, such that the size of the bands specified the oligonucleotide(s) that was ligated to the telomere, which in turn identified the terminal nucleotide for the C strand. The intensity of each band determined the proportion of the oligonucleotide that was successfully ligated to the telomere.

This was validated on an artificial telomeric overhang ending in one specific base, which gave the correct single band on the gel (Figure 1B, top). A plasmid that contained a mix of six terminal nucleotides generated six bands of equal intensities (Figure 1B, bottom). By using DNA from cultured human BJ fibroblasts, most of the primer ligation products represented the oligonucleotide 3′-CCAATC, which corresponded to a C strand ending in CCAATC-5′ (Figure 1C). After randomization with the 5′→3′ T7 exonuclease, the preference for the C strand terminus CCAATC-5′ was completely lost (Figure 1Da and 1Db), whereas the small enrichment for the top and bottom oligonucleotides (Figure 1Da) was preserved. Moreover, treating with Exonuclease I (Exol) to digest the overhang did not abolish the enrichment for the top and bottom oligonucleotides (See Figure S1 in the Supplemental Data available with this article online), suggesting that the slight bias for these oligonucleotides was not significant.

Mammary epithelial cells (HMEC), breast tumor cells (MCF7), mouse cells (3T3), and human lung fibroblasts (WI38) all displayed the same pattern, with most ligation products corresponding to telomeric C strands ending in CCAATC-5′ (Figure S1).

We confirmed the above results by using a modification of single-strand telomere length analysis (STELA) (Baird et al., 2003) to look at individual chromosomes and to identify their terminal nucleotides(s). In this assay, “C telorette” oligonucleotides that contained seven nucleotides of telomeric repeats followed by a unique sequence were ligated to genomic DNA. The DNA was then diluted until only a few amplifiable molecules were present and PCR amplified by using XpYp chromosome-specific subtelomeric primers and “teltail” primers complementary to the unique sequence of the telorette. A Southern blot utilizing XpYp subtelomeric probes revealed individual bands that represented individual telomeric molecules. We performed six ligation reactions, each containing a telorette with one specific base at the 3′ end representing a permutation of the telomeric repeat (AATCCC) (Figure 2A). The number of bands amplified with each C telorette reflected the proportion of telomeres that were ligated to the 3′ end of the specific telorette, which in turn defined the end nucleotide of the C strand.

Multiple amplification reactions were performed for each ligation reaction by using BJ human foreskin fibroblast DNA. Nonspecific ligations to all six telorettes were obtained by using 0.9 μM of C telorettes (as in Baird et al., 2003) (Figure S2). This represents an ~4 x 107-fold excess of oligonucleotide to telomeres. Specific ligations were seen over a broad range of concentrations when the input telorettes were diluted 100- to 100,000-fold. Approximately 80% of the amplification products represented the C telorette that ends in 3′-CCAATC (Figure 2B). This confirmed the primer-ligation assay results showing that the terminus of most C-rich strands is CCAATC-5′. Randomizing the terminal nucleotide by treating the DNA with the 5′→3′ T7 Exonuclease abolished the preference for C telorette 3′-CCAATC (Figure 2C).

Cells with telomerase activity exhibited the same end nucleotide (data not shown), suggesting that telomerase did not alter the nucleotide specificity of C strand processing. The terminal nucleotide of other chromosomes (7p and 10q) gave the same results (data not shown). Dephosphorylating and re-phosphorylating the telomeres prior to ligation did not alter the outcome, indicating that the preference for CCAATC-5′ did not represent a bias for in vivo 5′ phosphorylation of selected ends (data not shown).

We next investigated whether leading and lagging daughter strands end in the same terminal nucleotide(s). Cells were fed the thymidine analog BrdU for one doubling, during which leading-strand synthesis (using AATCCC as the template) incorporated twice as much BrdU as lagging-strand synthesis (with TTAGGG as the template). The different densities of the strands allowed their separation on a CsCl gradient (W.C., Q. Du, J.W.S., and W.E.W., unpublished data). STELA ligations and PCR showed that both leading and lagging strands exhibited the same major preference for the sequence CCAATC-5′ at the 5′ end of the C-rich strand (Figure 2D). This suggests that the overhang on the lagging strand is not solely produced by removing the last Okazaki primer; most likely it is subjected to the same final C strand nuclease processing step as the leading strand.

The Terminal Nucleotide of the G-Rich Telomeric Strand in Human Cells Is Less Precisely Regulated and Is Slightly Altered by Telomerase Expression

To determine the terminal nucleotide of the G-rich strand, we first annealed an oligonucleotide (3′-AATCCC-5′) to the overhang. This “platform” produced a 5′ overhang that guided the ligation of G teleroettes to the G-rich strand terminal nucleotide (Figure 3A). After ligating six individual G teleroettes, each with a different permutation of the TTAGGG repeat, individual telo-rettes-tagged telomeric molecules were amplified and detected on a Southern blot. All six G teleroettes were able to ligate to some G ends from BJ fibroblasts (Figure 3B), suggesting that G strand processing of mammalian telomeres is less precise than reported for Tetrahymena and Euplotes. Nevertheless, G strand processing was not totally random; ~70% of 990 telomeric molecules analyzed ended in GGTTAG-3′, suggesting that the C strand most frequently ends in CCAATC-5′.
Figure 2. Identifying the C Strand Terminal Nucleotide by STELA on the XpYp Telomere

(A) Strategy: Six individual C telorettes are ligated in separate reactions to the same amount of DNA and then amplified by using a forward XpYp chromosome-specific subtelomeric primer and a reverse Teltail primer. Only the C telorette annealing adjacent to the last base of the C strand will be ligated to the telomere end and can produce a PCR product.

(B) STELA adaptations with telorette concentration ranging from $10^{-2}$ to $10^{-5}$ M were individually ligated to DNA from BJ cells. Multiple amplifications using the XpYp E2 forward primer and the Teltail reverse primer show that 80% of the bands appear when using telorette 3 (ends in 3'-CCAATC). The first three nucleotides of the labels represent the terminal nucleotides of the C-rich telomeric strand with the last base in bold and underlined, whereas the boxed sequence represents the sequence of the ligated oligonucleotide.

(C) STELA was performed on BJ DNA that was pretreated with T7 exonuclease. This randomized the terminal nucleotide, and all telorettes ($10^{-3}$ M) gave similar numbers of amplification products. The panel numbers (1–6) match the numbers in Figure 2B.

(D) STELA was applied to leading and lagging DNA by using six individual ligation reactions possessing different C telorettes at $10^{-3}$ M concentration. Both daughter telomeres have a major preference for C strands terminating in CCAATC-5'.
Figure 3. The Identity of the G Strand Terminal Nucleotide for Individual Chromosomes

(A) Annealing a guide template to the overhang provides a platform for subsequently ligating G telorettes to the 3' end of the G-rich strand. Six individual telorettes are ligated in separate reactions to the same amount of DNA and then PCR amplified by using a forward chromosome-specific primer and a reverse Teltail primer.

(B) STELA adaptation with G telorettes individually ligated to DNA from BJ cells. Multiple amplifications show that G telorettes that ligate in register to AGGGTT-3', GGTTA-3', and GGTTAG-3' generate 70% of the total number of bands. The first three nucleotides of the labels represent the terminal nucleotides of the G-rich strand with the last base in bold and underlined, whereas the boxed sequence represents the sequence of the ligated nucleotide.

(C) STELA adaptation was applied to DNA from telomerase positive HeLa cells. Results show further enrichment for the G-telorette ligation to overhangs ending in GGTTAG-3'.

(D) Percentage for each terminal nucleotide from 3B (BJ cells) and 3C (HeLa cells). 990 telomeric molecules were analyzed for BJ cells and 330 telomeres for HeLa cells. The dotted line shows what would be expected for a random distribution. The difference in the frequency of the GGTTAG-3' ends in HeLa versus BJ cells is highly significant by the Student's t test.

(E) The pattern of variability observed for the G-terminal nucleotide is consistent with the replication complex sometimes generating a blunt end (leaving a GGTTAG-3' end) or dissociating one or two nucleotides prior to the terminus.
GGGTTA-3', or AGGGTT-3' (Figure 3D). To confirm that this bias was not just a ligation or PCR artifact, we randomized the 3' end. Treatment with a combination of ExoI exonuclease (3'→5') and T7 Exonuclease (5'→3') successfully produced a total randomization of ligation products (data not shown).

We next determined whether the enzymatic activity of telomerase altered the G terminus. A clear shift in the distribution pattern of G terminal nucleotides was observed in telomerase-positive HeLa cells, resulting in a greater enrichment (~40% of total telomeres) for GGTTAG-3' (Figures 3C and 3D). The G terminal nucleotide GGTTAG-3' matches the last base of the hTR (telomerase RNA) template region and is the pause site following which telomerase translocates prior to the next cycle of repeat synthesis. A similar increased bias in favor of GGTTAG-3' was seen in BJ and MRC5 fibroblasts expressing exogenous hTERT (data not shown).

**Discussion**

Overhangs are a critical component of the telomere-end structure. Their importance exceeds simple telomere capping functions since they establish telomere shortening rates and influence replicative senescence. Our observations define the products of the final steps of telomere end processing for both the C-rich and G-rich strands in human cells. Although the detailed mechanisms remain to be determined, the results have important implications for understanding the generation of telomeric overhangs. We show that the vast majority of telomeric C strands end with the sequence CCAATC-5', thus C strand processing is very tightly specified. This is in agreement with results from ciliates (Euplotes and Tetrahymena) that possess a specific terminal nucleotide, suggesting that the mechanism of C strand resection and the factors involved in the processing events are most likely conserved.

Leading-daughter overhangs are only about 60% as long as lagging-strand daughters (W.C., Q. Du, J.W.S., and W.E.W., unpublished data), indicating some difference in length control during overhang processing. Our present observations do not distinguish between extensive C strand processing of both daughters when replication is completed versus the simple removal of the RNA primer from the last Okazaki fragment of the lagging-strand daughter with extensive resection of the C strand on leading daughter. Nevertheless, our results imply that although there might be differences in the steps involved in producing these different-sized overhangs, the mechanism specifying the terminal nucleotide and the final step in C strand processing is likely to be shared.

What determines the precision of the C strand resection and the identity of the terminal nucleotide is still unknown; it could be due to a base-specific endonuclease/helicase that cleaves exclusively between two C residues or to a telomere binding protein that provides the nucleotide-specific boundary for 5'→3' exonuclease trimming. Such a protein could be a single-stranded binding protein or a protein that binds to double-stranded DNA and physically hinders further resection by the nuclease. A set of telomere binding proteins bind to the telomere end with great specificity and cap the chromosome end. One such protein is POT1, an overhang binding protein that binds with great specificity (Kelleher et al., 2005; Lei et al., 2004; Loayza et al., 2004). Its role in overhang processing is yet to be determined.

That the identity of the C strand nuclease is unknown is a fundamental gap in the understanding of telomere replication. Artemis, a 5'→3' nuclease involved in nonhomologous end joining has been linked to telomeric function (Rooney et al., 2003). Artemis deficiency results in increased telomere end-end fusion. We found the same preference for CCAATC-5' in Artemis-deficient cells (data not shown). Our results do not rule out the role of Artemis in overhang generation but show that Artemis resection is not required for the accuracy of the processing.

We found a nonrandom mixture of terminal G strand nucleotides with marked preference for GGTTAG-3', GGGTTA-3', and AGGGTT-3'. This contrasts with the highly specified G-terminal nucleotides of model organisms. If the leading-strand replication machinery is able to accurately synthesize DNA all the way to the final C-rich template nucleotide (CCAATC-5'), then G strands should end in GGTTAG-3'. Variability could occur if replication sometimes generated a blunt end (leaving a GGTTAG-3' end) but also dissociated one or two nucleotides prior to the terminus, leaving GGGTTA-3' or AGGGTT-3' ends (Figure 3E). In this scenario no processing of the human G strand would take place, and the terminal nucleotide would simply represent the fail-

![Figure 4. Comparison of the Telomeric-End Structures](image)
ure of the replication complex to always copy the final one or two nucleotides. The significance of an abundance of TTAGGG-3’ termini greater than predicted by this model remains to be determined. The results are also consistent with some imprecise nuclease processing of the G strand that yields less nucleotide specificity than C strand processing. Such processing would be regulated by a set of proteins that bind the overhangs with some flexibility to generate a less precise terminal nucleotide.

Interestingly, we found some change in the distribution of G ends in the presence of telomerase. Although our results do not exclude telomerase participating in a processing complex (Oulton and Harrington, 2004), the increased preference for GGTAGQ-3’ termination in cells expressing telomerase is fully consistent with the hypothesis that the altered distribution is due to telomerase dissociation without further processing to generate the terminal nucleotide on some of the telomeres.

Figure 4 compares current knowledge of the overhangs and terminal nucleotides of some model organisms and humans. Saccharomyces cerevisiae has overhangs that are 12–14 nt in length (Larivée et al., 2004). Their G strand termini show no base specificity (Forstemann et al., 2000), whereas their C strand termini have not been directly determined. Euplotes and Tetrahymena continuously express telomerase and have very short and uniform telomeric overhangs. The two ciliates are known to exhibit precise processing to generate specific G-rich and C-rich terminal nucleotides (Fan and Price, 1997; Jacob et al., 2003, 2001).

Human telomeres have variable overhang lengths ranging from 35–600 nt (Makarov et al., 1997; Stewart et al., 2003; Wright et al., 1997). The present report demonstrates that end processing at human telomeres has some aspects that are common to model organisms and others that are different, probably reflecting these different telomeric dynamics. Understanding the mechanisms of end processing in human cells may permit interventions to accelerate the loss of telomere length during telomerase-inhibition-based cancer therapy or to reduce the rate of telomere shortening to retard replicative senescence.
CGTGCGCTGCT-3'; G teltail (reverse primer), 5'-ACGAGGCACGTA
GACCGTAG-3'.

Supplemental Data
Supplemental Data including two figures are available online with
this article at http://www.molecule.org/cgi/content/full/18/1/131/
DC1/.

Acknowledgments
We thank Susanne Steinert and Nicholas Forsyth of the Shay/
Wright Lab for helpful advice and discussions. This work was sup-
ported by grant BC031037 from the Department of Defense (A.J.S.)
and AG01228 from the National Institutes of Health (W.E.W.).
W.E.W. is an Ellison Medical Foundation Senior Scholar.

Received: December 14, 2004
Revised: February 17, 2005
Accepted: February 28, 2005
Published: March 31, 2005

References

(2003). Extensive allelic variation and ultrashort telomeres in senes-

Cimino-Reale, G., Pascale, E., Battiloro, E., Starace, G., Verna, R.,
and D'Ambrosio, E. (2001). The length of telomeric G-rich strand
3'-overhang measured by oligonucleotide ligation assay. Nucleic
Acids Res. 29, E35.

21, 532–540.

C strand length during new telomere synthesis. Mol. Biol. Cell 8,
2145–2155.

Forstemann, K., Hoss, M., and Lingner, J. (2000). Telomerase-
dependent repeat divergence at the 3' ends of yeast telomeres.
Nucleic Acids Res. 28, 2690–2694.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A.,
large duplex loop. Cell 97, 503–514.

Jacob, N.K., Skopp, R., and Price, C.M. (2001). G-overhang dy-
namics at Tetrahymena telomeres. EMBO J. 20, 4299–4308.

Jacob, N.K., Kirk, K.E., and Price, C.M. (2003). Generation of tel-
omeric G strand overhangs involves both G and C strand cleavage.

telomeres 1 (POT1) is a negative regulator of telomerase activity

Larrivee, M., Lebel, C., and Wellinger, R.J. (2004). The generation
of proper constitutive G-tails on yeast telomeres is dependent on
the MRX complex. Genes Dev. 18, 1391–1396.

POT1 bound to telomeric single-stranded DNA provides a model
for chromosome end-protection. Nat. Struct. Mol. Biol. 11, 1223–
1229.

Loayza, D., Parsons, H., Donigian, J., Hoke, K., and de Lange, T.
(2004). DNA binding features of human POT1: a nonamer 5'-
TAGGGTTAG-3' minimal binding site, sequence specificity, and in-
ternal binding to multimeric sites. J. Biol. Chem. 279, 13241–13248.

both ends of human chromosomes suggest a C strand degradation


Rooney, S., Alt, F.W., Lombard, D., Whitlow, S., Eckersdorff, M.,
Fleming, J., Fugmann, S., Ferguson, D.O., Schatz, D.G., and Seki-
guchi, J. (2003). Defective DNA repair and increased genomic insta-

Smogorzewska, A., and de Lange, T. (2004). Regulation of telom-
erase by telomeric proteins. Annu. Rev. Biochem. 73, 177–208.

Stansel, R.M., de Lange, T., and Griffith, J.D. (2001). T-loop assem-
ly in vitro involves binding of TRF2 near the 3' telomeric overhang.
EMBO J. 20, 5532–5540.

Stewart, S.A., Ben-Porath, I., Carey, V.J., O'Connor, B.F., Hahn,
W.C., and Weinberg, R.A. (2003). Erosion of the telomeric single-
strand overhang at replicative senescence. Nat. Genet. 33, 492–
496.

Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D., and Shay,
J.W. (1997). Normal human chromosomes have long G-rich tel-
omeric overhangs at one end. Genes Dev. 11, 2801–2809.