The generation of long telomere overhangs in human cells: a model and its implication

Yu-hua Hao¹ and Zheng Tan¹, ²,*

¹College of Life Sciences, Wuhan University, Wuhan 430072, People’s Republic of China and ²Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, People’s Republic of China

Received on September 21, 2001; revised on November 26, 2001; accepted on December 6, 2001

ABSTRACT
Motivation: Linear chromosomes carry on both ends repetitive DNA sequences called telomere. In the conventional model of semi-conservative DNA replication, the 3′-end of a linear DNA strand cannot be fully replicated, resulting in a single-stranded 3′ overhang at one end of the double-stranded DNA product. In this model, the length of the overhang is expected to be about the size of an RNA primer (about nine nucleotides for human cells). However, it has been found that the telomere overhangs in human cells can be as long as several hundred nucleotides. At present, the opinion regarding how such long overhangs are produced is controversial.

Results: In order to gain insight into the mechanism by which long telomere overhangs are produced, we derived a mathematical model that can perfectly describe the length distribution of telomere overhangs in several human cell strains. The model suggests that the production of telomere overhangs can be explained by three contributions corresponding to three regions on the G-rich telomere template strand, namely, the region occupied by the last primer, that missed out by this primer at its 5′-side and the 3′-terminus of the template strand that is inaccessible to primase. The model can also be used to simulate incomplete telomere replication.

Contact: tanclswu@public.wh.hb.cn

INTRODUCTION
Telomere is the G-rich repetitive DNA sequence that caps both ends of a chromosome. In human cells, the repeat unit is composed of six nucleotides (TTAGGG). Telomere plays important role in maintaining chromosome stability by protecting chromosome ends from degradation and end-to-end fusion (Bouffler et al., 1996). DNA replication is achieved by a semi-conservative mechanism that requires RNA primer to initiate the unidirectional 5′ to 3′ synthesis of new DNA strand. When the primer at the 3′ end of the linear DNA template is removed, a piece of sequence on the template is left unreplicated and forms a single-stranded protruding 3′-overhang (Figure 1a). This incomplete replication, referred to as the ‘end-replication problem’ in DNA synthesis, results in telomere shortening at each cell division and has been suggested as a method of counting the number of divisions a cell can undergo (Harley et al., 1992). In immortalized cells, such as cancerous cells and germ lines, telomere length is stabilized by telomerase that adds telomeric repeats to telomere ends. The single-stranded telomere 3′-overhang functions as substrate for telomerase (Greider, 1990). The overhang is also involved in forming higher order structure of telomere in mammalian cells (Greider, 1999; Griffith et al., 1999; Han and Hurley, 2000).

In the conventional model of the end-replication problem (Olovnikov, 1973), a chromosome is expected to have a 3′-overhang at one end and be blunt at the other (Figure 1a). The length of the overhang is expected to be about as long as that of the last RNA primer bound to the extreme 3′-end of the template strand. In human cells, the size of RNA primers is about 9 ± 1 nucleotides (DePamphilis and Wassarman, 1980). However, measurements revealed that the telomere overhangs in human cells can be as long as several hundred nucleotides (Cimino-Reale et al., 2001; Huffman et al., 2000; Makarov et al., 1997; Wright et al., 1997). The discrepancy between the expected and actual telomere overhang length suggests a more complicated mechanism in telomere end replication in human cells. Currently there are two proposals to explain this phenomenon. Makarov et al. have estimated that, in the cells they examined, more than 80% of the telomeres have long 3′-overhangs in the range of 130–210 nucleotides (Makarov et al., 1997). They proposed that telomeres might experience post-replication processing by exonuclease that degrades telomere strand at the 5′-end. Consequently, long overhangs can be produced at both ends of a chromosome (Figure 1b). On the other hand, Wright et al. found that only half of the telomeres have long overhangs of about 200 nucleotides long in BJ
cells (Wright et al., 1997). They speculated that during DNA replication, a primer may be randomly positioned and the last primer may not occupy the very end of the 3′-terminus of a parental strand. Therefore, an overhang is the sum of two contributions, one from the sequence occupied by the primer and one from the sequence that is missed out by the primer at the extreme 3′-end of the template strand (Figure 1c).

Although the mechanisms regarding the production of telomere overhangs are different in the above two proposals, they both explain why the length of an overhang can be longer than that of an RNA primer. Direct measurements of telomere overhang length have been made in several human cell strains using electron microscopy (Huffman et al., 2000; Wright et al., 1997). These data provide an opportunity to gain insight into the mechanism involved in the generation of telomere overhangs in human cells. In the present work, we analyze these data mathematically and propose a three-component model for the generation of telomere overhangs in normal human cells. We also discuss the biological implications of the model.

**MODEL OF OVERHANG GENERATION AND ITS COMPARISON WITH EXPERIMENTAL DATA**

Our model is based on the proposal of random positioning of RNA primer (Wright et al., 1997) and the experimental data of electron microscopic measurements of telomere overhang length in five human cell strains (Huffman et al., 2000; Wright et al., 1997). The measurements show that the length of telomere overhangs varies within a range of a few hundred nucleotides. The data seem to share a common characteristic in that they are all bi-phasic with a break at the peak value. This characteristic implies that an overhang is produced through at least two contributions and is compatible with the random primer-positioning model. Based on this fact, we assume that an overhang can be represented by two stochastic components, i.e. a primer-occupied region ($d_1$) and a primer-missed region ($d_2$) at the 3′-end of the template strand (see Figure 1c). Let $d$ stands for the length of an overhang after the removal of the primer, i.e. the number of nucleotides that are not replicated on the parental strand, then we have

$$d = d_1 + d_2. \quad (1)$$

In order to get the density function of $d$, two further assumptions are made as follows partly based on the shape of the experimentally observed distribution. First, the length of primers is various and uniformly distributed over a fixed length between $k_1$ and $k_2$ nucleotides, i.e.

$$p_1(d_1) = \frac{1}{k_2 - k_1}, \quad (k_1 \leq d_1 \leq k_2) \quad (2)$$

Fig. 1. Model of overhang generation at telomere ends in normal human cells. During DNA replication, synthesis of new DNA strand are initiated from RNA primers and elongated in the 5′ to 3′ direction (filled arrow) using a parental DNA strand as template, followed by primer removal and Okazaki fragment ligation to form an intact strand in complementary to the parental one. (a) Conventional model: removal of the last primer (hatched block) at the very end of the 3′-terminus of a parental strand results in a protruding single-stranded overhang (dotted block) at one end of each chromosome. The size of the overhang produced is that of the primer. (b) Exonuclease degradation model: after DNA replication, strand-specific exonuclease trims off nucleotides at the 5′-end of each DNA strand. Therefore, both ends of a chromosome will have an overhang, one contributed by exonuclease hydrolysis (crosshatched block) and one by the combination of exonuclease hydrolysis (crosshatched block) and primer occupancy (dotted block). (c) Random primer-positioning model: during DNA synthesis, the last primer may not be placed at the extreme 3′-end of a template strand. After the removal of this primer (hatched block), an overhang is composed by the nucleotides that were either occupied (dotted block) or missed out by the primer at its 3′-side (vertically lined block) on the parental strand. The other end of the chromosome is blunt. The overhang size in both models (b) and (c) can be larger than that of the primer.
with a mean of \((k_1 + k_2)/2\). Second, the synthesis of a primer on to the template strand is random and the probability for each nucleotide at the template strand to be missed out by the primer is the same, then an exponential distribution

\[
p_2(d_2) = \frac{e^{-d_2/m}}{m}, \quad (k_2 \leq d \leq \infty)
\]

is used for \(d_2\). Here \(m\) is the mean of \(d_2\). By the following integration to combine the above two components:

\[
p(d) = \int_{k_1}^{d} p_1(x) p_2(d - x) \, dx, \quad (k_1 \leq d \leq k_2),
\]

\[
p(d) = \int_{k_1}^{k_2} p_1(x) p_2(d - x) \, dx, \quad (k_2 \leq d \leq \infty),
\]

the density function for the length distribution of telomere overhangs is then obtained as

\[
p(d) = \begin{cases} 
\frac{1}{k_2 - k_1} [1 - e^{-(d-k_1)/m}], & (k_1 \leq d \leq k_2), \\
\frac{1}{k_2 - k_1} (e^{2d/m} - e^{k_1/m}) e^{-d/m}, & (k_2 \leq d \leq \infty),
\end{cases}
\]

where \(p(d)\) represents the probability of producing an overhang of \(d\) nucleotides. The mean of \(d\) is \([(k_1 + k_2)/2 + m]\).

In order to find out how well the model describes the distribution of overhangs in human cells, (6) was fitted to the experimental data and the results are given in Figure 2. It can be seen that, although the distribution of the overhangs is different in different cell strains, the model is able to fit them all satisfactorily.

**INTERPRETATION AND IMPLICATION OF THE MODEL**

The perfect fitting of (6) to the experimental data demonstrates the rationality of the model and supports the random primer-positioning proposal. However, when the value of the parameters is examined, it is noticed that \(k_1\) and \(k_2\), which is designated to represent the minimal and maximal length of primers respectively, is far greater than the size of primers for human cells, which is about nine nucleotides (DePamphilis and Wassarman, 1980). For instance, the \(k_1\) and \(k_2\) for human umbilical vein endothelial cells are 170 and 290 nucleotides respectively (see the legend of Figure 2). Such a large primer is very unlikely.

For the above reason, a modification of the model is required. Mathematically speaking, the component \(d_1\) originally assigned to represent primer occupancy can be decomposed into two components, one with a fixed length of \(k_0\) and one with length uniformly distributed between \((k_1 - k_0)\) and \((k_2 - k_0)\). Let us redefine \(d_0\) and \(d_1\) to be the length of these two components respectively, then their length distributions are given by the following function respectively:

\[
p_0(d_0) = 1, \quad (d_0 = k_0), \quad (7)
\]

\[
p_1(d_1) = \frac{1}{(k_2 - k_0) - (k_1 - k_0)} \frac{1}{k_2 - k_1}, \quad [(k_1 - k_0) \leq d_1 \leq (k_2 - k_0)]. \quad (8)
\]

This modification does not alter the density function for the telomere overhang distribution given in (6), but provides an alternative interpretation for it. Now the interpretation is that a telomere overhang might be generated through a combination of three components represented by \(d_0\), \(d_1\) and \(d_2\) respectively. The length distribution of each component is given by (7), (8) and (3) respectively.
Now we have to find out what biochemical event or structure \(d_0\) and \(d_1\) may represent respectively. Human cells have primers of well-defined size of 9(±1) nucleotides (DePamphilis and Wassarman, 1980). From (8) and Figure 2, it can be seen that the value of \(d_1\) spans over a large range within a few hundreds nucleotides. Taking human umbilical vein endothelial cells as an example, the difference between the minimum and maximum of \(d_1\) is 120 nucleotides. This component can not be the primer occupancy because its value is far larger than the primer size. Therefore, the primer occupancy can only be represented by \(d_0\), and \(d_1\) should represent a biochemical event or structure that is currently unclear. The component \(d_1\) is unlikely a contribution of exonuclease hydrolysis of telomere at the 5'-end. The fact that overhang distribution is perfectly described by (6) suggests that the overhang pool is a homogeneous population produced through a same mechanism. If nuclease hydrolysis makes such a considerable contribution, then the whole overhang pool detected will be a heterogeneous mixture of two populations of overhangs produced through two different mechanisms, i.e. one with involvement of primer and one without. In this case, the overhang distribution will not be explained by (6). Because \(d_0\) and \(d_2\) together represent the region on the template strand that is freely accessible to primase, \(d_1\) then can only be explained by a region at the 3'-terminus of template strand that may not be accessible to primase during the replication of telomere ends. A schematic illustration of how an overhang can be produced by this modified model is given in Figure 3 and the correspondent parameters obtained by fitting the model to experimental data are provided in Table 1.

Table 1. Parameters of the three-component model of telomere overhang generation obtained by fitting the model to experimental data of telomere length distribution in five human cell strains

<table>
<thead>
<tr>
<th>Parameter</th>
<th>BI foreskin fibroblasts</th>
<th>HME31 human mammary epithelial cells</th>
<th>IMR90 lung fibroblasts</th>
<th>Human umbilical vein endothelial cells</th>
<th>BI foreskin fibroblasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_0)</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>(m)</td>
<td>90</td>
<td>75</td>
<td>30</td>
<td>140</td>
<td>95</td>
</tr>
<tr>
<td>(k_1)</td>
<td>6</td>
<td>161</td>
<td>111</td>
<td>161</td>
<td>0</td>
</tr>
<tr>
<td>(k_2)</td>
<td>136</td>
<td>236</td>
<td>187</td>
<td>281</td>
<td>171</td>
</tr>
</tbody>
</table>

The graphic presentations of the fittings are identical to those shown in Figure 2. Experimental data are from Huffman et al. (2000), Wright et al. (1997). Here \(k_1' = (k_1 - k_0)\) and \(k_2' = (k_2 - k_0)\).

![Fig. 3. Interpretation of the three-component model of telomere overhang generation. A telomere overhang is produced by a combination of three contributions during the replication of telomere end. They can be assigned to three regions on the parental template strand: (1) the region (dotted block) bound by the last RNA primer (hatched block). (2) The region (vertically lined block) missed out by the last primer at its 5'-side. (3) The region at the 3'-end (horizontally lined block) that is protected and inaccessible to primase. Here \(k_1' = (k_1 - k_0)\) and \(k_2' = (k_2 - k_0)\), which are the minimum and maximum values of \(d_1\). The former two components are produced by the random positioning of the last primer onto the accessible region adjacent to the blocked region at the template. Therefore, a long telomere overhang can be produced through two distinct activities, i.e. the random positioning of the last RNA primer onto the G-rich telomere template strand \(d_0\) and \(d_2\) and the blockage of the 3'-terminus at this telomere strand \(d_1\).](image-url)
factors, TRF1 and TRF2, have been shown to regulate telomere length negatively, possibly by making the 3′-terminus inaccessible to telomerase through promoting formation of a closed t-looop (Smogorzewska et al., 2000). These facts indicate that the accessibility of the single-stranded telomere overhang plays a role in the homeostasis of telomere length.

Perhaps such a mechanism could also be functional during the replication of telomere ends in immortalized cells that are telomerase-positive. It can be seen that during DNA replication, telomerase should distinguish the 3′-terminus at the template strands from those at the newly synthesized ones. The 3′-terminus of the template strand should not be elongated in order to maintain telomere length stability. If it is, then the telomere length will expand gradually over successive cell divisions. In our opinion, this can be prevented by hiding the 3′-terminus of the template strand so that it is not accessible to telomerase. Under this circumstance, this piece of sequence may also be inaccessible to primase; therefore, possibly contributing to the production of telomere overhangs (Figure 3).

DISCUSSIONS

This work is based on the electron microscope measurements of telomere overhang length in several human cells (Huffman et al., 2000; Wright et al., 1997). Recently, Cimino-Reale et al. measured telomere overhang length distribution in two normal human cells and two cancerous cells by oligonucleotide ligation assay (Cimino-Reale et al., 2001). The distributions in these cells are dramatically different from those obtained by electron microscopy. The reason for the difference is not clear. A major concern with the oligonucleotide ligation measurement is that the distributions in four different cells are virtually identical to each other. It has often being seen that different type of cells loss different amount of telomeric sequence in cell division. To mention a few example, human lymphocytes loss 120 (Vaziri et al., 1993), human fibroblasts 40 (Levy et al., 1992) or 139 (Allsopp and Harley, 1995), BJ foreskin fibroblasts 49, IMR90 lung fibroblasts 64, HME31 human mammary epithelial cells 89, human umbilical vein endothelial cells 101 (Huffman et al., 2000) base pairs per cell division. Because it has been shown telomere shortening is proportional to the size of the telomere 3′-overhang (Huffman et al., 2000), it is expected that different type of cells may have different telomere overhang length distributions. This is what was observed in the measurement of telomere overhang length by electron microscopy (Huffman et al., 2000; Wright et al., 1997). Perhaps the identical overhang length distributions in different cells obtained by the oligonucleotide ligation measurement may have something to do with the technique used.

Our work shows that with an additional contribution, the random primer-positioning model explains quantitatively the length distribution of long telomere overhangs in several normal human cells. It suggests that long telomere overhangs in human cells can be produced through two distinct activities, i.e. random positioning of the last RNA primer onto the G-rich telomere template strand and blockage of the 3′-terminus at this telomere strand. The perfect fitting of our model to the length distribution of five human cell strains suggests that the long overhangs in different type of cells are produced through a common mechanism although the distribution pattern can be different. It does not exclude the possibility of exonuclease processing of the 5′-end in the generation of 3′-overhangs, because the techniques used in measuring overhang length by microscopy can not detect very short overhangs (Wright et al., 1997).

Our work also suggests that a small piece (within a few hundreds nucleotides) of sequence at the 3′-terminus of the parental telomere strand may be blocked and inaccessible during the replication of telomere ends. In order to maintain telomere length stability in telomerase-positive cells, telomerase may not work on the 3′-terminus of the template telomere strand but only on that of the newly synthesized one. The inaccessibility of the 3′-terminus at the template telomere strand is a possibility to achieve this goal. In the meantime, this may also contribute to the formation of telomere overhangs. The exact biochemical nature of this possibility is currently not certain. We speculate this may involve specific protein binding and/or higher order structures of telomere formed by telomerase overhang.

The size of long overhang is directly proportional to the rate of telomere shortening in human cells (Huffman et al., 2000). This fact suggests that the long overhang is the primary determinant of the rate of telomere shortening. The length of an overhang reflects the number of nucleotides that are not replicated or lost at the 5′-end of its complementary strand. In principle, these can be caused by either incomplete NDA replication or post-replication hydrolysis of the 5′-end of telomere strand by exonuclease. Our model does not support an involvement of exonuclease processing in the generation of long telomere overhangs. Therefore, exonuclease processing of the 5′-end of telomeres is unlikely a primary contribution to the telomere shortening in normal human cells.

ACKNOWLEDGEMENTS

Supported by grant nos G2000057001, 39925008 and 39970365 from the NSFC.

REFERENCES


