STRUCTURAL ANALYSIS OF TELOMERE BINDING FACTORS AND REPETITIVE DNAs

Nicole Fouché

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics.

Chapel Hill
2006

Approved by:

Jack D. Griffith

Shawn Ahmed

Michael B. Jarstfer

Barry R. Lentz

Brian D. Strahl

ABSTRACT

NICOLE FOUCHÉ: Structural Analysis of Telomere Binding Factors and Repetitive DNAs
(Under the direction of Jack D. Griffith)

Telomeres are typically long nucleoprotein structures that protect the ends of linear chromosomes from exonucleolytic attack and from being recognized as DNA double strand breaks. Telomeric DNA is lost either gradually, as a consequence of cellular ageing, or sometimes rapidly, when the telomere structure is disrupted or due to errors in telomere maintenance. When greatly shortened, telomeres will signal a DNA damage message that results in an active growth arrest and eventually, cell death. Telomeres are thus molecular clocks that present a formidable barrier against cancer, because functional telomeres are essential for continued cell proliferation. Thus an important strategy for cancer therapeutics lies in understanding and targeting telomere-lengthening mechanisms used by cancer cells. However, telomere length is determined by the intricate interplay of the components that lengthen telomeres and those that shorten them. Therefore, a detailed understanding is also needed of the structures and functions of telomere factors present in normal cells, and of how changes in these can lead to cancer. Presented here are structural analyses of the telomeric DNA itself; the yeast RNA component of the telomere-lengthening protein, telomerase; and the DNA

binding properties of telomerase and another telomere-specific protein, TRF2. Together, the data adds to a more comprehensive structural model for the manner in which telomeres function *in vivo*.

For Michael and Carys:

My Family, My Joy,

I thank you.

ACKNOWLEDGEMENTS

I first want to thank my husband Michael, without whose love, strength and constant support, I could not have stayed the course through the most difficult time of my life. I only hope that I can return the favor during this lifetime that we share.

To my daughter Carys, thank you. For the laughter and the joy that has made this process so easy. Thank you also for the perspective. You brought everything into focus in a way that allowed me to finish easily and well.

I am grateful for the unvarying love, conversation, insight and encouragement that I received from my Mom, Dad (Deon), Danièlle, David, André, John, Geraldine, Jay and Michelle. You have allowed this time to be about personal growth, rather than merely intellectual growth, so that I feel that I have learned the true lesson that I came here to learn.

Thanks to Sarah, Tony and Deepa for their honest (sometimes too honest) friendship and understanding. A lab can be very close quarters when spent with the same people for so many hours, so thanks for the laughter and conversations – I enjoyed them all.

I would also like to thank the following people for their advice and mentorship during the past five years: Jack Griffith, Barry Lentz, Michael

Jarstfer, Shawn Ahmed, Brian Strahl, Aziz Sancar and Sasha (Alexander) Makhov.

Finally, a note of thanks to Barry Lentz and Lisa Phillippie for providing us with such a fantastic Biophysics program. I am honored that I was allowed to participate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Telomeres are protein-DNA structures that regulate the action of cellular proteins at the ends of linear chromosomes. They consist of long arrays of repeated DNA elements, bound by specific DNA binding proteins, and which terminate in a single-stranded (ss) 3' overhang (1,2). In all vertebrates, telomeric DNA is comprised of the tandem repeat $TTAGGG_n$ (3,4). In humans, this non-coding DNA extends from 5 to 15 kilobase pairs, depending on the tissue and donor age, ending in an overhang that ranges from ~60 to 200 nucleotides in length (5-9). Telomeric chromatin is thought to be organized into lariat-like structures (t-loops) found at the ends of the telomeres of humans, mice, plants, yeasts and protozoa (10-15). Human telomeric DNA *in vivo* is complexed by the components of the shelterin complex, which are unique to the telomere and include the ds DNA telomere repeat binding factors 1 and 2 (TRF1 and TRF2), the ssDNA telomere factor Protection of Telomeres 1 (POT1), and the proteins that bind to them, Rap1, Tin2 and Tpp1 (16-18). A host of other cellular proteins are found at the telomere, including DNA-repair factors such as Ku70/80 and the Mre11-Rad50-NBS1 complex required for the S-phase DNA damage checkpoint, the homologous recombination (HR) protein RAD51D and the histones (19-21).

TELOMERE LENGTH MAINTENANCE

Most eukaryotic telomeres shorten by ~50-200 nt during each cell cycle, due to the 'end-replication problem', oxidative damage, and nucleolytic processing (7,22-28). In cells with indefinite replicative potential, leading strand telomeric repeats can be restored by the reverse transcriptase telomerase, while the lagging strand is concurrently elongated by the standard replicative machinery (29,30). However, the activity of telomerase is negligible in human somatic cells and in culture these cells will cease to divide when their telomere length decreases to ~3 kb (31). This arrested state has been termed replicative senescence, or mortality stage 1 (M1), and is further characterized by an enlarged cell size, expression of a pH-dependent β-galactosidase activity and an altered pattern of gene and protein expression (32). In the absence of p53 and/or p16/Rb, cells can bypass M1 senescence and continue to divide. Thus, telomeres continue to shorten and this will usually result in mortality stage 2 (M2) characterized by many 'uncapped' chromosome ends, mitotic catastrophe and a high fraction of apoptotic cells (33).

Evidence for the direct involvement of telomere length in replicative senescence, is that this process can be prevented by the expression of hTERT, the catalytic component of the ribonucleoprotein telomerase (34-36). Normal human cells stably expressing transfected telomerase either before M1 or in between M1 and M2 can divide indefinitely. Telomeres are thus a

potent tumor-suppressor mechanism in the cell, acting as molecular clocks that count the number of times a cell has divided.

Telomerase

The telomerase complex consists of a telomere-specific reverse transcriptase protein, a template RNA and accessory factors (37-39). The protein subunit (TERT in humans and ciliates and EST2 in yeast) is able to maintain telomere length by using the 3′ single-strand overhang as a primer and a defined region of its integral RNA component (hTERC in humans and TLC1 in yeast) as a template for the de novo synthesis of telomeric repeats (40-43). The exact mechanism used is not yet understood, but we present evidence in chapter 2 that supports a model in which two cooperating telomerase enzymes that bind to and extend a single DNA substrate. Also, we present evidence in chapter 3 that is consistent with a model for an additional function of the RNA component, that is, to act as a flexible scaffold for the telomerase protein subunits (44,45).

In telomerase-positive human cells, telomere length is partly controlled by a feedback mechanism in which telomere elongation by telomerase is limited by the accumulation of the TRF1 complex at chromosome ends (reviewed in (46). More specifically, the amount of TRF1 present at telomeres is dependent on their lengths, the TRF1 complex interacts with and regulates POT1 binding, and human POT1 controls telomerase access to telomeres (47-49). *In vitro* EM studies have shown that TRF1 binds to the ds

telomeric tract in long filamentous arrays, and in the presence of TRF1, loosely-formed telomeric chromatin becomes tightly compacted (Nicole Fouché, unpublished data). Also, TRF1 specifically recognizes nucleosomal binding sites and alters nucleosome structure (50). These data suggest that TRF1 may be involved in the formation of a complex folded structure at the telomere that restricts telomerase access to the overhang within the larger context of telomeric chromatin. Perhaps it is this structure that requires a minimum of ~3 kb of telomeric DNA in humans, below which the resultant change in telomere state (uncapping) leads to growth arrest.

Alternative lengthening of telomeres

In the absence of telomerase, chromosome ends can also be maintained by recombination (51-53). Between 5 and 10% of human cancers maintain their telomeres via a telomerase-independent recombination pathway called alternate lengthening of telomeres (ALT) (54). Interestingly, mammalian ALT cells often contain some very long telomeres up to ~50 kb. Because ALT is a cancer-associated phenotype, it arises during escape from crisis when most telomeres are very short. Thus it is perplexing that these cells are able to use recombination to generate longer telomeric repeats. A possible solution to this problem has been proposed, in which rolling circle replication of telomeric repeat-containing circles is combined with the standard mechanisms of sister chromatid exchange (55-57). Specifically, the ALT mechanism is thought to involve loss of a repressor that controls

inappropriate HR at the telomere, resulting in a resolution event at the t-loop junction (t-loop HR) and the formation of t-loop sized telomeric circles (t-circles). Consistent with this idea, ALT cells contain abundant telomeric circles that are generated as a result of recombination events, because knockdown of XRCC3 or NBS1, two proteins associated with t-loop HR, abolished these circles in ALT cells (S. Compton unpublished data and (58,59).

STRAND-SPECIFIC TELOMERE PROTECTION

Telomere "end-capping" serves to hide the chromosome ends from factors in the cell that would otherwise process them as double-strand breaks or HR intermediates. Telomere protection involves multiple pathways, including repression of non-homologous end joining (NHEJ), repression of a DNA damage signal, protection from inappropriate HR, and control of telomere end processing after replication. When telomeres become "uncapped" due to a change in the telomeric DNA structure or sequence, or when telomere proteins are depleted or mutated, a rapid induction of growth arrest ensues. This growth arrest, termed STASIS (STress or Aberrant Signaling Induced Senescence) is similar to replicative senescence, but it is thought to be brought about by a different signaling pathway (60,61).

Control of telomere end processing after replication

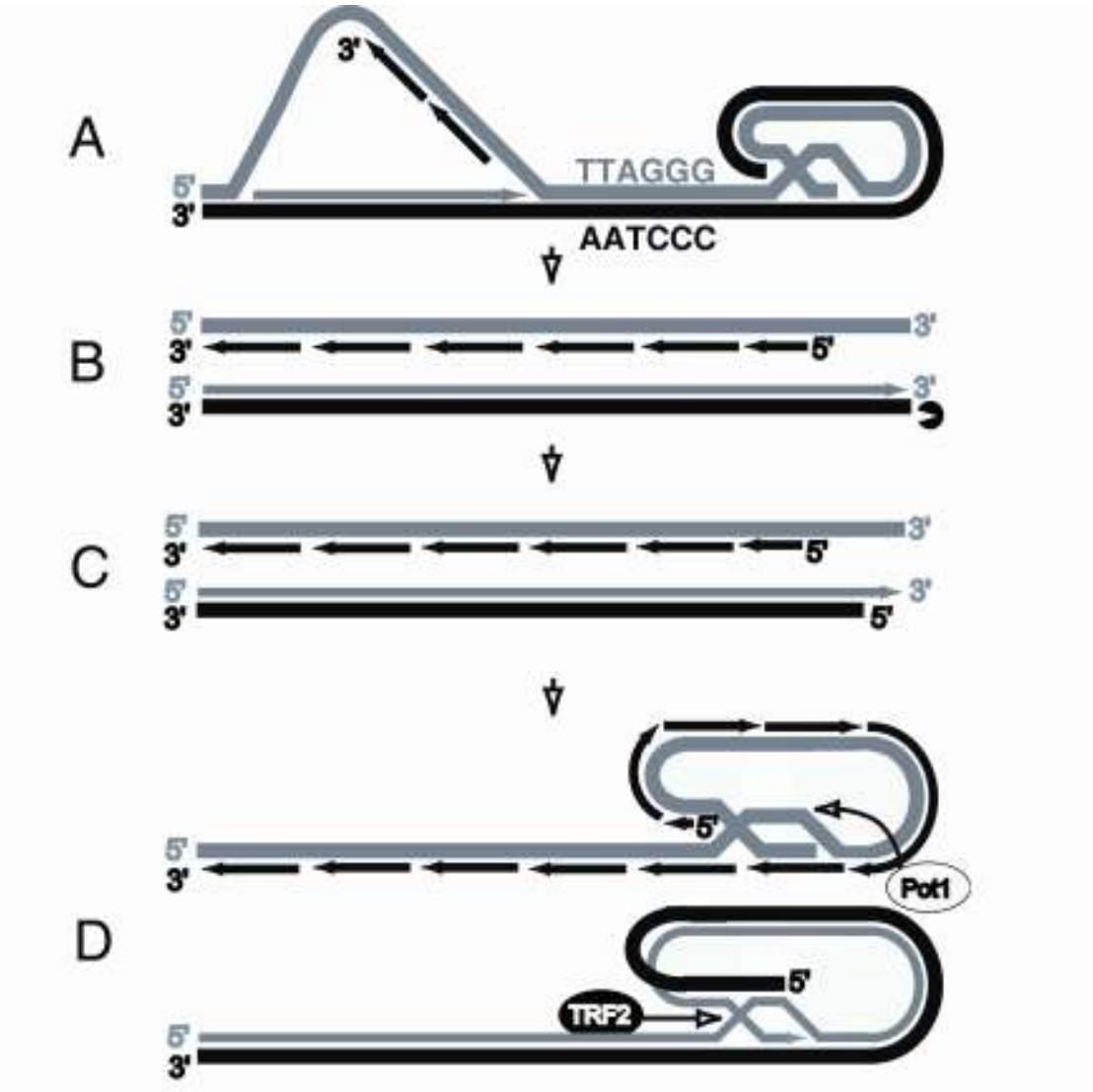
Upon the completion of replication in telomerase-negative cells, leading-strand telomere ends are blunt, because DNA synthesis is continuous on this strand. Conversely, lagging-strand telomeres have a 3′ overhang, owing to removal of the last RNA primer plus the possible failure of the last priming event (end replication problem, Figure 1.1 A-B, (22,23). Because overhangs are present on both chromosome ends however, it is presumed that a nuclease acts on the 5′ C strands of at least the leading-strand telomeres (28). The 5′ to 3′ nuclease Artemis may be this C strand nuclease, because Artemis deficiency results in increased telomere end-to-end fusions, a hallmark of missing overhangs (62).

Interestingly, the terminal nucleotides at C strands are the same on both leading and lagging daughter telomeres, ending at 5′-ATC (28). This suggests that a similar mechanism acts to resect the C strands of both the lagging and leading daughter telomeres. This mechanism of C strand resection is likely conserved, because *Euplotes* and *Tetrahymena* also possess precise nucleotide sequences at C-strand termini (28,63-65).

Recently, the shelterin component Protection of Telomeres 1 (POT1) was found to dictate the precise sequence of the 5′ C strand terminus (62). Thus, POT1 is thought to constrain the action of the nuclease involved, providing the nucleotide-specific boundary for 5′ to 3′ exonuclease trimming.

Figure 1.1: Model of strand-specific telomere protection.

Schematic of the processes involved in telomere end-capping following replication. (A). Strands containing TTAGGG repeats are depicted in grey; strands containing AATCCC repeats are depicted in black. Nascent leading strand indicated by solid grey arrow, nascent lagging strand indicated by broken black arrows. Replication machinery and shelterin components omitted. (B - C). Lagging strand replication generates a natural 3′ overhang (top strand) whereas leading strands are blunt-ended after telomere replication has completed (bottom strand). In this case, overhang generation seems to require nuclease activity. (D). T-loops on lagging (top) and leading (bottom) daughters may have different structures at the sites of strand-invasion, and they may be stabilized by different proteins.


A
3'
TTAGGG
5'
3'
AATCCC
B
5'
3'
3'
5'
5'
3'
3'
5'
C
5'
3'
3'
5'
5'
3'
3'
5'
D
5'
5'
3'
Pot1
5'
TRF2
5'
3'

Indeed, the absence of POT1 at telomeres results in 3′ overhang elongation (66-68).

Repression of NHEJ

Telomere overhangs are presumably required for telomere end capping via t-loop formation (10). The crucial feature of t-loops is thought to be the invasion of the 3' ss overhang into the duplex repeat tract (Figure 1.1 D). Consistent with this idea, POT1, which binds directly to the G-strand overhang, is not required to protect the overhang from degradation and has a relatively minor role in preventing telomere fusions (66-68). On the other hand, loss of TRF2, which can remodel linear telomeric DNA into t-loops *in vitro*, leads to a dramatic telomere-fusion phenotype (69-71). The fusions arise because the nucleotide excision repair nuclease ERCC1/XPF removes the G-overhangs from the uncapped telomeres, allowing ligation of the chromosome ends by the NHEJ machinery (72,73). Thus, TRF2 but not POT1, is required for overhang protection and repressing NHEJ, presumably by facilitating t-loop formation at the chromosome ends.

Protection from inappropriate HR

The exact structure of the strand invasion site of the t-loop is not known. A widely proposed model suggests limited strand invasion of just the G-rich strand, to form a displacement loop (D-loop) (Figure 1.1 D, lagging strand (10). EM studies incorporating *E. coli* single-stranded binding protein

(SSB) staining of ssDNA at the t-loop junctions indicated the presence of 75 to 200 nt of ssDNA. This suggests that overhangs of a similar length had inserted into the duplex repeat tract, resulting in a stretch of displaced G strand DNA that could be coated by the SSB.

Another proposed model for the strand invasion structure involves branch migration of the D-loop and more extensive strand invasion involving both the G and C strands, resulting in a Holliday junction resembling a DNA recombination intermediate (Figure 1.1 D, leading strand, (69). Consistent with this idea, EM studies using a linear model telomere with a very short (6 nt) TTAGGG-3′ tail generated a t-loop that was covalently fixed with psoralen and UV. Thus, it seems probable that in the case of a short overhang, this four-stranded structure is more stable, and thus more energetically favorable, than the D-loop.

Evidence suggests that the structures at the ends of leading- versus lagging-strand telomeres may be different (see model in Figure 1.1 D). In normal cells, leading strand overhangs are significantly shorter than lagging strand overhangs (60 nt versus 105 nt, (5). Also, overexpression of a mutant TRF2 protein ($TRF2^{\Delta B}$) results in preferential deletion of leading strand t-loops by HR, resulting in telomere shortening, a DNA damage response, t-circles and induction of senescence (59). In chapter 5 we will show that this mutant of TRF2 has lost the ability to bind to DNA junctions, specifically four-stranded junctions, even though it retains DNA binding activity (59). On the other

hand, loss of the murine POT1 paralog (POT1a) results in preferential loss of lagging strand telomeres, a DNA damage response, t-circles, and aberrant HR at telomeres (66,67). Furthermore, leading daughter overhangs are preferentially extended in telomerase expressing cells (5). Because POT1 negatively regulates telomere extension by telomerase, it is possible that POT1 restricts telomerase access to the overhangs at lagging strand t-loops, but not leading strand t-loops (48,74). Thus, it seems plausible that the structure at the base of the t-loop in the leading strand may be the more energetically favored four-stranded Holliday junction, stabilized by TRF2, whereas lagging strand overhangs may be sufficiently long for stable D-loop formation, stabilized by POT1 binding to the displaced ssDNA.

Interestingly, it appears that the ds telomeric DNA binding activity of the TRF2 Myb domain alone, presumably in the context of a shelterin complex, is necessary for t-loop formation. Whereas the basic domain was required for t-loop formation *in vitro* (discussed in Chapter 5), it is dispensable for t-loop formation *in vivo*, as evidenced by the presence of t-circles and the lack of NHEJ in cells expressing TRF2$^{\Delta B}$ (59). T-circles are also seen in mouse knockouts of POT1, suggesting that although this protein may be needed for stabilizing the (lagging strand) D-loop, it is not necessary for t-loop formation. Conversely, t-circles are absent after conditional deletion of the TRF2 gene in mouse embryo fibroblasts or upon expression of the dominant-negative mutant that removes endogenous TRF2 from telomeres (TRF2$^{\Delta B\Delta M}$)

even when the NHEJ pathway is inactive (59,70,71,73,75).

Repression of a DNA damage signal

It appears that in addition to t-loop formation, other factors are required to suppress the DNA damage signal. In the above-mentioned POT1 deletion and TRF2$^{\Delta B}$ phenotypes, a DNA damage response was detected despite the presence of t-loops (59,62,66,67). In addition, an apoptotic signal is generated after TRF2 inhibition, even without concomitant overhang degradation, telomere replication or chromosome segregation (73,76). Indeed, subunits of the Mre11 complex (a DNA damage sensor of the ATM pathway) are present at human telomeres throughout the cell cycle (20,77-79). This suggests that the potential for damage recognition is ever-present at the telomere.

One solution to this problem may lie in the ability of TRF2 to bind to the Ataxia-telangiectasia mutated (ATM) kinase and inhibit the ATM-dependent DNA damage response (80). This would be especially important during replication and prior to t-loop formation, when the telomere is unfolded and the overhang is exposed. Indeed, ATM is recruited to telomeres in late S or G2 phase, where it initiates a transient DNA damage response that is subsequently quenched, presumably by refolding of the telomere by the shelterin complex (81-84). However, in the absence of ATM, TRF2 inhibition still results in Telomere Dysfunction-Induced Foci (TIFs) and senescence,

pointing to a second ATM-independent damage-signaling pathway (81).

In human cells, ssDNA alone is able to promote checkpoint activation, by recruiting replication protein A (RPA), which in turn recruits the ataxia-telangiectasia Rad3-related protein kinase (ATR), a key player in the cellular response to DNA damage (85-87). RPA is present at the telomeres of the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, and it is directly involved in telomere maintenance (88-90). RPA is also associated with the telomere maintenance phenotype Alternative Lengthening of Telomeres (ALT) in some human cancers (91,92). Thus it may be that, at a minimum, preventing inappropriate binding of ssDNA binding proteins (e.g. RPA) to ssDNA at the telomere (e.g., at the D-loop), is needed to suppress the telomere-induced DNA damage signal. This role may be attributed to the ss telomeric DNA binding protein, POT1, as was previously suggested, as this would be consistent with the loss of POT1 leading to a telomeric DNA damage response without disruption of the core telomere protein complex or telomere uncapping (68).

STRAND-SPECIFIC TELOMERE LOSS

The mechanism underlying stochastic telomere loss is not known, but may be a result of deletion of t-loops by homologous recombination, or the failure to unwind or correctly process higher-order structures of G-rich

telomeric DNA during telomere replication (59,66,93,94). However, the conventional models do not explain the strand-specific phenotypes that arise from telomere replication. We have discussed the idea of strand-specific deletion of t-loops by HR that could result from differential t-loop formation during postreplicative “recapping” of the leading and lagging strands (above). Taking into account the data presented in chapters 4 and 5, we now propose an additional model for replication-coupled telomere shortening that would also account for these strand-specific telomere deletions.

Problems associated with telomere replication

The telomere is thought to be replicated by the conventional semi-conservative polymerase machinery, including polymerases α and δ (30,63,95). However, the repetitive nature of the telomeric DNA likely presents inherent difficulties for the DNA synthesis machinery. In addition to the presence of G-quartets or tetrameric structures known to form on both the leading and lagging strands, telomeres are thought to accumulate oxidative DNA damage as a result of their high triple-G content (25,96-98). Another impediment to the moving replication fork is the t-loop that must be unfolded for the replication machinery to reach the very end of the telomere, as well as the proteins that are needed to protect the ends from DNA repair factors during this time. Furthermore, in chapter 4, we present data that suggests that stalled replication forks at telomeric repeats have a higher propensity to regress, forming stable four-stranded chickenfoot structures (resembling

Holliday junctions) that would also need to be resolved before replication could continue (Figure 1.2 B).

Telomere-specific replication factors

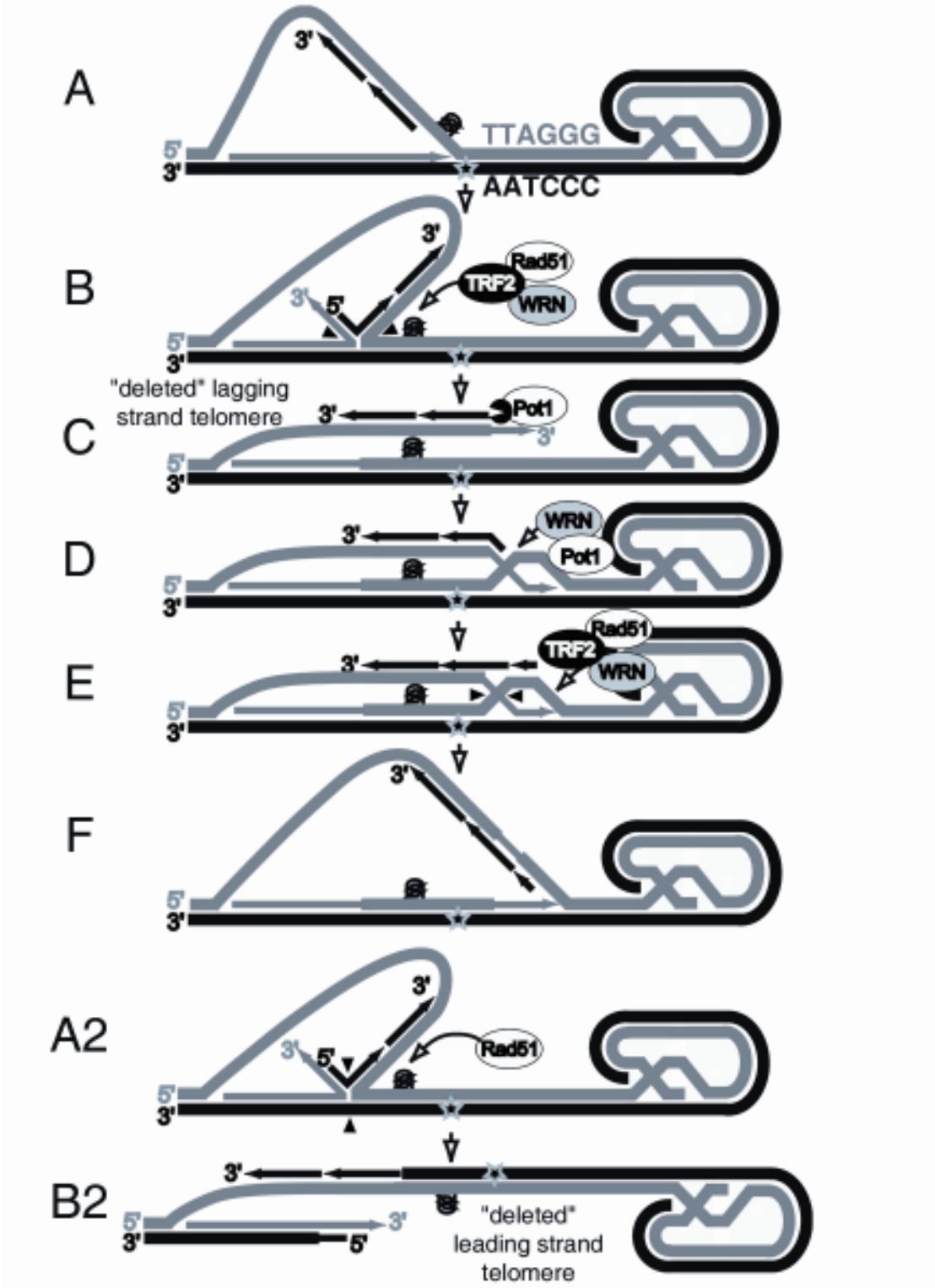
It is possible then that telomere-specific factors facilitate replication of the telomere (99). Telomeres are replicated as rapidly as the bulk DNA, and recently, the telomere binding protein in *S. pombe* (Taz1) was determined to be necessary for efficient replication of yeast telomeres (100,101). Also, Bloom's (BLM) and Werner's (WRN) syndrome helicases, have been shown to be important for proper telomere replication and maintenance in human cells (94,102-106). These RecQ helicases have been shown to unwind G quartets and D-loop structures, and to promote branch migration of 4-stranded junctions similar to chickenfoot structures in telomeric DNA templates (107-114). Moreover, the requirement for telomere-specific helicase activity appears to be conserved, because at least four different helicases have been reported to be involved in telomere function in *S. cerevisiae*, and the putative DNA helicase, Rtel, is a putative regulator of telomere length in mice (93,115-117).

Replication restart

Although factors at the telomere are able to facilitate unwinding of G quartets and even the D-loop, the replication fork is still expected to stall or collapse at sites of DNA damage known to be present at the telomere

Figure 1.2: Model of strand-specific replication restart

Schematic of the potential problems encountered during telomere replication and replication restart that could lead to preferential deletion of daughter strands. (A). Stalled replication due to a problem such as a site of DNA damage (star) or a G-quartet (quadruplex symbol). Strands containing TTAGGG repeats are depicted in grey; strands containing AATCCC repeats are depicted in black. Nascent leading strand indicated by solid grey arrow, nascent lagging strand indicated by broken black arrows. (Replication machinery omitted). (B). Chickenfoot structure and the proteins that may control replication fork regression versus unwinding of problematic structures (i.e. G-quartets). Black arrows indicate sites of G strand cleavage. (C). G strand cleavage by Holliday junction resolvases would lead to a double strand break on the lagging strand telomere. Subsequent exonuclease activity may be required to generate a 3′ overhang for replication restart. (D). Replication restart may require the activities of WRN and POT1 for D-loop formation. (E). Strand migration or lagging strand elongation of the complementary strand of the D-loop are required to form another Holliday junction structure. Black arrows again indicate sites of G strand cleavage by Holliday junction resolvases. (F). The replication fork is re-established. (A2 - B2). Alternate cleavage (black arrows) of the C strands of the chickenfoot structure and then subsequent failure of the replication fork to restart would lead to the loss of leading strand telomeres.


A
3'
TTAGGG
5'
3'
AATCCC
B
3'
Rad51
TRF2
WRN
3'
5'
5'
3'
"deleted" lagging
strand telomere
C
3'
Pot1
3'
5'
3'
D
3'
WRN
Pot1
5'
3'
E
Rad51
TRF2
3'
WRN
5'
3'
F
3'
5'
3'
A2
3'
3'
5'
Rad51
5'
3'
B2
3'
5'
3'
3'
5'
"deleted"
leading strand
telomere

(25,26,118). Nevertheless, replication forks are endowed with an extraordinary potential to coordinate fork stalling with fork resumption processes. Stalled forks generate single-stranded DNA that activates the replication checkpoint, which in turn functions to protect the stability of the fork until replication can resume (118). In the absence of converging forks at the telomere, however, recombination-mediated replication restart and/or damage-bypass mechanisms are expected to facilitate completion of replication, because they do not require replication initiation functions (118,119).

The first step in recombination-mediated replication restart involves fork reversal to form a four-stranded chickenfoot intermediate (Figure 1.2 B). In chapter 5, we will present data that shows that TRF2 recognizes and binds with high affinity to these structures *in vitro*, suggesting that they do in fact occur at telomeres *in vivo*. Indeed, because repetitive DNA has a natural tendency to form chickenfoot structures, it seems plausible that telomere replication would have evolved in order to take advantage of this trait, selecting for replication restart via homologous recombination over a damage-bypass mechanism.

Consistent with this idea, the HR protein RAD51D associates with telomeres and is required for proper telomere maintenance in mouse embryonic fibroblasts (120). This Rad51 paralog forms a complex with RAD51B, RAD51C and XRCC2 (BCDX2) that is able to promote

formation of chickenfoot structures on synthetic replication forks, and which can bind to and facilitate cleavage of Holliday junctions *in vitro* (S. Compton, unpublished data and (121). In mammalian cells, RAD51C is required both for branch migration and HJ resolution (122). Observations in yeast have also suggested an active Rad51-dependent process of four-way junction formation at stalled forks deprived of the replisome (123). It is possible, then, that TRF2-shelterin, in complex with some Rad51 paralogs, associates with stalled fork intermediates to direct fork regression and subsequent cleavage of the G strands by Holliday junction resolvases (black arrows, Figure 1.2 B, chapter 5).

This would effectively release the lagging strand telomere as a free end (Figure 1.2 C). As previously discussed, the Mre11 complex is associated with human telomeres (20). However, the regulatory subunit of Mre11, Nbs1, joins TRF2 at telomeres only in S phase, suggesting a role for this complex in telomere replication (20). Mre11 is capable of tethering DNA fragments, such as the free end of the lagging strand telomere, and it was recently suggested that the essential role of the Mre11 complex in vertebrates might be attributed to the restart of collapsed replication forks (124-127). Interestingly, the association of ATM with TRF2, which did not block ATM kinase activity *in vitro*, would further stimulate this activity of Mre11, because ATM and ATR are able to promote Mre11-dependent restart of collapsed replication forks (80,127). ATM-like kinases are also necessary for telomere

maintenance in *S. cerevisiae* (128) and *S. pombe* (129).

The next steps in recombination-mediated replication restart would presumably involve strand resection of the lagging C strand 5′ terminus (Figure 1.2 C), strand invasion of the resultant 3′ G overhang into the leading strand duplex repeats (Figure 1.2 D) and strand migration or lagging strand elongation of the complementary strand of the D-loop, to form a Holliday junction structure (Figure 1.2 E). The final step would be relatively identical to the first, involving Holliday junction resolution by an activity that again cleaves the G strands (black arrows, Figure 1.2 E). Evidence for this mechanism of replication restart along tracts of repetitive DNA has come from studies of long tracts of CTG repeats, associated with myotonic dystrophy, that undergo deletion at a high rate in plasmids in *Escherichia coli* (130,131).

Because the above-mentioned steps resemble t-loop formation and t-loop HR, these processes would no doubt require the actions of several proteins already implicated in telomere end capping or t-loop HR. Thus the 5' C strand exonuclease would likely be the same protein involved in C strand resection during telomere end processing and would likely be regulated by POT1 binding to the overhang (Figure 1.2 C). Strand invasion during HR is typically catalyzed by the combined actions of RAD51, RAD52 and RPA, of which RAD51 and RPA both have telomere phenotypes (88-90,120,132). However, WRN also naturally co-localizes with RPA upon replication arrest and is able to prevent aberrant recombination events at sites of

stalled replication (108). Since POT1 cooperates with WRN to efficiently unwind telomeric substrates *in vitro*, both proteins may also be required for formation of the strand-invasion D-loop structure at telomeres (Figure 1.2 D and (133). POT1 may also be required to stabilize the D-loop structure by binding to the displaced ss G rich DNA (Figure 1.2 D). Moreover, WRN may be required for branch migration of the D-loop or lagging strand synthesis required to establish a bona fide recombination substrate between the two replicating strands (Figure 1.2 E). Consistent with this idea, deletion of POT1 in mice leads to the preferential loss of lagging strand telomeres, and human cells lacking WRN also exhibit deletion of telomeres replicated by lagging strand synthesis (66,94). Because telomeric G- and C strand synthesis are coordinately regulated in *Euplotes crassus*, this suggests that in eukaryotes, leading strand replication is unlikely to continue in the absence of lagging strand replication (63). Thus, a possible explanation for the above-mentioned phenotypes could be that collapsed replication forks that had already been cleaved as in Figure 1.2 C, failed to restart in these cells. Indeed, if this occurred close to the sites of replication initiation, the lagging strand telomeres would have been effectively removed. In this case, lack of replication of the lagging strand telomeres is expected to mimic telomere loss.

At telomeres, the key regulatory factor involved in replication appears, however, to be TRF2. In human cells, TRF2 co-localizes and physically interacts with WRN (134), and it binds to and stimulates the activities of both the WRN and BLM helicases *in vitro* (105). Also, the dominant negative

mutant of TRF2 ($TRF2^{\Delta B\Delta M}$) is able to bind to POT1 and prevent it from localizing to telomeres, suggesting that TRF2 binding to telomeric DNA may be required for POT1 recruitment to, for example, recombination intermediates with little to no ssDNA present (Figure 1.2 C, (135). However, it is the separation of function mutant of TRF2 ($TRF2^{\Delta B}$) that may have revealed the most about TRF2 function during telomere replication. When $TRF2^{\Delta B}$ was expressed in cells, it induced catastrophic deletions of telomeric DNA in which both the C-rich and the G-rich strands became shortened (59). However, as previously mentioned, $TRF2^{\Delta B}$ expression resulted in the preferential deletion of leading strand telomeres, suggesting problems during telomere replication (59). This was dependent on XRCC3 (another Rad51 paralog associated with Holliday junction resolution activity) and Nbs1, suggesting problems during replication restart via the recombination pathway.

Although $TRF2^{\Delta B}$ is unable to bind to the chickenfoot intermediates of fork regression, it is still able to bind to the duplex telomeric repeats and thereby recruit the proteins, such as POT1 and WRN that may be needed for efficient replication. Thus, the loss of leading strand telomeres may be limited to problems encountered during telomere end-capping, as previously discussed. However, another possible explanation exists if we consider a possible novel function for TRF2 at stalled replication forks. That is, if TRF2 was needed to direct the preferential cleavage of the G rich strands of chickenfoot intermediates, then this could also account for the postreplicative loss of leading strand telomeres in the $TRF2^{\Delta B}$ phenotype (Figure 1.2 A2,

B2). Evidence suggests that the C rich strands may indeed be the default during Holliday junction resolution. In humans, the branch migration-associated resolvase (of which RAD51C and XRCC3 are components) cleaves Holliday junctions preferentially between two cytosine residues (122,136).

SCOPE OF DISSERTATION

Presented here are structural analyses of three of the key components of the telomere: telomeric DNA, telomerase and TRF2. The primary goal was to reveal mechanistic clues that would contribute to the broader understanding of the way in which telomeres regulate, and are regulated by, processes in the cell.

In Chapter 2, we present data obtained in collaboration with the laboratory of Michael B. Jarstfer, that was published in the journal *Biochemistry*, and which is reproduced here with permission (137). The binding of the telomerase ribonucleoprotein from the ciliate *Euplotes aediculatus* to telomeric DNA *in vitro* is examined by electron microscopy (EM). Visualization of the structures that formed revealed a globular protein complex that localized to the DNA end containing the *E. aediculatus* telomere consensus 3'-single-strand $T_4G_4T_4G_4T_4G_2$ overhang. Gel filtration confirmed that purified *E. aediculatus* telomerase is an active dimer in solution, and

comparison of the size of the DNA-associated complex with apoferritin suggests that it binds to a single telomeric 3'-end as a dimer. Up to 43% of the telomerase-DNA complexes appeared by EM to involve tetramers or larger multimers of telomerase in association with two or more DNA ends. These data provide the first direct evidence that telomerase is a functional dimer and suggest that two telomerase ribonucleoprotein particles cooperate to elongate each *Euplotes* telomere *in vivo*.

Recently, a detailed working model for the secondary structure of the *S. cerevisiae* telomerase RNA, TLC1, was proposed by the laboratory of Thomas R. Cech, using *mfold* predictions of energetically favorable RNA conformations coupled with genetic, biochemical and phylogenetic data from four species of *Saccharomyces* (44). This model proposes that the telomerase RNA consists of a central catalytic core containing the RNA template and Est2p-binding region, plus three long quasihelical RNA arms that bind to telomerase accessory factors. In chapter 3 we have used EM, in collaboration with the Cech laboratory, to test this *mfold* structure prediction. We directly visualized a full-length TLC1 RNA, as well as a deletion mutant (ΔΔ) proposed to have truncation of the longest RNA helix. Our results were consistent with the existence of three RNA arms in the full-length molecule. Also, a nano-scale 'biopointer' confirmed that in the majority of molecules, a biotinylated oligonucleotide that was complementary to the RNA template region was most often located at the center of the RNA.

Chapter 4 describes a study examining two of the repetitive sequences found in the human genome: the telomeric repeat, necessary for the protection of chromosome termini, and the disease-associated triplet repeat (CTG)•$(CAG)_n$. Evidence suggests that replication of both types of repeats is problematic, and that a contributing factor is the repetitive nature of the DNA itself. Here we have used EM to investigate DNA structures formed at replication forks on large model DNAs containing these repeat sequences, in an attempt to elucidate the contributory effect that these repetitive DNAs may have on their replication. Visualization of the DNA revealed that there is a high propensity for a paused replication fork to spontaneously regress when moving through repetitive DNAs, and that this results in a four-way chickenfoot intermediate that could present a significant block to replication *in vivo*, possibly leading to unwanted recombination events, amplifications or deletions.

The replication of long tracts of telomeric repeats may therefore require specific factors to avoid fork regression. In Chapter 5 we show that TRF2 binds to model replication forks and 4-way junctions *in vitro*, in a structure-specific, but sequence-independent manner. A synthetic peptide encompassing the TRF2 basic domain also binds strongly to DNA 4-way junctions, whereas the TRF2 truncation mutant ($TRF2^{\Delta B}$) and a mutant basic domain peptide do not. In the absence of the basic domain, the ability of TRF2 to localize to model telomere ends and facilitate t-loop formation *in vitro* is greatly diminished. We propose that TRF2 plays a key role during

telomere replication in binding chickenfoot intermediates of telomere replication fork regression. Junction-specific binding would also allow TRF2 to contribute to the stability of a three- or four-stranded DNA structure that may exist at the strand invasion site of the t-loop.

CHAPTER 2: ELECTRON MICROSCOPIC VISUALIZATION OF TELOMERASE FROM *EUPLOTES AEDICULATES* BOUND TO A MODEL TELOMERE DNA[1]

INTRODUCTION

Telomeres are nucleoprotein structures essential for chromosome stability in eukaryotes and regulation of the replicative lifespan of somatic cells. Encompassing the termini of all linear chromosomes, the telomeric DNA typically consists of long arrays of short tandem repeats *(4)* bound by specific DNA binding proteins, and terminates in a single-stranded 3′ overhang (1,2). In the absence of telomerase or another mechanism to maintain telomere length, most eukaryotic telomeres shorten by 50-200 nucleotides (nt) during each cell cycle, due in part to the 'end replication problem' that results from the inability of the lagging strand to be replicated to the very end of the chromosome (22,23).

---

[1] This chapter was published in Fouché N. et al. (2006) *Biochemistry* **45** (31), 9624-31.

The telomere-specific reverse transcriptase telomerase is able to maintain telomere length by using the single strand overhang as a primer and a defined region of its integral RNA component as the template for the *de novo* synthesis of telomeric repeats (37,138). One unique feature of telomerase is its ability to add multiple copies of the repeat to a DNA substrate following a single initial binding event (138,139). This telomerase processivity depends on two types of translocation, type I or nucleotide addition processivity and type II or repeat addition processivity (140,141). Nucleotide addition processivity involves simultaneous movement of the RNA–DNA duplex relative to the active site after each nucleotide addition. Repeat addition processivity involves unpairing of the RNA–DNA hybrid after repeat addition, followed by translocation and re-alignment of the DNA substrate relative to the 3' region of the RNA template. It has been proposed that telomerase can also associate with its DNA substrate via template-independent interactions that are regulated by a protein-dependent anchor site (142) and that these interactions may determine whether the product can remain bound to telomerase when released from the template site (143).

A feature of telomerase that may contribute to telomerase processivity is its ability to form dimers or multimers. Human TERT, the reverse transcriptase component of telomerase, as well as telomerase RNA form functionally cooperative oligomers in cell lysate or when reconstituted *in vitro* (144,145) and a mutation that weakens human telomerase RNA dimerization preferentially impairs type II processivity (146). *Tetrahymena*

thermophila telomerase eluted from a gel filtration column at the size of a monomeric complex (147), but in *Saccharomyces cerevisiae* the telomerase ribonucleoprotein particle (RNP) contains at least two active sites that both act as templates for DNA polymerization (148). Also, while glycerol gradient centrifugation of purified *Euplotes aediculatus* telomerase suggested an active RNP monomer in solution (149), gel filtration chromatography of *E. aediculatus* nuclear extracts revealed a functional telomerase complex in agreement with telomerase dimer formation (150). Indeed, *Euplotes crassus* telomerase complexes contain at least two active sites and the telomerase catalytic subunit *Ec*TERT undergoes multimerization *in vitro* (151).

The existence of telomerase multimers suggests that two or more telomerase RNP's may cooperate during processive elongation to simultaneously extend one or more DNA substrates. Three models have thus been proposed for such cooperation in a coordinated dimer (145,148,152). The parallel extension model consists of two active sites within two different but associated telomerase RNP's simultaneously extending two separate chromosome 3′ ends (145,148). It has been proposed that this type of coordinated extension could exist for elongation of both leading and lagging strand telomeres after DNA replication of sister chromatids. The template switching model evokes two catalytic sites within a dimer of telomerase acting sequentially during processive telomere synthesis to elongate a single telomere 3′ end (145). After addition of a repeat by the first telomerase RNP and upon translocation, the DNA substrate is re-aligned relative to

the 3′ region of the RNA template within the second telomerase RNP. The DNA anchor site model proposes one telomerase RNP template stabilizing the interaction with a single telomere, while the other template is used for reverse transcription of the 3′ end (152). The template of one telomerase subunit is thus used primarily for substrate binding, presumably interacting with nucleotides upstream of the 3′ end of the DNA substrate, while the other is copied during telomere repeat addition. It is possible that these models are not mutually exclusive, but that combinations of all three may exist, with the exact mechanics yet to be revealed.

E. aediculatus is a hypotrichous ciliate with a polyploid macronucleus containing millions of gene sized chromosomes (153). The abundance of telomerase in each *E. aediculatus* cell thus makes it possible to routinely purify telomerase from crude extracts. *E. aediculatus* telomerase contains three subunits including *Ea*TR, the 64 kDa RNA subunit, *Ea*TERT, the 123 kDa catalytic subunit, and a 43 kDa telomerase accessory protein, a La homolog that is important for nuclear retention and anchorage to an apparent end-replication complex (149,154). Together, these proteins form the functional ~230 kDa telomerase RNP *in vivo.*

Here, we used gel filtration and EM to examine purified *E. aediculatus* telomerase RNP and its binding to a model chromosome containing the *E. aediculatus* consensus 3′ overhang $T_4G_4T_4G_4T_4G_2$ (1,155). Size exclusion chromatography of telomerase compared to proteins with known

molecular weights suggests that purified *E. aediculatus* telomerase is active as a dimer in solution. EM examination of binding reactions revealed mostly globular telomerase complexes bound to DNA. Comparison of particle size (projected area) to the globular protein apoferritin suggests that telomerase binds to telomeric 3′ ends as a dimer, and that higher order multimerization of these bound telomerases occurs, accordingly associating free DNA ends *in vitro*.

MATERIALS AND METHODS

Growth of E. aediculatus and Preparation of Nuclear Extracts

E. aediculatus was grown under nonsterile conditions using Chlorogonium as the food source as previously described. Cultures were grown in 5 gallon flasks with continuous aeration (156). Nuclei were isolated by sucrose cushion gradient centrifugation, and nuclear extracts were prepared by Dounce homogenization, as previously described (155).

Purification of E. aediculatus telomerase

Telomerase was purified from nuclear extracts following the procedure described by Lingner and Cech (149). In short, nuclear extracts (15 ml from 1 $\times$ 10^9 cells) were chromatographed over a heparin sepharose column (Amersham Biosciences, Piscataway, NJ) using a linear gradient of

increasing potassium glutamate, and the telomerase containing fractions were pooled and concentrated using an Amicon stirred cell concentrator (Millipore, Billerica, MA). Telomerase was purified from these enriched fractions by affinity chromatography using a bait oligonucleotide, 5′-biotin-TAGACACCTGTTA-(rmeG)$_2$-(rmeU)$_4$- (rmeG)$_4$- (rmeU)$_4$-(rmeG)-3′ to trap telomerase onto Ultralink Neutravidin beads (Pierce Biotechnology, Inc., Rockford, IL). Telomerase was displaced by a chase oligonucleotide that is complementary to the bait, and the chase was removed by extensive dialysis against reaction buffer (20 mM Tris Acetate (pH 7.5), 10 mM $MgCl_2$, 50 mM potassium acetate, and 1 mM DTT) and concentrated.

Gel Filtration

An AKTA FPLC system equipped with a Superdex 200 10/300 GL column (Amersham) was used. The column was equilibrated with 20 mM Tris-HCl pH 7.5, 200 mM potassium acetate, 10 mM $MgCl_2$, 1 mM EDTA, 10% glycerol and 1 mM DTT. Affinity-purified telomerase was injected and the column was run at a rate of 0.4 ml/min at 4 ºC. After 7 ml of void volume had passed through the column, 200 µl fractions were collected. The column was calibrated three times using the high molecular weight calibration kit from Amersham (R^2 = 0.9978). Aldolase (158 kDa), Ferritin (440 kDa) and Thyroglobulin (669 kDa) were run at the same time on the Superdex 200 column and their retention volumes were determined twice, in two separate runs, and fit to the curve y = -148.18x + 2124.4. The marker protein

Catalase (232 kDa) was run as a separate control to validate the calibration curve, since it has a retention volume that would be similar to that of a telomerase monomer.

Telomerase Quantification

The telomerase was quantified by analysis of the RNA subunit by solution hybridization with a ^{32}P-labeled probe for the *E aediculatus* telomerase RNA. Quantities were normalized to RNA standards and the amount of telomerase RNP was determined according to the calculation: ng RNP = ng RNA / 64 kDa x 230 kDa, where the mass of the RNA subunit is known to be 64 kDa and the mass of the telomerase RNP monomer is 230 kDa (149,154).

DNA Manipulation

Model telomeres were synthesized by a modification of the method described by Stansel et al. (69). In short, the plasmid pRST5 (10 µg), was digested with BsmB1 to generate a 5′ overhang with the sequence TCCC. The linearized DNA was incubated with the Klenow fragment of DNA polymerase 1, 33 µM dTTP, and 0.4 mM dCTP to create a blunt end on one side of the linear DNA. A 3′ overhang was generated by ligating either a telomeric DNA Ea_CM_22 ($AGGGT_4G_4T_4G_4T_4G_2$) or non-telomeric DNA (AGGGATTGAATGACTACGAAGATGAA) oligomer onto linearized pRST5. Ligation reactions contained 5 µg linearized pRST5, 5-fold molar excess of 5′

phosphorylated oligonucleotide, T4 DNA ligase (200 U, New England Biolabs, Inc., Ipswich, MA) and T4 DNA ligase buffer (New England Biolabs) and were incubated at 25º C for 30 min. Excess oligonucleotide was removed by size exclusion chromatography using Sephacryl-400 (Promega). DNA was deproteinized by the addition of 80 µg/ml proteinase K in 1% SDS followed by extraction with phenol/chloroform and concentration by ethanol precipitation. DNA was resuspended in 10 mM Tris-HCl (pH 7.6), 1 mM EDTA to give a final concentration of ~20 ng/µl. Ligation efficiency was determined by treating ligated and unreacted linearized pRST5 with the Klenow fragment of DNA polymerase 1 and [α-^{32}P]-dGTP and dATP and comparing the percent of incorporated dGTP. Typically, we achieved 70% ligation efficiency.

Telomerase Activity Assay and Catalyzed Extension of Model Telomeres

The ability of telomerase to extend telomere model DNAs was determined by a modification of a primer extension assay (157). Each 10 µl DNA extension assay contained 20 ng of DNA (see Figure 2.2B for details), 2 fmol of purified telomerase, 2 µCi [α-^{32}P]-dGTP (3000 Ci/mmol), 10 µM dGTP, 50 µM dTTP, 20 mM Tris-acetate (pH 7.5), 50 mM potassium acetate, 10 mM $MgCl_2$, and 1 mM DTT. Reaction mixtures were incubated at 25º C for 30 min and quenched by the addition of 20 mM Tris-acetate (pH 7.5), 10 mM EDTA, 1% SDS and 80 µg/ml proteinase K. Extension products were recovered by ethanol precipitation and analyzed by electrophoresis on a 10 cm x 10 cm 4-12% acrylamide gradient gel containing 7 M urea. Dried gels

were imaged by phosphorimaging (Molecular Storm 860) and quantified using ImagQuant version 5.2. The activity of telomerase in fractions eluted from the gel filtration column was similarly determined using a primer extension assay, where 50 µl of each fraction was reacted with 1 µM of the primer 5′-AATGAATGACTACGATTTT-3′ at 25º C for 20 h. The radiolabeled primer [α-^{32}P]-T_{10} was added to the quenching solution as a loading control and extension products were analyzed by electrophoresis on a 20 cm x 20 cm denaturing PAGE gel.

Determination of Primer K_m

K_m for telomerase binding to the telomere model DNA Ea_CM_22 or to a short, single stranded DNA primer, pEA22 (5′-$T_4G_4T_4G_4T_4G_2$) were determined by a DE81 filter-binding assay. Affinity purified telomerase (0.5 nM) was incubated with varying concentrations of primer (either pEA22, or Ea_CM_22), 50 µM dTTP, 10 µM dGTP, and 0.33 µM [α-^{32}P]-dGTP (3000 Ci/mmol) in telomerase reaction buffer for 30 min at 25º C. Reactions were quenched by the addition of 50 mM EDTA and spotted onto DE81 filter paper. Unincorporated [α-^{32}P]-dGTP was washed away with 0.5 M sodium phosphate, pH 7.0. Filters were counted by liquid scintillation counting and the data was corrected by the subtraction of background binding to the filter using a telomerase negative reaction. Data were fit to the Michaelis-Menten equation using Sigma Plot.

Electron Microscopy

Model telomeric DNA was diluted to 1 µg/ml in 10 mM HEPES pH 7.5, 50 mM potassium acetate, 8% PEG, 2 mM magnesium acetate followed by addition of *E. aediculatus* telomerase to 0.7 µg/ml for 10 min at 37º C. Proteins were fixed onto the DNA with 0.6% glutaraldehyde and the mixtures were filtered through 2 ml columns of Bio-Gel A-5m (Bio-Rad Laboratories, Inc., Hercules, CA) that had been equilibrated with 0.01 M Tris-HCl (pH 7.6), 0.1 mM EDTA. Addition of spermidine to 2.5 mM and $MgCl_2$ to 1 mM allowed for sample adsorption onto copper grids supporting thin, glow-charged carbon foils. The samples were washed stepwise with 25%, 50%, 75% and 100% ethanol for 5 min followed by air-drying and rotary shadowcasting with tungsten (158). An FEI Tecnai 12 electron microscope was used at an accelerating voltage of 40 kV to photograph images on plate film or a Gatan 4K x 4K CCD camera. Micrographs for publication were captured from plate film negatives using a Nikon SMZ1000 digital camera and morphometry measurements were done using ImageJ version 1.29 (National Institutes of Health).

RESULTS

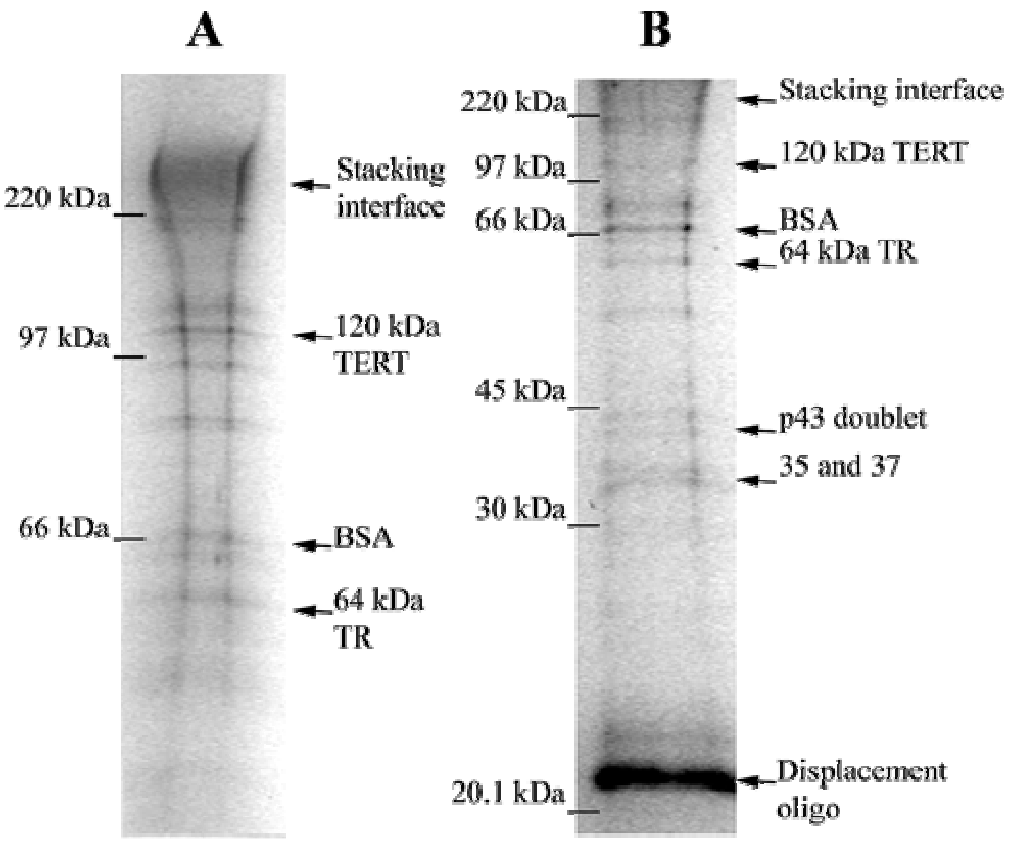
Gel Filtration Suggests Telomerase Dimers

E. aediculatus telomerase was purified by the method of Lingner and Cech (149). The total fold purification of the affinity-purified fractions could not be determined due to very low yields as well as the large amount of BSA used in the purification procedure. However, we estimated the telomerase to be approximately 50% pure, as reported previously for the purification of a much larger culture with much higher yields (149).

The quality of the purified telomerase was analyzed by silver staining of SDS-PAGE gels (Figure 2.1). Bands corresponding to all known *E. aediculatus* subunits were seen (120 kDa TERT, 64 kDa TR, and p43 doublet) as well as bands for known polypeptide contaminants (BSA and polypeptides 35 and 37) (149). The interface between the stacking and separating gels (stacking interface) is shown as a possible site of protein precipitation. Additional bands were seen that were likely polypeptide contaminants, but because samples were very dilute, it was not possible to determine the stoichiometry of these contaminating proteins with respect to telomerase. Nevertheless, these proteins did not appear to interfere with telomerase activity, which was robust in all assays, and because there was no protein observed in the binding of EM to nontelomeric DNA templates, we do not believe that these contaminants interfered with the telomerase EM binding

Figure 2.1: Affinity-purified telomerase fractions were separated on 10% polyacrylamide gels and stained with silver.

The sample in lane A was run twice as long as the sample in lane B. Sizes of molecular mass markers that were run with the samples are given in kilodaltons at the left. Locations of known polypeptides, DNA, and RNA are indicated (arrows). The interface between the stacking and separating gels (stacking interface) is also indicated as a possible site of protein precipitation.

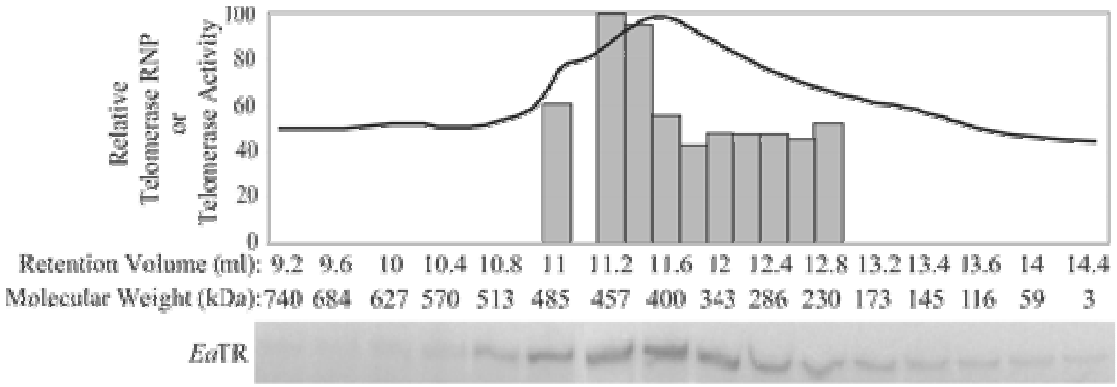

A
220 kDa
97 kDa
66 kDa
Stacking interface
120 kDa TERT
BSA
64 kDa TR
B
220 kDa
97 kDa
66 kDa
45 kDa
30 kDa
20.1 kDa
Stacking interface
120 kDa TERT
BSA
64 kDa TR
p43 doublet
35 and 37
Displacement oligo

experiments (results discussed below). Indeed, when Hammond et al. cross-linked partially purified *E. aediculatus* to ^{32}P-radiolabeled single-stranded telomeric DNA, only one type of protein-primer product was observed (159).

The oligomeric state of the telomerase was determined by gel filtration on a Superdex 200 column. The fractions were assayed for *E. aediculatus* telomerase RNA (*Ea*TR) content and for telomerase activity (Figure 2.2). The retention volume of telomerase was compared to that of four marker proteins with known molecular weights. The *Ea*TR content peaked in retention volume 11.6 ml, corresponding to an approximate molecular weight of 400 kDa, whereas the telomerase activity peaked in retention volume 11.2 ml, corresponding to an approximate molecular weight of 457 kDa. If the shape of the telomerase particle does not greatly deviate from spherical, then this result would be consistent with the predicted mass of a telomerase dimer of ~460 kDa (the molecular weight of telomerase is ~230 kDa), a result that is in agreement with gel filtration chromatography of *E. aediculatus* nuclear extracts done previously by Aigner, et al. (150). The slight offset between the peak RNA content and the peak telomerase activity could then be due to some dissociation of telomerase subunits during the gel filtration experiments, which could also account for the lower activity of the fractions with smaller apparent molecular weight. These data suggest, therefore, that *E. aediculatus* telomerase may be an active dimer in solution.

Figure 2.2: Telomerase profile which suggests a dimer in solution.

Elution profile of affinity-purified *Euplotes* telomerase chromatographed on a Superdex 200 gel filtration column. The amount of telomerase (line graph) in each fraction collected was determined by detection of the RNA subunit *Ea*TR. Fractions that contained telomerase were then assayed for telomerase activity (bar graph), as described in Experimental Procedures. Only fractions with activity above background are shown. Amounts and activities are shown relative to the highest value, which was set to 100%.


Relative Telomerase RNP or Telomerase Activity
100
80
60
40
20
0
Retention Volume (ml): 9.2 9.6 10 10.4 10.8 11 11.2 11.6 12 12.4 12.8 13.2 13.4 13.6 14 14.4
Molecular Weight (kDa): 740 684 627 570 513 485 457 400 343 286 230 173 145 116 59 3
EaTR

Synthesis of a Model Telomere

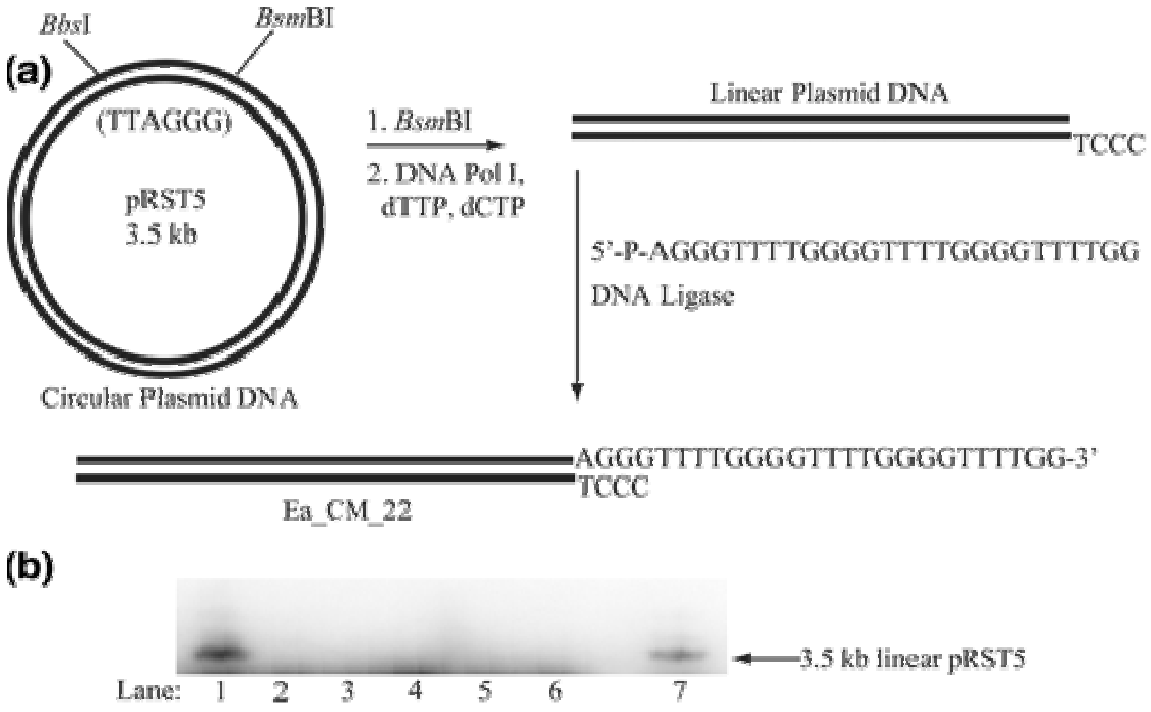
A synthetic model telomere was prepared by linearizing the plasmid pRST5 to generate a 3.5 kb DNA with one end terminating in 550 bp of (TTAGGG) repeats and a 5′ overhang. A single strand oligonucleotide was ligated to this end to generate a 22 nt 3′ overhang consisting of the *E. aediculatus* repeat $T_4G_4T_4G_4T_4G_2$ (Ea_CM_22, Figure 2.3A) (Methods and Materials). A control DNA with a non-telomeric overhang was similarly prepared using the random sequence single strand oligonucleotide ATTGAATGACTACGAAGATGAA. In these reactions, we routinely achieved 65-70% ligation efficiency.

Telomerase can Extend the Model Telomere

To determine if the model telomere Ea_CM_22 could act as a telomerase primer, we examined the ability of telomerase to extend the model DNA using [α-^{32}P]-dGTP in the reaction mixture. Reaction products were analyzed by electrophoresis on an acrylamide gradient gel for the incorporation by telomerase of [α-^{32}P]-dGTP into the substrate DNAs. We found that Ea_CM_22 was an efficient substrate while linear pRST5 DNA that did not contain a telomeric overhang was not a substrate (Figure 2.3B), and determined that the K_m for telomerase binding to Ea_CM_22 was 4 nM as compared to 14 nM for a 22-nucleotide primer used as a control, , which is comparable to the reactivity of primers with any of the possible permutations of the telomeric sequence (157,160).

Figure 2.3: Telomerase can extend a model chromosome.

(A) Synthesis of a *Euplotes* model telomere DNA (Ea_CM_22). Details of the enzymatic steps are in Methods and Materials. (B) Affinity-purified telomerase was incubated with dTTP and [α-^{32}P]-dGTP and either Ea_CM_22 (lane 1), Ea_CM_22 and RNase A (lane 2), linear pRST5 with a nontelomeric end (lane 3), linear pRST5 with a nontelomeric end and RNAse (lane 4), linear pRST5 (lane 5) or linear pRST5 with RNase (lane 6). Lane 7 is a marker for linear pRST5.


(a)
BbsI
BsmBI
(TTAGGG)
pRST5
3.5 kb
Circular Plasmid DNA
1. BsmBI
2. DNA Pol I,
dTTP, dCTP
Linear Plasmid DNA
TCCC
5'-P-AGGGTTTTGGGGTTTTGGGGTTTTGG
DNA Ligase
AGGGTTTTGGGGTTTTGGGGTTTTGG-3'
TCCC
Ea_CM_22
(b)
3.5 kb linear pRST5
Lane: 1 2 3 4 5 6 7

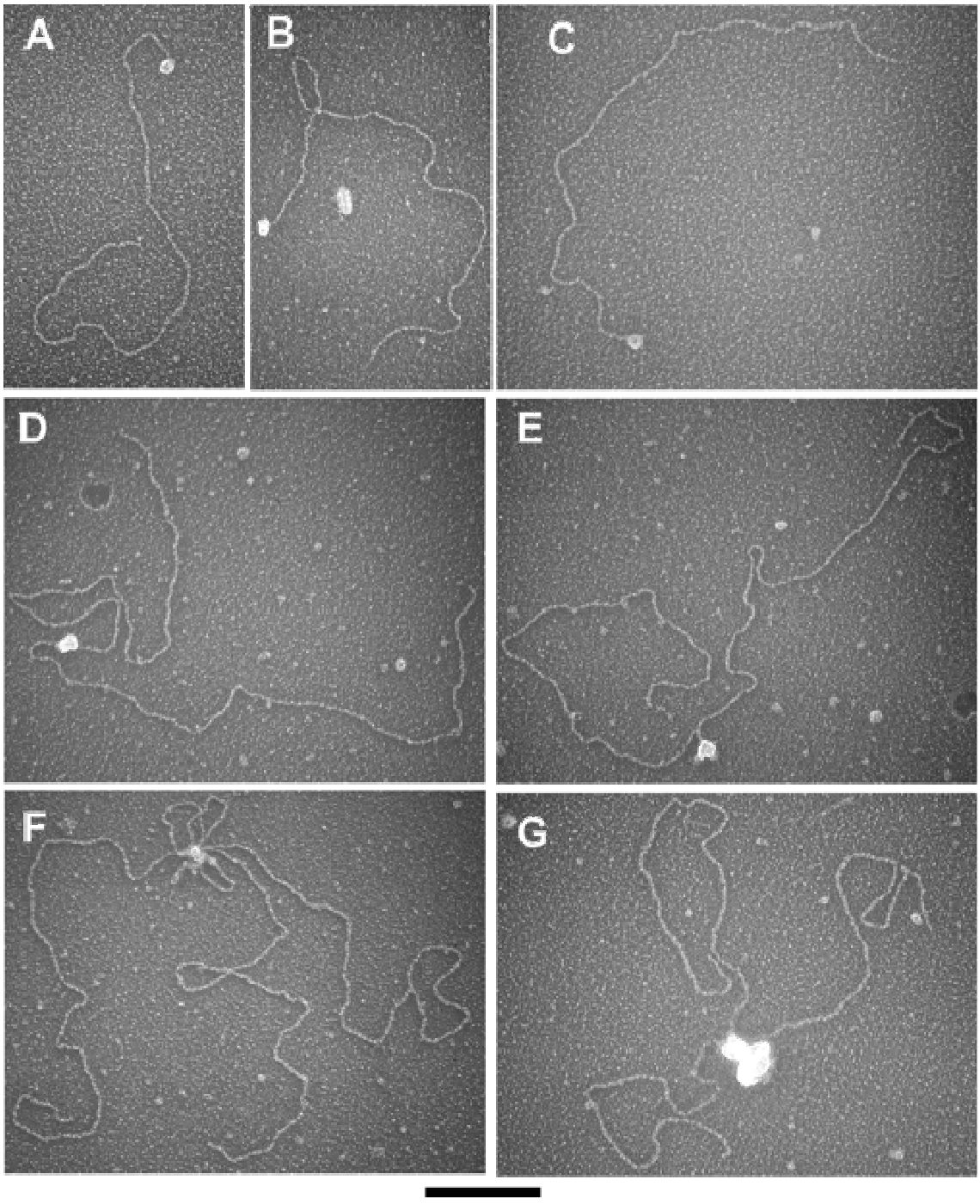
Visualization of Telomerase Bound to Telomere Models

Purified *E. aediculatus* telomerase was incubated with the 3.5 kb model telomere DNA Ea_CM_22 or the linear control DNA with the random sequence overhang. The reaction conditions used were consistent with those used to show that the purified telomerase was able to extend the model telomere, but were varied slightly to achieve the best conditions for visualizing protein-DNA complexes by EM. Alternative permutations of the telomerase repeat sequence were not studied because it has been previously shown that, while telomerase processivity is affected, telomerase is able to bind to and utilize primers containing very few or no telomeric nucleotides at the 3' end (161). Thus, binding reactions were performed for different reaction times (1 to 20 min) at varying temperatures (4º C, 25º C, 37º C) and protein:DNA molar concentrations (5:1 to 14:1) and in the presence or absence of PEG, a molecular crowding agent. Optimum conditions included using a 7.5:1 molar concentration of telomerase to DNA in the presence of 8% PEG for 10 min at 37º C. The resulting complexes were fixed with glutaraldehyde and examined by EM.

When *E. aediculatus* telomerase was incubated with the model telomere and prepared for EM, an array of telomerase-DNA complexes were observed (Figure 2.4). The most common species consisted of a single model telomere DNA bound at one end by telomerase (Figure 2.4A-C). Also present were complexes containing two or more model telomeres, each

Figure 2.4: Visualization of *Euplotes* telomerase bound to a model telomere substrate.

E. aediculatus telomerase was incubated with a 3 kb model telomere DNA (Figure 2.3), then prepared for EM by fixation with glutaraldehyde, mounting on carbon coated EM grids, and rotary shadowcasting with tungsten (Methods and Materials). Telomerase bound at one end of the model telomere (A-C*).* Examples of two (D, E*)* or three (F,G) model telomere DNAs bound together at one end by a telomerase complex. Images shown in reverse contrast. Bar is equivalent to 500 bp.


A
B
C
D
E
F
G

with one end synapsed to the other DNA by a large commonly-bound telomerase complex (Figure 2.4D-G). Less frequently seen were DNA molecules containing one or more internally-bound protein complexes (not shown).

Five separate reactions were scored with an average of 71 molecules counted per experiment (Table 1). In total, 46% of all telomere model DNAs showed a telomerase complex bound at one end. When the non-telomeric overhang was employed, only 4% of the ends showed protein bound. Thus the telomeric 3′ end of the model telomere is bound specifically by one or more telomerase molecules. Quantification of the protein-bound DNA molecules revealed an average of 45% with a telomerase complex bound to a single model telomere end, whereas 22% and 21% had two and three DNAs respectively associated at their ends through a large telomerase complex, likely a telomerase oligomer. Telomerase was observed bound non-specifically to internal DNA sequences within the model telomere in 12% of the DNAs. Occasionally, very large DNA-protein aggregates were seen, and were not included in the scoring. Using a lower concentration of telomerase in the binding reactions alleviated this problem somewhat, although conditions for binding required that the telomerase concentration could not be too low.

From these observations we conclude that *E. aediculatus* telomerase binds to the 3′ consensus overhang of the model telomere and that it is also

Table 2.1: Percent of model telomere DNAs bound at their end by a telomerase complex.

Reaction	Molecules counted	End-bound complexes
1	104	52%
2	109	47%
3	45	40%
4	43	42%
5	54	50%
Average	71	46%

capable of self-association, consequently bringing two or more DNA ends together *in vitro*. This, however, should not necessarily be taken as suggesting that telomerase is able to synapse two chromosome ends *in vivo*.

Telomerase Oligomerization

The degree of oligomerization of the bound telomerase complexes was determined using a variation of a method used in our previous studies (162,163). A large protein standard of known mass is mixed with the sample and the size (projected area in the micrographs) of the standard is compared to the projected area of the protein bound to the DNA. If the standard and sample are of similar size and shape, molecular weight estimates can be derived that can differentiate with certainty between different oligomeric states of the protein bound to DNA. In this study since there was a significant amount of free telomerase in the background on the grids, the molecular weight standard was adsorbed to separate grids and the standard and sample processed for EM side-by-side. This can be expected to add some but not a significant additional measurement error.

The telomerase complexes on DNA appeared mostly globular, and inspection showed that apoferritin (443 kDa) was close in size (projected area) to the class of telomerase particles most frequently seen at the ends of a single model telomere DNA, suggesting that these particles may consist of telomerase dimers. Photographs of fields of apoferritin molecules were taken and the mean projected area of 30 examples was measured (Figure

2.5E). The average projected area of the apoferritin was set to an arbitrary value of 100 units. On the basis of the mass of a telomerase monomer, the predicted projected area relative to apoferritin should be 64 units as determined by the formula: mass of telomerase/mass of apoferritin = (area of telomerase/area of apoferritin)$^{3/2}$.

Similarly, the predicted projected areas of a dimer, trimer, tetramer and pentamer would be 102, 134, 162, and 188 units respectively. These sizes were then compared to measurements of projected areas of telomerase complexes bound to the end of a single model telomere DNA (Figure 2.5A-D). A histogram representing the calculated mass derived from such measurements is shown in Figure 2.6. The mean projected area of the complex was thus 101(± 23) units and the mean calculated mass value was 455(± 160) kDa (n = 34). Assuming the telomerase complex binds only to the consensus 3′ overhang, we subtracted ~8.6 kDa for the single strand sequence $[T_4G_4]_2$TTTTGG which resulted in a mass of 447(±151) kDa. Although the distributions may have been broadened by the slightly oblong shape of the telomerase and thus by variations in projections of the telomerase particles, these data suggest that the telomere end was most commonly bound by a dimer of telomerase. This finding is consistent with the size of telomerase determined by gel filtration here and previously reported (150).

Figure 2.5: Estimation of the oligomeric state of DNA-bound telomerase by direct size comparison.

Telomerase complexes at the ends of single model telomeres (A-D) or joining two model telomeres (F-H) were compared to the size of apoferritin particles (E). The telomerase DNA complexes and apoferritin were prepared for EM side-by-side on separate EM supports. Shown in reverse contrast, bar is equivalent to 500 bp.

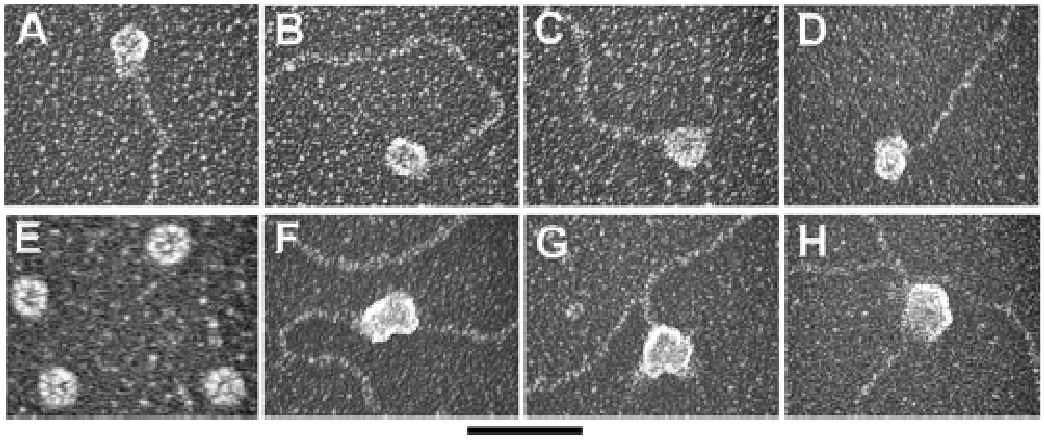
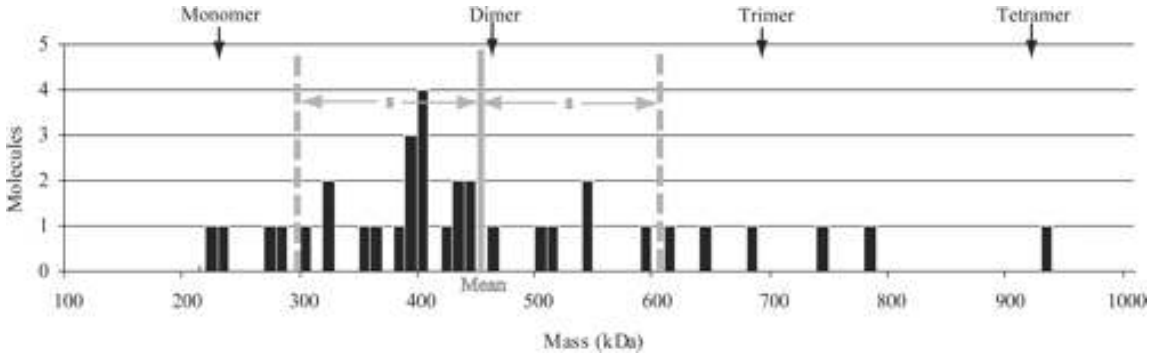

A
B
C
D
E
F
G
H

Figure 2.6: Histogram of the calculated mass of telomerase at the end of a single model telomere.

The calculated mass values derived from measurements of the projected areas of telomerase complexes bound to the end of a single model telomere DNA are shown. The mean calculated mass value (solid grey line) was 447 kDa with a standard deviation of 151 kDa (s, dotted grey lines). Molecular weights of a monomer (230 kDa), dimer (460 kDa), trimer (690 kDa) and tetramer (920 kDa) of telomerase are indicated by arrows.


Monomer
Dimer
Trimer
Tetramer
Molecules
0
1
2
3
4
5
100
200
300
400
500
600
700
800
900
1000
Mean
Mass (kDa)

For the class of telomerase particles bound at the junction of two model telomeres, the protein mass was frequently much larger than apoferritin. The projected areas of a subset (n = 19) of the smallest particles was measured, however, revealing a broad distribution of calculated mass values (Figure 2.5F-H). When compared to apoferritin (set to 100 units) the mean projected area of these particles was 176(± 64) units. After subtracting ~17 kDa for the DNA content this resulted in a mean calculated mass of 1061(± 569) kDa. Although this value is sufficiently broad to encompass a dimer (460 kDa) or trimer (690 kDa) of telomerase, the data seem more likely to suggest that, at a minimum, two dimers of telomerase, each associated with a model telomere, bind to form a tetramer (920 kDa) *in vitro*. Indeed, for this class of particles one could occasionally distinguish a region of separation between two globular particles, each the size of a dimer, that is consistent with a dimer-of-dimers (Figure 2.4D-G).

DISCUSSION

In this study, *E. aediculatus* telomerase was isolated from nuclear extracts using affinity chromatography with an antisense oligonucleotide. Although electrophoretic analysis of fractions revealed several contaminants, these did not interfere with telomerase activity, which was robust, or with our EM binding experiments, as seen by the lack of protein binding to control

DNA templates.

The molecular mass of telomerase, determined by gel filtration, was consistent with the mass of a dimer. When the telomerase particle was visualized by EM, its shape was seen to be mostly globular, and inspection of the telomerase bound to its substrate showed that it bound predominantly as a dimer. Taken together, we find these observations to be convincing evidence that *E. aediculatus* telomerase dimerizes in solution and that it binds to telomeric DNA as a multimer, most likely as a dimer.

It was also shown by EM and a modified primer extension assay that telomerase specifically binds to and extends the consensus telomeric 3'-overhang *in vitro*. Further, when end-to-end pairing of two DNA molecules occurred, the telomerase complex was likely a tetramer or larger oligomer.

These findings are consistent with a model of processive telomere reverse transcription consisting of two cooperating telomerases that bind to and extend a single DNA substrate. The data do not support an obligate parallel extension model of telomerase cooperation, where two active sites within two different but associated telomerase RNPs simultaneously extend two separate chromosome 3'-ends (145,148), although it is still possible that two dimers of telomerase can coordinate to extend two separate but adjacent chromosome ends, such as sister chromatids.

Rivera et al. showed that a dimeric human telomerase can

processively utilize a single template (164), which argues against the template switching model in humans, where the two catalytic sites would act sequentially to elongate a single telomere 3'-end (145). Nevertheless, *Euplotes* is an evolutionarily very distant species, and it is still possible that the template switching model or the DNA anchor site model, where one telomerase RNP template anchors the dimer while the other template is used for reverse transcription (152), or combinations of these models may exist.

This study is the first visualization of intact telomerase bound to a DNA substrate, and it provides the first direct proof of telomerase RNP multimerization in *E. aediculatus*.

CHAPTER 3: ELECTRON MICROSCOPIC VISUALIZATION OF YEAST TELOMERASE RNA SUPPORTS *MFOLD* PREDICTIONS OF SECONDARY STRUCTURE

INTRODUCTION

Telomere sequences are lost during cell division, due to the 'end-replication problem' (22,23), oxidative damage (25), and possible nucleolytic processing (24). In cells with indefinite replicative potential, leading strand telomeric repeats can be restored by the ribonucleoprotein telomerase (29), while the lagging strand is concurrently elongated by the standard replicative machinery (30). In *Saccharomyces cerevisiae*, the catalytic protein subunit of telomerase (EST2) uses the template sequence in the telomerase RNA (TLC1) to synthesize telomeric DNA by reverse transcription (37,41).

Whereas the sequences of the catalytic subunits are conserved among ciliates (138), yeast and mammals (37), the sequences and sizes of the RNA components are highly variable (41,43,44,165,166). In budding yeasts, there is a striking divergence in the sequence and size of the telomerase RNAs. Consequently, the lack of information has led to complications deducing their secondary structure. Recently, a detailed working model for the secondary structure of the *S. cerevisiae* telomerase RNA, TLC1, was proposed (44,165)

using *mfold* predictions of energetically favorable RNA conformations (167) coupled with genetic and phylogenetic data from four species of *Saccharomyces*. This model proposed that the telomerase RNA consists of a central catalytic core containing the RNA template and Est2p-binding region, plus three long quasihelical RNA arms that bind Est1p (168), the Ku heterodimer (169), and the Sm heteroheptamer (170) (Figure 3.1A).

To test this model, we used EM to directly visualize a 1261 nt TLC1 RNA, as well as a 1133 nt deletion mutant (ΔΔ) proposed to have truncation of the longest RNA helix (44), Figure 3.1). Consistent with the *mfold* predictions, most of the RNA molecules that were visualized contained two or three 'arms'. In addition, a nano-scale 'biopointer' was used to 'point to' a biotin moiety attached to an oligonucleotide (oligo) that was complementary to the RNA template region (171). We found that in the majority of molecules, the biotinylated oligo was most often located at the center of the RNA.

In a subsequent study, deletion mutagenesis that reduced the 1,157-nucleotide (nt) TLC1 RNA to a size smaller than the 451-nt human RNA, but retained its function *in vivo*, demonstrated that the RNA serves as a flexible scaffold for the protein subunits, described as 'beads' (proteins) on a 'string' (RNA), and that the long protein-binding arms are not essential for the RNA to serve its scaffolding function (45). This differs markedly from the best-studied ribonucleoprotein enzyme, the ribosome, where the RNAs have specific three-dimensional structures that orient the functional elements (172).

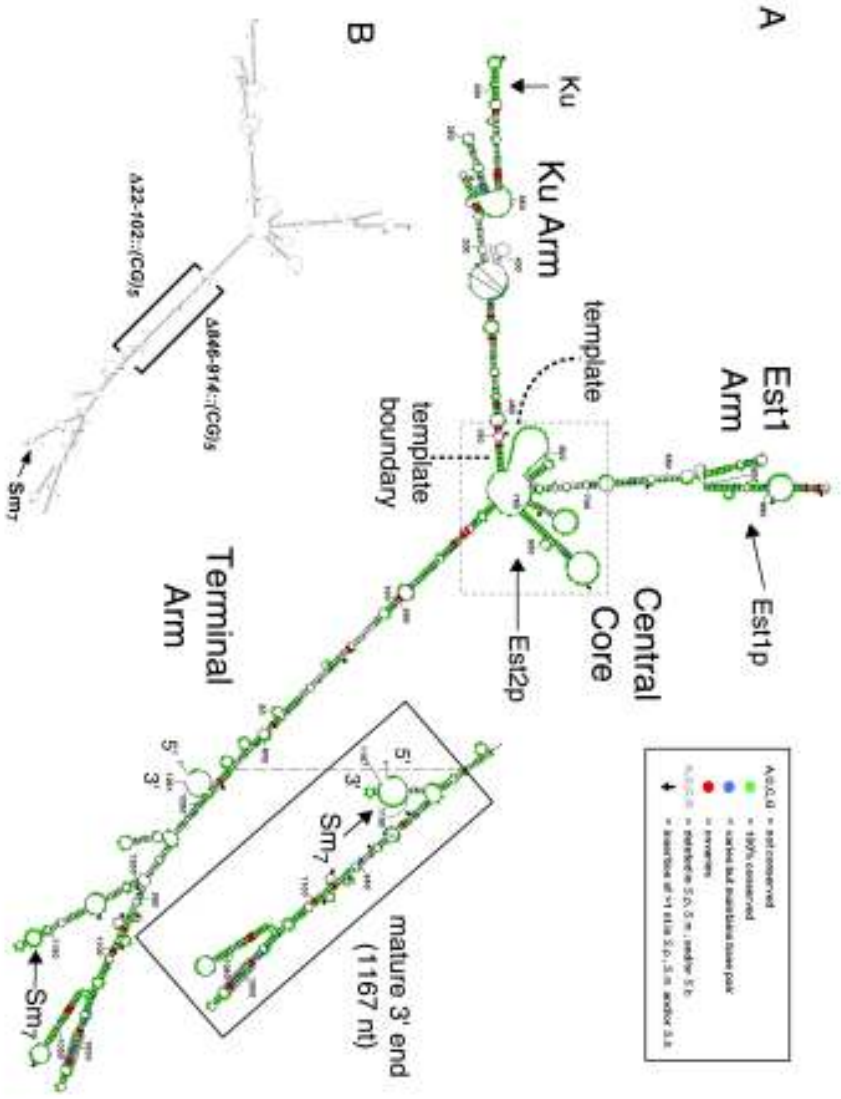
Figure 3.1 A model of S. cerevisiae telomerase RNA secondary structure[2].

(A). Two forms of TLC1 with different 3' ends (~1261 nt and 1167 nt (boxed)) are hypothesized to be precursor and mature forms, respectively (170,173,174). Phylogenetic data for each nucleotide (legend, boxed) were derived from alignment of TLC1 from *Saccharomyces cerevisiae, paradoxus, mikatae* and *bayanus*. Proteins that bind to TLC1 RNA are indicated and arrows indicate positions within the RNA to which they bind. A dashed box indicates the central core region containing the template, template boundary, and proposed Est2p binding regions, necessary for telomerase RNA enzymatic function. The template used for reverse transcription and the Sm7 complex binding site are known to be single-stranded RNA regions. (B). Schematic showing the nucleotide replacements in the TLC1 truncation mutant (ΔΔ). Nucleotides 22–102 were replaced with the sequence $(CG)_5$, predicted to pair with the $(CG)_5$ that replaced nucleotides 846–914 on the other side of the predicted terminal arm.

---

[2] Adapted from Zappulla D. C, et al. (2004) *Proc. Natl. Acad. Sci. U. S. A.***101** (27) 10024-29.

3



[3] Adapted from Zappulla D. C, et al. (2004) *Proc. Natl. Acad. Sci. U. S. A.***101** (27) 10024-29.

MATERIALS AND METHODS

In vitro TLC1 RNA Expression and Purification.

The TLC1 RNA was transcribed *in vitro* from plasmids pT7-TLC1 (1261 nt; a gift from Art Zaug) and pDZ118 (Δ22-102::$(CG)_5$ Δ846-914::$(CG)_5$) using the following transcription conditions in a 1 ml reaction at 37 ºC for 1 hour: 20 µg/ml plasmid DNA template (linearized with *Fok*I), 40 mM Tris pH 7.5, 12 mM $MgCl_2$, 4 mM NTPs, 2 mM spermidine, 10 mM DTT, 0.4 µg/ml inorganic phosphatase and T7 RNA polymerase (~10 µg/ml). The full length RNA was then purified by continuous elution electrophoresis through a tube- shaped, denaturing 4% polyacrylamide gel (2.8 cm diameter x 6 cm long) using a Model 491 Prep Cell electroelution apparatus (Bio-Rad, Hercules, CA) run at 12 W of constant power. Elution buffer consisted of 3 M urea, 0.5X TBE buffer and the flow rate was 1 ml/min. TLC1 RNA fractions were then concentrated to approximately 300 µg/ml (0.5 µM) by centrifugal filtration in 10 mM HEPES pH 7.0 and stored at -80 ºC. For RNA refolding, TLC1 was heated to 94 ºC for 5 min in 10 mM HEPES pH 7.0, 1 mM EDTA and 150 mM NaCl, put into a 0ºC water bath and subsequently stored at -80ºC in aliquots for EM.

Synthesis of biopointers.

Nano-scale biopointers were prepared as previously described (171). In brief, a 179 bp region of pBluescript ® II S/K (+) Phagemid Vector

(Stratagene, Inc.) was amplified between the T7 and T3 promoters by PCR with one of the two primers biotinylated. The DNA was purified by gel exclusion chromatography and mixed with a 10-fold excess of streptavidin. The streptavidin-bound DNAs (biopointers) were banded in CsCl (density 1.6 g/ml) to equilibrium and the monomer biopointers collected.

For the binding of biopointers to TLC1 RNA, folded TLC1 was incubated with a biotinylated DNA oligonucleotide complementary to the template, unbound oligo was removed from the solution by passing the complex though a Sephadex G-25 spin column (Roche), and then biopointers were added to the mixture. This allowed visualization of the TLC1-RNA:biotinylated-oligo complex by EM.

Single-molecule Electron Microscopy.

TLC1 RNA was diluted to 1 μg/μl in 10 mM Tris-HCl pH 7.5, 125 mM NaCl, 0.1 mM EDTA. When required, biopointers were added to 0.5 μg/μl. Spermidine was added to sample mixtures to 2.5 mM and $MgCl_2$ to 1 mM followed by adsorption onto copper grids supporting thin, glow-charged carbon foils. The grids were washed stepwise with 50% ethanol in water for 10 min and then 75% and 100% ethanol for 5 min followed by air-drying and rotary shadowcasting with tungsten (158). Samples were visualized on a FEI Tecnai 12 electron microscope using an accelerating voltage of 40 kV. Plate and digital photographs were taken and molecules measured, counted and

categorized using Gatan Digital Micrograph software.

RESULTS

Synthesis and folding of TLC1 RNA.

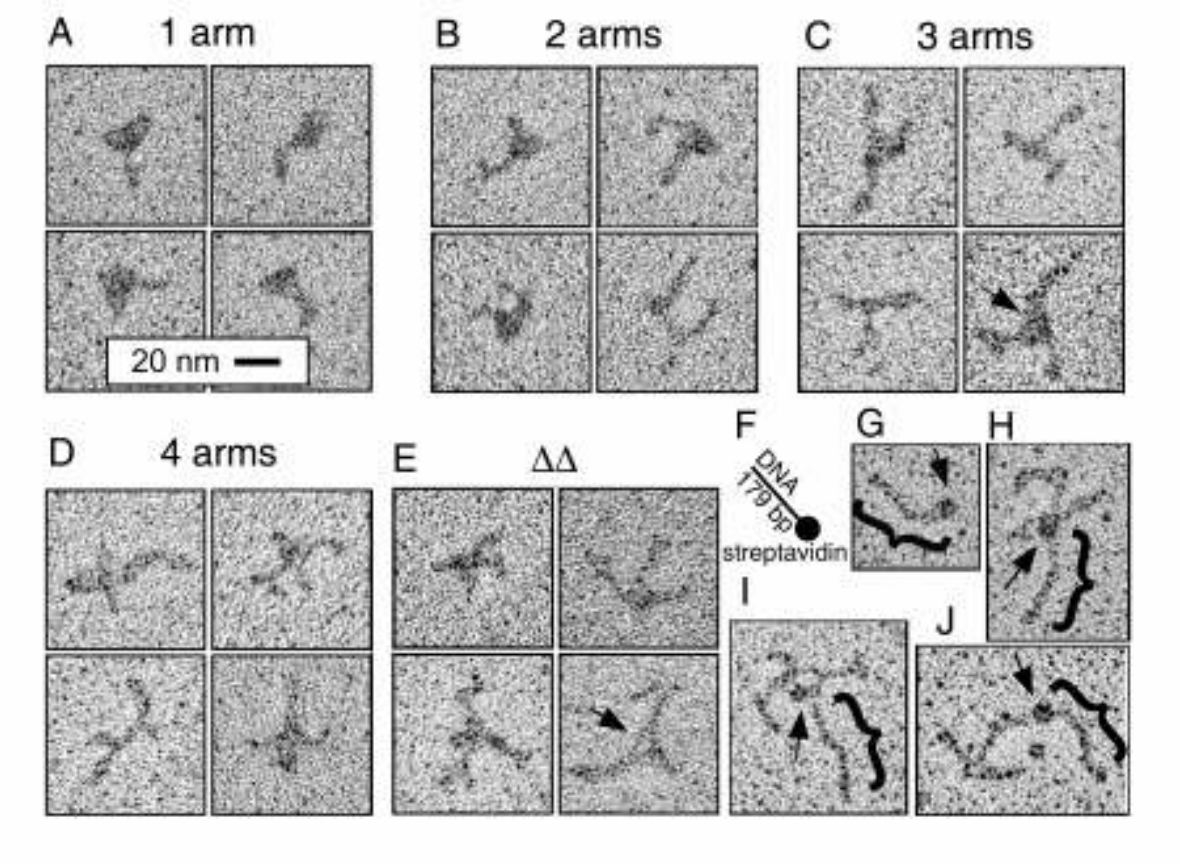
A 1261 nt TLC1 RNA transcription product was synthesized *in vitro* using T7 RNA polymerase. This RNA was purified by continuous elution electrophoresis on a denaturing polyacrylamide gel, and the RNA was then refolded by heat denaturation and quick or slow refolding in the presence of 150 mM NaCl. Native gel electrophoresis of refolded RNA demonstrated two predominant bands of similar mobility for the RNAs folded in all of these conditions (data not shown). The quickly refolded RNA was used for the experiments presented here.

Single-molecule electron microscopy of TLC1 RNA.

Previous work has shown that long helices of RNA can be readily observed by EM (175). Since the TLC1 RNA secondary structure model proposes a molecule with three long, largely helical arms, TLC1 was examined by single-molecule EM (Figure 3.2). An array of RNA conformations were seen, where the most common species consisted of two to four quasihelical arms protruding from an unstructured site on the RNA, giving it a 'spider' like appearance (Figure 3.2B-D). Also present were

Figure 3.2. Electron microscopy of TLC1 RNA.

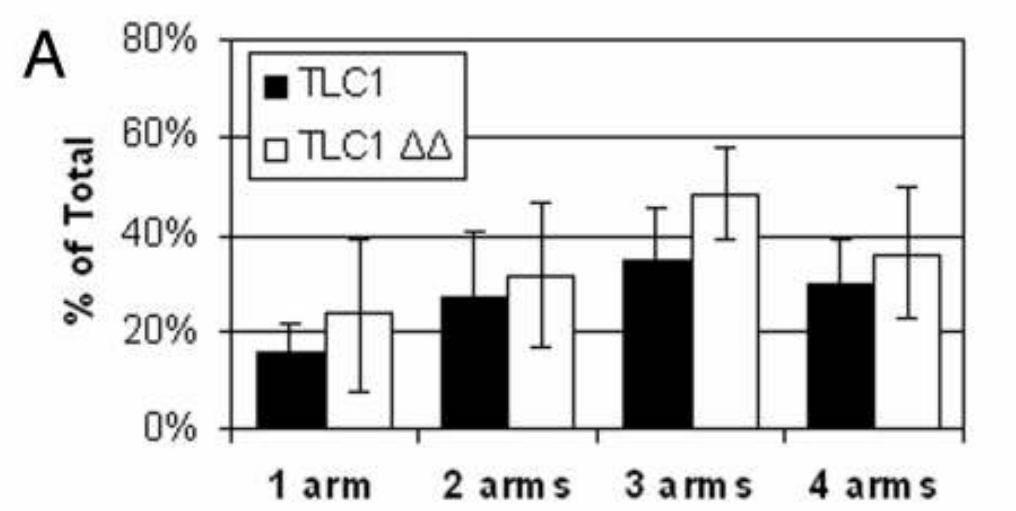
(A-D) Four representative examples each of TLC1 RNA molecules with 1,2,3, or 4 arms. Arrow in C indicates electron-dense RNA core. (E). Representative structures observed from analysis of TLC1 harboring the truncation in the terminal arm (ΔΔ, as described in Figure 3.1b). Arrow indicates lower electron density in the RNA core. (F) Diagram of a biopointer consisting of a 179 bp DNA bound to a molecule of streptavidin protein. (G-J). Representative examples of a biopointer alone (G), or bound to the center (H-I) or to the helical arm (J) of the TLC1 RNA. Arrows indicate the 53 kDa streptavidin protein.

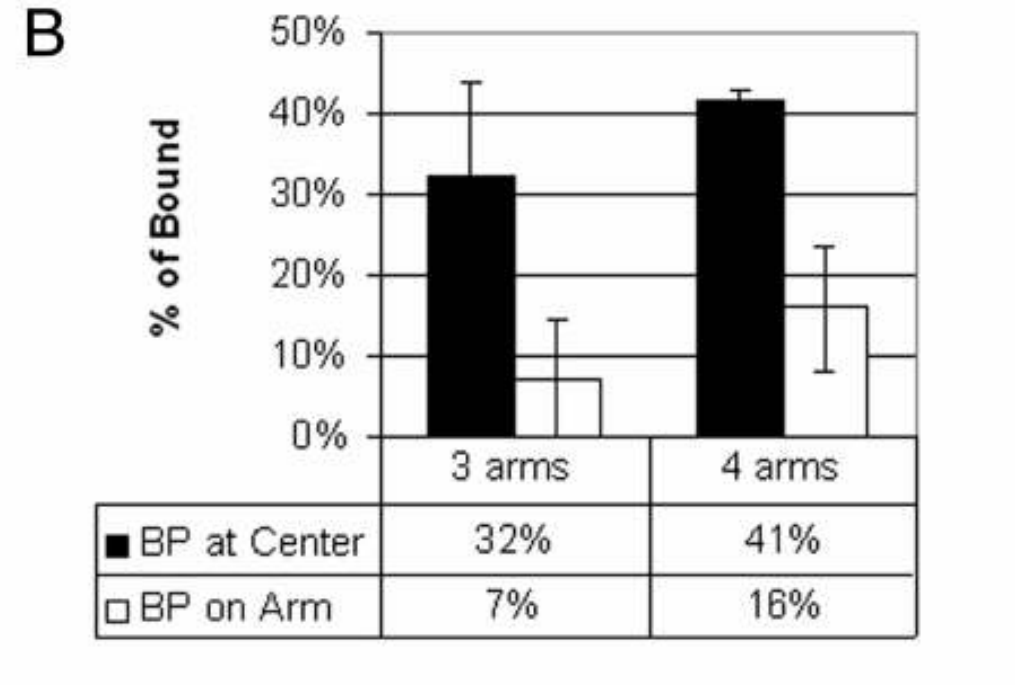

A
1 arm
B
2 arms
C
3 arms
20 nm
D
4 arms
E
ΔΔ
F
DNA
179 bp
streptavidin
G
H
I
J

molecules containing one arm (Figure 3.2A) or no arms (data not shown). Three separate preparations were scored, with more than 100 molecules counted per experiment (Figure 3.3A). The observed distribution may underestimate the true number of arms in the RNA, since molecules may land on the grid with arms overlapping, sticking up so that they do not appear as arms, or kinked back on themselves in such a way that they are no longer discernable. Molecules that had an ambiguous structure or appeared globular (and could not be confirmed to be RNA) were not counted. The RNA construct harboring truncation of the terminal arm (ΔΔ), which is functional *in vivo* (44), Figure 3.1B), was also visualized by EM (Figure 3.2E). The distribution of apparent arms was slightly shifted towards a greater number in this mutant molecule (Figure 3.3A). Thirty 3-armed ΔΔ TLC1 RNA molecules were measured from the center of the RNA mass to the tip of each arm. The longest arm had a mean length of 30.5 ± 4.3 nm with the other arms being 24.5 ± 3.7 and 20.2 ± 3.6 nm. Using 0.27 nm as the axial rise per nucleotide (176), this correlates with arms of 113, 91 and 75 bp. In remarkable agreement, the model of Figure 3.1B predicts the equivalent of 114 bp in the truncated terminal arm, 99 bp in the Ku arm, and 74 bp in the Est1p arm. (We count an unpaired base in an internal loop as contributing the equivalent of one bp in length. Because the structures and lengths of such loops cannot be predicted, this represents a source of uncertainty in the correlation of arm length measured by EM with our model). Surprisingly, measurement of thirty-five 3-armed molecules of full-length TLC1 gave arm length

Figure 3.3. Histograms summarizing the EM data.

(A). The number of apparent arms per molecule of TLC1 were counted and compared to the number of apparent arms per molecule of the deletion mutant (ΔΔ). (B). Histogram indicating the percentage of instances when biopointers (BP) were observed at the center of RNA molecules or arms.


A
80%
60%
40%
20%
0%
% of Total
TLC1
TLC1 ΔΔ
1 arm
2 arms
3 arms
4 arms


B
50%
40%
30%
20%
10%
0%
% of Bound
3 arms
4 arms
BP at Center
32%
41%
BP on Arm
7%
16%

measurements of 29.1 ± 4.1 nm, 23.4 ± 3.7 nm, and 18.3 ± 3.6 nm, indistinguishable from those of the ΔΔ molecules. This result is addressed in the Discussion.

'Biopointers' indicate the template is at the center of the RNA.

The nano-scale 'biopointer' used here consists of a short, stiff 179 bp DNA rod with a single streptavidin tetramer bound to one end (171), Figure 3.2F). Biopointers are easily seen in the EM preparative method used here (Figure 3.2G), yet are small enough to not obscure the molecule or complex being tagged. The streptavidin tetramer can then bind and 'point' to biotin molecules present on the macromolecules.

In this study, we have used biopointers to identify a biotinylated DNA oligonucleotide bound to the template region of TLC1 RNA, with the aim of elucidating the relative position of the template. When the biotinylated template oligo was annealed to TLC1, biopointers were seen bound to the RNA 30% of the time. In control experiments, where TLC1 was incubated with a biopointer in the absence of the biotinylated template oligo, biopointers were seen bound to the RNA only 5% of the time.

The vast majority of biopointers were observed bound to the centers of the RNAs (Figure 3.2H-I). Less often, biopointers were observed bound to the ends of the RNA arms (Figure 3.2J). Specifically, 32 ± 11% of 3-armed RNAs and 41 ± 1% of 4-armed RNAs had a biopointer pointing to the center

of the molecule, whereas 7 ± 7% of 3-armed RNAs and 16 ± 8% of 4-armed RNAs had a biopointer pointing to the end of an arm (Figure 3.3B). Interestingly, there seemed to be a moderate increase in binding of the biopointers to the 4-armed RNAs than to the 3-armed RNAs, both at the center of the molecules and at the ends of the RNA arms. A possible explanation for this may be that the 4-armed RNAs were more loosely structured than the 3-armed structures, such that the template region was able to migrate; or alternatively that it is more accessible for binding *in vitro*. In summary, the biopointers show the template to be generally within the central core of the multi-armed RNA molecules visualized by EM.

DISCUSSION

Electron microscopy supports the existence of several helical arms protruding from a central core. The mean distribution of the number of arms seen supports the prediction of three long quasi-helical RNA arms, although the results are not conclusive. Since RNA molecules with long quasi-helical arms can be trapped on the grid with arms pointing out in three dimensions or with arms overlapping, the number of arms observed were likely less than what would have been seen if the TLC1 RNA was properly stretched out over the two-dimensional surface. Thus, the abundance of two-armed molecules for TLC1 may be overestimated in this data. Alternatively, it could reflect the incomplete re-folding of the RNA molecules from the denatured state.

The latter may also be the case for the 4-armed molecules observed.

The observation that truncation of the terminal arm had no net effect on the length of any of the TLC1 arms is curious. It is possible that the base of the longest arm collapsed into the central body of the RNA, giving rise to an electron-dense unstructured region (arrow, Figure 3.2C as contrasted to arrow, Figure 3.2E); or that the longest arm was able to fold back so that its terminus interacted with the central core, resulting in an arm half as long, but twice as thick (which we were not able to measure). Such compaction may have caused the apparent underestimation of the terminal arm length. In the terminal arm truncation mutant, nucleotides that are proposed to form an interrupted helix were replaced by 10 Watson-Crick C-G base pairs. The shortening and/or insertion of this sequence may have caused the dynamics of the terminal arm to be affected such that it became more readily extended. This would explain the excellent agreement between expectation and measurement for the longest arm of the ΔΔ mutant RNA. Such a trend for the more frequent extension of arms of this mutant is also supported by the fact that more arms were observed in the terminal arm truncation mutant (Figure 3.3A) than in the wild-type RNA.

In summary, these results support the overall model of TLC1 structural flexibility, such that the yeast telomerase RNP, in contrast to the ribosome, is a rather loosely ordered complex of RNA and protein subunits (44,45)..

CHAPTER 4: REPLICATION FORK REGRESSION IN REPETITIVE DNA's

INTRODUCTION

Repetitive DNA sequences found in the human genome consist of repeat units ranging from mono-, di-, and tri-nucleotide repeats to long repeating units found in Alu and LINE elements. Overall, repetitive DNA makes up approximately 30% of the human genome with the Alu and LINE elements constituting the greatest amount (177). The short sequence units which include the triplet and telomeric repeats are of particular interest due to the high number of repeats per unit length of DNA which may bestow unique biological and physical properties.

The triplet repeats which include $(CGG)_n$•$(CCG)_n$, $(CAG)_n$•$(CTG)_n$ and $(GAA)_n$•$(TTC)_n$ have been implicated in numerous human hereditary diseases, a hallmark of which is the appearance of disease pathology when the repeat blocks expand beyond certain tight length thresholds generally exceeding 35 repeats (178). In addition, tetrameric $(CCTG)_n$•$(CAGG)_n$ (179), pentameric $(AATCT)_n$•$(AGATT)_n$ (180) and dodecameric $(C_4GC_4GCG)_n$•$(CGCG_4CG_4)_n$ (181) repeats have been linked to the genetic diseases myotonic dystrophy type 2 (DM2), spinocerebellar ataxia type 10 (SCA10) and progressive myoclonus epilepsy, respectively. The length of disease-related repeats can

vary from as little as a few repeats in normal individuals to up to 40 kb in the SCA10 expansions (182). Whereas the exact mechanism of repeat expansion in humans remains unknown, one feature common to all expanded repeats is that they are highly unstable above a threshold of approximately 100–200 bp (183).

Telomeric repeats which are composed of the hexameric unit TTAGGG in all mammals and many animals (TTTAGGG in plants) are essential for chromosome stability and regulating the replicative lifespan of somatic cells (184). These repeats comprise the DNA component of the telomere (3,4), a nucleoprotein structure which protects the ends of chromosomes and enables cells to distinguish telomeric ends from random double-strand (ds) break ends (16,185).

Telomeric repeats can reach lengths of 15 kb in humans and as much as 150 kb in plants. In the absence of telomerase, a telomere reverse transcriptase, telomeric repeat sequences are gradually lost during cell division, due in part to the ‘end replication problem’ that results from the inability of the lagging strand to be replicated to the very end of the chromosome (22,23). Large blocks of telomere repeat sequences can also be lost stochastically when the proteins required for end protection functions are disrupted, or problems are encountered during DNA replication or repair (184). In the absence of telomerase, certain human cancer cells have been shown to exhibit highly unstable telomeres (ALT phenotype) with

rapid increases or decreases in telomere lengths (54,186,187).

Evidence suggests that the nature of repetitive DNA may itself be a causative factor in mutagenesis (188-191). The relative instability of long blocks of short repeats may also be related to inherent difficulties of the DNA synthesis machinery in replicating through this type of DNA. A large body of evidence shows that there is frequent polymerase pausing in triplet blocks, that both the lagging and leading strands may form hairpins, G-quartets or triplex structures when comprised of certain repeats, that the polymerase can slip during synthesis through repeat tracts, and that primer template misalignment can occur as a result of hairpins in the template strands (reviewed in (183,192-194). Also, telomeric sequences undergoing replication have the potential of forming G-quartets. These impediments to the replication fork could then give rise to repeat expansions or deletions due to reiterative DNA synthesis or replication restart via recombination intermediates. Indeed, recent evidence in *E. coli* suggests that a major mechanism for $(CAG)_n \cdot (CTG)_n$ repeat instability is replication restart, via a Holliday junction 'chickenfoot' intermediate, after DNA polymerase pausing and the resultant collapse of the replication fork (131,195).

Relatively little is known about the replication of mammalian telomeric DNA. Over expression of the telomeric binding proteins TRF1 and TRF2 has been shown to lead to replication fork stalling *in vivo* (196), and *in vitro* human telomeric DNA is replicated much less efficiently than non-telomeric DNA

(196) and unpublished data). Further in addition to the normal replicative helicases present at forks, the RecQ helicases WRN and BLM, implicated in premature ageing diseases, have been shown to be important for proper telomere replication and maintenance in human cells (94,102,103). These helicases have been shown to unwind G quartets and four-stranded junctions similar to chickenfoot structures (107-109,111,113,197) suggesting that these or other secondary structures unique to telomeric repeats may form and present barriers to replication.

These observations suggest that the polymerase machinery is much more prone to pause or stall during replication of long blocks of short repeats. In the absence of stabilizing proteins, this could lead to fork regression and the generation of four-stranded chickenfoot molecules (pictured in Figure 4.2B). Resolution of these chickenfoot intermediates could lead to restoration of replication but inappropriate resolution could lead to expansion or contraction of the DNA tracts or possibly the generation of extrachromosomal repeat DNA.

To begin testing this hypothesis, we generated model replication fork templates which mimic a replication fork that has transited a long block of either telomeric repeats or triplet (CTG) repeats. Examination of the structure of this stalled fork by EM provided a means of detecting forks which may have either fully regressed or partially regressed leading to a chicken foot structure. Previously we used this model system with sequences which were not

repetitive and we showed that chickenfoot structures could be observed when p53 was present to trap these forms during fork regression (198). In this paper, we show that nearly half of all repeat-containing templates had spontaneously regressed, with a large percentage of molecules forming chickenfoot intermediates. These four-way junction molecules were not observed in the absence of p53 in our previous study of non-repetitive DNA (198). We discuss a model in which repetitive DNA is highly slippery with a high tendency to generate long-lived chickenfoot structures and the implicit implications related to triplet and telomere DNA stability.

MATERIALS AND METHODS

Construction of plasmids

Complementary oligonucleotides (oligos) 1) 5′-TCGAAGACTTAGGGGCTGAGGTT-3′ and 2) 5′-CCCTAACCTCAGCCCCTAAGTCT-3′ containing a site for the nicking endonuclease N.BbvC IA (New England Biolabs, Beverley, MA) were annealed and cloned into the XhoI and BbsI sites of the plasmid pRST5 (69), to create pRST5NICK. Dr. Y. H. Wang (UMDNJ) generously provided the plasmid pGEM$(CTG)_{130}$, which was constructed by digesting the pSH2 plasmid (199) with restriction endonucleases SacI and HindIII (New England Biolabs) to isolate the CTG fragment, and then cloning this fragment into the

same restriction sites in the plasmid pGEM3zf(+) (Promega Corporation, Madison, WI) (unpublished data). Site-directed mutagenesis was performed on the plasmid pGEM$(CTG)_{130}$, resulting in the creation of two sites for the restriction endonuclease BsmI (New England Biolabs) directly adjacent to, and on either side of, the CTG repeat tract. During this process, some CTG repeats were lost, resulting in a tract length of [CTG]110. The complementary oligos 1) 5′-TCAGCCAGGCCGAAAGAAAGAAAAGGACAGAGAAAGCC-3′ and 2) 5′-CTTTCTCTGTCCTTTTCTTTCTTTCGGCCTGGC-3′ were annealed and cloned into the BsmI and BbvCI sites of this plasmid, downstream of the CTG repeat tract, generating a 221 bp region devoid of A's followed by a site for the nicking endonuclease N.BbvC IB (New England Biolabs). During this process, more CTG repeats were lost, generating the plasmid p$(CTG)_{60}$NICK with a final tract length of $[CTG]_{60}$. Plasmid constructs were confirmed by DNA sequencing (The DNA Facility, Office of Biotechnology, Iowa State University).

Construction of replication fork templates

Model replication forks were synthesized by a modification of the method described by Subramanian et al. (198). In short, the plasmid pRST5NICK was digested with N.BbvC IA to generate a nick in the G-rich strand at the start of the telomeric tract $[TTAGGG]_{96}$. Similarly, the plasmid p$(CTG)_{60}$NICK was digested with N.BbvC IB to generate a nick in the A-rich strand at the start of the 221 bp region containing the CTG repeat tract.

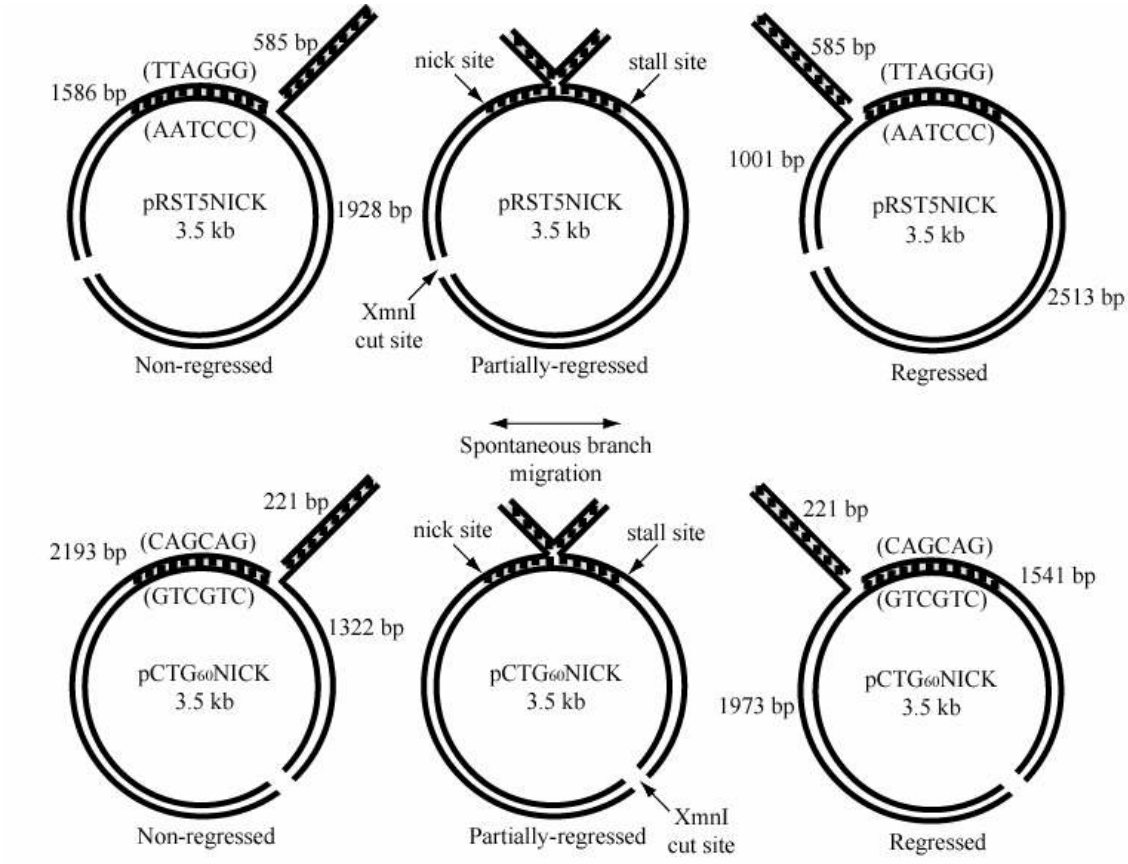
The nicked DNA was incubated with the Klenow fragment (exo$^-$) of DNA polymerase 1 (New England Biolabs), and 0.5 mM each dTTP, dATP, dGTP (telomeric template) or dGTP, dCTP, dATP (CTG template) to generate a single strand (ss) tail by strand displacement of the repeat tract. The ss tail was then converted to a double strand (ds) tail by annealing a 228-fold molar excess of the oligo 5′-CCCTAACCCTAACCCTAACCCTAA-3′ (telomeric template) or a 120-fold molar excess of the oligo 5′-CTGCTGCTGCTGCTG-3′ (CTG template) to each template for 30 min at 37 ºC in 100 mM NaCl, then ligating with T4 DNA Ligase (400 U, New England Biolabs) at 16 ºC overnight in a buffer containing 50 mM NaCl, 10 mM Tris-HCl (pH 7.9) and 1 mM dithiothreitol. The plasmids were then linearized with XmnI (New England Biolabs), allowing detection of the replication fork junction relative to the DNA ends, and measurement of the asymmetrically-generated long linear segments of the plasmid, to occur (Figure 4.1).

Electron Microscopy

Replication fork template DNA was diluted to 1 µg/ml in 10 mM HEPES pH 7.5, 0.1 mM EDTA and prepared for EM visualization as previously described (158). An FEI Tecnai 12 instrument was used at 40 kV to photograph images on plate film or a Gatan US4000SP Ultrascan camera. Micrographs for publication were captured from plate film negatives using an Imacon Flextight 848 scanner. Molecule lengths (in nm) were measured from Gatan digital images using Gatan Digital Micrograph software and

Figure 4.1. Schematic representation of replication fork templates.

Details of the synthesis steps are in Materials and Methods. The telomeric replication fork template was constructed on the plasmid pRST5NICK, and the CTG repeat template was made using the $pCTG_{60}$NICK plasmid. The lengths of the long linear segments of the plasmid are indicated in both the non-regressed and fully regressed forms of molecules containing only a single ds tail. Patterned region indicates repetitive DNA. Positions of the nicking site, the site of replication stalling, and the XmnI restriction site used to linearized the plasmid DNA are shown in the center panel of each template diagram.


585 bp
(TTAGGG)
1586 bp
(AATCCC)
pRST5NICK
3.5 kb
1928 bp
Non-regressed
nick site
stall site
pRST5NICK
3.5 kb
XmnI
cut site
Partially-regressed
585 bp
(TTAGGG)
(AATCCC)
1001 bp
pRST5NICK
3.5 kb
2513 bp
Regressed
Spontaneous branch
migration
221 bp
2193 bp
(CAGCAG)
(GTCGTC)
pCTG60NICK
3.5 kb
1322 bp
Non-regressed
nick site
stall site
pCTG60NICK
3.5 kb
Partially-regressed
XmnI
cut site
221 bp
(CAGCAG)
1541 bp
(GTCGTC)
1973 bp
pCTG60NICK
3.5 kb
Regressed

converted to bp.

The mean and standard deviation of the total DNA length (nm) of all measured molecules was determined, per experiment, and only molecules falling within the range: mean ± standard deviation, were used to determine percent of molecules regressed or non-regressed.

RESULTS

Synthesis of replication fork templates

Synthetic model replication forks were prepared by nicking the repeat-containing plasmids pRST5NICK and p$(CTG)_{60}$NICK adjacent to the repeat tract and replicating in the absence of one of the four nucleotides. Replication through the repeats stalled at the end of the tract, generating a ss tail that was converted to a ds tail by annealing and ligating complementary oligos along the length of the displaced strand. In both cases, the positions of the nicking sites and sites of replication fork stalling were sufficiently close to (within 2-4 bp of) the repeat tract that there was very little non-repeat DNA present in the ss tail and most of the displaced ss DNA could be converted to ds DNA by ligating complementary oligos. The plasmids were then linearized so that the position of the replication fork junction relative to this restriction site could be determined by measuring the length of each of the longest segments of the

replication fork template from the DNA end up to the fork junction (Figure 4.1).

Visualization of replication fork templates

Visualization of the replication fork molecules containing telomeric or CTG repeat tracts revealed an array of DNA configurations (Figure 4.2). For both model DNAs the most common species consisted of a linear replication fork template (Figure 4.2A, C, E, G) containing a single ds tail. Also present were molecules containing two shorter ds tails (shown at higher magnification in Figure 4.2B, D, F, and H). These are typical of chickenfoot intermediates generated by fork regression as seen by EM (198). Six separate preparations of telomeric DNAs and three separate preparations of the CTG triplet repeat template were scored, with more than 200 molecules counted per experiment. On average, 32 ± 10% of all telomere model DNAs contained a chickenfoot structure within the repeat tract while 15 ± 7% of the CTG repeat containing DNAs contained such structures (Figure 4.3). The substantial presence of these four-way junctions is highly significant, since these structures were absent in similar preparations of non-repeat containing replication fork templates synthesized under the same conditions and prepared for EM using the identical protocol (198). Thus the appearance of these chickenfoot structures must reflect the result of spontaneous fork migration caused by features unique to repetitive DNA.

Figure 4.2 Visualization of DNA configurations by EM.

Model replication forks were prepared for EM by mounting on carbon coated EM grids, and rotary shadowcasting with tungsten (Materials and Methods). Examples of linear molecules seen include replication fork templates comprising telomeric *(A, B, C, D)* or CTG *(E, F, G, H)* repeats, and containing only a single ds tail *(A, C, E, G)* or two shorter ds tails *(B, D, F, H)*. Images shown in reverse contrast. Bar is equivalent to 150 bp in panels showing full-length molecules.

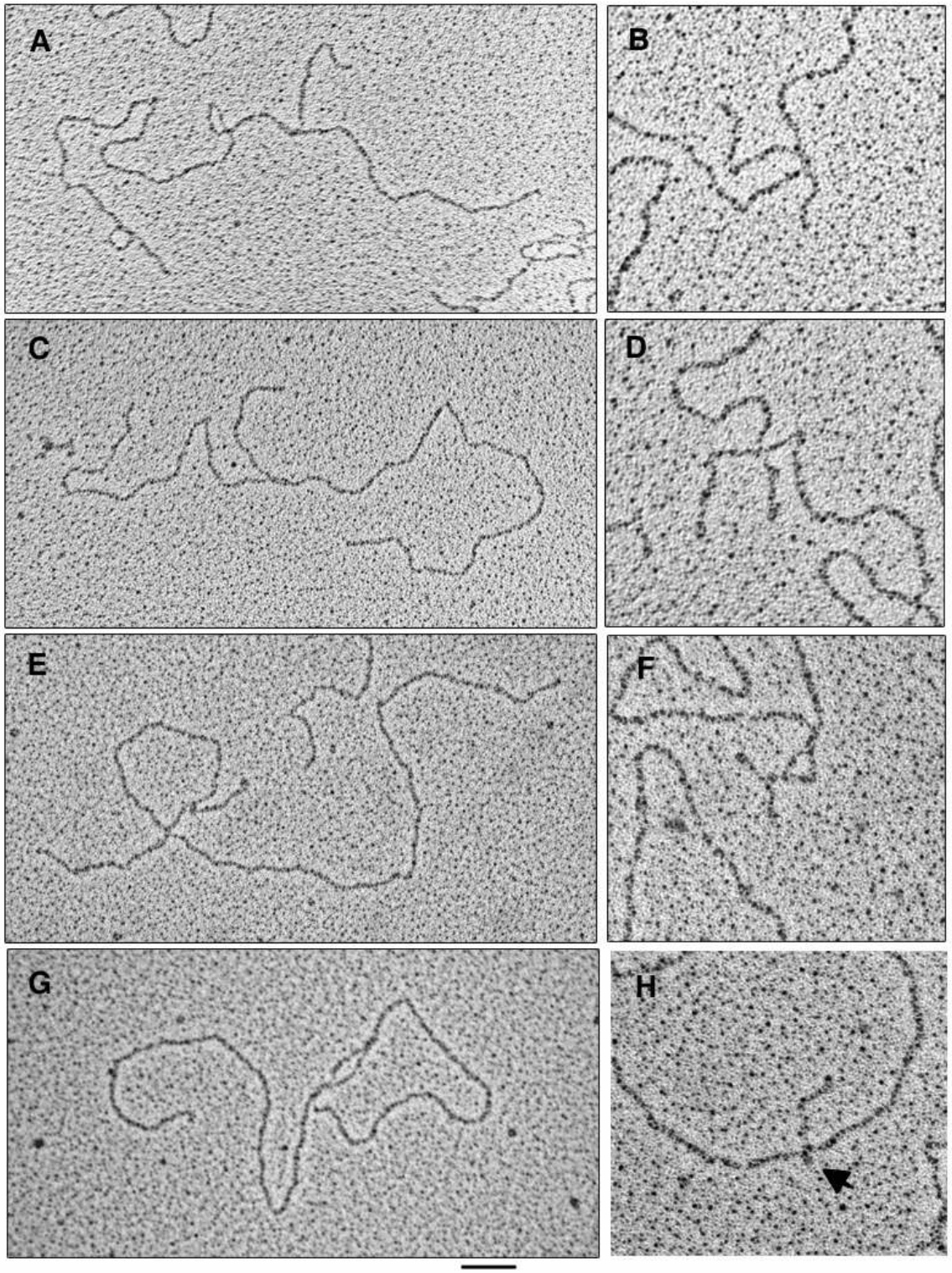
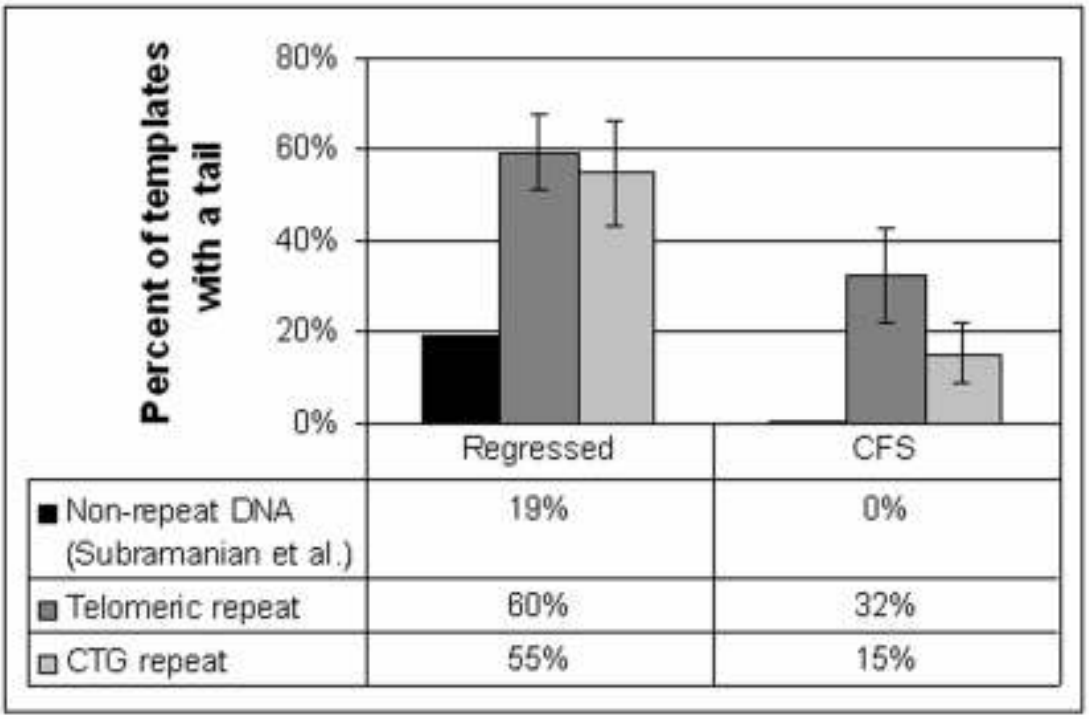

A
B
C
D
E
F
G
H

Figure 4.3. Graph of spontaneous regression of replication forks *in vitro*.

All tailed molecules were counted and measured, per experiment, and the average fraction of these molecules that were determined to have regressed was calculated (lane 1). Previously-reported values for regression of non-repeat DNA (198) to allow for comparison to the repeat DNA results. The average fraction of chickenfoot structures visualized by EM was also graphed as a percentage of all tailed molecules seen (lane 2).



	Regressed	CFS
■ Non-repeat DNA (Subramanian et al.)	19%	0%
■ Telomeric repeat	60%	32%
□ CTG repeat	55%	15%

Spontaneous replication fork regression in repetitive DNAs

Template DNAs were linearized so that the repeat block measured from 28% to 45% (telomere repeat) or 38% to 44% (CTG repeat) of the total DNA length from the nearest end, providing a means to uniquely determine the position of the replication fork (Figure 4.1). Molecules containing a single ds tail were photographed, and a minimum of 75 molecules for each template were analyzed to determine the position of the replication fork junction (see "Materials and Methods"). This was necessary because molecules containing a single ds tail could represent one of three types of replication fork templates – unregressed; fully regressed; or partially regressed chickenfoot molecules in which one of the ds tails was too short to be visualized by EM (Figure 4.2H, arrow). This was particularly important in the case of the CTG template, where the repeat tract was relatively short and the likelihood was high that one of the two ds tails of the chickenfoot intermediate would go unobserved.

Telomeric forks were considered to have begun regressing when the longer segment of the plasmid was greater than 1928 bp plus one standard deviation of the mean total DNA length, per experiment. Similarly the CTG repeat forks were considered to have begun regressing when the shorter segment of the plasmid was greater than 1322 bp plus one standard deviation of the mean total DNA length, per experiment. On the average, 60 ± 8% of telomeric replication forks and 55 ± 12% of CTG repeat forks had regressed to some degree (Figure 4.3).

Previously, when replication fork molecules which lacked any repetitive sequences at or near the fork and which contained only a single ds tail were examined, only 19% of all molecules were found where measurement of the long linear segments indicated that the replication fork had begun to regress (198). Spontaneous replication fork regression in repetitive DNA was therefore determined to be 41% and 36% higher in telomeric and CTG repeat DNA, respectively, than in non-repeat DNA. These results suggest that there is a high propensity for a paused replication fork to spontaneously regress when moving through repetitive DNA, resulting in a chickenfoot intermediate that would present a significant block to replication, requiring the action of recombination proteins to restart replication.

DISCUSSION

In this study we used EM to visualize stalled replication forks containing long runs of repetitive DNA sequences. We have shown that these forks have a much greater tendency to spontaneously regress, resulting in four-stranded chickenfoot intermediates, than non-repeat containing DNA. Specifically, 60% of telomeric and 55% of CTG repeat forks had regressed to some degree, in contrast to 19% regression seen in non-repeat DNA. Of particular interest were the greatly increased fractions of chickenfoot molecules seen in the repeat-containing DNAs: 32% of the telomeric templates and 15% of the CTG repeat templates were in 4-stranded

chickenfoot forms. In contrast, our previous study of non-repeat containing forms revealed that these 4-stranded intermediates were absent unless p53 was present to trap them (198).

The accumulation of chickenfoot structures is intriguing, given that four-way junctions have a higher number of broken base pairs and are likely less energetically favorable than three-way junctions. A possible explanation may be that chickenfoot structures carrying repetitive runs may be stabilized by additional secondary structures or repeat slippages in each of the repeat-containing DNA arms. However, we did not see any T- or Y-shaped protrusions in the slipped DNA arms, arguing that if they are present, they are not large. These DNAs seem inherently more “slippery” than non repeat-containing DNA, such that the replication forks are able to more easily transition back and forth between non-regressed and fully regressed states. During replication, repeat-containing DNA could therefore spend a significantly larger fraction of time in the partially-regressed state than other DNAs, and this could possibly account for the large percentage of chickenfoot structures seen. This model assumes that fork regression would be equal, regardless of the orientation of the repeats. However, since our studies concentrated on just one orientation for both repeats studied, we cannot rule out the possibility of orientation dependence on replication fork regression in these DNAs.

This study has thus revealed a new feature of repetitive DNA that

could present a significant barrier to replication. Furthermore, we believe that the four-stranded chickenfoot structures could present a significant problem to the cell, resulting in recruitment of unwanted recombination factors or leading to deleterious recombination events if repaired. We have shown in our laboratory that p53 will bind to chickenfoot structures with great affinity (198) and it greatly increases the rate of Holliday junction cleavage by resolvase enzymes (200) *in vitro*. Recently, similar results have been shown for the homologous recombination DNA repair protein XRCC3 in complex with Rad51C (S. Compton, unpublished data). Thus, an abundance of chickenfoot structures in the cell may have significant downstream consequences for cellular signaling and DNA repair.

Although the number of chickenfoot structures present in CTG repeat-containing DNA was approximately half of that seen in the telomeric samples, we believe that this number would have been higher if the repeat tract in the plasmid was longer. From measurements of the long linear segments of the plasmids, it was determined that approximately the same number of molecules had regressed to some degree in the telomeric samples as in the CTG-repeat samples. The triplet repeat DNA therefore appears to be as slippery as the telomeric DNA. Thus it seems probable that a large percentage of the chickenfoot molecules in the CTG-repeat samples contained one ds tail that was too short to be visualized by EM (Figure 4.2H, arrow). Also, we found that annealing of the oligos along the ss tail to make the ds tail in these samples was somewhat inefficient, resulting in

replication fork templates containing a ds tail shorter than the expected 221 bp. Although the presence of ssDNA regions in the displaced tail may have had the ability to bind back to the template, resulting in replication forks with a loop at the fork junction, no evidence of these structures was seen. More significantly, ssDNA regions resulting from poor oligo annealing did not seem to interfere with the ability of the replication forks to regress to a high degree, and binding back of the displaced ss tail to the template could not account for the chickenfoot structures seen.

Because of the considerable instability of repeat tracts in bacteria, this study was limited to the telomeric repeat TTAGGG and the triplet repeat $(CAG)_n$•$(CTG)_n$. However, in the future studies with replication fork templates containing sufficiently long stretches of the repeats $(CGG)_n$•$(CCG)_n$, $(GAA)_n$•$(TTC)_n$, $(CCTG)_n$•$(CAGG)_n$, $(AATCT)_n$•$(AGATT)_n$ and $(C_4GC_4GCG)_n$•$(CGCG_4CG_4)_n$ will be important to extend and generalize these observations to all of the known disease-related repeats.

Our observations are therefore consistent with repetitive DNA being a poor substrate for replication *in vitro* (192,196,201,202). The existence of stable four-stranded chickenfoot structures may explain the need for additional helicases such as BLM and WRN for efficient replication through telomeres *in vivo* (93,94,102,103,116). The data also favor a model for expansion of disease-related repeats that involves replication restart via chickenfoot intermediates, particularly in human cells where the repeat

blocks can be much longer than those investigated in the bacterial model study (131).

Interestingly, whereas replication of telomeric DNA tends to stall *in vitro* ((196) and unpublished data) and the G-rich strand may form G-quartets *in vivo* (203), human telomeres are replicated as rapidly as bulk DNA (100,204,205). Our data therefore also raise the possibility that factors at the telomere are actively involved in recognizing regressed-fork chickenfoot structures and rapidly resolving them in a tightly regulated process to restart replication, without allowing significant changes to the length of the telomere. The recent report that the telomeric binding protein Taz1 is required for replication of telomeres in the fission yeast Schizosaccharomyces pombe (101) further suggests that factors such as the telomere binding proteins TRF1 or TRF2 might play such a role in human cells.

CHAPTER 5: THE BASIC DOMAIN OF TRF2 DIRECTS BINDING TO DNA JUNCTIONS IRRESPECTIVE OF THE PRESENCE OF TTAGGG REPEATS

INTRODUCTION

Telomeres are nucleoprotein structures that protect the ends of chromosomes and are essential for regulating the replicative lifespan of somatic cells. The DNA component of the mammalian telomere consists of long double-stranded (ds) tracts of the hexameric repeat unit TTAGGG (3) that ends with a G-rich 3′ single stranded (ss) overhang (2). Telomeric DNA is thought to be organized into a t-loop "end-capping" structure by the telomere binding proteins TRF1, TRF2 and POT1 and the proteins that bind to them, TIN2, TPP1 and Rap1 (10,16). This higher-order structure may enable cells to distinguish chromosome ends from random double-strand breaks. Large blocks of telomere repeat sequences can be lost when these end-capping proteins are disrupted, or problems are encountered during DNA replication or repair (reviewed in (206). This typically results in p53- and Rb-mediated senescence or cellular crisis, as evidenced by end-to-end fusions of chromosomes, ATM dependent activation of p53 and apoptosis (reviewed in (185).

Much has been learned about the properties of TRF1 and TRF2 including their binding to DNA and the effects of their ablation or overexpression in the cell. We observed that TRF1 forms filamentous structures on long tracts of telomeric DNA *in vitro* (162), whereas TRF2 binds preferentially to the telomeric DNA at the junction between the duplex repeats and the ss overhang (69). Both TRF1 and TRF2 contain a similar Myb domain at their C-terminus that mediates their binding to ds telomeric DNA (207). TRF1 and TRF2 differ in their N-termini however, which are rich in either acidic residues in TRF1, or basic residues in TRF2. The function of the basic domain of TRF2 is poorly understood. Deletion of this domain ($TRF2^{\Delta B}$) does not impede the DNA binding activity of TRF2 or its localization to telomeres *in vivo*, but expression of $TRF2^{\Delta B}$ resulted in stochastic deletions of telomeric DNA, generation of t-loop-sized telomeric circles (t-circles), cell cycle arrest and induction of senescence in human cells (59,70). In addition, recent evidence suggested that the basic domain, but not the Myb domain, was required for TRF2 association with photo-induced double strand breaks in non-telomeric DNA in human fibroblasts (208).

Relatively little is known about the replication of mammalian telomeric DNA *in vivo*, however experiments in ciliates and budding yeast have provided insight into how this occurs in other eukaryotes. During each round of replication, all but the very end of the telomere is replicated by the conventional semi conservative polymerase machinery (95). Leading strand sequences eroded in the last round of replication (end replication problem)

can be restored by the reverse transcriptase telomerase (29), while the lagging strand is concurrently elongated by polymerases α and δ using the newly formed G strand as the template (30).

A possible complication of replication at the telomere is the requirement for protecting the DNA ends from recognition by DNA repair factors, while still allowing the DNA to be accessible to the replication machinery. Also, replication of telomeric DNA tends to stall *in vitro* (196) and long blocks of telomeric repeats are highly unstable when transformed into *E. coli* cells that lack recombination enzymes, suggesting difficulties with DNA replication through the telomeric tract (unpublished data). Furthermore, the G-rich strand of telomeric DNA has the tendency to form G-quartets (97) and the complementary cytosine-rich strand can fold into an intercalated tetramer called the i-motif (98).

The fact that human telomeres are replicated as rapidly as the bulk DNA (100) suggests that in addition to the standard replicative machinery, telomere-targeted factors may exist to actively facilitate its rapid replication. Direct evidence for the requirement of such telomeric factors at the replication fork was recently discovered in the fission yeast *Schizosaccharomyces pombe*, where the telomere binding protein Taz1 is required for efficient replication of telomeres (101). Also, in addition to the normal replicative helicases present at forks, the RecQ helicases WRN and BLM, implicated in premature ageing diseases, have been shown to be important for proper

telomere replication and maintenance in human cells (94,102,103).

The RecQ helicases have been shown to unwind G quartets (107) and promote branch migration of 4-stranded junctions similar to chickenfoot structures (108,109). In human cells, TRF2 co-localizes and physically interacts with WRN (134), and it binds to and stimulates the activities of both the WRN and BLM helicases *in vitro* (105). TRF1 and POT1 have also been shown to regulate WRN and BLM unwinding of telomeric substrates *in vitro* (110,133). Furthermore, *in vitro*, overexpression of TRF1 and TRF2 led directly to replication fork stalling (196), suggesting that the telomere binding factors also have a direct effect on the replication machinery.

We recently discovered a new feature of telomeric DNA that may explain this requirement for the RecQ helicases during telomere replication. Using EM and model replication fork templates that mimic a replication fork that had transited a long block of telomeric repeats, we discovered that telomeric DNA is inherently more slippery than non-repeat-containing DNA, such that the replication forks are able to more easily transition back and forth between non-regressed and fully regressed states (unpublished data). During replication, repeat-containing DNA could therefore spend a significantly larger fraction of time in the partially-regressed state, characterized by a Holliday junction or “chickenfoot” structure, than other DNAs. We believe that this presents a significant problem to the cell, where these 4-stranded structures could result in recruitment of unwanted recombination factors or lead to

deleterious recombination events if repaired. p53 will also bind to stalled chickenfoot structures with great affinity, suggesting that it may have the ability to halt excessive fork regression (198). These observations led us to ask whether one or both of the primary ds telomere binding proteins TRF1 and TRF2 might also show some unusual binding with regards to 3- and 4-way DNA junctions, when they occur within telomeric tracts.

To test this hypothesis, we generated a set of DNA templates including telomeric and non-telomeric replication forks, Holliday junctions, and model telomeres containing 3′ overhangs. Using EM and polyacrylamide gel-shift assays we evaluated binding to these templates by TRF1, TRF2 and TRF2$^{\Delta B}$ as well as a peptide encompassing the basic domain of TRF2 and another similar “mutant” peptide containing a rearrangement of four amino acids (aa).

In this paper we show that TRF2, but not TRF1 nor TRF2$^{\Delta B}$, is able to target the junctions of replication forks, chickenfoot structures and Holliday junctions. Junction binding occurred irrespective of the presence of TTAGGG repeats, and a bias for 4-stranded junctions was detected. The peptide mimicking the basic domain of TRF2 recapitulated this DNA-junction binding, whereas the mutant peptide could not. Furthermore, in the absence of other telomere binding proteins, TRF2 lacking the basic domain was not able to target the end of the large model telomeres and facilitate t-loop formation *in vitro*. We therefore suggest a novel role for the previously uncharacterized basic domain of TRF2, which is to facilitate TRF2 binding to

chickenfoot intermediates of telomere replication fork regression, presumably preventing their recognition by Holliday junction resolvases. The data are also the first direct demonstration of TRF2 binding specifically and selectively to DNA junctions, irrespective of the presence of telomeric repeats.

MATERIALS AND METHODS

DNA probes and templates.

[γ-^{32}P]ATP end-labeled J12 4-way junction probes (200); large Holliday-junction DNA templates (HJ575; (209); model non-telomeric replication forks (198); and model telomere DNA (69) was synthesized as previously described. A telomeric probe with a 14 nt overhang was prepared by annealing the [γ-^{32}P]ATP end-labeled oligonucleotide (oligo) 5′-CTAACCCTAACCCTGTCCTAGCAATGTAATCGTCTATGAGTCTG-3′ to the oligo 5′-CAGACTCATAGACGATTACATTGCTAGGACAGGGTTAGGGTTAGGGTTAGGGTTAGGG-3′. A hairpin probe consisting of a 7 nt linker and a 21 bp stem was prepared by heating and step-wise cooling the [γ-^{32}P]ATP end-labeled oligo 5′-CTTATTCACAGACCACGACTCAAAAAAAGAGTCGTGGTCTGTGAATAAG-3′. All annealed products were purified on 10% non-denaturing polyacrylamide gels.

Telomeric replication forks were created using a variant of pRST5 (69) containing a nicking site directly adjacent to the telomeric repeat tract (unpublished data). The plasmid was nicked with N.*Bbv*C IA (New England Biolabs (NEB), Ipswich, MA) and then incubated with the Klenow fragment (exo^-) of DNA polymerase 1 (NEB), and 0.5 mM of each of dTTP, dATP, dGTP to generate a ss tail by strand displacement of the repeat tract. The ss tail was converted to a ds tail by annealing a 228-fold molar excess of the oligo 5′-CCCTAACCCTAACCCTAACCCTAA-3′ to the template for 30 min at 37 ºC in 100 mM NaCl, and ligating with T4 DNA Ligase (400 U, NEB) at 16 ºC overnight in 50 mM NaCl, 10 mM Tris-HCl (pH 7.9) and 1 mM dithiothreitol. When required, the replication templates were linearized with *Xmn*I (NEB) for 1 h at 37 ºC.

Proteins and Peptides

NH_2-terminal $[His]_6$-tagged human TRF1, TRF2 and $TRF2^{\Delta B}$ were purified to homogeneity from baculovirus-infected insect cells by the method of Bianchi et al., except that a Talon™ Metal affinity resin (Clontech, Palo Alto, CA) was employed instead of Ni-NTA (210). p53 as well as the carboxy-terminal domain of the p53 protein, comprising amino acid residues 311 to 393, were purified as previously described (211).

Two peptides, each containing an NH_2-terminal biotin motif, were synthesized by the UNC Micro-Protein Facility, UNC School of Medicine, Chapel Hill, NC (Figure 5.1B). The first peptide (Peptide_B) consisted

of aa 2-31 of wild type human TRF2 (207), whereas the second peptide (Peptide_M) comprised a four aa sequence rearrangement of the first peptide.

Electron Microscopy

Binding assays of TRF2 and TRF2$^{\Delta B}$ to the model telomere were done as previously described (69). Complexes of p53, TRF1, TRF2, TRF2$^{\Delta B}$ and both peptides with Holliday junction or replication fork DNA were formed by incubating a 25:1 molar ratio of protein monomer:DNA in a 20 µl volume of EM buffer (20 mM HEPES pH 7.5, 0.1 mM EDTA, 0.5 mM DTT, 75 mM KCl) for 20 min at room temperature. A 5X molar excess of streptavidin (Molecular Probes, Eugene, OR) to peptide was added to samples containing biotin-tagged peptide, for 5 min at room temperature. Samples were fixed with 0.6% (w/v) glutaraldehyde for 5 min at room temperature followed by filtration through 2 ml columns of 2% agarose beads (50-150 µ, Agarose Bead Technologies, Tampa, FL) pre-equilibrated with 0.01 M Tris-HCl (pH 7.6), 0.1 mM EDTA. The purified samples were prepared for EM by rotary shadowcasting with tungsten as previously described (158). An FEI Tecnai 12 electron microscope equipped with a Gatan 4K x 4K CCD camera were used to photograph images.

Mobility shift assays

Reaction mixtures (10 µl) containing probes (10 nM) and the proteins (see figure legends for details) were incubated at room temperature for 20

min in EM buffer. The mixtures were adjusted to 10% glycerol and loaded on 3.5% non-denaturing polyacrylamide gels in 45 mM Tris–borate, 1 mM EDTA. The gels were run at 140 V for 1 h at 4°C, dried, analyzed by autoradiography and quantified using a Storm 840 phosphorimager (Molecular Dynamics).

Calculating Kd

GraphPad Prism (GraphPad Software, Inc., San Diego, CA) was used for non-linear regression of the data obtained from the mobility shift assays. The one site hyperbolic binding equation used was: % Probe Shifted = (B_{max} x [nM Protein])/([n MProtein] + Kd). In experiments with the Holliday junction probe, dissociation constant (Kd) values were converted to association constant (Ka) values: Ka = 1/Kd; then multiplied by the EM specific binding factor (see Results): Ka, apparent = Ka x (% EM junction binding); and again converted to a dissociation constant: Kd, apparent = 1/Ka, apparent (nM).

RESULTS

TRF2, but not TRF1, binds to DNA Junctions in vitro.

To examine the binding of TRF1 and TRF2 to replication forks and Holliday junctions, and to compare this with previous studies of p53 (198), a series of large model templates were constructed for EM. Previously we described the generation of a model replication fork consisting of a ~500 bp

arm extended from a 3 kb linear or circular DNA at a unique site which is at the end of a 500 bp cassette consisting of a random but G-less sequence (198). We constructed a new replication fork template based on this design from a plasmid containing a 560 base pair (bp) telomeric cassette (Figure 5.1A). In this DNA, two arms of the Y-fork molecule contain telomere repeats leading up to the fork junction. In addition, a Holliday junction template containing 500 bp arms of plasmid-derived DNA extended from the well known J12 junction was prepared as previously described (200).

Purified TRF1 and TRF2 were incubated with the template DNAs and prepared for EM. EM binding experiments were conducted in parallel with purified p53. Examination of fields of molecules from incubations of TRF2 with the model telomere replication fork revealed several DNA-protein configurations. When a 25:1 molar ratio of TRF2 monomers to DNA template was used, approximately one third of all molecules contained a particle of TRF2 bound at the center of the stalled fork junction (Figure 5.2A). Of these molecules, a subset of the forks had regressed, generating chickenfoot structures, and TRF2 was also observed bound at these junctions (insert, Figure 5.2A). The remainder of the DNA consisted of DNA templates with no protein bound (the majority), a fewer number of DNAs containing TRF2 protein bound elsewhere on the DNA template but not at the fork junction, and more frequently, aggregates of two or more DNAs bound by a large mass of

Figure 5.1: DNA, protein and peptide constructs used in this study.

(*A*). Telomeric replication fork template and chickenfoot structure, patterned region indicates TTAGGG repeats. (*B*). TRF1 and TRF2 have similar TRFH dimerization and Myb domains. TRF2 has a unique N-terminal basic domain (B, aa 13-30) with conserved residues (indicated by a *). TRF2$^{\Delta B}$ is a deletion mutant of TRF2 lacking the first 44 aa. The basic (Peptide_B) and mutant (Peptide_M) peptides have an N-terminal biotin moeity (bt), and they differ in their sequences (grey lettering) by a re-arrangement of four aa (black lettering).

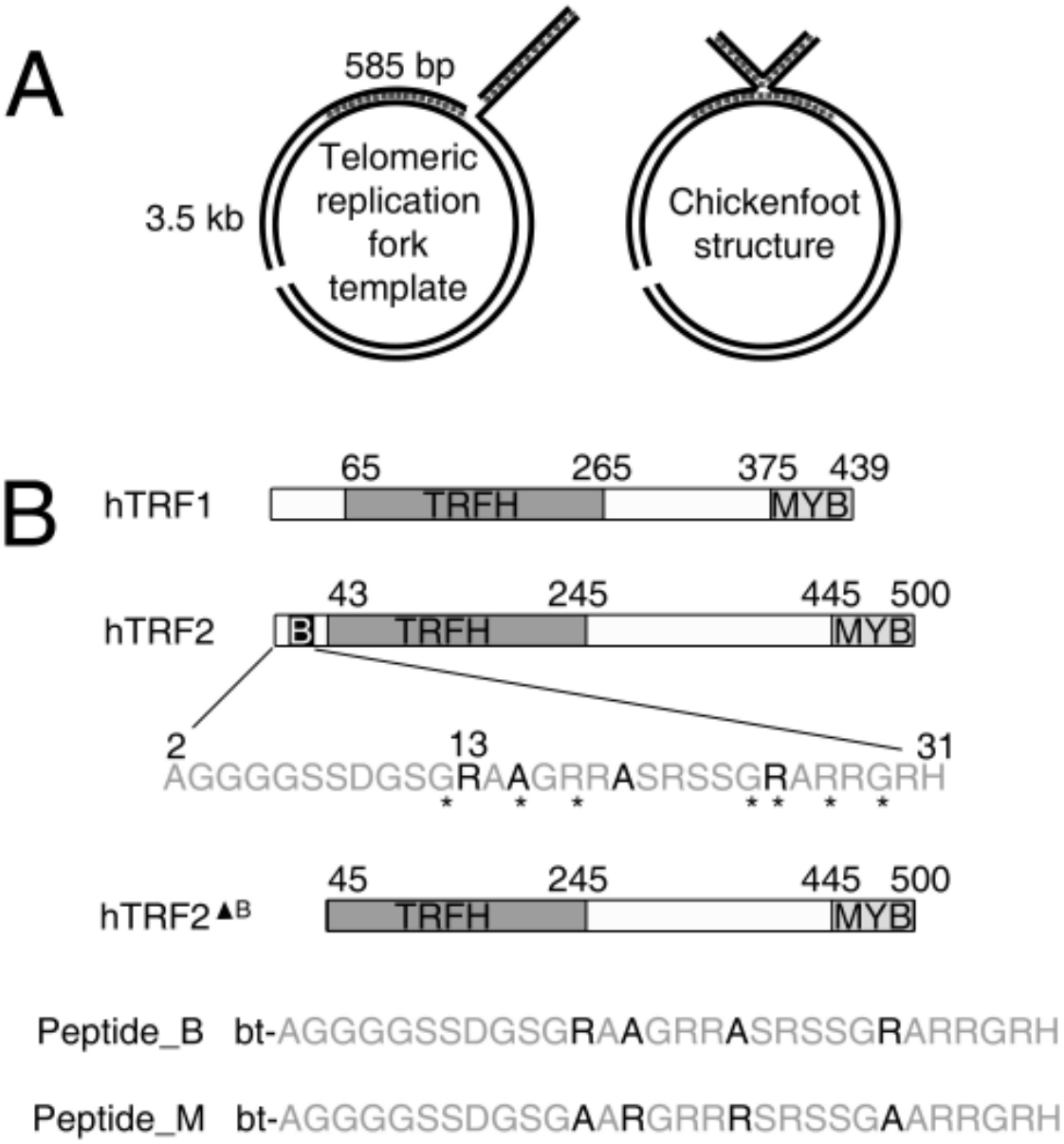
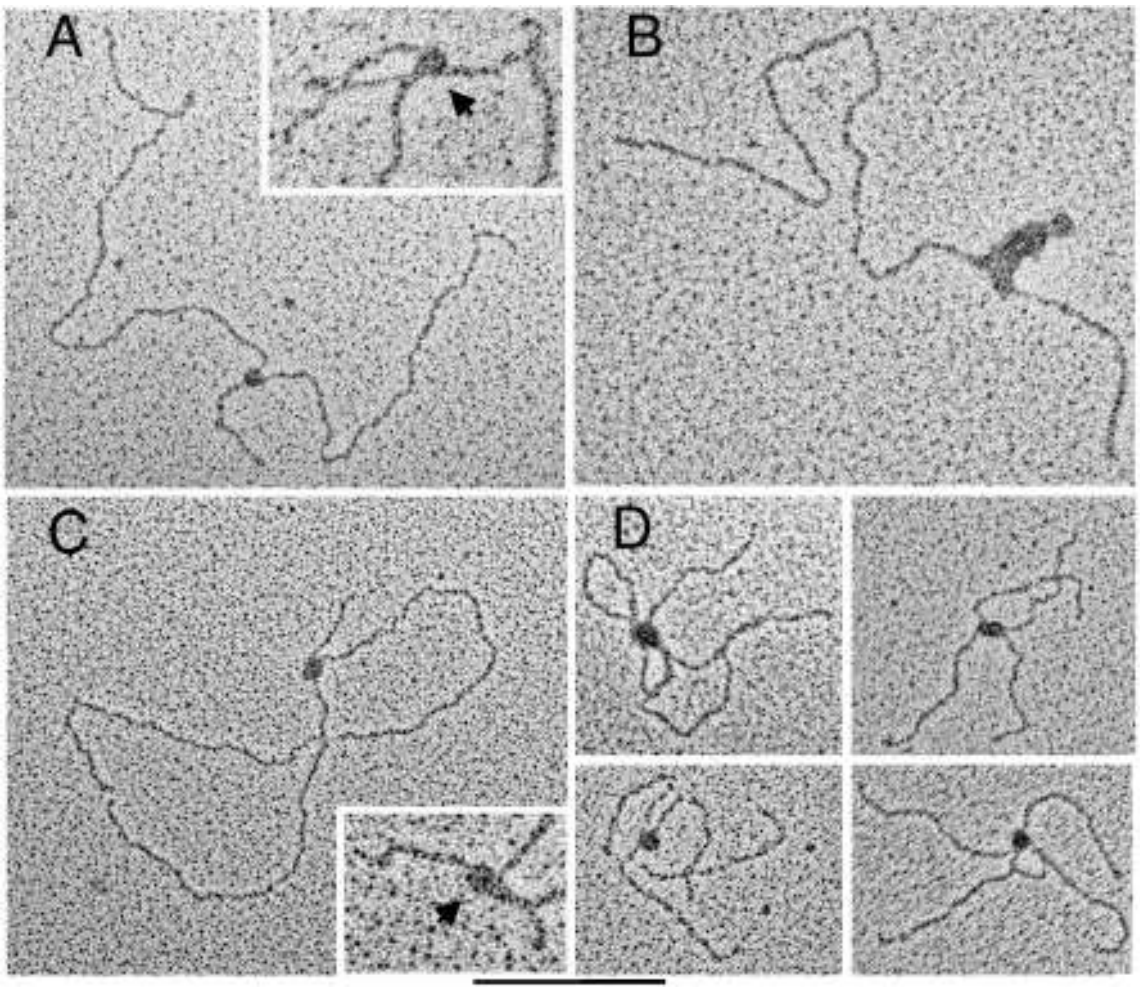

A
585 bp
3.5 kb
Telomeric replication fork template
Chickenfoot structure
B
hTRF1
65
265
375
439
TRFH
MYB
hTRF2
43
245
445
500
B
TRFH
MYB
2
13
31
AGGGGSSDGSGRAAGRRASRSSGRARRGRH
hTRF2▲B
45
245
445
500
TRFH
MYB
Peptide_B bt-AGGGGSSDGSGRAAGRRASRSSGRARRGRH
Peptide_M bt-AGGGGSSDGSGAARGRRRSRSSGAARRGRH

Figure 5.2: TRF2 binds DNA junctions in vitro.

(*A, C*). Discrete complexes of TRF2 bound to the junctions of the 3-stranded replication forks and the 4-stranded chickenfoot structures (arrow, insert); (*A,* telomeric template; *C*, non-telomeric template). (*B*). Extensive, filamentous binding of TRF1 to a model telomeric replication forks. (*D*). TRF2 bound to the center of Holliday junction template DNAs. Bar is equivalent to 450 bp in panels showing full-length molecules.


A
B
C
D

TRF2 (data not shown). These aggregates became more abundant in binding preparations containing molar ratios of TRF2 monomers to DNA template greater than 25:1. Of the TRF2-bound molecules where TRF2 was not located at the fork junction, the majority had TRF2 bound sufficiently close to the fork that we presumed it was bound along the TTAGGG tracts. Quantification of the results from scoring hundreds of molecules is described below. While the TRF2 particles varied in size, in most cases the size range was much less variable when bound to the center of the 3-way or 4-way junction than in the instances when TRF2 was observed bound along the duplex TTAGGG tracts (Figure 5.2A, insert) and the particle size was suggestive of TRF2 dimers or tetramers. When parallel experiments were performed with TRF1, long arrays of bound TRF1 were seen localized along the telomeric repeats (Figure 5.2B) resembling the TRF1 filaments formed along duplex TTAGGG tracts as previously described (162).

For comparison, TRF1 and TRF2 were incubated with the random sequence replication fork template lacking TTAGGG repeats. The binding of TRF1 to this template was low (quantified below), as expected from the lack of telomeric repeats. Of great interest however was the observation that incubation of TRF2 with this template led to discrete TRF2 complexes at the 3-way or 4-way junctions (Figure 5.2C). In molecules in which the forks had regressed into 4-way chickenfoot forms, TRF2 complexes were more frequently observed bound to the DNA and almost always at the 4-way junction (insert, Figure 5.2C and quantified below). This led us to

examine TRF2 binding to Holliday junction DNA containing 500 bp arms (Figure 5.2D). Indeed, by EM we found that TRF2 bound well to these structures, localizing to the center of the 4-way junction. By comparison, TRF1 bound poorly to these structures, and when it did, the binding appeared to be random with only a small percent of bound TRF1 found at the center of the template.

Quantification of TRF1 and TRF2 binding to 3-way and 4-way junctions: comparison with p53 binding.

We previously examined the binding of p53 to random sequence replication fork templates (198) as well as to large Holliday junction templates (200). We were thus able to combine these results in a comparison of TRF1 and TRF2 binding to these templates (Figure 5.3A, B and C). An average of 179 molecules over three experiments was counted per experimental condition.

Incubation of p53 (25:1) with the telomeric replication fork template resulted in observation of p53 at the fork and the frequency (69 ± 3% versus 8 ± 1% elsewhere on the DNA, Figure 5.3A) was comparable to previous observations with non-telomeric forks (56% at the fork versus 15% bound elsewhere; (198). We could not assess TRF1 binding to telomeric replication fork templates because of its extensive binding to the telomeric tract. Also, since the number of DNA molecules contained within the TRF2 aggregates could not be determined by EM, they were not included in the total

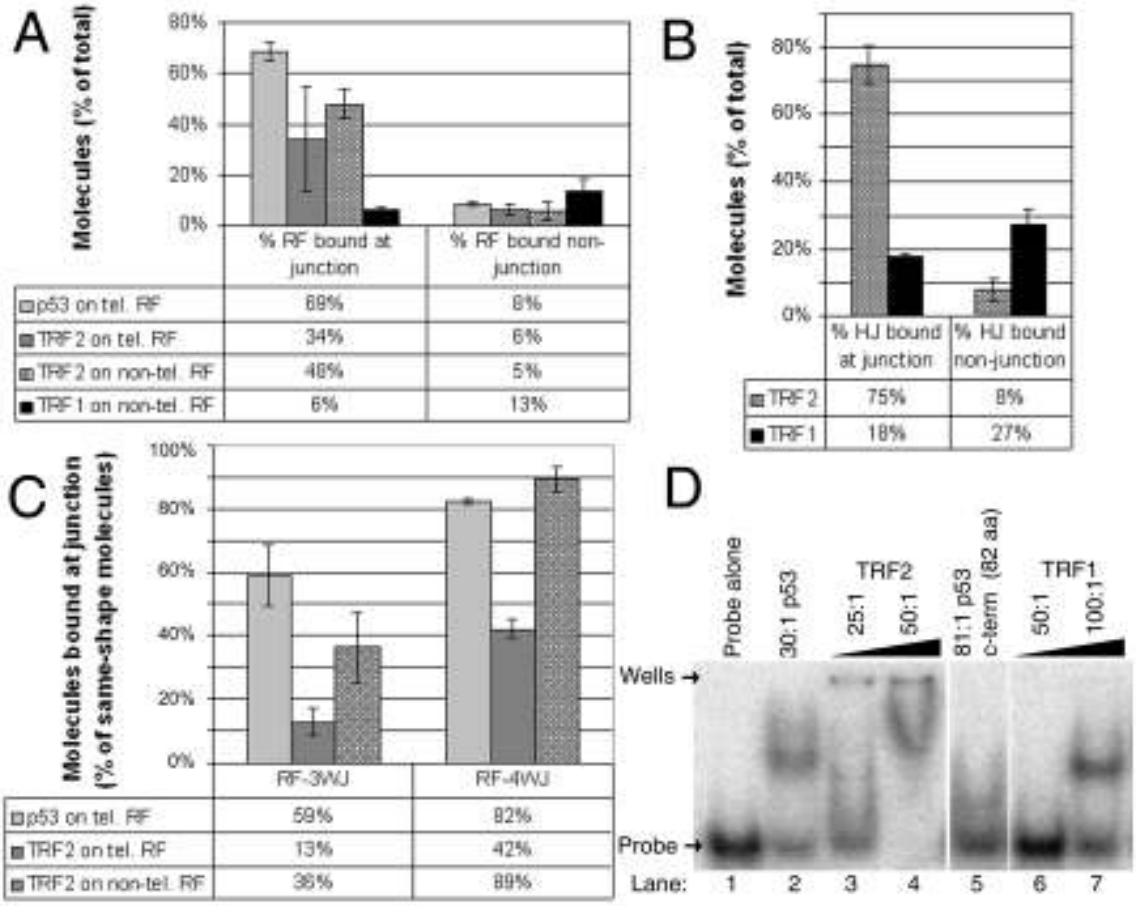
number of DNAs counted. Thus, the actual level of TRF2 binding to the telomeric replication fork is likely to be higher than the value cited here (34 ± 21% versus 6 ± 2% elsewhere on the DNA, Figure 5.3A) since the values were calculated as the percentage of the individual (non-aggregated) DNAs only. The relatively wide spread in calculated TRF2 binding is most likely due to a variation in the abundance of TRF2 aggregates in these samples, because when TRF2 binding in these samples was evaluated as a percentage of individually bound molecules, the standard deviation was small (76 ± 5% bound at the junction versus 24 ± 5% bound elsewhere).

A more pronounced preference for the DNA junction was revealed for TRF2 binding to non-telomeric replication forks, where protein-DNA aggregates were less abundant, and TRF2 bound to the fork junction 48 ± 6% of the time, and elsewhere on the DNA only 5 ± 3% of the time. This was in contrast to TRF1 binding to the non-telomeric replication forks, which was low and seemingly random; only 6 ± 1% of molecules had TRF1 bound at the fork junction, whereas 13 ± 5% had TRF1 bound somewhere else on the DNA.

TRF2 bound even more strongly to the Holliday junction templates, targeting the center of the 4-way junctions 75 ± 5% of the time and binding elsewhere on the DNA only 8 ± 4% of the time (Figure 5.3B). Previously, when p53 binding to Holliday junctions was studied by EM, 59% of all molecules were bound at their center by p53 (200). When TRF1 binding to the Holliday junction templates was examined, we found that it bound

Figure 5.3: TRF2 junction binding is biased towards 4-way junctions.

(*A*). Binding of p53, TRF2 and TRF1 to the telomeric replication fork templates (tel. RF) and the non-telomeric replication fork templates (non-tel. RF) was visualized by EM and quantified. Percentages are calculated as a fraction of all replication forks counted (3-way and 4-way junctions combined). Only molecules with protein bound at the junction of the replication fork or chickenfoot structure were considered junction-bound molecules. Data are represented as mean ± SD. (*B*). Holliday junction (HJ) templates bound by TRF1 or TRF2 were similarly quantified. (*C*). Percentages of TRF2 and p53 binding to replication forks are calculated as a fraction of molecules with the same shape. (*D*). Mobility shift assay of the [γ^{-32}P]-labeled J12 junction probe alone (lane 1) or bound by p53 (at a molar ratio of 30:1 protein:probe, lane 2); TRF2 (25:1, lane 3; 50:1, lane 4); an 82 aa C-terminal fragment of p53 containing the basic domain (81:1, lane 5); and TRF1 (50:1, lane 6; 101:1, lane 7).


A
Molecules (% of total)
80%
60%
40%
20%
0%
% RF bound at junction
% RF bound non-junction
p53 on tel. RF 69% 8%
TRF2 on tel. RF 34% 6%
TRF2 on non-tel. RF 48% 5%
TRF1 on non-tel. RF 6% 13%
B
Molecules (% of total)
80%
60%
40%
20%
0%
% HJ bound at junction
% HJ bound non-junction
TRF2 75% 8%
TRF1 18% 27%
C
Molecules bound at junction (% of same-shape molecules)
100%
80%
60%
40%
20%
0%
RF-3WJ
RF-4WJ
p53 on tel. RF 59% 82%
TRF2 on tel. RF 13% 42%
TRF2 on non-tel. RF 36% 89%
D
Probe alone
30:1 p53
TRF2
25:1
50:1
81:1 p53 c-term (82 aa)
TRF1
50:1
100:1
Wells
Probe
Lane: 1 2 3 4 5 6 7

considerably less well (18 ± 6% bound at the junction) than TRF2, and that most binding occurred on one or more of the arms of the Holliday junction, rather than at the DNA junction (27 ± 5%). When binding to non-telomeric replication forks was examined in more detail it became apparent that TRF2 has a greater affinity for the 4-stranded replication fork structures (Figure 5.3C). Specifically, we considered TRF2 binding to the Y-shaped 3-way junctions as a fraction of all 3-stranded junctions and similarly, TRF2 binding to the chickenfoot 4-way junctions as a fraction of all 4-way junctions. A clear bias was observed, where TRF2 bound to 89 ± 4% of all 4-way junctions versus 36 ± 11% of all 3-way junctions. When the DNA template contained telomeric repeats, this preference was not as strong (42 ± 3% of all 4-way junctions versus 13 ± 4% of all 3-way junctions), suggesting that the sequence-specific binding of TRF2 to the surrounding sequences led to the apparent lower fractional localization to the fork. p53 showed a slight bias for chickenfoot structures when bound to the telomeric replication fork (82 ± 1% of all 4-way junctions versus 59 ± 10% of all 3-way junctions) and no bias when binding to non-telomeric replication forks (approximately 52 ± 8% of all 4-way junctions versus 52 ± 7% of all 3-way junctions; (198). These data suggest, therefore, that TRF2 has a considerable and specific binding affinity for DNA junctions, with a bias for 4-stranded DNA junctions resembling Holliday junctions.

Mobility shift analysis of TRF1 and TRF2 binding to 3- and 4- armed DNA templates.

To further examine the binding of these proteins to fork junctions, mobility shift assays were carried out using TRF2, TRF1, p53 and the 82 aa C-terminal fragment of p53 with a 100 bp [γ-^{32}P] end-labeled Holliday junction probe (Figure 5.3D). TRF2 shifted the Holliday junction probe equally as well as p53, confirming our EM results. In contrast, much higher levels of TRF1 were required to observe any shift of the probe suggesting that this interaction with the probe is non-specific for the DNA strands (as was visualized by EM). As a control, we duplicated these binding assays with a small [γ-^{32}P] end-labeled hairpin probe, and in this case neither TRF1 nor TRF2 were able to bind to and shift the probe (data not shown).

The TRF2 basic domain targets DNA junctions in vitro.

These results suggest that TRF2 and p53 share common features in binding to 3-way and in particular 4-way junctions in a non-sequence-specific manner. This activity resides in the C-terminal basic domain of p53. To determine if the basic domain of TRF2 is involved in targeting TRF2 to DNA junctions, we purified an NH_2-terminal deletion mutant of TRF2 lacking the TRF2-specific basic domain (TRF2$^{\Delta B}$, Figure 5.1B). We also had a peptide synthesized (Peptide_B, Figure 5.1B) which consists of the amino acid sequence from aa 2-31 of TRF2, and thus encompasses the basic domain (aa 13-31) of TRF2. In addition, we had a “mutant” peptide synthesized

(Peptide_M, Figure 5.1B) comprised of the same number and composition of amino acids as Peptide_B, such that the overall positive charge of both peptides was the same, but containing a rearrangement of four amino acids, two of which are conserved among the species *Gallus gallus* (Chicken), *Muntiacus reevesi* (Chinese muntjak), *Muntiacus muntjak vaginalis* (Muntjak), *Mus musculus* (Mouse) and *Homo sapiens* (Human) (Homologous vertebrate genes database, HOVERGEN).

EM binding experiments were conducted with these proteins using the same conditions as used for TRF2 and TRF1, except that streptavidin was added to each of the peptide-containing samples just prior to sample preparation for EM. Both peptides contained an NH_2-terminal biotin moeity to which streptavidin could bind, thereby increasing the overall molecular weight of the peptides from 3 kDa to 56 kDa (equivalent to a monomer of TRF2 of 55 kDa) and allowing them to be visualized by EM.

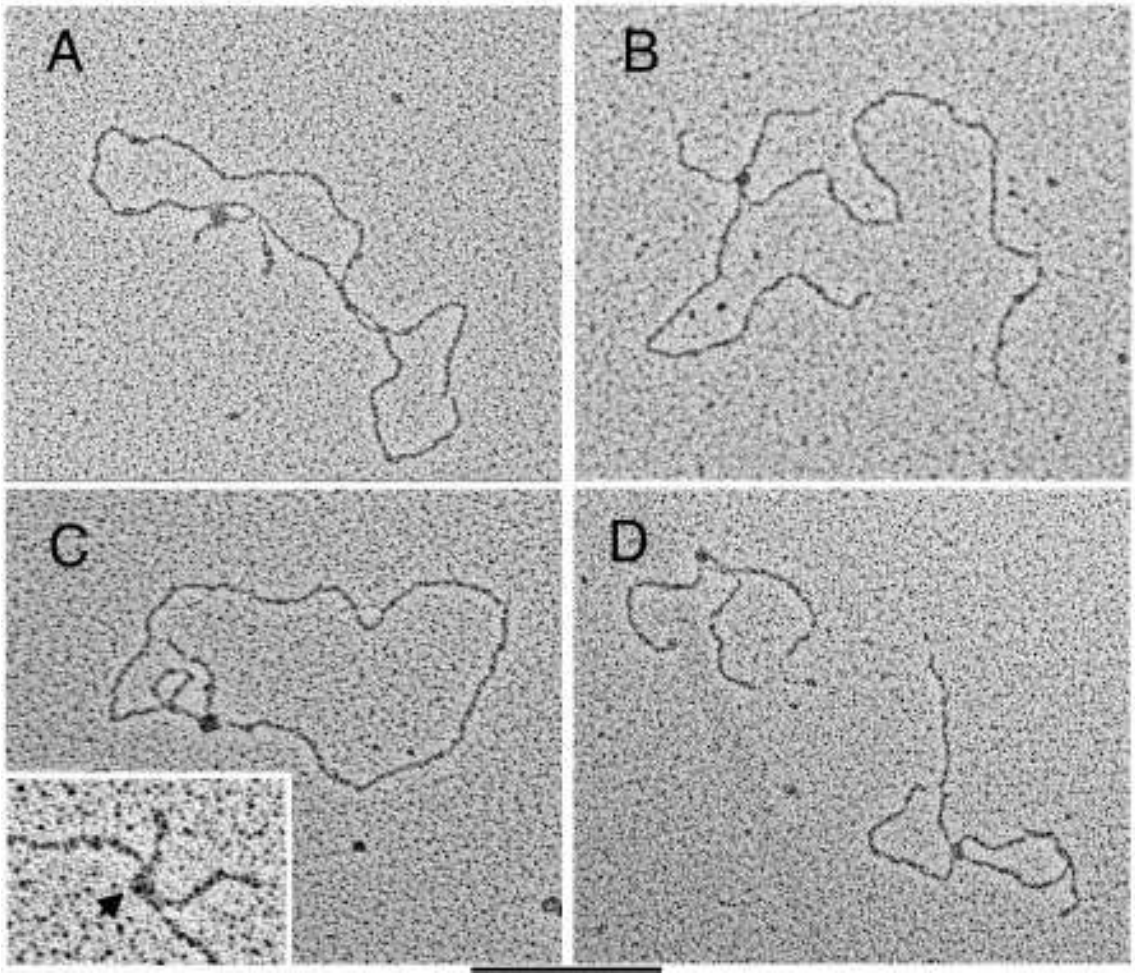
In experiments examining TRF2$^{\Delta B}$ binding to telomeric replication fork templates, the large majority of molecules consisted of aggregates of two or more DNAs bound by a large mass of TRF2$^{\Delta B}$ protein, although in some cases individual molecules of DNA could be seen. Frequently, these individual DNAs contained long arrays or masses of bound TRF2$^{\Delta B}$, although less often we were able to visualize a discrete particle of TRF2$^{\Delta B}$ bound to the DNA (Figure 5.4A). In such instances, TRF2$^{\Delta B}$ was localized to the telomeric repeats (on the displaced strand or within a distance equivalent to 500

bp from the fork junction) but it was rarely observed bound to the replication fork junction. Because most of the TRF2$^{\Delta B}$ protein in the sample was contained within the aggregates of TRF2$^{\Delta B}$, we were not able to quantify TRF2$^{\Delta B}$ binding to this template DNA. Also, we did not separately quantify binding of the peptides to the telomeric replication fork templates. The peptides do not have a telomere-repeat recognition motif and were observed to have similar binding to these templates (Figure 5.4B) as to the non-telomeric replication fork templates (Figure 5.4C).

As with TRF2, in experiments containing Peptide_B and the non-telomeric replication fork template, 41 ± 1% of all molecules consisted of a complex of streptavidin and Peptide_B bound at the center of the stalled fork junction (Figure 5.4C, non-telomeric replication fork with insert showing chickenfoot structure; see also Figure 5.5A). Protein bound elsewhere on the DNA templates 8 ± 1% of the time. No protein-bound complexes were seen in samples containing DNA and streptavidin alone (data not shown). Intriguingly, the mutant Peptide_M displayed 30 ± 5 % and 35 ± 6% binding to the junction and DNA arms of the non-telomeric replication forks, respectively. This reflects an increase in the overall binding from 49% (Peptide_B) to 65% (Peptide_M), with a seemingly small difference in junction binding (41% versus 30% respectively) between the two peptides. However, when we examined the junction binding in more detail, we noticed that the mutant Peptide_M had significantly decreased binding to 4-way junctions, with

Figure 5.4: The TRF2 basic terminus binds DNA junctions in vitro whereas TRF2$^{\Delta B}$ protein does not.

(*A*). TRF2$^{\Delta B}$ bound to the TTAGGG repeats of the telomeric replication fork template. (*B, C, D*). Discrete molecules of streptavidin-bound Peptide_B at the junctions of the telomeric chickenfoot structure (CFS, *A*); the non-telomeric replication fork (*C*) or CFS (arrow, insert); and the Holliday junction template DNAs (*D*). Bar is equivalent to 450 bp in panels showing full-length molecules.

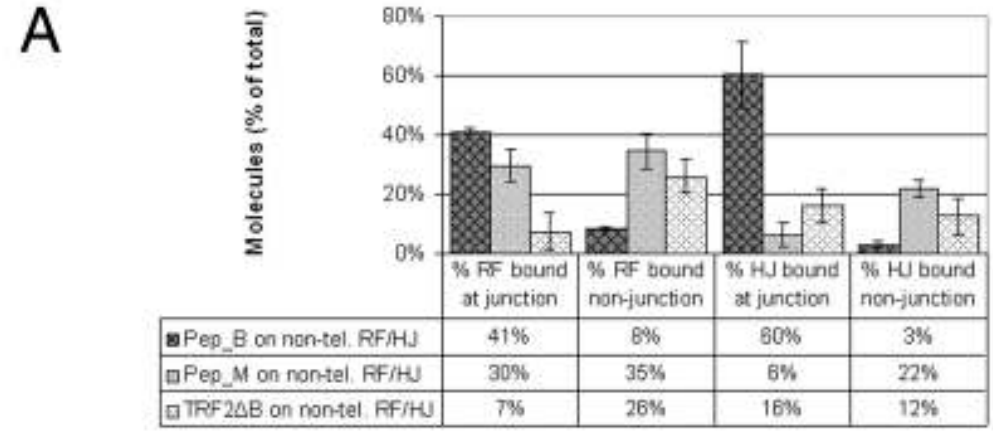

A
B
C
D

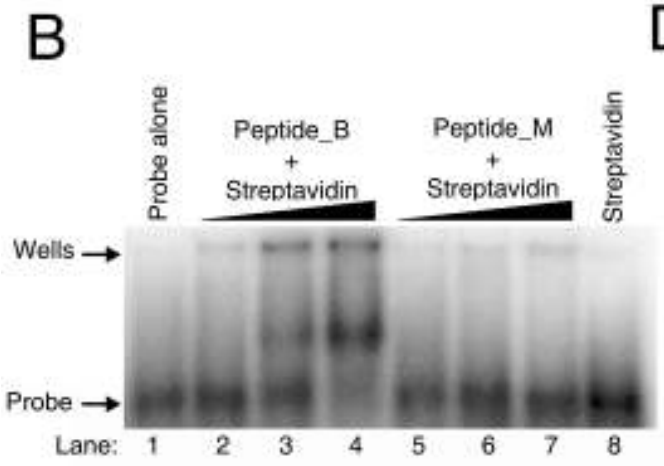
little change in binding to 3-way junctions. Specifically, when Peptide_B-bound molecules were evaluated as a fraction of the total DNA molecules with the same shape, Peptide_B bound to 94 ± 9% of all 4-stranded replication fork junctions versus 32 ± 8% of all 3-stranded junctions. In contrast, Peptide_M bound to only 33 ± 0% 4-stranded junctions (a reduction in binding of 61%) whereas it was still able to bind to 29 ± 6% of the 3-stranded junctions (a reduction in binding of only 3%). Therefore, it appears that a rearrangement of two conserved amino acids with two non-conserved amino acids specifically disrupted 4-way junction binding of the peptide mimicking the TRF2 basic domain, even though overall (non-specific) binding increased, and binding to the 3-way junctions remained largely the same.

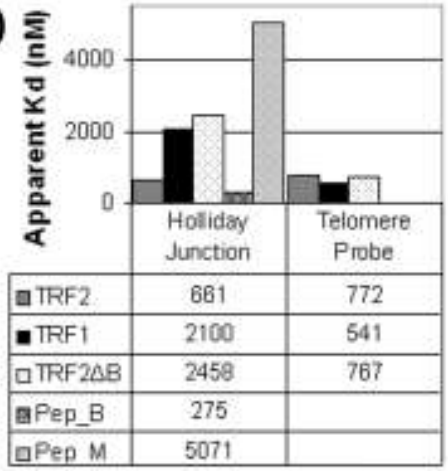
The TRF2 truncation mutant, $TRF2^{\Delta B}$, showed only 7 ± 6% binding to the non-telomeric replication fork junctions, preferring to bind somewhere else on the template 26 ± 6% of the time. Therefore it seems that the loss of the basic domain in the truncated TRF2 construct greatly attenuated its ability to target replication fork junctions, where junction-binding was reduced from 48% to 7%, and non-junction binding was increased from 5% to 26% (compare Figures 5.2A and 5.5A). These results were even more pronounced in EM experiments done with the Holliday junction template (Figure 5.4D). Junction binding for the Peptide_B, Peptide_M and $TRF2^{\Delta B}$ constructs was 60 ± 11%, 6 ± 4% and 16 ± 6% respectively, and non-junction binding was 3 ± 1%, 22 ± 3% and 12 ± 6% respectively (Figure 5.5A).

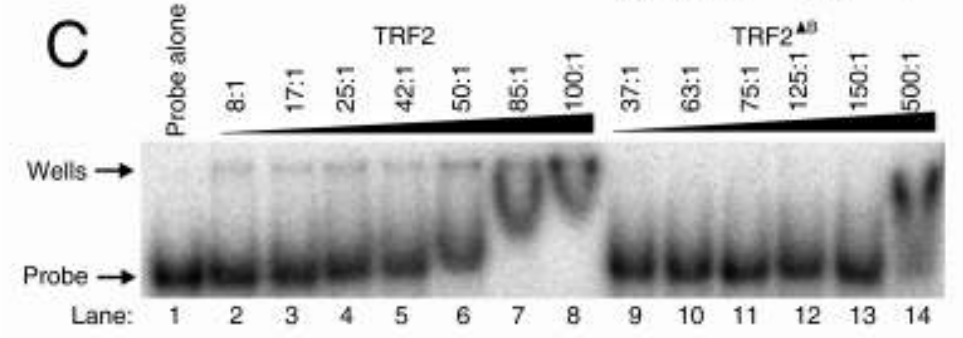
Figure 5.5: The basic domain of TRF2 is required for binding to 4-way junctions.

(*A*). Binding of the TRF2 basic peptide (Peptide_B), the mutant peptide (Peptide_M) and the TRF2 truncation mutant TRF2$^{\Delta B}$ to the non-telomeric replication fork templates (non-tel. RF) and the Holliday junction templates (HJ) was visualized by EM and quantified. Percentages are calculated as a fraction of all molecules counted. Only molecules with protein bound at the replication fork junction or at the center of the four strands of the Holliday junction were considered junction-bound molecules. Data are represented as mean ± SD. (*B*, *C*). Mobility shift assays of the [γ-^{32}P]-labeled J12 junction probe alone (lanes 1) or bound by the basic peptide (*B*, lanes 2-4); the mutant peptide (*B*, lanes 5-7); TRF2 (*C*, lanes 1-8) and TRF2$^{\Delta B}$ (*C*, lanes 9-14). Increasing peptide:probe molar ratios used were 1:1, 5:1 and 10:1, with each sample containing a 5:1 molar ratio of streptavidin protein:peptide. (*D*). Apparent dissociation constants (Kd) for TRF2, TRF1, TRF2$^{\Delta B}$, the basic peptide (Pep_B) and the mutant peptide (Pep_M) binding to the [γ-^{32}P] end-labeled Holliday junction probe (J12) or the small [γ-^{32}P] end-labeled telomere probe.


A
Molecules (% of total)
80%
60%
40%
20%
0%
% RF bound at junction
% RF bound non-junction
% HJ bound at junction
% HJ bound non-junction
Pep_B on non-tel. RF/HJ 41% 8% 60% 3%
Pep_M on non-tel. RF/HJ 30% 35% 6% 22%
TRF2ΔB on non-tel. RF/HJ 7% 26% 16% 12%


B
Probe alone
Peptide_B + Streptavidin
Peptide_M + Streptavidin
Streptavidin
Wells
Probe
Lane: 1 2 3 4 5 6 7 8


D
Apparent Kd (nM)
4000
2000
0
Holliday Junction
Telomere Probe
TRF2 661 772
TRF1 2100 541
TRF2ΔB 2458 767
Pep_B 275
Pep_M 5071


C
Probe alone
TRF2
8:1 17:1 25:1 42:1 50:1 85:1 100:1
TRF2ΔB
37:1 63:1 75:1 125:1 150:1 500:1
Wells
Probe
Lane: 1 2 3 4 5 6 7 8 9 10 11 12 13 14

Thus, TRF2 binding to the center of Holliday junction templates was reduced by 59% (from 75%) when the basic region was absent, and Peptide_B binding was reduced by 56% (from 60%) in the mutant Peptide_M samples. All of these data suggest, therefore, that the junction-binding activity (specifically the 4-way junction-binding activity) of TRF2 is mediated by the NH_2-terminal amino acids containing the basic domain.

Junction-binding evaluated using an apparent dissociation constant.

To compare the proteins used in this study with respect to Holliday junction binding, mobility shift assays were performed using a constant amount (10 nM) of the small Holliday junction probe in each sample. The concentration was varied for each protein, in increasing increments, until no more probe could be shifted into the gel. Two examples of these experiments are shown in Figures 5.5B and C. All experiments were performed a minimum of three times and averages were taken for each data point. Binding isotherms were generated with % probe shifted (y-axis) as a function of protein concentration (x-axis). In all instances, non-linear regression of the data using a one site hyperbolic binding equation gave a goodness of fit (R^2) ≥ 0.9. The apparent Kd values obtained were therefore: TRF2 (614 nM), TRF1 (819 nM), $TRF2^{\Delta B}$ (1401 nM), Peptide_B (261 nM) and Peptide_M (1065 nM). However, because gel-shift analysis does not distinguish junction binding from non-specific ds DNA binding, we adjusted the apparent Kd values we obtained by the following "specific binding factors"

(corresponding to the % junction-specific binding observed for each protein via EM): TRF2 (93%), TRF1 (39%), TRF2$^{\Delta B}$ (57%), Peptide_B (95%) and Peptide_M (21%). The resultant "apparent Kd" values (Figure 5.5D) clearly highlight the strong junction binding of TRF2 (661 nM) and the basic domain peptide (275 nM) versus the weaker, more random binding of TRF1 (2100 nM), TRF2$^{\Delta B}$ (2458 nM) and the mutant peptide (5071 nM). Interestingly, the apparent Kd values for TRF1 and TRF2$^{\Delta B}$ are very similar. This result is not surprising considering that both of these proteins have very similar Myb DNA binding domains that would likely bind to non-telomeric DNA with a similar low-level affinity.

To test the activity of the telomere-binding proteins under the binding conditions used, the mobility shift binding assays and non-linear regressions of the data were duplicated using a small [γ-^{32}P] end-labeled telomeric probe. In all cases, the one site binding hyperbola again gave a decent fit ($R^2 \geq 0.91$). Binding of all three proteins to the telomeric probe appeared to be equally as good, with TRF1 binding the strongest (Figure 5.5D). Most interestingly, TRF2 and TRF2$^{\Delta B}$ had an almost identical affinity for the telomere probe, suggesting that deletion of the basic domain had not destabilized the TRF2 core structure. Therefore, the lower junction binding seen in the TRF2$^{\Delta B}$ constructs were not a result of destabilization of the Myb DNA-binding domain. Moreover, under the experimental conditions used TRF2 appeared to have a slightly smaller Kd when binding to the Holliday junction probe than to the telomeric probe, suggesting that the basic domain of TRF2 has a

similar, if not greater, affinity for DNA 4-way junctions as the Myb domain has for the consensus telomeric sequence. The same mobility shift assays were also performed with streptavidin binding to the Holliday junction probe, as well as all proteins binding to a small single-stranded DNA probe. In both cases, none of the probe was shifted in the gels, indicating the absence of binding (data not shown).

TRF2$^{\Delta B}$ exhibits diminished targeting of the ds/ss telomere overhang junction.

Large model telomeres were constructed from a 3.5 kb telomere-repeat containing vector such that 560 bp of telomeric repeats were positioned at the end of the linearized DNA. A 54 nt overhang was ligated onto the end containing the telomere repeats, while the opposite end was kept blunt. TRF2 and TRF2$^{\Delta B}$ were simultaneously incubated with the DNA and prepared for EM (69).

TRF2 and TRF2$^{\Delta B}$ appeared to have the same activity and overall affinity for the model telomeres, binding (at all positions) 73 ± 11% (TRF2) and 70 ± 12% (TRF2$^{\Delta B}$) of the time, respectively. However, we noticed a difference in the position of each of these proteins on the model telomere template. In experiments containing TRF2, when we considered protein-bound molecules, 50 ± 10% consisted of a particle of TRF2 bound at the end of the model telomere (Figure 5.6A, Figure 5.7) and 34 ± 6% had TRF2 bound within 560 bp of the end of model telomere (presumably within the

telomeric repeats). Also observed were molecules where it appeared that the ss overhang had invaded the duplex repeat to form a t-loop, with 15 ± 2% bound by a molecule of TRF2 at the junction of the invasion site (Figure 5.6B and C). TRF2$^{\Delta B}$ bound to the end of the model telomere in 21 ± 1% of the DNAs, within the telomere repeats 69 ± 1% (Figure 5.6D and insert), and at the junction of a t-loop 3 ± 4% of the time. Aggregates of two or more DNAs bound by a large mass of TRF2 or TRF2$^{\Delta B}$ were observed, but not scored (data not shown). Our results are therefore comparable with previous experiments in our laboratory that were optimized for t-loop formation, where TRF2 bound to t-loop junctions 19% of the time (69). We show therefore that the basic domain of TRF2 is required for it to localize to model telomere ends and to facilitate t-loop formation *in vitro*.

DISCUSSION

In this study we used EM and gel-shift assays to evaluate the binding of several telomeric factors to an array of DNA templates. We found that TRF2 was able to bind to the junctions of replication forks, chickenfoot structures and Holliday junctions, whereas TRF1 and the TRF2 truncation mutant, TRF2$^{\Delta B}$, could not. TRF2 binding to the DNA junctions did not depend on telomere sequences being present in the DNA, suggesting that the binding was not related to the telomere-specific Myb domain. Also, a

Figure 5.6: Diminished t-loop formation and telomere ss/ds junction binding by TRF2$^{\Delta B}$.

(*A, B C*). Discrete complexes of TRF2 bound to the ends of the model telomeres (*A*) or at the strand invasion sites of the t-loop (*B, C*). (*D* and insert). TRF2$^{\Delta B}$ typically bound within the TTAGGG repeat tract but not at the ends of the model telomere. Bar is equivalent to 450 bp in panels showing full-length molecules.

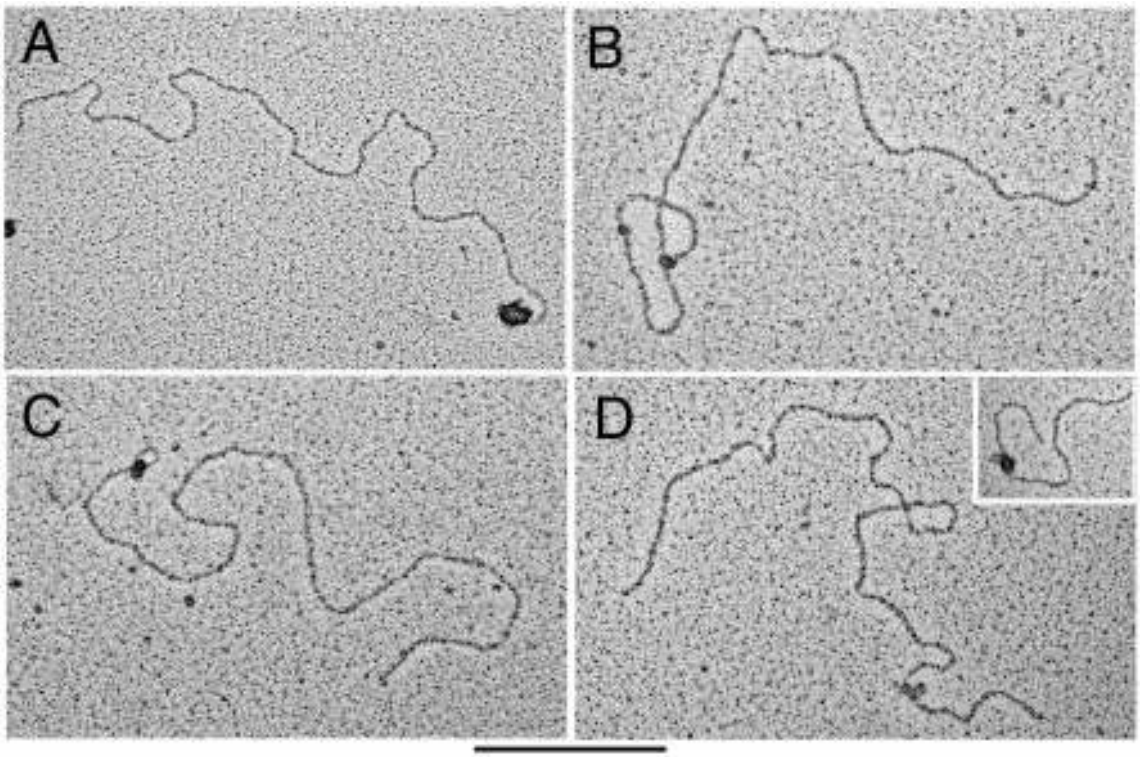
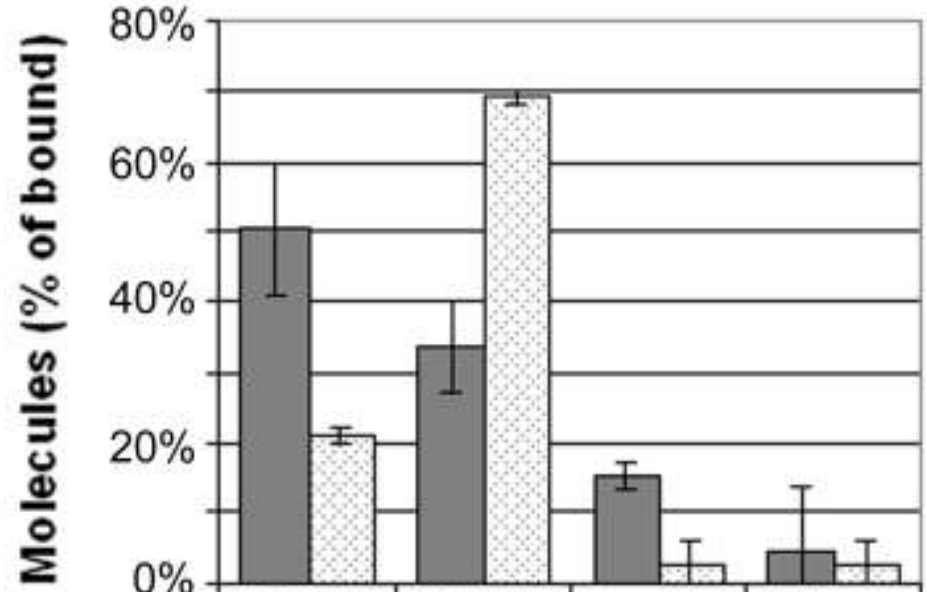

A
B
C
D

Figure 5.7: The basic domain of TRF2 promotes end-binding and t-loop formation on model telomeres in vitro.

Binding of TRF2 and TRF2$^{\Delta B}$ to the large model telomeres was visualized by EM and quantified. Percentages of molecules with protein bound at the DNA end (end-bound), bound within 560 bp from the end (within TTAGGG), bound at the t-loop invasion site (t-loops), or bound more than 560 bp from the DNA end (non-specific) were calculated as a fraction of all protein-bound molecules scored. Data are represented as mean ± SD.


Molecules (% of bound)
80%
60%
40%
20%
0%
End-bound
Within TTAGG
t-loops
Non-specific
TRF2
50%
34%
15%
4%
TRF2ΔB
21%
69%
3%
3%

significant bias for 4-stranded junctions was observed, especially in non-telomeric substrates, suggesting that these are a major target of TRF2 structure-specific binding.

Of particular interest were the different binding affinities of the two peptides used in this study. The peptide encompassing the basic domain of TRF2 bound well to DNA junctions, with a much higher affinity for 4-stranded chickenfoot structures than the 3-stranded replication fork templates. In contrast, while the mutant peptide had the same net charge and amino acid composition of the TRF2 basic domain, binding to 4-stranded DNA junctions was almost completely disrupted. Thus, the two amino acids which we rearranged in the mutant peptide and which are conserved within the TRF2 basic domains of some mammals and birds, must be critical for TRF2 binding to 4-way DNA junctions. Although the mutant peptide showed some DNA binding, this likely reflects an electrostatic interaction between the positively charged peptide and negatively charged DNA.

The relatively small dissociation constants that were calculated for the proteins binding to radiolabeled Holliday junction probes reflected the strong, specific binding of TRF2 and the basic peptide, whereas TRF1, TRF2$^{\Delta B}$, and the mutant peptide all had much higher dissociation constants, reflective of their weaker binding. Also, when TRF2 binding to large model telomeres was examined by EM, we found that in the absence of the basic domain (TRF2$^{\Delta B}$) the ability of TRF2 to target the end of the large model telomeres was

greatly diminished, as was its ability to facilitate t-loop formation.

In gel-shift assays with a small telomeric probe, TRF1, TRF2 and TRF2$^{\Delta B}$ all shifted equally as well, with TRF2 and TRF2$^{\Delta B}$ having an almost identical affinity for the telomere probe. Hence, deleting the basic domain from TRF2 does not appear to destabilize the Myb domain, which remained intact in the mutant. The apparent Kd values also suggest that TRF2 DNA-structure-binding via the basic domain is equally as strong as telomere-DNA-binding via the Myb domain.

The ability of TRF2 to bind telomeric DNA in a sequence-specific manner, but to have structure-dependent recruitment to the telomeric ds/ss junction as well as a possible function that is independent of its Myb-dependent binding, is reminiscent of the multiple DNA-binding properties of p53. p53 has a sequence-specific DNA binding domain as well as an additional domain at its extreme C terminus shown to bind in a sequence nonspecific manner to a wide variety of DNA targets (198,200,212). These targets include unusual DNA structures such as the 4-stranded DNA Holliday junction, the Y-shaped replication fork, and chickenfoot intermediates of fork regression. Intriguingly, this structure-specific binding domain of p53 is also characterized by basic residues, and it has the same length and overall positive charge as the TRF2 basic domain.

Based on these findings we suggest a dual role for the TRF2 basic domain at the telomere. The first role involves a role for TRF2 in t-loop

formation and stabilization. We propose that the junction-binding property of the basic domain enhances the ability of TRF2, in conjunction with other telomere-associated proteins, to target chromosome ends and facilitate t-loop formation. Junction binding capacity would also make it possible for TRF2 to contribute to the stability of the t-loop by binding to junctions formed at the strand invasion site of the t-loop. The second role for TRF2 involving the basic domain would be a critical function during telomere replication in mammals, similar to that of Taz1 in the fission yeast (101). We propose that if replication at the telomere stalls because of an impediment to the polymerase machinery, for instance a G-quartet in the unwound G-rich strand, TRF2 would be able to recognize and bind to the chickenfoot intermediates produced by fork regression, thereby preventing the activity of Holliday junction resolvases. TRF2 can then recruit factors like the RecQ helicases to unwind the chickenfoot structures, allowing replication to restart. Indeed this may explain why TRF2 is able to interact with and stimulate the RecQ helicases, known to be important for proper telomere replication and maintenance in human cells (105,134).

Both of these proposed roles for TRF2 are consistent with the TRF2$^{\Delta B}$ phenotype previously observed, where deletion of the basic domain does not impede the DNA binding activity of TRF2 or its localization to telomeres *in vivo*, yet massive losses of telomeric DNA are seen, accompanied by the appearance of t-circles, a DNA damage response, and induction of senescence. Specifically, Wang et al. (2004) showed that cells

expressing the same level of TRF2$^{\Delta B}$ as control cells, but with impaired function of XRCC3 protein do not show telomere deletions. We have shown in our laboratory that XRCC3 will bind to Holliday junctions as well as chickenfoot structures *in vitro* in complex with Rad51C (unpublished data). These observations suggest that without the basic domain, TRF2 is unable to protect complex DNA structures at the telomere, such as those that have been seen at model telomere replication forks or the junction structure at the t-loop, from being recognized and processed by DNA repair proteins, including XRCC3. Nevertheless, because TRF2$^{\Delta B}$ is still able to suppress non-homologous end joining (NHEJ) events, we believe that the loss of junction binding ability does not directly lead to telomere uncapping, characterized by de-protection and loss of the 3′ overhang, as seen in cells expressing the dominant-negative allele of TRF2 (TRF2$^{\Delta B\Delta M}$) (70).

We previously observed that psoralen cross-linking stabilized t-loop formation *in vitro* on DNAs with very short overhangs, suggesting that more than just the nucleotides of the ss tail are inserted into the duplex to form the D-loop (69). Based on our findings, we would further suggest that the more energetically favorable structure at the strand invasion site is a 4-way junction (preferred by TRF2 to the 3-way junction) formed by branch migration of the invasion site. The data are also the first direct demonstration of TRF2 binding specifically and selectively to non-telomeric DNA *in vitro*, and may explain the ability of TRF2 to localize to sites of DNA damage in irradiated human

fibroblasts (208).

CHAPTER 6: CONCLUSIONS AND FINAL THOUGHTS

The primary goal of this study has been to examine some of the key components of the telomere by using electron microscopy and other biochemical techniques, in order to gain insight about the mechanisms in which they function. In this regard, I believe that we have been overwhelmingly successful.

Our examination of *Euplotes aediculatus* telomerase binding to telomeric DNA is the first visualization of telomerase from any source bound to a model telomere DNA. We show that affinity-purified telomerase exists as a dimer in solution, that it binds to a model telomere as a homodimer and further that it has a strong propensity to bring together two model telomere DNAs in a complex which contains two DNAs, and at least four telomerase monomers. The work will thus be of great interest to the field of telomeres and telomerase, as the possibility that telomerase binds to its substrate as a dimer has been proposed, but no direct data has been presented to support this model.

In the study of replication fork regression in the two types of repetitive DNAs: the telomeric repeat $TTAGGG_n$ and the disease-associated triplet repeat $(CTG)\cdot(CAG)_n$, we describe a new feature of repetitive DNA that could

present a significant barrier to replication. That is, the spontaneous and unassisted regression of model stalled replication forks consisting of repetitive sequences, to form stable chickenfoot structures that resemble recombination intermediates. The work will be of great interest to two fields interested in repetitive DNA. In the field of telomeres, little is known about replication of telomeric DNA. Our data offer a possible explanation for the puzzling observation that helicases such as BLM and WRN are required for efficient replication through telomeres. The data will also be of great interest to the field of triplet repeat diseases, where it supports a model for expansion of disease-related repeats that involves recombination-mediated replication restart.

In our examination of TRF2 binding to DNA templates, we discovered that TRF2 has a high affinity for binding to DNA junctions, irrespective of the presence of telomeric repeats, and that this is facilitated by the basic domain of TRF2. This work will be of great interest to the field of telomeres for several key reasons. TRF2 has been shown to be central to telomere protection and loss of its function leads to telomere fusions, apoptosis, and cell death. Further, loss of the basic domain is also highly deleterious to the cell but why this is so has not been discovered in spite of many detailed studies by the de Lange laboratory and others. We feel that the work presented in chapter 5 provides critical insights into this question and also opens the door for considering a panoply of additional functions for TRF2 in the cell. Finally the data suggests a novel role for TRF2 at telomeric

replication forks, and it is the first direct demonstration of TRF2 binding selectively to non-telomeric DNA.

When considering future directions of the work presented, several ideas come to mind. In the study of *Euplotes* telomerase binding to model telomeric DNA, it would be interesting to determine whether the RNA component was involved in telomerase dimer formation. Functional dimerization of the RNA component of human telomerase (hTR) has previously been shown to be dependent on the catalytically essential pseudoknot and CR4/CR5 domains (145,213,214). Oligonucleotide probes that selectively base pair with conserved regions within the *Ea*TR could thus be used to block potential RNA-RNA base pairing interactions, and complexes could then be visualized by EM to evaluate whether dimer formation has been disrupted. This oligonucleotide approach has been used previously to identify an RNA-RNA dimerization site in hTR, as well as to locate RNA recognition elements in HIV genomic RNA and tRNA (213,215-217).

Another experiment which we originally attempted, but which proved to be more difficult than expected, was the examination of *E. aediculatus* telomerase binding to the model telomeres under elongation conditions (i.e. in the presence of dTTP and dGTP). This would still be an interesting experiment to do, however the conditions for binding would have to be

refined, to avoid obtaining the large aggregates of telomerase and DNA that were previously seen.

In the study of replication fork regression in repetitive DNAs, only two types of repeats were studied, and only in one orientation each. This study could therefore be expanded to include more of the disease-associated repeats as well as an examination of the orientation dependence (or lack thereof) of the spontaneous fork regression that was seen. Furthermore, as previously discussed, a hallmark of the triplet repeats is the appearance of disease pathology when the repeat blocks expand beyond certain tight length thresholds (178). We have proposed that the resolution of the chickenfoot intermediates of fork regression could lead to this expansion, possibly through the generation of extrachromosomal repeat DNA. In support of this idea, ALT cells and some yeast mutants lacking telomerase are able to use recombination to generate very long telomeres, presumably through rolling circle replication of telomeric repeat-containing circles (51-53,55-57). These t-circles are thought to originate from aberrant homologous recombination at the t-loop junction. However, one cannot rule out that another source of these circles may be DNA cleaved out during aberrant recombination-mediated replication restart of stalled forks at the telomere. Indeed, extrachromosomal linear telomeric DNA (that can easily be ligated to form t-circles) is found in ALT-associated PML bodies (APBs) (187). Therefore, it would be interesting to determine whether repeat-containing DNA circles are also present in cells

cultured from individuals that have repeat-associated diseases.

TRF2 and p53 have very similar characteristics in that they to bind DNA in a sequence-specific manner with one domain, and to DNA structures such as 4-stranded DNA Holliday junctions or chickenfoot structures with another domain (198,200,212). In TRF2, the structure-specific binding is dependent on the N-terminal basic domain, whereas in p53, it is a C-terminal basic domain. Interestingly, both of these domains are characterized by the same length and overall positive charge. Circular dichroism spectroscopy has showed that both domains assume unordered structures in aqueous solutions, and the p53 basic domain has been shown to adopt a helical structure upon binding to its substrate (218) and data not shown). Similar structural data of the conformation adopted by the TRF2 basic domain when bound to DNA has not been found, although it has a medium-to-high helical propensity when modeled on a helical wheel (data not shown). Thus, it would be interesting to determine whether the p53 basic domain, when fused on to the N-terminus of the TRF2$^{\Delta B}$ protein (albeit with a sufficiently long linker region to allow for correct folding) could rescue the function of this mutant protein in the cell, thereby preventing aberrant t-loop HR, a DNA damage response, and senescence.

REFERENCES

1. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3015-3019

2. Henderson, E. R., and Blackburn, E. H. (1989) *Mol. Cell. Biol.* **9**, 345-348

3. Moyzis, R. K., Buckingham, J. M., Cram, L. S., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L., and Wu, J. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 6622-6626

4. Meyne, J., Ratliff, R. L., and Moyzis, R. K. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 7049-7053

5. Chai, W., Du, Q., Shay, J. W., and Wright, W. E. (2006) *Mol. Cell* **21**, 427-435

6. Huffman, K. E., Levene, S. D., Tesmer, V. M., Shay, J. W., and Wright, W. E. (2000) *J. Biol. Chem.* **275**, 19719-19722

7. Makarov, V. L., Hirose, Y., and Langmore, J. P. (1997) *Cell* **88**, 657-666

8. McElligott, R., and Wellinger, R. J. (1997) *EMBO J.* **16**, 3705-3714

9. Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D., and Shay, J. W. (1997) *Genes Dev.* **11**, 2801-2809

10. Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., and de Lange, T. (1999) *Cell* **97**, 503-514

11. Nikitina, T., and Woodcock, C. L. (2004) *J. Cell. Biol.* **166**, 161-165

12. Cesare, A. J., Quinney, N., Willcox, S., Subramanian, D., and Griffith, J. D. (2003) *Plant J* **36**, 271-279

13. Groff-Vindman, C., Cesare, A. J., Natarajan, S., Griffith, J. D., and McEachern, M. J. (2005) *Mol. Cell. Biol.* **25**, 4406-4412

14. Munoz-Jordan, J. L., Cross, G. A., de Lange, T., and Griffith, J. D. (2001) *EMBO J.* **20**, 579-588

15. Murti, K. G., and Prescott, D. M. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14436-14439

16. de Lange, T. (2005) *Genes Dev.* **19**, 2100-2110

17. Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., and de Lange, T. (1995) *Science* **270**, 1663-1667

18. Bilaud, T., Brun, C., Ancelin, K., Koering, C. E., Laroche, T., and Gilson, E. (1997) *Nat. Genet.* **17**, 236-239

19. Bianchi, A., and de Lange, T. (1999) *J. Biol. Chem.* **274**, 21223-21227

20. Zhu, X. D., Kuster, B., Mann, M., Petrini, J. H., and de Lange, T. (2000) *Nat. Genet.* **25**, 347-352

21. Rossetti, L., Cacchione, S., Fua, M., and Savino, M. (1998) *Biochemistry* **37**, 6727-6737

22. Olovnikov, A. M. (1973) *J. Theor. Biol.* **41**, 181-190

23. Watson, J. D. (1972) *Nat. New Biol.* **239**, 197-201

24. Lydall, D. (2003) *J. Cell Sci.* **116**, 4057-4065

25. von Zglinicki, T. (2002) *Trends Biochem. Sci.* **27**, 339-344

26. von Zglinicki, T., Pilger, R., and Sitte, N. (2000) *Free Radic. Biol. Med.* **28**, 64-74

27. Wellinger, R. J., Ethier, K., Labrecque, P., and Zakian, V. A. (1996) *Cell* **85**, 423-433

28. Sfeir, A. J., Chai, W., Shay, J. W., and Wright, W. E. (2005) *Mol. Cell* **18**, 131-138

29. Greider, C. W., and Blackburn, E. H. (1987) *Cell* **51**, 887-898

30. Diede, S. J., and Gottschling, D. E. (1999) *Cell* **99**, 723-733

31. von Zglinicki, T., Petrie, J., and Kirkwood, T. B. (2003) *Nat. Biotechnol.* **21**, 229-230

32. Shelton, D. N., Chang, E., Whittier, P. S., Choi, D., and Funk, W. D. (1999) *Curr. Biol.* **9**, 939-945

33. Wright, W. E., and Shay, J. W. (1992) *Exp. Gerontol.* **27**, 383-389

34. Bodnar, A. G., Ouellette, M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., Harley, C. B., Shay, J. W., Lichtsteiner, S., and Wright, W. E. (1998) *Science* **279**, 349-352

35. Halvorsen, T. L., Leibowitz, G., and Levine, F. (1999) *Mol. Cell. Biol.* **19**, 1864-1870

36. Steinert, S., Shay, J. W., and Wright, W. E. (2000) *Biochem. Biophys. Res. Commun.* **273**, 1095-1098

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

38. Shippen-Lentz, D., and Blackburn, E. H. (1990) *Science* **247**, 546-552

39. Lin, J. J., and Zakian, V. A. (1995) *Cell* **81**, 1127-1135

40. 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

41. Singer, M. S., and Gottschling, D. E. (1994) *Science* **266**, 404-409

42. Feng, J., Funk, W. D., Wang, S. S., Weinrich, S. L., Avilion, A. A., Chiu, C. P., Adams, R. R., Chang, E., Allsopp, R. C., Yu, J., and et al. (1995) *Science* **269**, 1236-1241

43. Greider, C. W., and Blackburn, E. H. (1989) *Nature* **337**, 331-337

44. Zappulla, D. C., and Cech, T. R. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10024-10029

45. Zappulla, D. C., Goodrich, K., and Cech, T. R. (2005) *Nat. Struct. Mol. Biol.* **12**, 1072-1077

46. Smogorzewska, A., and de Lange, T. (2004) *Annu. Rev. Biochem.* **73**, 177-208

47. Loayza, D., and De Lange, T. (2003) *Nature* **423**, 1013-1018

48. Kelleher, C., Kurth, I., and Lingner, J. (2005) *Mol. Cell. Biol.* **25**, 808-818

49. van Steensel, B., and de Lange, T. (1997) *Nature* **385**, 740-743

50. Galati, A., Rossetti, L., Pisano, S., Chapman, L., Rhodes, D., Savino, M., and Cacchione, S. (2006) *J. Mol. Biol.* **360**, 377-385

51. Murnane, J. P., Sabatier, L., Marder, B. A., and Morgan, W. F. (1994) *EMBO J.* **13**, 4953-4962

52. Dunham, M. A., Neumann, A. A., Fasching, C. L., and Reddel, R. R. (2000) *Nat. Genet.* **26**, 447-450

53. Bechter, O. E., Zou, Y., Walker, W., Wright, W. E., and Shay, J. W. (2004) *Cancer Res.* **64**, 3444-3451

54. Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A., and Reddel, R. R. (1997) *Nat. Med.* **3**, 1271-1274

55. Natarajan, S., and McEachern, M. J. (2002) *Mol. Cell. Biol.* **22**, 4512-4521

56. Bailey, S. M., Brenneman, M. A., and Goodwin, E. H. (2004) *Nucleic Acids Res.* **32**, 3743-3751

57. Londono-Vallejo, J. A., Der-Sarkissian, H., Cazes, L., Bacchetti, S., and Reddel, R. R. (2004) *Cancer Res.* **64**, 2324-2327

58. Cesare, A. J., and Griffith, J. D. (2004) *Mol. Cell. Biol.* **24**, 9948-9957

59. Wang, R. C., Smogorzewska, A., and de Lange, T. (2004) *Cell* **119**, 355-368

60. Drayton, S., and Peters, G. (2002) *Curr. Opin. Genet. Dev.* **12**, 98-104

61. Shay, J. W., and Wright, W. E. (2005) *Carcinogenesis* **26**, 867-874

62. Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E., and de Lange, T. (2005) *EMBO J.* **24**, 2667-2678

63. Fan, X., and Price, C. M. (1997) *Mol. Biol. Cell* **8**, 2145-2155

64. Jacob, N. K., Skopp, R., and Price, C. M. (2001) *EMBO J.* **20**, 4299-4308

65. Jacob, N. K., Kirk, K. E., and Price, C. M. (2003) *Mol. Cell* **11**, 1021-1032

66. Wu, L., Multani, A. S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J. M., Bachilo, O., Pathak, S., Tahara, H., Bailey, S. M., Behringer, R. R., and Chang, S. (2006) *Cell* **126**, 49-62

67. Hockemeyer, D., Daniels, J. P., Takai, H., and de Lange, T. (2006) *Cell* **126**, 63-77

68. Churikov, D., Wei, C., and Price, C. M. (2006) *Mol. Cell. Biol.* **26**, 6971-6982

69. Stansel, R. M., de Lange, T., and Griffith, J. D. (2001) *EMBO J.* **20**, 5532-5540

70. van Steensel, B., Smogorzewska, A., and de Lange, T. (1998) *Cell* **92**, 401-413

71. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S., and de Lange, T. (1999) *Science* **283**, 1321-1325

72. Zhu, X. D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H., and de Lange, T. (2003) *Mol. Cell* **12**, 1489-1498

73. Celli, G. B., and de Lange, T. (2005) *Nat. Cell Biol.* **7**, 712-718

74. Bunch, J. T., Bae, N. S., Leonardi, J., and Baumann, P. (2005) *Mol. Cell. Biol.* **25**, 5567-5578

75. Smogorzewska, A., and de Lange, T. (2002) *EMBO J.* **21**, 4338-4348

76. Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A., and de Lange, T. (2002) *Curr. Biol.* **12**, 1635-1644

77. Carson, C. T., Schwartz, R. A., Stracker, T. H., Lilley, C. E., Lee, D. V., and Weitzman, M. D. (2003) *EMBO J.* **22**, 6610-6620

78. Petrini, J. H., and Stracker, T. H. (2003) *Trends Cell. Biol.* **13**, 458-462

79. Uziel, T., Lerenthal, Y., Moyal, L., Andegeko, Y., Mittelman, L., and Shiloh, Y. (2003) *EMBO J.* **22**, 5612-5621

80. Karlseder, J., Hoke, K., Mirzoeva, O. K., Bakkenist, C., Kastan, M. B., Petrini, J. H., and de Lange, T. (2004) *PLoS Biol.* **2**, E240

81. Takai, H., Smogorzewska, A., and de Lange, T. (2003) *Curr. Biol.* **13**, 1549-1556

82. Takata, H., Kanoh, Y., Gunge, N., Shirahige, K., and Matsuura, A. (2004) *Mol. Cell* **14**, 515-522

83. Takata, H., Tanaka, Y., and Matsuura, A. (2005) *Mol. Cell* **17**, 573-583

84. Verdun, R. E., Crabbe, L., Haggblom, C., and Karlseder, J. (2005) *Mol. Cell* **20**, 551-561

85. Zou, L., and Elledge, S. J. (2003) *Science* **300**, 1542-1548

86. Zhou, B. B., and Elledge, S. J. (2000) *Nature* **408**, 433-439

87. Abraham, R. T. (2001) *Genes Dev.* **15**, 2177-2196

88. Smith, J., Zou, H., and Rothstein, R. (2000) *Biochimie* **82**, 71-78

89. Schramke, V., Luciano, P., Brevet, V., Guillot, S., Corda, Y., Longhese, M. P., Gilson, E., and Geli, V. (2004) *Nat. Genet.* **36**, 46-54

90. Ono, Y., Tomita, K., Matsuura, A., Nakagawa, T., Masukata, H., Uritani, M., Ushimaru, T., and Ueno, M. (2003) *Nucleic Acids Res.* **31**, 7141-7149

91. Yeager, T. R., Neumann, A. A., Englezou, A., Huschtscha, L. I., Noble, J. R., and Reddel, R. R. (1999) *Cancer Res.* **59**, 4175-4179

92. Wu, G., Lee, W. H., and Chen, P. L. (2000) *J. Biol. Chem.* **275**, 30618-30622

93. Ding, H., Schertzer, M., Wu, X., Gertsenstein, M., Selig, S., Kammori, M., Pourvali, R., Poon, S., Vulto, I., Chavez, E., Tam, P. P., Nagy, A., and Lansdorp, P. M. (2004) *Cell* **117**, 873-886

94. Crabbe, L., Verdun, R. E., Haggblom, C. I., and Karlseder, J. (2004) *Science* **306**, 1951-1953

95. Greider, C. W. (1996) *Annu. Rev. Biochem.* **65**, 337-365

96. Henle, E. S., Han, Z., Tang, N., Rai, P., Luo, Y., and Linn, S. (1999) *J. Biol. Chem.* **274**, 962-971

97. Wang, Y., and Patel, D. J. (1993) *Structure* **1**, 263-282

98. Phan, A. T., Gueron, M., and Leroy, J. L. (2000) *J. Mol. Biol.* **299**, 123-144

99. Karlseder, J. (2006) *Nat. Struct. Mol. Biol.* **13**, 386-387

100. Ten Hagen, K. G., Gilbert, D. M., Willard, H. F., and Cohen, S. N. (1990) *Mol. Cell. Biol.* **10**, 6348-6355

101. Miller, K. M., Rog, O., and Cooper, J. P. (2006) *Nature* **440**, 824-828

102. Lillard-Wetherell, K., Machwe, A., Langland, G. T., Combs, K. A., Behbehani, G. K., Schonberg, S. A., German, J., Turchi, J. J., Orren, D. K., and Groden, J. (2004) *Hum. Mol. Genet.* **13**, 1919-1932

103. Bai, Y., and Murnane, J. P. (2003) *Hum. Genet.* **113**, 337-347

104. Schulz, V. P., Zakian, V. A., Ogburn, C. E., McKay, J., Jarzebowicz, A. A., Edland, S. D., and Martin, G. M. (1996) *Hum. Genet.* **97**, 750-754

105. Opresko, P. L., von Kobbe, C., Laine, J. P., Harrigan, J., Hickson, I. D., and Bohr, V. A. (2002) *J. Biol. Chem.* **277**, 41110-41119

106. Hickson, I. D. (2003) *Nat. Rev. Cancer* **3**, 169-178

107. Yang, Q., Zhang, R., Wang, X. W., Spillare, E. A., Linke, S. P., Subramanian, D., Griffith, J. D., Li, J. L., Hickson, I. D., Shen, J. C., Loeb, L. A., Mazur, S. J., Appella, E., Brosh, R. M., Jr., Karmakar, P., Bohr, V. A., and Harris, C. C. (2002) *J. Biol. Chem.* **277**, 31980-31987

108. Constantinou, A., Tarsounas, M., Karow, J. K., Brosh, R. M., Bohr, V. A., Hickson, I. D., and West, S. C. (2000) *EMBO Rep* **1**, 80-84

109. Karow, J. K., Constantinou, A., Li, J. L., West, S. C., and Hickson, I. D. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6504-6508

110. Opresko, P. L., Otterlei, M., Graakjaer, J., Bruheim, P., Dawut, L., Kolvraa, S., May, A., Seidman, M. M., and Bohr, V. A. (2004) *Mol. Cell* **14**, 763-774

111. Mohaghegh, P., Karow, J. K., Brosh Jr, R. M., Jr., Bohr, V. A., and Hickson, I. D. (2001) *Nucleic Acids Res.* **29**, 2843-2849

112. Opresko, P. L., Cheng, W. H., von Kobbe, C., Harrigan, J. A., and Bohr, V. A. (2003) *Carcinogenesis* **24**, 791-802

113. Sun, H., Karow, J. K., Hickson, I. D., and Maizels, N. (1998) *J. Biol. Chem.* **273**, 27587-27592

114. Huber, M. D., Lee, D. C., and Maizels, N. (2002) *Nucleic Acids Res.* **30**, 3954-3961

115. Schulz, V. P., and Zakian, V. A. (1994) *Cell* **76**, 145-155

116. Ivessa, A. S., Zhou, J. Q., Schulz, V. P., Monson, E. K., and Zakian, V. A. (2002) *Genes Dev.* **16**, 1383-1396

117. Choe, W., Budd, M., Imamura, O., Hoopes, L., and Campbell, J. L. (2002) *Mol. Cell. Biol.* **22**, 4202-4217

118. Branzei, D., and Foiani, M. (2005) *Curr. Opin. Cell Biol.* **17**, 568-575

119. Wang, X., Ira, G., Tercero, J. A., Holmes, A. M., Diffley, J. F., and Haber, J. E. (2004) *Mol. Cell. Biol.* **24**, 6891-6899

120. Tarsounas, M., Munoz, P., Claas, A., Smiraldo, P. G., Pittman, D. L., Blasco, M. A., and West, S. C. (2004) *Cell* **117**, 337-347

121. Thacker, J. (1999) *Trends Genet.* **15**, 166-168

122. Liu, Y., Masson, J. Y., Shah, R., O'Regan, P., and West, S. C. (2004) *Science* **303**, 243-246

123. Lopes, M., Cotta-Ramusino, C., Liberi, G., and Foiani, M. (2003) *Mol. Cell* **12**, 1499-1510

124. de Jager, M., van Noort, J., van Gent, D. C., Dekker, C., Kanaar, R., and Wyman, C. (2001) *Mol. Cell* **8**, 1129-1135

125. Costanzo, V., Paull, T., Gottesman, M., and Gautier, J. (2004) *PLoS Biol.* **2**, E110

126. Hopfner, K. P., Karcher, A., Craig, L., Woo, T. T., Carney, J. P., and Tainer, J. A. (2001) *Cell* **105**, 473-485

127. Trenz, K., Smith, E., Smith, S., and Costanzo, V. (2006) *EMBO J.* **25**, 1764-1774

128. Craven, R. J., Greenwell, P. W., Dominska, M., and Petes, T. D. (2002) *Genetics* **161**, 493-507

129. Matsuura, A., Naito, T., and Ishikawa, F. (1999) *Genetics* **152**, 1501-1512

130. (1998) *Genetic Instabilities and Hereditary Neurological Diseases* (Wells, R. D., and Warren, S. T., Eds.), Academic Press

131. Kim, S. H., Pytlos, M. J., and Sinden, R. R. (2006) *Mutat. Res.* **595**, 5-22

132. West, S. C. (2003) *Nat. Rev. Mol. Cell. Biol.* **4**, 435-445

133. Opresko, P. L., Mason, P. A., Podell, E. R., Lei, M., Hickson, I. D., Cech, T. R., and Bohr, V. A. (2005) *J. Biol. Chem.* **280**, 32069-32080

134. Machwe, A., Xiao, L., and Orren, D. K. (2004) *Oncogene* **23**, 149-156

135. Yang, Q., Zheng, Y. L., and Harris, C. C. (2005) *Mol. Cell. Biol.* **25**, 1070-1080

136. Constantinou, A., Chen, X. B., McGowan, C. H., and West, S. C. (2002) *EMBO J.* **21**, 5577-5585

137. Fouche, N., Moon, I. K., Keppler, B. R., Griffith, J. D., and Jarstfer, M. B. (2006) *Biochemistry* **45**, 9624-9631

138. Greider, C. W., and Blackburn, E. H. (1985) *Cell* **43**, 405-413

139. Greider, C. W. (1991) *Mol. Cell. Biol.* **11**, 4572-4580

140. Lue, N. F. (2004) *Bioessays* **26**, 955-962

141. Peng, Y., Mian, I. S., and Lue, N. F. (2001) *Mol. Cell* **7**, 1201-1211

142. Harrington, L. A., and Greider, C. W. (1991) *Nature* **353**, 451-454

143. Collins, K. (1999) *Annu. Rev. Biochem.* **68**, 187-218

144. Beattie, T. L., Zhou, W., Robinson, M. O., and Harrington, L. (2001) *Mol. Cell. Biol.* **21**, 6151-6160

145. Wenz, C., Enenkel, B., Amacker, M., Kelleher, C., Damm, K., and Lingner, J. (2001) *EMBO J.* **20**, 3526-3534

146. Moriarty, T. J., Marie-Egyptienne, D. T., and Autexier, C. (2004) *Mol. Cell. Biol.* **24**, 3720-3733

147. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2003) *Mol. Biol. Cell* **14**, 4794-4804

148. Prescott, J., and Blackburn, E. H. (1997) *Genes Dev.* **11**, 2790-2800

149. Lingner, J., and Cech, T. R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 10712-10717

150. Aigner, S., Postberg, J., Lipps, H. J., and Cech, T. R. (2003) *Biochemistry* **42**, 5736-5747

151. Wang, L., Dean, S. R., and Shippen, D. E. (2002) *Nucleic Acids Res.* **30**, 4032-4039

152. Kelleher, C., Teixeira, M. T., Forstemann, K., and Lingner, J. (2002) *Trends Biochem. Sci.* **27**, 572-579

153. Prescott, D. M. (1994) *Microbiol. Rev.* **58**, 233-267

154. Aigner, S., Lingner, J., Goodrich, K. J., Grosshans, C. A., Shevchenko, A., Mann, M., and Cech, T. R. (2000) *EMBO J.* **19**, 6230-6239

155. Lingner, J., Hendrick, L. L., and Cech, T. R. (1994) *Genes Dev.* **8**, 1984-1998

156. Swanton, M. T., Heumann, J. M., and Prescott, D. M. (1980) *Chromosoma* **77**, 217-227

157. Jarstfer, M. B., and Cech, T. R. (2002) *Biochemistry* **41**, 151-161

158. Griffith, J. D., and Christiansen, G. (1978) *Annu. Rev. Biophys. Bioeng.* **7**, 19-35

159. Hammond, P. W., Lively, T. N., and Cech, T. R. (1997) *Mol. Cell. Biol.* **17**, 296-308

160. Hammond, P. W., and Cech, T. R. (1997) *Nucleic Acids Res.* **25**, 3698-3704

161. Hammond, P. W., and Cech, T. R. (1998) *Biochemistry* **37**, 5162-5172

162. Griffith, J., Bianchi, A., and de Lange, T. (1998) *J. Mol. Biol.* **278**, 79-88

163. Allen, D. J., Makhov, A., Grilley, M., Taylor, J., Thresher, R., Modrich, P., and Griffith, J. D. (1997) *EMBO J.* **16**, 4467-4476

164. Rivera, M. A., and Blackburn, E. H. (2004) *J. Biol. Chem.* **279**, 53770-53781

165. Dandjinou, A. T., Levesque, N., Larose, S., Lucier, J. F., Abou Elela, S., and Wellinger, R. J. (2004) *Curr. Biol.* **14**, 1148-1158

166. Tzfati, Y., Knight, Z., Roy, J., and Blackburn, E. H. (2003) *Genes Dev.* **17**, 1779-1788

167. Zuker, M. (2003) *Nucleic Acids Res.* **31**, 3406-3415

168. Seto, A. G., Livengood, A. J., Tzfati, Y., Blackburn, E. H., and Cech, T. R. (2002) *Genes Dev.* **16**, 2800-2812

169. Peterson, S. E., Stellwagen, A. E., Diede, S. J., Singer, M. S., Haimberger, Z. W., Johnson, C. O., Tzoneva, M., and Gottschling, D. E. (2001) *Nat. Genet.* **27**, 64-67

170. Seto, A. G., Zaug, A. J., Sobel, S. G., Wolin, S. L., and Cech, T. R. (1999) *Nature* **401**, 177-180

171. Chastain, P. D., 2nd, Bowers, J. L., Lee, D. G., Bell, S. P., and Griffith, J. D. (2004) *J. Biol. Chem.* **279**, 36354-36362

172. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science* **289**, 905-920

173. Chapon, C., Cech, T. R., and Zaug, A. J. (1997) *RNA* **3**, 1337-1351

174. Bosoy, D., Peng, Y., Mian, I. S., and Lue, N. F. (2003) *J. Biol. Chem.* **278**, 3882-3890

175. Nakamura, T. M., Wang, Y. H., Zaug, A. J., Griffith, J. D., and Cech, T. R. (1995) *EMBO J.* **14**, 4849-4859

176. Wang, Y. H., Howard, M. T., and Griffith, J. D. (1991) *Biochemistry* **30**, 5443-5449

177. Hancock, J. M. (1996) *Bioessays* **18**, 421-425

178. Wells, R. D., and Warren, S. T. (1998) *Genetic Instabilities and Hereditary Neurological Diseases*, Academic Press

179. Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., and Ranum, L. P. (2001) *Science* **293**, 864-867

180. Matsuura, T., Yamagata, T., Burgess, D. L., Rasmussen, A., Grewal, R. P., Watase, K., Khajavi, M., McCall, A. E., Davis, C. F., Zu, L., Achari, M., Pulst, S. M., Alonso, E., Noebels, J. L., Nelson, D. L., Zoghbi, H. Y., and Ashizawa, T. (2000) *Nat. Genet.* **26**, 191-194

181. Lalioti, M. D., Scott, H. S., Buresi, C., Rossier, C., Bottani, A., Morris, M. A., Malafosse, A., and Antonarakis, S. E. (1997) *Nature* **386**, 847-851

182. Handa, V., Yeh, H. J., McPhie, P., and Usdin, K. (2005) *J. Biol. Chem.* **280**, 29340-29345

183. Mirkin, S. M. (2006) *Curr. Opin. Struct. Biol.* **16**, 351-358

184. Blackburn, E. H. (2005) *FEBS Lett.* **579**, 859-862

185. de Lange, T. (2002) *Oncogene* **21**, 532-540

186. Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S., and Reddel, R. R. (1995) *EMBO J.* **14**, 4240-4248

187. Henson, J. D., Neumann, A. A., Yeager, T. R., and Reddel, R. R. (2002) *Oncogene* **21**, 598-610

188. Wang, G., and Vasquez, K. M. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13448-13453

189. Wang, G., Christensen, L. A., and Vasquez, K. M. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2677-2682

190. Lin, Y., Dion, V., and Wilson, J. H. (2006) *Nat. Struct. Mol. Biol.* **13**, 179-180

191. Bacolla, A., Jaworski, A., Larson, J. E., Jakupciak, J. P., Chuzhanova, N., Abeysinghe, S. S., O'Connell, C. D., Cooper, D. N., and Wells, R. D. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101**, 14162-14167

192. Krasilnikova, M. M., and Mirkin, S. M. (2004) *Mol. Cell. Biol.* **24**, 2286-2295

193. Hile, S. E., and Eckert, K. A. (2004) *J. Mol. Biol.* **335**, 745-759

194. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M., and Son, L. S. (2005) *Nucleic Acids Res.* **33**, 3785-3798

195. Hashem, V. I., Rosche, W. A., and Sinden, R. R. (2004) *Mutat. Res.* **554**, 95-109

196. Ohki, R., and Ishikawa, F. (2004) *Nucleic Acids Res.* **32**, 1627-1637

197. Fry, M., and Loeb, L. A. (1999) *J. Biol. Chem.* **274**, 12797-12802

198. Subramanian, D., and Griffith, J. D. (2005) *J. Biol. Chem.* **280**, 42568-42572

199. Murphy, F. L., Wang, Y. H., Griffith, J. D., and Cech, T. R. (1994) *Science* **265**, 1709-1712

200. Lee, S., Cavallo, L., and Griffith, J. (1997) *J. Biol. Chem.* **272**, 7532-7539

201. Pelletier, R., Krasilnikova, M. M., Samadashwily, G. M., Lahue, R., and Mirkin, S. M. (2003) *Mol. Cell. Biol.* **23**, 1349-1357

202. Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997) *Nat. Genet.* **17**, 298-304

203. Jonsson, F., Postberg, J., Schaffitzel, C., and Lipps, H. J. (2002) *Chromosome Res.* **10**, 445-453

204. Wright, W. E., Tesmer, V. M., Liao, M. L., and Shay, J. W. (1999) *Exp. Cell. Res.* **251**, 492-499

205. Hultdin, M., Gronlund, E., Norrback, K. F., Just, T., Taneja, K., and Roos, G. (2001) *Exp. Cell. Res.* **271**, 223-229

206. Lansdorp, P. M. (2005) *Trends Biochem. Sci.* **30**, 388-395

207. Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997) *Nat. Genet.* **17**, 231-235

208. Bradshaw, P. S., Stavropoulos, D. J., and Meyn, M. S. (2005) *Nat. Genet.* **37**, 193-197

209. Alani, E., Lee, S., Kane, M. F., Griffith, J., and Kolodner, R. D. (1997) *J. Mol. Biol.* **265**, 289-301

210. Bianchi, A., Smith, S., Chong, L., Elias, P., and de Lange, T. (1997) *EMBO J.* **16**, 1785-1794

211. Wu, L., Bayle, J. H., Elenbaas, B., Pavletich, N. P., and Levine, A. J. (1995) *Mol. Cell. Biol.* **15**, 497-504

212. Lee, S., Elenbaas, B., Levine, A., and Griffith, J. (1995) *Cell* **81**, 1013-1020

213. Ren, X., Gavory, G., Li, H., Ying, L., Klenerman, D., and Balasubramanian, S. (2003) *Nucleic Acids Res.* **31**, 6509-6515

214. Ly, H., Xu, L., Rivera, M. A., Parslow, T. G., and Blackburn, E. H. (2003) *Genes Dev.* **17**, 1078-1083

215. Lanchy, J. M., Rentz, C. A., Ivanovitch, J. D., and Lodmell, J. S. (2003) *Biochemistry* **42**, 2634-2642

216. Monie, T., Greatorex, J., and Lever, A. M. (2001) *Virus Res.* **78**, 45-56

217. Wittenhagen, L. M., and Kelley, S. O. (2002) *Nat. Struct. Biol.* **9**, 586-590

218. Hoffmann, R., Craik, D. J., Pierens, G., Bolger, R. E., and Otvos, L., Jr. (1998) *Biochemistry* **37**, 13755-13764