THE ROLE OF SHELTERIN COMPONENTS ON THE STRUCTURAL DYNAMICS OF TELOMERIC DNA

NEZAHAT ÖZLEM ARAT

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

Chapel Hill
2013

Approved by:

Jack D. Griffith, Ph. D.

Dale A. Ramsden, Ph. D.

Jean G. Cook, Ph. D.

Shawn Ahmed, Ph. D.

Michael B. Jarstfer, Ph. D.

ABSTRACT

NEZAHAT ÖZLEM ARAT: THE ROLE OF SHELTERIN COMPONENTS ON THE
STRUCTURAL DYNAMICS OF TELOMERIC DNA
(Under the direction of Jack D. Griffith)

Telomeres are the nucleoprotein caps found at chromosome ends. They help to
differentiate the chromosome ends from DNA double strand breaks and define the molecular
clock of a cell. The molecular clock determines the number of times a cell can divide and
when telomeres significantly shorten, cellular senescence or growth arrest occurs. Cancer
cells reverse this process either by activating telomerase or by utilizing Alternative
Lengthening of Telomeres (ALT). Suppression of ALT, regulation of telomerase, and end
capping are all performed by the shelterin complex. The shelterin complex is composed of
six proteins; TRF1, TRF2, Rap1, TIN2, TPP1 and Pot1. Rap1, TRF2 and Pot1 are important
for the suppression of ALT and regulate the topology of the telomeric DNA. Telomeric DNA
is composed of G-rich tandem repeats and this unique character facilitates the formation of
higher order structures such as t-loops and G-quadruplexes. Eventhough both t-loops and G-
quadruplexes are important for end protection, they need to be resolved for
replication/transcription complexes to progress.

Presented here is an investigation into how Rap1, Pot1 and their associated complexes
TRF2/Rap1 and Pot1/TPP1 mediate structural changes of telomeric DNA and whether t-
circles could act as a substrate for the telomere extension through ALT. We observed that
hRap1 directly interacts with DNA and modulates the localization of TRF2 at telomeric ends

by increasing its affinity and specificity to telomeric sequences and telomeric junction sites. As a result, the TRF2/Rap1 complex can form more t-loops than TRF2 alone, presenting one possible mechanism of end protection and explanation of hRap1's role. We also demonstrated that G-rich ss telomeric DNA exists as a beads-on-a-string conformation with each bead size ~500 nucleotides. Furthermore, we showed that Pot1 and the Pot1/TPP1 complex can form filaments along G-rich ss telomeric DNA as they open up the G-quadruplexes. These findings can explain how the telomeric DNA structure is modulated during replication. Additionally, we showed that only C-rich t-circles can act as a substrate for telomere extension by a rolling circle mechanism. These information all together demonstrate the importance of Rap1, TRF2 and Pot1 and can be used to explain how ALT initiates upon loss of the shelterin components.

For Annish, Babish, Vinusus, Cociks and Chipi

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

ALT: Alternative Lengthening of Telomeres

ATM: Ataxia Telangiectasia Mutated

ATR: Ataxia telangiectasia and Rad3 related

BFB: Breakage-fusion-bridge

BRCT: the C-terminal domain of a breast cancer susceptibility protein

CD: Circular dichroism

CHIP: Chromatin immunoprecipitation

DDR: DNA Damage Response

Ds: Double stranded

DSBR: Double Strand Break Repair

EM: Electron Microscopy

EMSA: Electrophoretic Mobility Shift Assay

Gap25: Replication Fork with 25 nt gap

G-beads: G-quadruplex beads

G4: G-tetrads G4

HJ: Holliday Junction

HDR: Homology Directed Repair

HR: Homologous Recombination

Ka: Association Constant

Kb: kilobases

Kd: Dissociation constant

MOI: Multiplicity of Infection

MEFs: Mouse embryonic fibroblasts

NHEJ: Non-Homologous End Joining

NLS: Nuclear localization signal

Nt: nucleotides

OB fold: Oligonucleotide/oligosaccharide folds

PAGE: Polyacrylamide Gel Electrophoresis

PML: Promyelocytic leukaemia bodies

PNK: Polynucleotide Kinase

Pot1: Protection Of Telomeres 1

Rap1: Repressor/Activator Protein 1

RCR: Rolling circle replication

RCT: C-terminal Rap1-specific protein-interaction domain

Ss: single stranded

t-circles: Telomeric circles

TIFs: Telomere dysfunction induced foci

TIN2: TRF2- and TRF1-Interacting Nuclear protein 2

TRF1/2: Telomeric Repeat binding Factor 1 and 2

TRF2ΔB: Basic terminus of human TRF2 deleted

T-SCE: Telomere Sister Chromatid Exchange

UTRs: Untranslated regions

2D: Two dimensional

CHAPTER 1: INTRODUCTION

Human chromosomes end in 4-12 kb tandem repeat tracts of 5'-TTAGGG-3' with a 3' overhang (1). This repetitive DNA is bound by sequence-specific proteins called shelterins. Overall, this DNA-protein complex is called a telomere (2). Telomeres evolved to protect against two challenges that linear chromosome ends in eukaryotic organisms pose for cells. First, cells need to have some dispensible DNA at the ends of chromosomes to prevent the loss of genomic information, which would occur as a result of the end replication problem (2,3). During cell division, the lagging strand shortens with each cell division by the size of an Okazaki fragment (~150 nucleotides) (3). Not only do telomeres protect against the loss of genetic information but they also set a limit to the number of divisions a cell can undergo before telomeres shorten significantly and activate senescence (3,4). This is known as the Hayflick Limit.

Second, linear chromosome ends need to be distinguished from double strand breaks to ensure chromosome stability. Telomeres hide the chromosome ends from DNA repair factors either by protein complexes present at chromosome ends or by forming higher order structures, such as telomere loops (t-loops). A t-loop forms by folding the single stranded 3' overhang back and strand invades the upstream double strand telomeric DNA (5). In humans, shelterin is thought to cap chromosome ends and maintain telomere length by regulating telomerase activity. More specifically, the shelterin component TRF2 has been shown to facilitate and stabilize the t-loop structures *in vitro* (6). However, little is known about the regulation of t-loops and their formation by other telomere-associated factors.

Shelterin Complex Proteins

Shelterin is composed of TRF1 and TRF2 (Telomeric Repeat binding Factor 1 and 2), Pot1 (Protection Of Telomeres 1), TIN2 (TRF2- and TRF1-Interacting Nuclear protein 2), Rap1 (the human ortholog of the yeast Repressor/Activator Protein 1) and TPP1 (also known as TINT1, PTOP, or PIP1) (7,8). These proteins are specific to and function at chromosome ends throughout the cell cycle (7). **(Figure 1)**

TRF1 and TRF2: TRF1 was the first mammalian telomeric protein isolated due to its specificity for the mammalian duplex repeat tract (9-11). Shortly after, TRF2 was found by a homology search using TRF1 as a reference (12). Both TRF1 and TRF2 bind to DNA either as homodimers or oligomers, which increase their affinity for repetitive DNA, and thus explains their abundance at the telomere termini (13,14). Specificity for telomeric DNA is achieved by the C-terminal SANT/Myb-type DNA binding domains of TRF1 and TRF2, which recognize the 5'-YTAGGGTTR-3' sequences in duplex telomeric DNA (13,15,16). TRF1 and TRF2 are highly homologous with the exception of their amino termini (14), in which TRF1's N terminus is acidic while TRF2's is basic (Gly/Arg-rich). This leads to different binding specificities. While TRF1 binds to and coats duplex repeats, binding of TRF2 is more structure specific and independent of the sequence (13,17,18). TRF2 prefers to bind to Holliday junctions, 3-way junctions and t-loops via its basic domain (16,17). Although both TRF1 and TRF2 have TRF homology domains (TRFH), which enables the oligomerization of TRFs, TRF1 and TRF2 do not interact directly (19).

Human Rap1 (hRap1) is an ortholog of budding yeast Rap1(20). Budding yeast Rap1 binds to ds telomeric DNA directly and acts as a transcription factor (21-24). Human Rap1 has an

N-terminal BRCT domain (the C_terminal domain of a breast cancer susceptibility protein), a central myb and coil-coil domain followed by a C-terminal RCT domain (C-terminal Rap1-specific protein-interaction domain) with Nuclear localization signal (NLS). hRap1 interacts with TRF2 through its RCT domain. This interaction is suggested to confer Rap1's localization to telomeres because hRap1's myb domain is neutral in charge (20). Recent studies demonstrated that hRap1 regulates transcription and prevents telomere fragility and recombination (25-27). How Rap1 exerts these functions remains to be solved.

Pot1, the highly conserved shelterin component, was identified by its sequence homology to telomere end binding factors in *Oxytricha nova* (28). While humans have one Pot1 gene, mice have two Pot1 genes (Pot1a and Pot1b). Mammalian Pot1 has three OB (oligonucleotide/oligosaccharide binding) folds, which allow it to bind to single stranded DNA (29). However, OB folds are not enough for localization of Pot1 to telomeres *in vivo* because RPA competes for the same binding sites (30). Pot1's interaction with TPP1 outcompetes RPA and TPP1 targets Pot1 to telomeric DNA (31).

TPP1 is the human homolog of the β subunit of protozoan telomere end binding complex (TEBPα-TEBPβ) suggesting that end capping by the Pot1-TPP1 complex is evolutionarily conserved (32). The Pot1-TPP1 complex regulates telomere length by acting as a processivity factor of telomerase (32). TPP1 forms the link between Pot1 and TIN2 through its central domain and C terminus (31,33)

TIN2's role at the telomeres is to essentially link the shelterin complex together. TIN2 can bind TRF1, TRF2, and TPP1, simultaneously and tether the double strand binding proteins to the ssDNA-binding protein POT1 (34). The interaction between TRF1 and TIN2 is through the TRFH domain of TRF1 and the FxLxP motif in the C terminus of TIN2 (35). TIN2 also

has two different sites in the N terminus that interact separately with the hinge domain of TRF2 and TPP1 (35,36).

The name shelterin specifically refers to the six telomere proteins described above and the various subcomplexes that are formed. A TRF2/TIN2/TPP1 triple complex has been reported in addition to a TIN2/ TRF1/ TRF2/ TPP1 quaternary complex (36). Moreover, TRF2/Rap1 as well as Pot1/TPP1 double complexes are found *in vivo* (37). The biochemical properties of complexes can be different than the individual components. Therefore, the DNA binding properties of the subcomplexes and how they modulate the topology of the telomeric DNA needs to be elucidated.
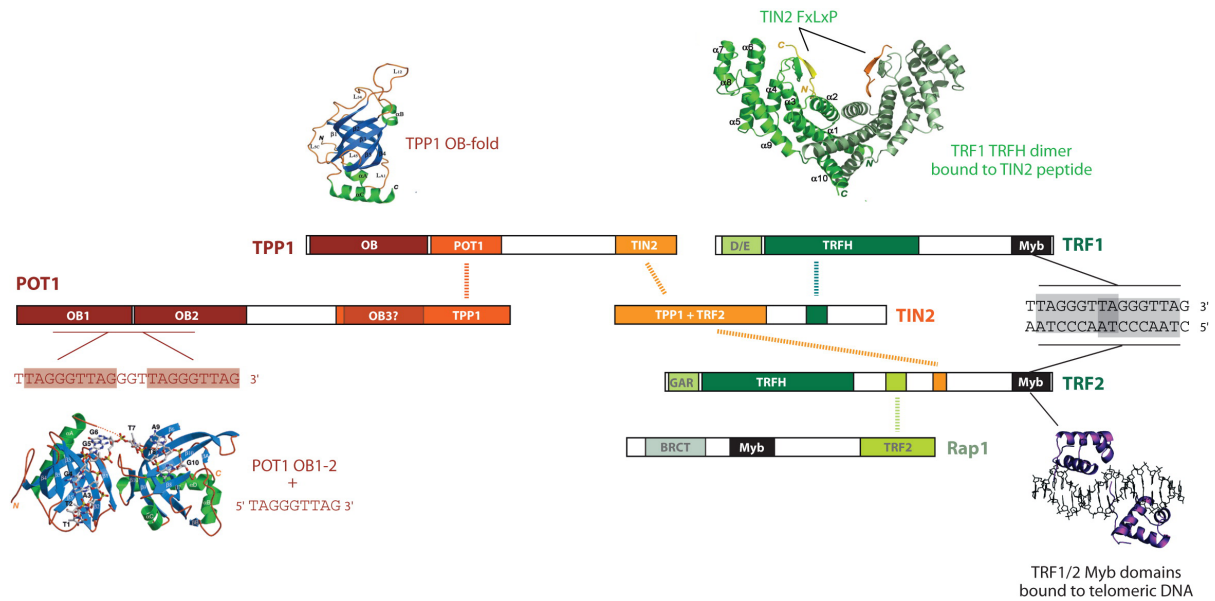
Figure 1: Shelterin complex components are shown with the domain and crystal structures. Domains with same colors are the ones for interaction among shelterin complex members. The lines denote the link between the interaction domains on different shelterin elements and the interactions of proteins with the telomeric DNA (7)

Shelterin and DNA Damage Response (DDR)

Integrity, length and structure of telomeres are very important for their function. The shelterin complex and other telomere associated proteins cap the telomeric ends to prevent unnecessary DSB repair (38,39). Critically short or unprotected ends can trigger the DNA Damage Response (DDR) either through activation of Ataxia Telangiectasia Mutated (ATM) or Ataxia Telangiectasia and Rad3 related (ATR) proteins (40,41). Eventually telomeric ends become prone to end processing by Non-homologous End Joining (NHEJ) and Homology Directed Repair (HDR) (42,43). HDR at telomeres causes an unequal exchange of telomeric sequences between two sister chromatids or different chromosomes. This exchange would result in telomere length changes, terminal deletions and short telomeres in the daughter cell with deleterious effects. NHEJ at telomeres results in chromosome end-to-end fusions between sister chromatids (or dicentric chromosomes) that are prone to initiate breakage-fusion-bridge (BFB) cycles in mitosis.

Critically short or uncapped telomeres fuse and can initiate BFB cycles, which are deleterious for the cells. During anaphase of the cell cycle, the fused ends break by the force applied from the centromeres resulting in terminal deletions, amplifications and inversions of repeat sequences (44). These gross chromosomal rearrangements are common features of carcinogenesis and consistent with the changes in the expression levels of some telomere associated factors such as TRF2 (44-47). Loss of TRF2 leads to formation of telomere dysfunction induced foci (TIFs) and chromosome end to end fusions (48). This indicates that TRF2 is necessary for end protection.

NHEJ can be either canonical (c-NHEJ) or alternative (alt-NHEJ) and depending on which shelterin component is dysfuntional, either one is activated (48). While TRF2 deletion activates c-NHEJ, alt-NHEJ is activated by the deletion of Pot1-TPP1 complex (48). The reason for that difference can be the DDR pathway activated upon their loss because deletion of TRF2 and Pot1 result in activation of ATM and ATR pathways, respectively (49). In the Ku70/80$^{-/-}$ background, TRF2 and Pot1 a/b deletions results in a ~10% exchange of telomeric sequences between sister telomeres (T-SCEs) (50-52). T-SCEs indicate the presence of HDR at telomeres (53). Therefore, this data indicates that both proteins are important in suppressing HDR.

So far, *in vivo* evidence revealing the function of hRap1 in suppression of unnecessary DNA repair at chromosome ends was impaired by the embryonic lethality of the Rap1 knockout mouse. This problem was overcome by a conditional deletion of the Rap1 gene from Mouse embryonic fibroblasts (MEFs), complementation of conditional TRF2 knockout MEFs with a TRF2 mutant that cannot bind to Rap1 and mice without a functional Rap1 gene (25,26,54). Unlike TRF2 and Pot1, deletion of Rap1 does not initiate DDR suggesting that Rap1 is dispensable for telomere capping (52). However, Rap1 deletions in Ku70/80 -/- results in higher number of T-SCEs (~10%) suggesting that mRap1 suppresses HDR (52). The exact mechanism of Rap1 mediated HDR suppression is unknown, yet there is evidence that Pot1 and Rap1 can cooperatively prevent initiation of HDR (55). Without Pot1a/b, Rap1 cannot repress HDR, and in Rap1$^{-/-}$/Ku$^{-/-}$ MEFs, Pot1 a/b alone cannot repress recombination (52,55). Similarly, TRF2 and Rap1 could work together to suppress recombination at telomeres because they exist as complex *in vivo* (20). *In vitro* evidence suggest that the TRF2/Rap1 complex is also important for the repression of NHEJ (27,56). This may involve

the recruitment and regulation of other protein(s). The TRF2/Rap1 complex is known to interact with DNA repair factors such as the MRN complex, Ku70/80 heterodimer and Slx4, which is a scaffold protein that binds to the HJ resolvase, nucleases and mismatch repair proteins (57-59). The exact mechanism of human Rap1 in DNA repair suppression at telomeres requires further study.

Telomere length maintenance

DNA polymerases can not fill in the last lagging strand gap, resulting in progressive shortening of chromosomes during cell division (60,61). In adults, somatic tissues have limited replicative capacity dependent on their telomere length (62). When cells reach their replicative capacity, they enter senescence and lose their ability to re-enter the active cell cycle (40,63,64). Lack of senescence and further cell divisions in the absence of functional tumor suppressors like p53 and retinoblastoma pushes the cells into a crisis phase (65). These cells have instability of karyotype, end-to-end fusions, visible genomic changes in the form of chromosome rearrangements, gain or loss of chromosome arms, deletions and amplifications due to the BFB cycles (44). To bypass crisis, cancer cells adopt two different mechanisms to increase telomere length: telomerase mediated telomere extension or alternative lengthening of telomeres by employing recombination.

Telomerase

Telomerase compensates for telomere attrition through de novo addition of TTAGGG repeats to the 3' end of the DNA strand during replication (66,67). Telomerase is a reverse transcriptase consisting of two subunits and accessory proteins such as NHP2, NOP10,

pontin-reptin, dyskerin and TCAB1 (68-71) . The catalytic subunit, TERT, adds the repeats onto the chromosome ends by using an associated RNA component as a template (TERC) while the accessory factors stabilize the TERC subunit and enable the assembly of the TERC-TERT complex (68,70,72). Telomerase activity is regulated by molecular chaperons like heat shock protein 90 (73) and some shelterin components (32). The activity and expression of telomerase is high in embryos and decreases as tissues differentiate, which correlates with the loss of proliferative potential (62,64). Adult tissues that have high levels of telomerase are dividing male germline cells, adult stem cells and activated immune cells (74,75). Also ~ 85% of cancer cells express telomerase at high levels to increase their replicative capacity (76). Therefore, it is important to understand how shelterin components regulate telomerase activity.

Recombination mediated telomere lengthening

Alternative lengthening of telomeres (ALT) was discovered in human cancer patients that are resistant to telomerase inhibitors, which constitutes ~ 10-15% of all cancers (76,77). ALT is observed in many cancer types but mainly in tumors of the mesenchymal origin such as glioblastoma multiforme, osteosarcoma and some soft tissue sarcomas as well as common cancers like breast carcinomas (78-84). There are no known therapeutic targets of ALT, therefore understanding its molecular mechanism is key for detection and treatment of cancers that extend their telomeres through ALT.

Common characteristics of ALT cells that are not present in normal human cells are heterogenous (85) and rapidly changing telomere length (86,87), a high rate of recombination at the telomeres (88), extrachromosomal telomeric DNA (89,90) and ALT-asociated

promyelocytic leukemia nuclear bodies (ALT-PML or APBs) (91). APBs can contain telomeric chromatin and telomere associated proteins such as shelterin components and DNA repair, replication and recombination factors including RAD52, RAD51, MRN, WRN, BLM, TRF1, TRF2, RAD17, RAD9-RAD1-HUS1 (91-96). A cell line lacking telomerase without APBs managed to maintain its length suggesting that APBs aren't required for ALT activity (97).

Telomeric Circles

Extrachromosomal telomeric DNA can be circular [double stranded (90,98) or partially single stranded (89,99)], linear (100,101) or the high molecular weight complex brached structure called a "t-complex"(89). T-circles were discovered in yeast in this laboratory by Lubomir Tomaska (102) and shown to be a hallmark of ALT by Anthony Cesare (90). The size of the double stranded circular DNA and t-circle, directly correlates with the size of the t-loops, thus t-circles are suggested to arise from the resolution of t-loops (98). Nbs1 and XRCC3 are required for the formation of t-circles in humans (103). Moreover, t-circle formation is elevated in cells that are lacking or are mutated in the TRF2 basic domain (98), Ku 86 (104), WRN helicase (98) or ORC2 (105). T-circles are also present in telomerase positive cells that have some telomerase components overexpressed (106). In that assay, telomere length reached a plateau after the cells were enriched in t-circles suggesting that t-circle formation is a mechanism to prevent the overlengthening of telomeres (106). However, at early stages of development, *X. laevis* embryo contains significant amounts of t-circles, which reduces substantially in later stages (107). This suggests that t-circles are not only

9

important for telomere trimming and to keep telomere length at a steady state but also provides advantages during development.

Single stranded C-rich circles (C-circles) are one marker of ALT cells and are formed of an intact C rich strand and a partial G rich strand (99). They are suggested to arise from nucleolytic degradation of t-circles but experimental evidence is lacking to prove this hypothesis. There are ~ 1000 C-circles per cell and these C-circles are 100 fold more abundant than G-circles (99). In contrast to t-circles there is a direct correlation between appearance of C-circles and onset of ALT activity (99). Furthermore, C-circles diminished when ALT was inhibited (99) suggesting that C-circles can be used in an assay for ALT prognosis but its exact role in telomere length maintenance by ALT needs further investigation.

Models for Telomere Length Maintenance by ALT

Unequal T-SCE model: Telomeric DNA contains nicks and gaps that would stall the replication forks (89) and may initiate recombinational repair. Therefore, telomeres may act as a substrate for T-SCEs (108). The frequency of telomere sister chromatid exchange events (T-SCEs) are higher in ALT cells than telomerase positive cells or somatic cells (88,109) providing evidence for this model. By this mechanism one daughter cell would have a longer telomere with a higher proliferative capacity while the other daughter cell would end up with a shorter telomere and lower proliferative capacity (110). However, this model accounts for prolonged proliferative capacity only if all the long telomeres are separated from all the short telomeres asymetrically, and that all the long telomeres are segregated into one daughter cell (111,112). Presence of such a telomere separation model has yet to be validated.

Homologous recombination dependent DNA replication model: In this model, the new telomeric DNA is synthesized by using the adjacent telomeric DNA as a template (77,113). The templates can be a t-loop, a sister chromatid, an adjacent chromosome, a linear extrachromosomal telomeric DNA, t-circles or C-circles. Experimental evidence for the use of an adjacent chromosome as a template comes from a study where a DNA tag was placed into the telomeres of ALT cells. Tagged telomere numbers increased, suggesting that tag was copied from one telomere to another (113). In yeast, telomeric circles can be used as substrates for telomere extension through Rolling Circle Replication (RCR), and thus in humans t-circles or C-circles can be substrates for RCR as well (77,99,114). Circles can anneal to the 3' G-rich overhang and get extended by RCR, which in turn can be a substrate for interchomosomal recombination. Moreover, this mechanism can explain the heterogenity and the rapid changes in ALT telomere length.
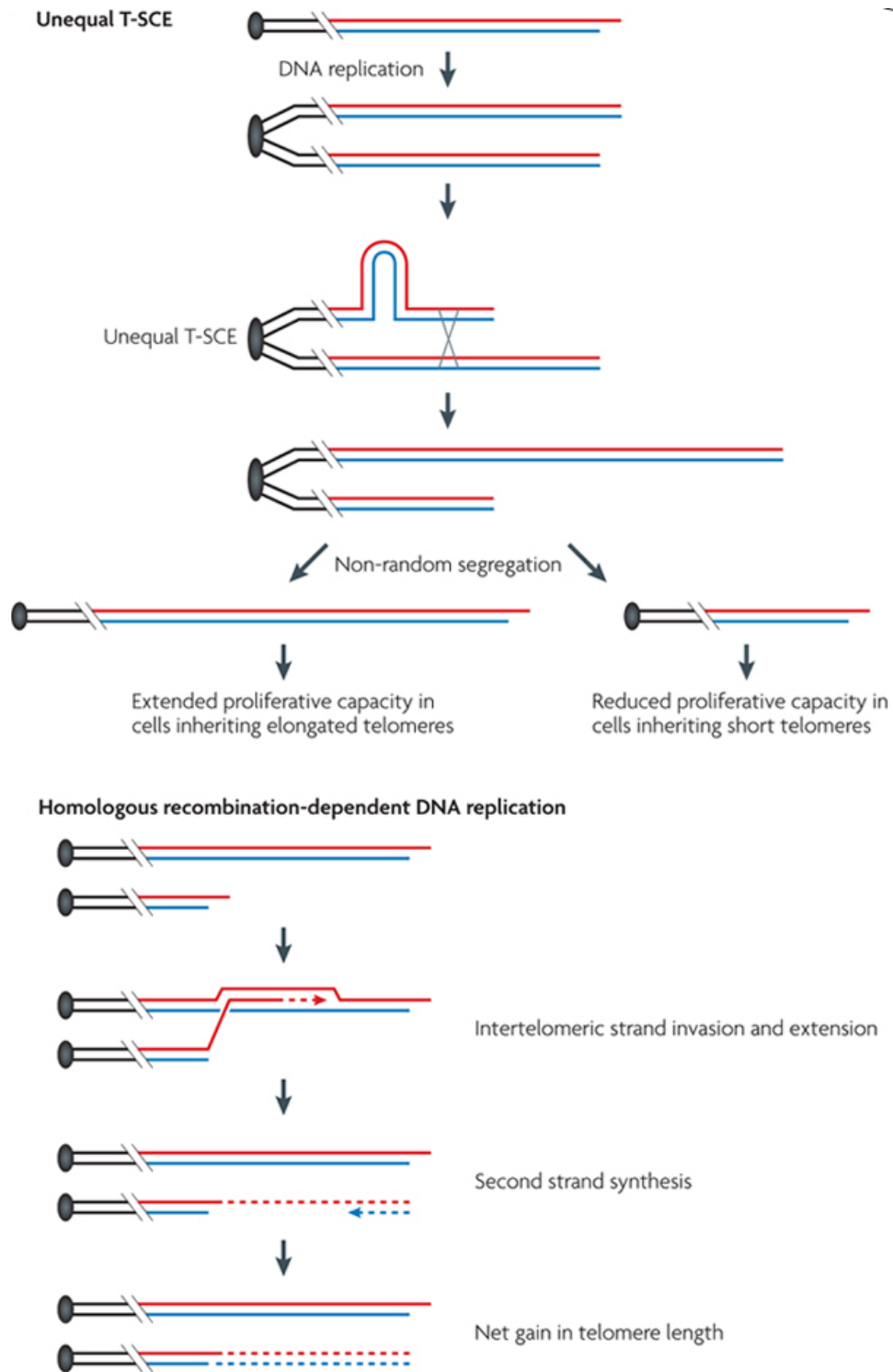
Figure 2: Mechanisms of telomere length regulation by ALT modified from (115).

Shelterin and telomere length regulation

Inhibition of yRap1 function results in increased telomere length, while over-expression or increased localization of yRap1 to telomeres results in telomere shortening. Therefore the accepted model for telomere length regulation in yeast is a negative feedback loop (116-120). Mammalian Rap1 has different effects on telomere length, depending on the cell type. Upon deletion of Rap1 in MEFs, there is no significant effect on telomere length (121). However, a 26% reduction in mean telomere length was observed in mouse epidermis lacking Rap1 (121). Similarly, overexpression of hRap1 or its mutants (122,123) results in telomere lengthening, indicating that a potential role of mammalian Rap1 exists in telomerase regulation.

Overexpression of TRF2 results in progressive shortening of telomeres, and this phenotype can be fully rescued by Xeroderma pigmentosum group F (XPF) deficiency (124). Moreover, TRF2 (ERCC1 and XPF1)(125) colocalize with factors in Nucleotide Excision Repair (NER) pathway indicating that TRF2 controls telomere length via recruitment of XPF to telomeric DNA. Moreover, TRF2 suppresses recombination at telomeres (126) and prevents enzymatic attack on four way junctions and t-loops (98). TRF2 also facilitates t-loop formation *in vitro* (127). Overexpression of a TRF2 mutant lacking the N terminal basic domain results in t-circle formation, which is a common feature of ALT cells (98). All together these data suggest that TRF2 can be a candidate protein in genesis of ALT by suppressing the resolution of telomeric recombination intermediates and by regulating telomeric recombination by t-loop formation.

While a double knockout of Pot1a/b in MEFs did not result in telomere length changes, (128), single knock out studies suggest that Pot1a deficient cells exhibit overall telomere lengthening and 3' overhang elongation (129). Moreover, overexpression of Pot1 mutant lacking the OB DNA binding domain in telomerase positive human cells resulted in telomere elongation when compared to wild type (130). The finding suggests that Pot1 inhibits telomerase action through its DNA binding ability. TPP1 is also a negative regulator of telomerase, potentially through its bridging funtion of Pot1 to TIN2 (131). Additionally, Pot1 may have a role in telomere length regulation in ALT cells through its ability to regulate the replication of the G-rich strand, because inhibition of Pot1 results in an increase in the number of broken replication forks and telomeric replication events such as T-SCEs (132). It remains unknown how ALT is suppressed in normal cells, despite the presence of most of the molecules necessary for ALT. Therefore, understanding the role of the shelterin components in ALT is a key to understand the initiation of ALT.

Table 1: Effect of shelterin components on telomere length regulation and the mutations applied.

	Mutation	Telomere Length	Mechanism
TRF2	Overexpression	Decrease	XPF dependent
	ΔB overexpression	t-circles Increase	ALT
Rap1	Deletion in NEF	No effect	
	Deletion in epidermis	Decrease	Telomerase
	Overexpression	Increase	Telomerase
Pot1a/b	k/o in MEF	No effect	
Pot1a	Deletion	Increase	Telomerase
Pot1	Mutated/OB overexpressed	Increase	Telomerase
Pot1/TPP1	Overexpression	Increased T-SCE	ALT/Telomerase?

Scope of Dissertation

Presented here is an investigation into the DNA binding properties of hRap1, hPot1 and their subcomplexes; TRF2/Rap1 and TPP1/Pot1 as well as an in vitro assay to test the role of C circles in ALT. The goal of the work presented here was to understand how shelterin components and subcomplexes modulate the higher order DNA structures formed at telomeres and how C-circles can mediate telomere extension.

Chapter 2 describes the DNA binding characteristics of hRap1 determined using EM and quantitative gel electrophoresis techniques. The role of Rap1 in modulation of hTRF2's DNA binding properties are also described. This discovery disproves the established knowledge about hRap1 that hRap1 cannot bind to DNA on its own. Moreover, the findings here provide a model for how hRap1 suppresses NHEJ and how the TRF2/Rap1 complex localizes to telomeres, regulates t-loop formation and suppresses NHEJ.

Chapter 3 provides structural information about the G quadruplexes formed during telomere extension and how Pot1 and Pot1-TPP1 open up these motifs. EM, CD spectrophotometry and gel electrophoresis analysis showed that the G-quadruplex motifs can be very complex uniform structures which are composed of ~500 nts in hybrid form. The very big complex character of G-quadruplex motifs can explain how telomeric G-quadruplexes can be a burden during DNA replication. We also showed that Pot1 and Pot1-TPP1 forms filaments along G motifs to open up the quadruplexes. These results provides a model how hPot1 relieves replicative stress and how Pot1-TPP1 increase telomerase processivity.

Chapter 4 describes development of a model replication system using nanocircles and model telomere DNA as a template. EM and gel electrophoresis showed that only C-rich

circles among extratelomeric DNA molecules can act as a template. This data is consistent with roll and spread mechanism of telomere extension and provides information about how C-circles are utilized in ALT cells for telomere maintenance. Moreover, the extension rate was ~ 130 bp/sec and the products ranged from 4 to 50 kb explaining why ALT cells have a big variation in their length and why telomere length in ALT cells change rapidly.

REFERENCES

1.   Wellinger, R. J., and Zakian, V. A. (1989) *Proc Natl Acad Sci U S A* **86**, 973-977

2.   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

3.   Autexier, C., and Greider, C. W. (1996) *Trends in biochemical sciences* **21**, 387-391

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

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

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

7.   Palm, W., and de Lange, T. (2008) *Annual review of genetics* **42**, 301-334

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

9.   McNeel, D. G., and Tamanoi, F. (1991) *Proc Natl Acad Sci U S A* **88**, 11398-11402

10.   Zhong, Z., Shiue, L., Kaplan, S., and de Lange, T. (1992) *Mol Cell Biol* **12**, 4834-4843

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

12.   Bilaud, T., Brun, C., Ancelin, K., Koering, C. E., Laroche, T., and Gilson, E. (1997) *Nature genetics* **17**, 236-239

13.   Bianchi, A., Stansel, R. M., Fairall, L., Griffith, J. D., Rhodes, D., and de Lange, T. (1999) *The EMBO journal* **18**, 5735-5744

14.   Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997) *Nature genetics* **17**, 231-235

15.   Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005) *EMBO Rep* **6**, 39-45

16.   Hanaoka, S., Nagadoi, A., and Nishimura, Y. (2005) *Protein science : a publication of the Protein Society* **14**, 119-130

17. Amiard, S., Doudeau, M., Pinte, S., Poulet, A., Lenain, C., Faivre-Moskalenko, C., Angelov, D., Hug, N., Vindigni, A., Bouvet, P., Paoletti, J., Gilson, E., and Giraud-Panis, M. J. (2007) *Nat Struct Mol Biol* **14**, 147-154

18. Fouche, N., Cesare, A. J., Willcox, S., Ozgur, S., Compton, S. A., and Griffith, J. D. (2006) *J Biol Chem* **281**, 37486-37495

19. Fairall, L., Chapman, L., Moss, H., de Lange, T., and Rhodes, D. (2001) *Molecular cell* **8**, 351-361

20. Li, B., Oestreich, S., and de Lange, T. (2000) *Cell* **101**, 471-483

21. Longtine, M. S., Wilson, N. M., Petracek, M. E., and Berman, J. (1989) *Current genetics* **16**, 225-239

22. Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell* **85**, 125-136

23. Buchman, A. R., Kimmerly, W. J., Rine, J., and Kornberg, R. D. (1988) *Mol Cell Biol* **8**, 210-225

24. Capieaux, E., Vignais, M. L., Sentenac, A., and Goffeau, A. (1989) *J Biol Chem* **264**, 7437-7446

25. Martinez, P., and Blasco, M. A. *Nat Rev Cancer* **11**, 161-176

26. Sfeir, A., Kabir, S., van Overbeek, M., Celli, G. B., and de Lange, T. *Science* **327**, 1657-1661

27. Sarthy, J., Bae, N. S., Scrafford, J., and Baumann, P. (2009) *EMBO J* **28**, 3390-3399

28. Baumann, P., and Cech, T. R. (2001) *Science* **292**, 1171-1175

29. Loayza, D., Parsons, H., Donigian, J., Hoke, K., and de Lange, T. (2004) *J Biol Chem* **279**, 13241-13248

30. Flynn, R. L., Centore, R. C., O'Sullivan, R. J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J., and Zou, L. (2011) *Nature* **471**, 532-536

31. Liu, D., Safari, A., O'Connor, M. S., Chan, D. W., Laegeler, A., Qin, J., and Songyang, Z. (2004) *Nat Cell Biol* **6**, 673-680

32. Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007) *Nature* **445**, 506-510

33. Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., O'Connor, M. S., and Songyang, Z. (2007) *Nature* **445**, 559-562

34. Ye, J. Z., Donigian, J. R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A. N., Chait, B. T., and de Lange, T. (2004) *J Biol Chem* **279**, 47264-47271

35. Chen, Y., Yang, Y., van Overbeek, M., Donigian, J. R., Baciu, P., de Lange, T., and Lei, M. (2008) *Science* **319**, 1092-1096

36. O'Connor, M. S., Safari, A., Xin, H., Liu, D., and Songyang, Z. (2006) *Proc Natl Acad Sci U S A* **103**, 11874-11879

37. Kim, S. H., Davalos, A. R., Heo, S. J., Rodier, F., Zou, Y., Beausejour, C., Kaminker, P., Yannone, S. M., and Campisi, J. (2008) *J Cell Biol* **181**, 447-460

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

39. Sfeir, A., and de Lange, T. (2012) *Science* **336**, 593-597

40. d'Adda di Fagagna, F., Reaper, P. M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N. P., and Jackson, S. P. (2003) *Nature* **426**, 194-198

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

42. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) *Cell* **91**, 25-34

43. Denchi, E. L. (2009) *DNA Repair (Amst)* **8**, 1118-1126

44. Murnane, J. P. (2010) *Cancer research* **70**, 4255-4259

45. Oh, B. K., Kim, Y. J., Park, C., and Park, Y. N. (2005) *The American journal of pathology* **166**, 73-80

46. Miyachi, K., Fujita, M., Tanaka, N., Sasaki, K., and Sunagawa, M. (2002) *Journal of experimental & clinical cancer research : CR* **21**, 269-275

47. Matsutani, N., Yokozaki, H., Tahara, E., Tahara, H., Kuniyasu, H., Haruma, K., Chayama, K., and Yasui, W. (2001) *International journal of oncology* **19**, 507-512

48. Rai, R., Zheng, H., He, H., Luo, Y., Multani, A., Carpenter, P. B., and Chang, S. (2010) *The EMBO journal* **29**, 2598-2610

49. Denchi, E. L., and de Lange, T. (2007) *Nature* **448**, 1068-1071

50. Celli, G. B., Denchi, E. L., and de Lange, T. (2006) *Nat Cell Biol* **8**, 885-890

51.   Palm, W., Hockemeyer, D., Kibe, T., and de Lange, T. (2009) *Mol Cell Biol* **29**, 471-482

52.   Sfeir, A., Kabir, S., van Overbeek, M., Celli, G. B., and de Lange, T. (2010) *Science* **327**, 1657-1661

53.   Bailey, S. M., Cornforth, M. N., Kurimasa, A., Chen, D. J., and Goodwin, E. H. (2001) *Science* **293**, 2462-2465

54.   Teo, H., Ghosh, S., Luesch, H., Ghosh, A., Wong, E. T., Malik, N., Orth, A., de Jesus, P., Perry, A. S., Oliver, J. D., Tran, N. L., Speiser, L. J., Wong, M., Saez, E., Schultz, P., Chanda, S. K., Verma, I. M., and Tergaonkar, V. (2010) *Nat Cell Biol* **12**, 758-767

55.   Kabir, S., Sfeir, A., and de Lange, T. (2010) *Cell cycle* **9**, 4061-4067

56.   Bae, N. S., and Baumann, P. (2007) *Mol Cell* **26**, 323-334

57.   O'Connor, M. S., Safari, A., Liu, D., Qin, J., and Songyang, Z. (2004) *J Biol Chem* **279**, 28585-28591

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

59.   Svendsen, J. M., Smogorzewska, A., Sowa, M. E., O'Connell, B. C., Gygi, S. P., Elledge, S. J., and Harper, J. W. (2009) *Cell* **138**, 63-77

60.   Olovnikov, A. M. (1973) *Journal of theoretical biology* **41**, 181-190

61.   Watson, J. D. (1972) *Nature: New biology* **239**, 197-201

62.   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

63.   Shay, J. W. (2003) *Clinical cancer research : an official journal of the American Association for Cancer Research* **9**, 3521-3525

64.   Harley, C. B., Futcher, A. B., and Greider, C. W. (1990) *Nature* **345**, 458-460

65.   Wright, W. E., Pereira-Smith, O. M., and Shay, J. W. (1989) *Mol Cell Biol* **9**, 3088-3092

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

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

68.   Fu, D., and Collins, K. (2007) *Molecular cell* **28**, 773-785

69.   Cohen, S. B., Graham, M. E., Lovrecz, G. O., Bache, N., Robinson, P. J., and Reddel, R. R. (2007) *Science* **315**, 1850-1853

70.   Venteicher, A. S., Meng, Z., Mason, P. J., Veenstra, T. D., and Artandi, S. E. (2008) *Cell* **132**, 945-957

71.   Venteicher, A. S., Abreu, E. B., Meng, Z., McCann, K. E., Terns, R. M., Veenstra, T. D., Terns, M. P., and Artandi, S. E. (2009) *Science* **323**, 644-648

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

73.   Holt, S. E., Aisner, D. L., Baur, J., Tesmer, V. M., Dy, M., Ouellette, M., Trager, J. B., Morin, G. B., Toft, D. O., Shay, J. W., Wright, W. E., and White, M. A. (1999) *Genes Dev* **13**, 817-826

74.   Flores, I., Canela, A., Vera, E., Tejera, A., Cotsarelis, G., and Blasco, M. A. (2008) *Genes Dev* **22**, 654-667

75.   Pacini, F., Cantara, S., Capezzone, M., and Marchisotta, S. (2011) *Nature reviews. Endocrinology* **7**, 420-430

76.   Shay, J. W., and Bacchetti, S. (1997) *European journal of cancer* **33**, 787-791

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

78.   Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A., and Reddel, R. R. (1997) *Nature medicine* **3**, 1271-1274

79.   Hakin-Smith, V., Jellinek, D. A., Levy, D., Carroll, T., Teo, M., Timperley, W. R., McKay, M. J., Reddel, R. R., and Royds, J. A. (2003) *Lancet* **361**, 836-838

80.   Henson, J. D., Hannay, J. A., McCarthy, S. W., Royds, J. A., Yeager, T. R., Robinson, R. A., Wharton, S. B., Jellinek, D. A., Arbuckle, S. M., Yoo, J., Robinson, B. G., Learoyd, D. L., Stalley, P. D., Bonar, S. F., Yu, D., Pollock, R. E., and Reddel, R. R. (2005) *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 217-225

81.   Costa, A., Daidone, M. G., Daprai, L., Villa, R., Cantu, S., Pilotti, S., Mariani, L., Gronchi, A., Henson, J. D., Reddel, R. R., and Zaffaroni, N. (2006) *Cancer research* **66**, 8918-8924

82.   Villa, R., Daidone, M. G., Motta, R., Venturini, L., De Marco, C., Vannelli, A., Kusamura, S., Baratti, D., Deraco, M., Costa, A., Reddel, R. R., and Zaffaroni, N.

(2008) *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 4134-4140

83. Subhawong, A. P., Heaphy, C. M., Argani, P., Konishi, Y., Kouprina, N., Nassar, H., Vang, R., and Meeker, A. K. (2009) *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **22**, 1423-1431

84. Jeyapalan, J. N., Mendez-Bermudez, A., Zaffaroni, N., Dubrova, Y. E., and Royle, N. J. (2008) *International journal of cancer. Journal international du cancer* **122**, 2414-2421

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

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

87. Perrem, K., Colgin, L. M., Neumann, A. A., Yeager, T. R., and Reddel, R. R. (2001) *Mol Cell Biol* **21**, 3862-3875

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

89. Nabetani, A., and Ishikawa, F. (2009) *Mol Cell Biol* **29**, 703-713

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

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

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

93. Wu, G., Jiang, X., Lee, W. H., and Chen, P. L. (2003) *Cancer research* **63**, 2589-2595

94. Nabetani, A., Yokoyama, O., and Ishikawa, F. (2004) *J Biol Chem* **279**, 25849-25857

95. Silverman, J., Takai, H., Buonomo, S. B., Eisenhaber, F., and de Lange, T. (2004) *Genes Dev* **18**, 2108-2119

96. Jiang, W. Q., Zhong, Z. H., Nguyen, A., Henson, J. D., Toouli, C. D., Braithwaite, A. W., and Reddel, R. R. (2009) *J Cell Biol* **185**, 797-810

97. Cerone, M. A., Autexier, C., Londono-Vallejo, J. A., and Bacchetti, S. (2005) *Oncogene* **24**, 7893-7901

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

99.   Henson, J. D., Cao, Y., Huschtscha, L. I., Chang, A. C., Au, A. Y., Pickett, H. A., and Reddel, R. R. (2009) *Nature biotechnology* **27**, 1181-1185

100.   Tokutake, Y., Matsumoto, T., Watanabe, T., Maeda, S., Tahara, H., Sakamoto, S., Niida, H., Sugimoto, M., Ide, T., and Furuichi, Y. (1998) *Biochemical and biophysical research communications* **247**, 765-772

101.   Ogino, H., Nakabayashi, K., Suzuki, M., Takahashi, E., Fujii, M., Suzuki, T., and Ayusawa, D. (1998) *Biochemical and biophysical research communications* **248**, 223-227

102.   Tomaska, L., Nosek, J., Makhov, A. M., Pastorakova, A., and Griffith, J. D. (2000) *Nucleic Acids Res* **28**, 4479-4487

103.   Compton, S. A., Choi, J. H., Cesare, A. J., Ozgur, S., and Griffith, J. D. (2007) *Cancer research* **67**, 1513-1519

104.   Wang, Y., Ghosh, G., and Hendrickson, E. A. (2009) *Proc Natl Acad Sci U S A* **106**, 12430-12435

105.   Deng, Z., Dheekollu, J., Broccoli, D., Dutta, A., and Lieberman, P. M. (2007) *Curr Biol* **17**, 1989-1995

106.   Pickett, H. A., Cesare, A. J., Johnston, R. L., Neumann, A. A., and Reddel, R. R. (2009) *The EMBO journal* **28**, 799-809

107.   Cohen, S., and Mechali, M. (2002) *EMBO Rep* **3**, 1168-1174

108.   Wilson, D. M., 3rd, and Thompson, L. H. (2007) *Mutation research* **616**, 11-23

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

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

111.   Muntoni, A., and Reddel, R. R. (2005) *Human molecular genetics* **14 Spec No. 2**, R191-196

112.   Blagoev, K. B., and Goodwin, E. H. (2008) *DNA Repair (Amst)* **7**, 199-204

113.   Dunham, M. A., Neumann, A. A., Fasching, C. L., and Reddel, R. R. (2000) *Nature genetics* **26**, 447-450

114. Tomaska, L., Nosek, J., Kramara, J., and Griffith, J. D. (2009) *Nat Struct Mol Biol* **16**, 1010-1015

115. Cesare, A. J., and Reddel, R. R. (2010) *Nat Rev Genet* **11**, 319-330

116. Conrad, M. N., Wright, J. H., Wolf, A. J., and Zakian, V. A. (1990) *Cell* **63**, 739-750

117. Lustig, A. J., Kurtz, S., and Shore, D. (1990) *Science* **250**, 549-553

118. Marcand, S., Gilson, E., and Shore, D. (1997) *Science* **275**, 986-990

119. McEachern, M. J., and Blackburn, E. H. (1995) *Nature* **376**, 403-409

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

121. Martinez, P., Thanasoula, M., Carlos, A. R., Gomez-Lopez, G., Tejera, A. M., Schoeftner, S., Dominguez, O., Pisano, D. G., Tarsounas, M., and Blasco, M. A. (2010) *Nat Cell Biol* **12**, 768-780

122. O'Connor, M. S., Safari, A., Liu, D., Qin, J., and Songyang, Z. (2004) *J Biol Chem* **279**, 28585-28591

123. Li, B., and de Lange, T. (2003) *Molecular biology of the cell* **14**, 5060-5068

124. Munoz, P., Blanco, R., Flores, J. M., and Blasco, M. A. (2005) *Nature genetics* **37**, 1063-1071

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

126. Blanco, R., Munoz, P., Flores, J. M., Klatt, P., and Blasco, M. A. (2007) *Genes Dev* **21**, 206-220

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

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

129. 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

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

131. Ye, J. Z., Hockemeyer, D., Krutchinsky, A. N., Loayza, D., Hooper, S. M., Chait, B. T., and de Lange, T. (2004) *Genes Dev* **18**, 1649-1654

132.	Arnoult, N., Saintome, C., Ourliac-Garnier, I., Riou, J. F., and Londono-Vallejo, A. (2009) *Genes Dev* **23**, 2915-2924

CHAPTER 2: HUMAN RAP1 INTERACTS DIRECTLY WITH TELOMERIC DNA AND REGULATES TRF2 LOCALIZATION AT THE TELOMERE

SUMMARY

The TRF2/Rap1 complex suppresses Non-homologous end joining and interacts with DNA-PKcs to prevent end joining. We previously demonstrated that hTRF2 is a double strand telomere binding protein that forms t-loops *in vitro* and recognizes 3- and 4- way junctions independent of DNA sequence. How the DNA binding characteristics of hTRF2 to DNA is altered in the presence of hRap1 however is not known. Here we utilized EM and quantitative gel retardation to characterize the DNA binding properties of hRap1 and the TRF2/Rap1 complex. Both gel filtration chromatography and mass analysis from 2D projections showed that the TRF2/Rap1 complex exists in solution and binds to DNA as a complex consisting of 4 monomer of each of hRap1 and hTRF2. EM revealed for the first time that hRap1 binds to DNA templates in the absence of hTRF2 with a preference for double strand-single strand junctions in a sequence independent manner. When hTRF2 and hRap1 are in a complex, its affinity for ds telomeric sequences is 2-fold higher than TRF2 alone and more than 10-fold higher for telomeric 3' ends. This suggests that as hTRF2 recruits hRap1 to telomeric sequences, hRap1 alters the affinity of hTRF2 and its binding preference on telomeric DNA. Moreover, the TRF2/Rap1 complex has a higher ability to re-model telomeric DNA than either component alone. This finding underlies the importance of complex formation between hRap1 and hTRF2 for telomere function and end protection.

INTRODUCTION

Eukaryotic chromosomes ends are protected by a combination of looped or fold-back structures and proteins that generate protective telomere-specific complexes (1,2). In higher eukaryotes, a set of six telomere-specific proteins termed the shelterins has been identified which are central to telomere maintenance (3). In addition to the shelterins, the reverse transcriptase telomerase and many general DNA replication and repair factors complete the repertoire of proteins protecting telomere ends (4-6). TRF1 and TRF2 bind directly to duplex telomeric repeats through their myb domains, while the other shelterins are suggested to scaffold onto these two proteins (7,8). TRF2 appears to be most central to these events. *In vitro* TRF2 is able to generate t-loop structures in which the 3' single strand (ss) overhang of the telomere invades the internal double strand (ds) telomere regions to generate a large duplex loop (1,9). Further, TRF2 is able to bind another shelterin factor, Rap1, as well as numerous DNA replication and repair factors including Ku 70, Apollo, MRN complex, WRN, Fen1, ORC1, PARP1 and PARP2 presumably to tether these factors near the chromosome end (10,11). The complex of TRF2 and Rap1 prevents NHEJ and works with DNA-PK to suppress joining of telomere ends and thus must be central to telomere maintenance and architecture (12,13).

The importance of TRF2 has been documented by studies involving over-expression of dominant negative alleles and more recently inducible knockouts of TRF2 (14,15). These studies show that TRF2 suppresses NHEJ and prevents chromosome end-to-end fusions (14,15). Additionally, down regulation or loss of TRF2 activates the ATM kinase pathway, induces the formation of DNA damage induced foci, up-regulates p53 and arrests cells at the

G1/S checkpoint (16-18). The interaction of TRF2 with telomeric DNA has been examined in detail *in vitro* in this and other laboratories (9,19,20). The protein contains a single myb domain as well as a dimerization domain and a highly basic N terminus (7). Since two myb domains are needed for stable DNA binding, the homodimer of TRF2 represents the minimal oligomer required for stable binding to duplex telomeric DNA (21). TRF2 however is also able to bind to DNA in a non-sequence but structure-specific manner, mediated by the basic N terminus (19). These studies described first in this laboratory showed that telomeric DNA is highly prone to replication fork slippage and thus the binding of TRF2 to slipped structures may be key in stopping further fork slippage and preventing the generation of deleterious structures (19). Moreover, TRF2 prevents NHEJ at non-telomeric sites through its basic domain (22).

hRap1 was first identified by yeast two hybrid analysis with hTRF2 as a bait (10). hRap1 binds directly to hTRF2 through its RCT domain and pull down experiments in human cell extracts have identified stable complexes of hTRF2 and hRap1 (2,10). hRap1 is more abundant in the cell than hTRF2 and immunodepletion of hRap1 results in a reduction of TRF2 levels (23). In contrast, lowering TRF2 with shRNA resulted in only a moderate change in the total Rap1 levels due to its overabundance (24). This may indicate that in contrast to mouse Rap1, hRap1 does not require hTRF2 for its stability in the cell. It was suggested that hRap1 does not bind DNA directly and is recruited to human telomeres by its binding partner hTRF2 (10). However, as observed by CHIP analysis, hTRF2 and hRap1 have distinct and overlapping binding sites along the chromosome (25,26). Further, mammalian Rap1 is found at non-telomeric loci in addition to telomeres (25). Some of the non-telomeric sites have consensus $(TTAGGG)_2$ motifs while others do not (25). The

29

localization of hRap1 to regions with TTAGGG sequences could be through its interaction with TRF2 but how it localizes to non-telomeric sites is not clear (26).

Budding yeast Rap1 plays a role in transcription, telomere length regulation and end capping. Rap1 is the most conserved shelterin component among different species (27), and recently, hRap1 was shown to have similar functions to its yeast counterpart. hRap1 regulates the expression level of genes that are in close proximity to its binding sites and regulates telomere length through its BRCT domain (11,26). Moreover, Rap1 alone suppresses NHEJ in human extracts and HR in mouse cells (23,28). However significantly less is known about the function and role of Rap1 in human or mouse cells in the suppression of the DNA Damage Response.

Our previous *in vitro* studies (9,19) of the interaction of TRF2 with telomeric DNA as well as Holiday junctions and replication forks have led to a better understanding of its role in telomere protection. To further understand the role of hRap1 alone or bound to hTRF2, it will be critical to determine how the binding of hRap1 to hTRF2 augments or modifies its binding to telomeric DNA and to unusual DNA structures. In this study we characterized the DNA binding ability of hRap1 to different model DNA templates in the absence of hTRF2 and then tested how the binding preference and affinity is altered when hRap1 is in a complex with hTRF2.

EXPERIMENTAL PROCEDURES

Preparation of DNA Molecules — A) Preparation of linear telomeric templates — The pRST5 plasmid contains ~575 bp of ds human telomeric DNA (9). A plasmid containing ~1.1 kb telomeric repeats, pOST6, was prepared from pRST5 with expansive cloning as described in Stansel et al (9). The minichromosome template consists of two model telomeres joined at their non-telomeric ends so that ds telomeric DNA with 3' ss overhangs are present at both ends. The minichromosome templates were prepared similarly to the model telomeres (9) by treating pRST5 with BsmBI and NotI followed by 3' overhang ligation with a 1:10 molar ratio of linear ds template to a 124 nt oligo consisting of 5' TTAGGG 3' repeats (IDT, Coralville, IA). The NotI digestion product has 5' GGCC 3'ends that facilitate dimerization at the nontelomeric ends upon ligation. Blunt end DNA molecules were obtained by treating BsmBI digested pOST6 with S1 nuclease as described in the manufacturer's protocol (Fermentas Inc., Glen Burnie, MD). To generate a template with a telomeric 5' overhang, BsmBI digested pOST6 was treated with ExoIII on ice for 5 minutes as described in the manufacturer's protocol (New England Biolabs, Ipswich, MA). A nontelomeric template with a 3' overhang was generated by treating EcoRI digested pGLGAP, which does not contain telomeric repeats (29), with T7 exonuclease at room temperature for 40 seconds followed by incubation on ice for 5 minutes (New England Biolabs).

B) Preparation of stalled replication forks and Holliday Junctions — To generate a circular ds DNA with a 400 nt single strand tail, pGLGAP plasmid was nicked using Nb.BbvCI (New England Biolabs) and the nicked strand was displaced by incubation with the Klenow fragment (exo–) of DNA polymerase I in the presence of dNTPs except for dCTP (29). To

make the tail ds with different gap sizes at the fork, primers were annealed to the ss tail and incubated with Klenow (exo-) polymerase in the absence of dGTP for 30 minutes at 37°C (30). Preparation of telomeric and nontelomeric Holliday Junctions were done as described (31).

C) Preparation of DNA for EMSA — pRST1 plasmid, which contains a 154 bp ds telomeric insert, was digested with BsmBI and HindIII and the telomeric insert isolated by gel electrophoresis (QIAGEN, Germantown, MD). The purified insert was used as a ds telomeric template for EMSA after the 5' ends were radiolabeled with [γ-^{32}P]ATP in a standard T4 Polynucleotide Kinase (PNK) reaction as described by the manufacturer (New England Biolabs). To generate a radiolabeled telomeric template with a 3' overhang, the gel isolated duplex DNA was ligated to a G rich oligo with 8 telomeric repeats after the oligo was phosphorylated with [γ-^{32}P]ATP from its 5' end as described above. A nontelomeric template with a 3' overhang was prepared from two oligos with sequences 5'ATAGCTAGACATAGACCTAGGATTCCGTAGCTAGCACTGGCATACTGCTAGAT CGCGATACTGGTCACTAGCTAGGCTACAGTCCTGACG 3' and 5' CGCGATCTAGCAGTATGCCAG TGCTAGCTACGGAATCCTAGGTCTATGTCTAGCTAT 3' (MWG Operon, Huntsville, AL). Only oligo 1 was phosphorylated at its 5' end with [γ-^{32}P]ATP as described by the manufacturer with T4 PNK (New England Biolabs) and annealed to oligo # 2 with a gradual cool down from 90°C.

Purification of proteins — The full-length hRap1 and hTIN2 genes were purchased from Open Biosystems (Fisher Scientific, Pittsburg, PA) and the full-length hTRF2 gene was a gift of Dr. Christopher Counter. Each gene was cloned into the pFastBacHtA plasmid. The full length NH_2-terminal His_6-tagged hTRF2 was purified with a Talon™ metal affinity resin (Clontech, Palo Alto, CA) from Sf21 extracts as described (32) and stored in 20 mM Hepes (pH 7.5), 300 mM NaCl, 20% glycerol, 3 mM $MgCl_2$, 1 mM DTT at -80°C. The full length NH_2-terminal His_6-tagged hRap1 was purified with Ni-NTA chromatography (QIAGEN, Germantown, MD) from Hi-5 cell extracts (10) and stored in 50 mM $NaPO_4$ (pH 7.5), 300 mM NaCl, 20% glycerol, 8 mM β-Mercaptoethanol (BME). Similarly, NH_2-terminal His_6-tagged hTIN2 was purified with Ni-NTA chromatography (QIAGEN, Germantown, MD) from Sf21 extracts (33) and stored at -80°C in 50 mM $NaPO_4$, 150 mM NaCl, 8 mM BME, 20% glycerol.

B) TRF2/Rap1 complex formation — TRF2/Rap1 complexes were formed by incubating 100 μg of hTRF2 and 100 μg hRap1 on ice for 30 minutes in 10 mM Hepes (pH 7.5), 25 mM $NaPO_4$, 300 mM NaCl, 4 mM BME, 1,5 mM $MgCl_2$, 0.5 mM DTT and 20% glycerol, followed by size exclusion chromatography on Sepharose 6 (GE Healthcare, Piscataway, NJ) using an elution buffer of 20 mM Hepes (pH 7.5), 300 mM NaCl, 20% glycerol, 3 mM $MgCl_2$ 8 mM BME. Fractions collected from the Sepharose 6 size exclusion column (GE Healthcare, Piscataway, NJ) were stored at -80°C. To determine the elution profile of the individual components, 100 μg of NH_2-terminal His_6-tagged hTRF2 or NH_2-terminal His_6-tagged hRap1 proteins were passed through the Sepharose 6 size exclusion column (GE

Healthcare, Piscataway, NJ) in a buffer of 300 mM NaCl, 20 mM Hepes (pH 8.75) and 8 mM BME.

EM analysis — A) Tungsten shadowcasting — DNA-protein complexes in the binding reaction mixture were cross-linked with 0.6% (w/v) gluteraldehyde for 5 minutes at room temperature and passed through 2-ml size exclusion columns with A5M beads (Agarose Bead Technologies, Spain) pre-equilibrated with 0.01 m Tris-HCl (pH 7.6), 0.1 mm EDTA. The cross-linked complexes were mixed with a buffer containing 2.5 mM spermidine, adsorbed to glow-charged carbon foil grids for 3 minutes and dehydrated with a series of water/ethanol washes, air-dried, and rotary shadowcast with tungsten at 1×10^{-6} torr as described (34). An FEI Tecnai 12 instrument (Hillsboro, OR) equipped with a Gatan Orius CCD camera (Gatan, Pleasanton, CA) at 40 kV was used to capture the images using Digital Micrograph software. Images for publication were arranged and contrast optimized using Adobe Photoshop CS5 (Adobe Systems, San Jose, CA). At least 100 molecules were scored in sequence as they were encountered at the EM and statistical analysis was done using Student's t-test (Graphpad Software, Inc., La Jolla, CA)

B) Negative staining and mass analysis — To generate the TRF2/Rap1 complex for negative staining, hTRF2 (200 ng) and hRap1 (200 ng) were incubated on ice for 30 minutes and then diluted to 20 ng/μl in 20 mM Hepes (pH 8.75) and 100 mM NaCl. To negative stain individual proteins, hRap1 or hTRF2 was diluted to 20 ng/μl as above and stained with 2% (w/v) uranyl acetate in water. Samples were examined in a Philips CM12 TEM at 80 kV and images captured on a Gatan First Light high sensitivity CCD camera (Gatan, Pleasanton,

CA). Negative stained images (1500-2000) for each protein were analyzed using ImageJ software (NIH, Bethesda, MD) for mass determination as described previously in numerous papers from this laboratory (31,35). Ferritin was used as a size standard for these experiments.

C) Mass analysis of proteins bound to Holliday Junctions — DNA-protein complexes were tungsten shadowcast as described above and a size standard (ferritin) was prepared side-by-side for each experiment (31). Images were captured as above (Tecnai 12, Hillsboro, OR). For each protein ~ 100 images were analyzed with ImageJ software to determine the area distribution (NIH, Bethesda, MD).

Electrophoretic Mobility Shift Assays — DNA binding reactions were done at room temperature for 20 minutes in 20 mM Tris (pH 8.0) at different protein concentrations and 5 nM DNA template. Reactions were quenched with 2.5 mM EDTA, 5% glycerol, 0.25 mg Bromophenol Blue and 0.25 mg Xylene Cyanol and loaded onto a 4% native polyacrylamide gel. The dried gel was analyzed by autoradiography and imaged with a Typhoon 9400 phosphoimager (Amersham Biosciences, Piscataway, NJ). Graphpad prism software was used for the nonlinear regression analysis to determine *Kd* values (GraphPad Software, Inc., La Jolla, CA). To verify that binding is specific to hRap1, a His-tagged antibody was used in the binding reactions as described above and the antibody amount in the reaction was determined according to manufacturer's protocol (ABGENT, San Diego, CA).

RESULTS

hRap1 binds to minichromosomes, Holliday Junctions and Replication forks

The DNA binding ability of hRap1 has been previously examined using gel shift assays employing short (72 bp) ds human telomeric DNA or ss and ds yeast telomeric DNA (10). By this approach, hRap1 did not show significant binding to telomeric DNA. However, short DNAs do not recapitulate the full telomere architecture, and thus we used EM to examine the DNA binding ability of hRap1 using DNA templates in the range of 0.7 to 7 kb including the minichromosomes, Holliday Junctions and replication forks illustrated in **Figure 2.1A-C**. The minichromosome template mimics a human chromosome with a 6 kb segment of plasmid DNA flanked by ~ 575 bp of repeating 5' TTAGGG 3' duplex repeats at both ends terminating in 120 nt, 3' TTAGGG overhangs **(Figure 2.1A)**. We also used a Holliday junction (HJ) with 175 bp intersecting arms, and a similar size telomeric HJ with 2 TTAGGG repeats at the intersection site **(Figure 2.1B)** (36). Finally, a replication fork was created from a 3.4 kb duplex circle that contains a ~ 400-bp displaced arm with different gap sizes at the fork (30), which enables us to test the ability of proteins to bind to junction sites through structure recognition as these replication forks do not contain telomere repeats **(Figure 2.1C)**.

Human Rap1 was purified from insect cells. To examine the binding preference of each protein on the different DNA templates, sub-saturating ratios of protein:DNA were used to gain evenly sized particles on the DNA and to avoid a significant number of aggregates consisting of multiple DNAs bound together by protein. To do this, different protein:DNA ratios were tested and the ones which yielded roughly one-half of the DNA bound by protein while less than 10% was in aggregates were selected. The optimal binding for each protein

was observed at 14:1 and 26:1 protein tetramers:DNA molar ratios for hTRF2 and hRap1 respectively. hTIN2 interacts with hTRF2 (37) but has not been shown to interact with telomeric DNA directly while hTRF2's behavior with similar templates has been studied previously (19,37). Thus hTIN2 and hTRF2 were used as controls in these assays. Supplementary figure 1 shows the binding of hTRF2 to the minichromosomes **(Supplementary 2.1A)**, a replication fork with a 25 nt gap (Gap25) **(Supplementary 2.1B)**, and telomeric HJs **(Supplementary 2.1C)**. In each case, hTRF2 bound to the 3- and 4- way junctions confirming previous findings (19). Under the same binding conditions even with as many as 568 hTIN2 protein monomers to DNA, TIN2 did not bind to the minichromosome (data not shown). hRap1 localized to the telomeric ends of the minichromosomes **(Figure 2.1D)**, to the crossover of the telomeric HJ **(Figure 2.1E)**, and to the fork junction of Gap25 DNA **(Figure 2.1F)**. In each case, the protein-bound DNA species showed a uniformly sized and shaped protein particle bound. hRap1 has an N terminal histidine tag and to confirm that the binding activity was specific to hRap1, we did an EMSA with an anti-His tag antibody. DNA template used was 57 nt long with 33 nt 3' overhang and did not contain any telomeric sequences (Experimental Procedures). The lack of a shift with antibody alone but the presence of a supershift with the antibody and hRap1 confirmed the binding activity to be specific to hRap1 **(Figure 2.1G)**. These data suggest that hRap1 has the ability to directly interact with DNA.

A.

B.

C.

D.

E.

F.

G.

Anti His	–	–	+	+
hRap1	–	+	+	–
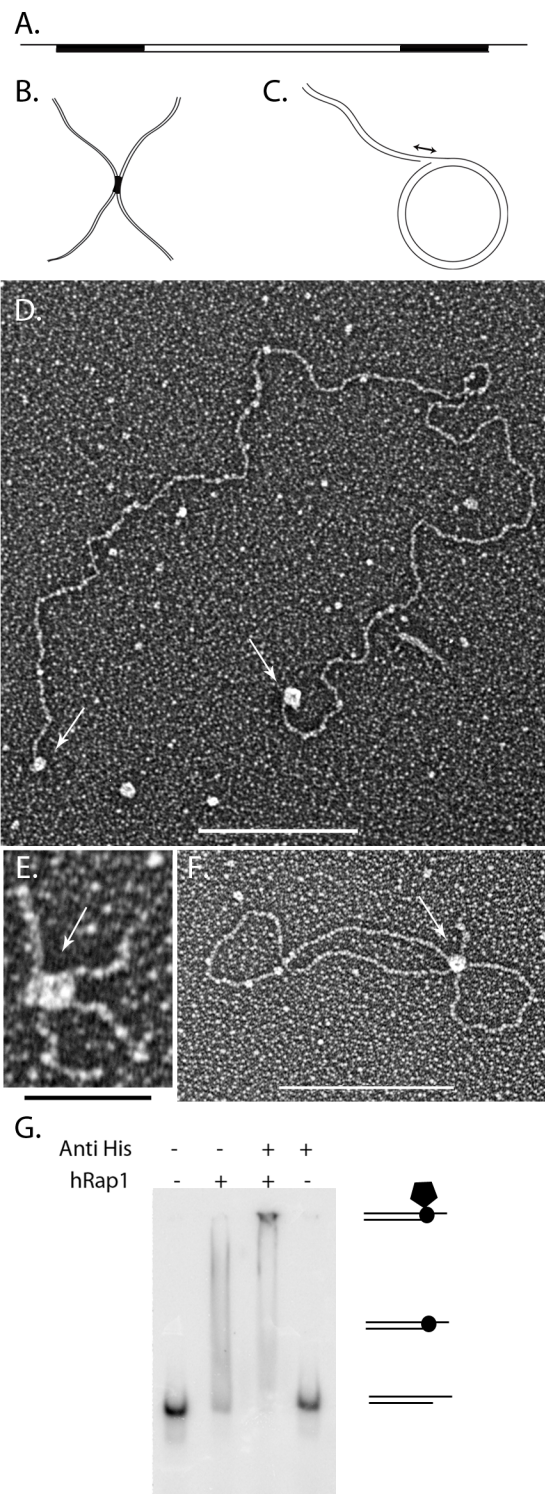
Figure 2.1: hRap1 binds to model DNA templates**.** Schematic representations of the minichromosome (A), replication fork with 25 nt gap (B) and Holliday Junction templates (C). Black regions correspond to the telomeric DNA. Tungsten shadowcast images of hRap1 illustrate protein bound to the minichromosome (D), replication fork (E) and telomeric Holliday junction template (F). Shown in reverse contrast. Size bar in (D) and (F) are 100 nm and is 50 nm in (E). EMSA with 4% PAGE demonstrates hRap1 bound to the nontelomeric template with a 3' overhang in the presence and absence of anti 6X his antibody (G).
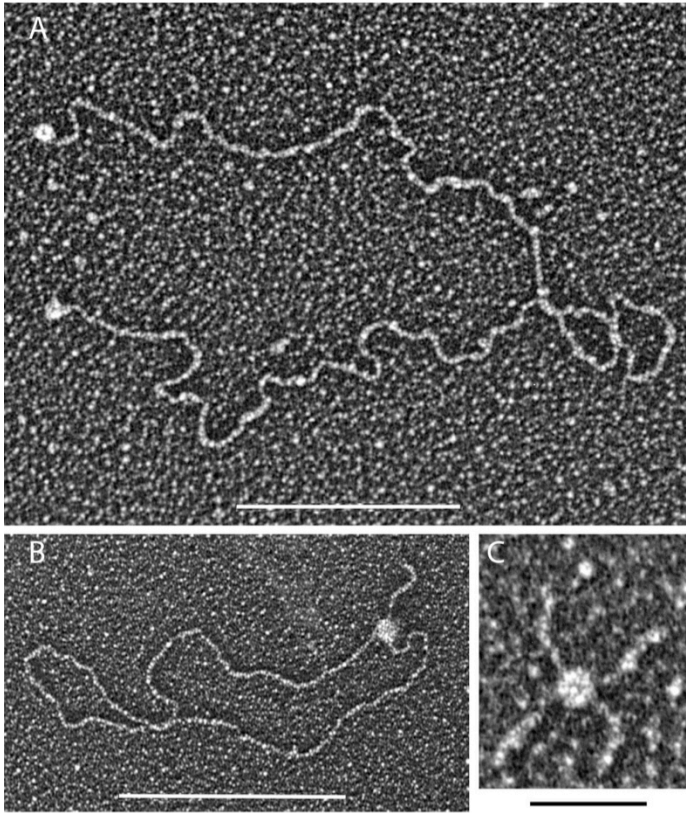
Supplementary Figure 2.1: hTRF2 localizes to the ds-ss junctions on different DNA templates. Tungsten shadowcast images of hTRF2 bound to the minichromosome (A), replication fork with a 25 nt gap (B), and the telomeric HJ template (C) shown in reverse contrast. Bars correspond to 100 nm in A and B and 50 nm in C.

hRap1 binds to ds-ss junctions independent of sequence but prefers 3' overhang structures

over 5' overhangs

We previously showed that hTRF2 binds to replication forks, HJs, and ds-ss overhangs independent of sequence (19). Because hRap1 bound to the same DNA templates, we wished to quantify binding of hRap1 with different DNA templates. On replication forks and Holliday junctions, the majority of hRap1 localized to the junction site **(Figure 2.2A)** at comparable levels to hTRF2 confirming that hRap1 recognizes 3- and 4- way junctions. Scoring the distribution of hRap1 molecules along the minichromosomes, we found that hRap1 localizes to the termini of the minichromosomes containing the 3' ds-ss junction site with a ~ 13 fold greater preference over the ds telomeric DNA **(Figure 2.2B)**. When the 3' ss overhang was removed, the total hRap1-binding dropped to background levels, confirming that hRap1 specifically recognizes the ds-ss junctions of minichromosomes **(Figure 2.2C)**.

We wanted to determine if localization of hRap1 to minichromosomes is sequence or structure dependent. Human telomeres engaged in Homologous recombination (HR) dependent telomere maintenance have 5' C rich overhangs (38). To further probe the structure specificity of hRap1 binding, we used a linear template containing a 1.1 kb telomeric tract at one end and then digested it with a 3' to 5' exonuclease to generate 5' tails up to 340 nt in length (as seen by EM). hRap1 was found to preferentially bind to DNA with 3' overhangs over 5' ends as does hTRF2 **(Figure 2.2D)**. To examine the sequence specificity of hRap1 binding at junction sites, a linear 3.5 kb nontelomeric plasmid DNA was digested with a 5' to 3' exonuclease to generate 3' overhangs of average ~ 560 nts. hRap1 localized to the 3' ends of the telomeric and nontelomeric linear templates equally with no statistical difference, while hTRF2 showed a preference for telomeric DNA with 3'

overhangs **(Figure 2.2E)**. There was no significant effect of the presence of telomeric sequences at HJs on hRap1 binding **(Supplementary Figure 2.2)**. These findings suggest that hRap1 has a strong preference for 3' ends on linear templates and has the ability to interact with the ds-ss junction sites independent of their sequence, in contrast to hTRF2. The 3' G rich overhang at telomeric ends is a requirement for the formation of a t-loop (1,9) and thus we addressed the t-loop forming ability of hRap1. Despite binding to junction sites, hRap1 formed t-loops 5-fold less well when compared to hTRF2 suggesting that this is not a strong activity for hRap1 alone **(Figure 2.2F)**.
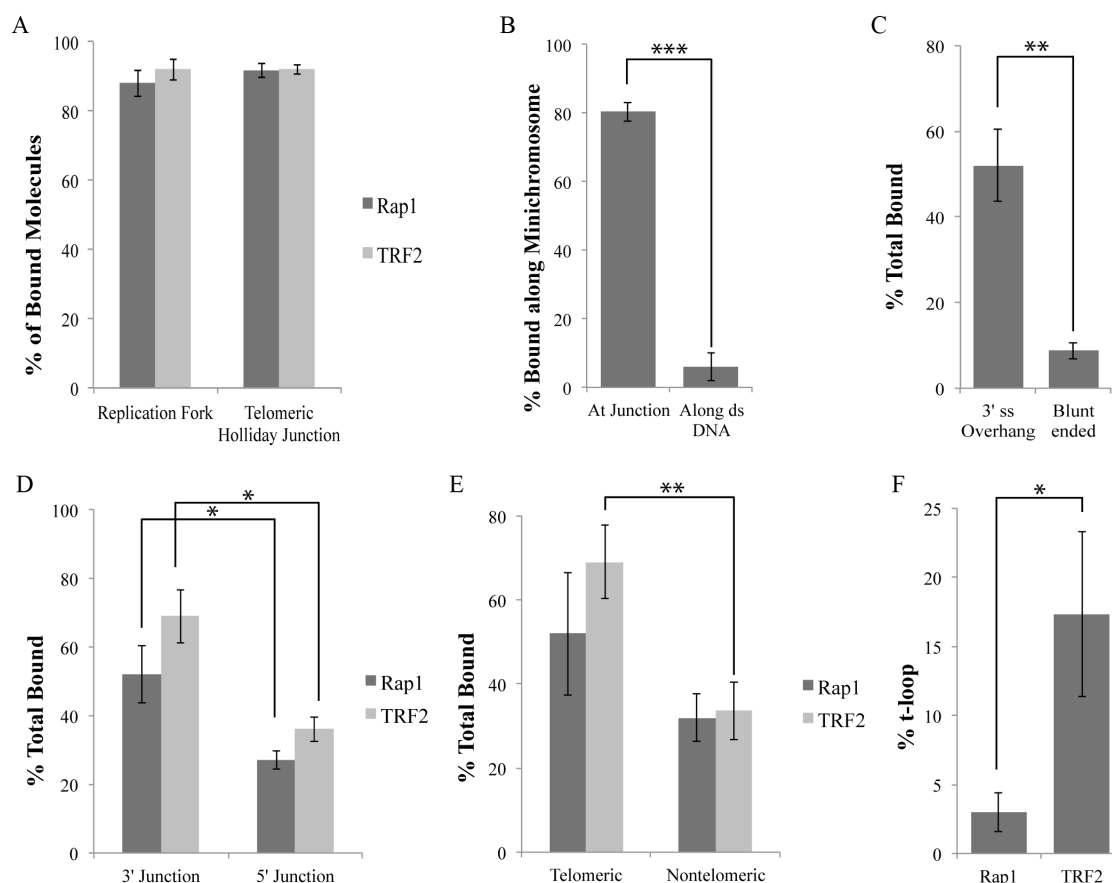
Figure 2.2: hRap1 recognizes the 3' ds-ss junction structures independent of sequence. (A) illustrates that both hRap1 and TRF2 have a strong and similar preference for binding to the ss-ds junction at replication fork with a 25 nt gap and the crossover at the HJ. (B) illustrates the strong preference for hRap1 binding to the end of the minichromosome containing a 3' ss extension as contrasted to binding internally along the ds telomeric segment. (C) showed the strong preference for hRAP1 binding to the minichromosome containing a 3' ss extension as contrasted to the same but blunt ended DNA. (D) shows that both hRap1 and TRF2 prefer to bind at the ss-ds junction of DNAs with a 3' overhang as contrasted to a 5' overhang. (E) compares the binding to DNAs containing 3' overhangs joined to either telomeric ds segments or non-telomeric. (F) compares the t-loop formation percentages of hRap1 and hTRF2 on the minichromosome. Each binding experiment was done in triplicate and at least 100 molecules were counted. * is $p<0.05$, ** is $p<0.01$ and $p<0.001$ is represented by ***.

43
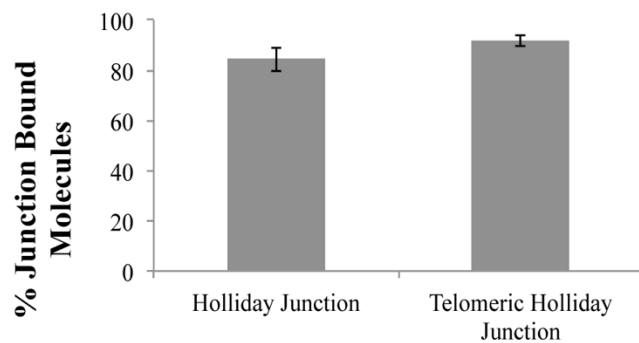
Supplementary Figure 2.2: Junction preference of hRap1 binding to telomeric and nontelomeric Holliday junctions. hRap1 binds telomeric and nontelomeric HJ's primarily at the junction site without a sequence preference. Binding experiments in quadruplicate with 100 molecules counted for each.

The hTRF2/Rap1 complex is formed from 4 molecules of hRap1 and 4 molecules of hTRF2

The oligomeric states of proteins can influence their binding characteristics. hTRF2 and hRap1 exist in a complex at human telomeres (2,39) and while it has been demonstrated that they exist in a 1:1 ratio (39), their oligomeric state is unknown. Therefore, we analyzed the mass of hRap1, hTRF2 and the TRF2/Rap1 complexes in the absence of DNA by EM. The TRF2/Rap1 complex was formed from the purified proteins (Experimental Procedures). In negative stained fields of hTRF2 **(Figure 2.3A)**, Rap1 **(Figure 2.3B)**, and the TRF2/Rap1 complex **(Figure 2.3C)** a variety of particle sizes and shapes were observed for all three, but in each case there was a predominant particle and there was a clear size difference between hTRF2, hRap1 and the TRF2/Rap1 complex.

To obtain estimates for the mass of the predominant species in each preparation, we compared their projected areas in the negative stained images to that of ferritin as carried out in previous studies (31). The projected area distributions of the TRF2/Rap1 complex showed a wider range than that of hTRF2 or hRap1 **(Figure 2.3D)** and presents a good example of a positively skewed Gaussian distribution **(Supplementary Figure 2.3D)** which could result, for example, from a cylindrical shape for the TRF2/Rap1 complex rather than spherical shape as observed in hTRF2 and hRap1 where the size distribution would be much more narrow. Analysis of 715-2135 particles (Experimental Procedures) yielded an estimated mass of 136 ± 0.4 kDa for hTRF2, 221 ± 0.3 kDa for hRap1 and 496 ± 1.0 kDa for the TRF2/Rap1 complex **(Table 2.1)**.

A.

B.

C.

Area Distribution

D.

Ferritin   TRF2   Rap1   TRF2/Rap1

E.

76-

-TRF2
- Rap1

52-
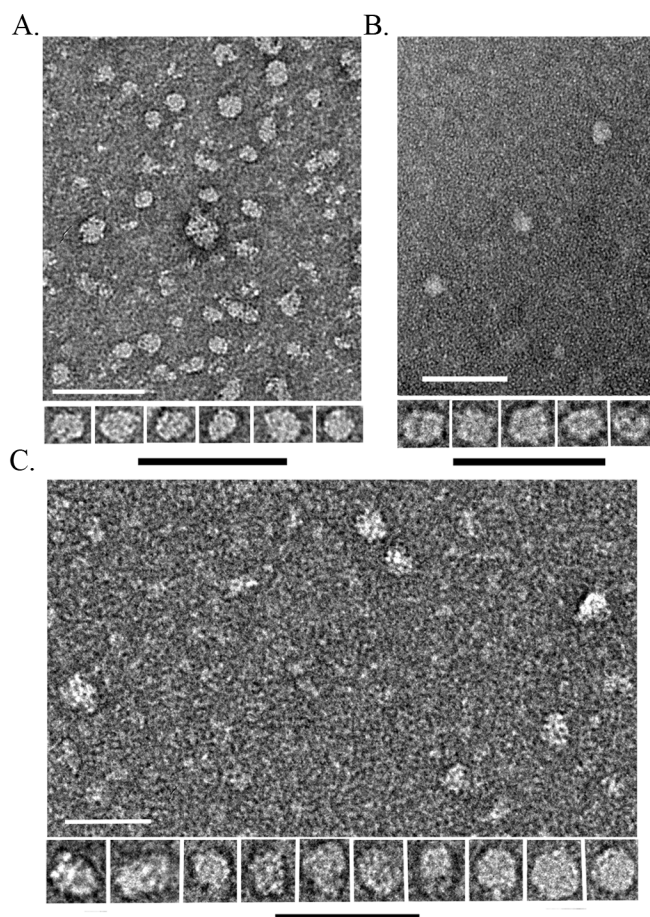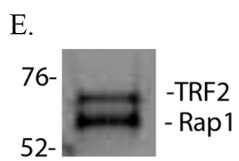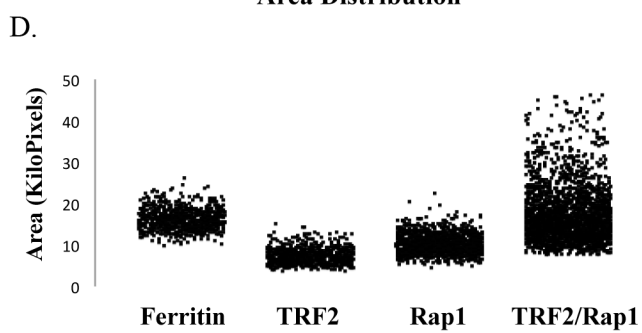
Figure 2.3: Mass and oligomeric state analysis of hRap1, hTRF2 and the hRap1/TRF2 complex. Representative negative stained images of hTRF2 (A), hRap1 (B) and the TRF2/Rap1 complex (C) are shown as fields and an array of selected single particles at higher magnification. Bars are equivalent to 200 nm. Area distributions of single protein particles and Ferritin as a size standard were calculated from 2D projections of the EM images (D). The TRF2/Rap1 complex was separated using 10% SDS-PAGE and stained with Coomassie Orange to determine the ratio of hRap1 to hTRF2 (E).

A. Ferritin

B. TRF2

C. Rap1

D. TRF2/Rap1

E.

	Ferritin	TRF2	Rap1	TRF2/Rap1
Mean (Area)	16423	7497	10378	17941
Standard error	88	72	64	174
95 % Confidence Interval	172	141	125	341
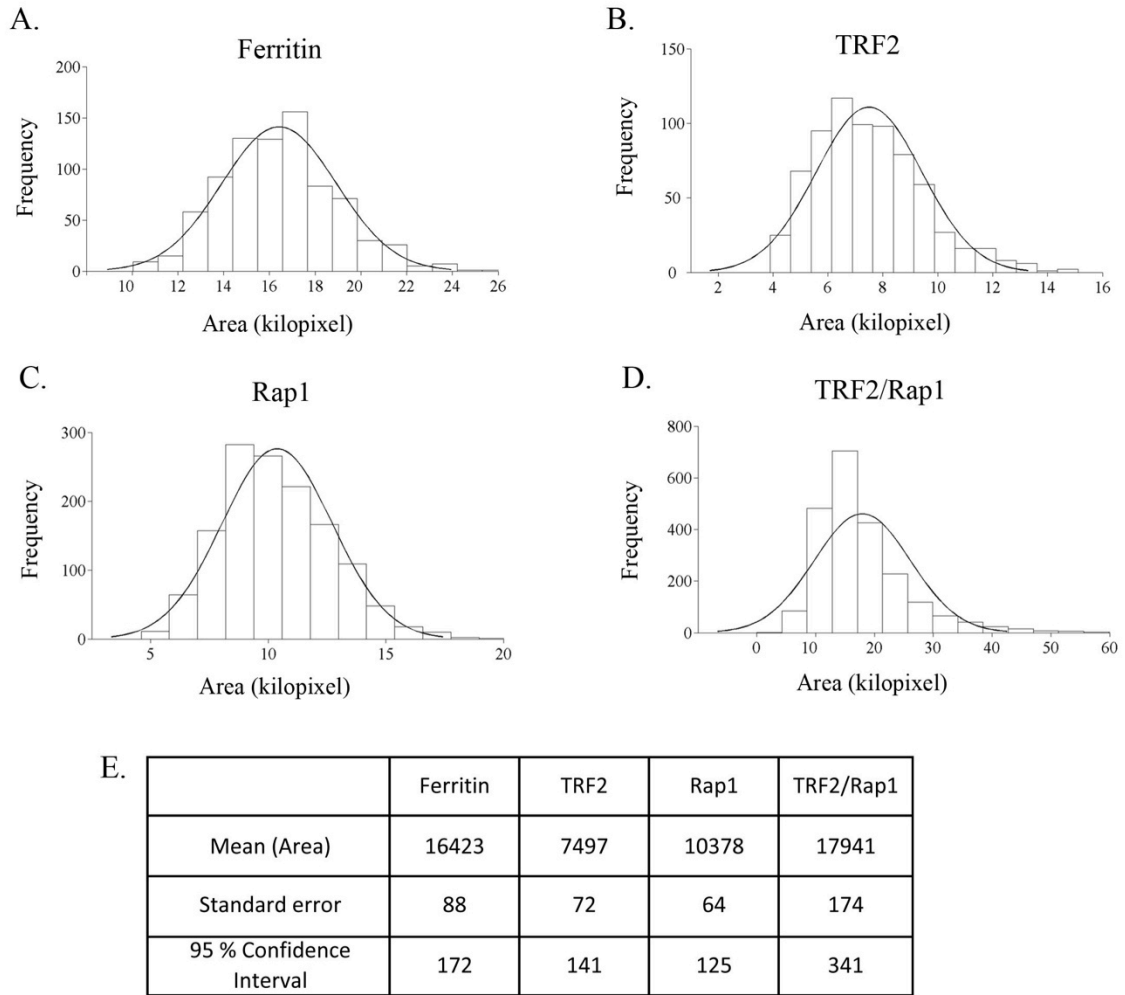
Supplementary Figure 2.3: Area distributions of hTRF2, hRap1 and the TRF2/Rap1 complex from 2D projections of negative stained EM images. The mean and standard error values derived from the frequency distributions of the areas with Gaussian fits of Ferritin (A), hTRF2 (B), hRap1 (C) and the TRF2/Rap1 complex (D) are summarized in table (E).

Size exclusion chromatography was used to confirm the EM results. The TRF2/Rap1 complex (Experimental Procedures), hRap1, and hTRF2 were passed through a Sepharose S6 matrix and each fraction analyzed by SDS-PAGE followed by Coomassie Blue or Coomassie Orange staining **(Supplementary Figure 2.4A and B)**. Based on parallel filtration of mass standards using the same column, hTRF2 eluted in a single peak equivalent to a mass of 121 kDa, and hRap1 eluted in a wider peak with a mass of 285 kDa. When hTRF2 was in a complex with hRap1 there was a significant shift of the peak in the elution profiles. The co-complex was found to elute as a wide peak with a mass of ~ 496-514 kDa likely due to the high glycerol concentration of the column buffer **(Supplementary Figure 2.4C)**. To determine the TRF2 to Rap1 ratio in the complex, SDS gel electrophoresis of the TRF2/Rap1 complex was carried out and the gel stained with a quantitative dye, Coomassie Orange. Analysis using Image J software (NIH) revealed that hTRF2 and hRap1 are in 1:1.17 ratio **(Figure 2.3E)** confirming previous findings (39). In summary, similar masses were obtained from the EM and gel filtration: 121 to 136 kDa for hTRF2, 221 to 285 kDa for hRap1, and 496 to 514 kDa for the TRF2/Rap1 complex. The mass estimates derived by EM and gel filtration are in good agreement, but in the future, other methods may be applied to further confirm these findings.

To deduce the oligomeric state corresponding to these masses, we used published protein values, determined by SDS-PAGE, as a reference (hTRF2 ~ 70 kDa, hRap1 ~ 60 kDa) (39). The calculated masses based on the DNA sequence do not take into account post-translational modifications and these can change the mass and shape of the protein and thus their mobility on size exclusion chromatography. Indeed, hTRF2 is known to migrate at a much higher apparent mass in gel filtration than its calculated mass and one explanation

offered is its flexible structure due to the presence of the long unstructured linker domain (21). The values equate to a dimer for TRF2, and a tetramer for hRap1. Even though various combinations are possible because hRap1 exists as a tetramer and hTRF2 is stable as a dimer in solution and they exist in 1:1 ratio, we conclude that the TRF2/Rap1 complex consists of a 4:4 complex consisting of one tetramer of hRap1 and either two dimers of hTRF2 or a tetramer of both proteins **(Table 2.1)**.

		Calculated Mass (kDa)	Expected Mass (kDa)	Oligomeric State
Size Exclusion Chromatography	hTRF2	121	140	Dimer
	hRap1	285	240	Tetramer
	TRF2/Rap1	514	520	Tetramer each of TRF2 and Rap1
Negative Staining	hTRF2	136	140	Dimer
	hRap1	221	240	Tetramer
	TRF2/Rap1	496	520	Tetramer each of TRF2 and Rap1

Table 2.1: Size analysis of hRap1, hTRF2 and the TRF2/Rap1 complex in solution. The estimated mass values obtained from the statistical analysis of the area distributions derived from negative staining as in Figure 3 and the size exclusion chromatography (supplementary figure 4) are shown with the corresponding oligomeric states.
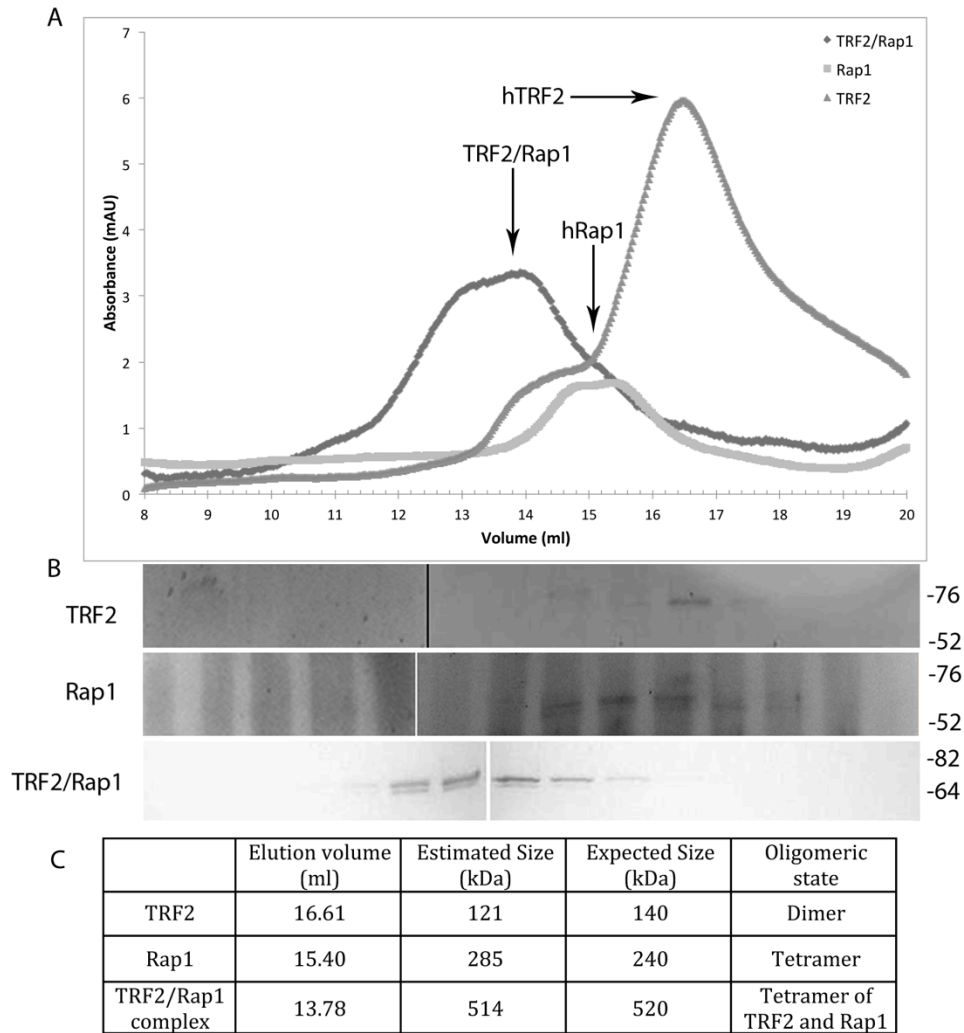
Supplementary Figure 2.4: Analysis of mass and oligomeric state of hTRF2, hRap1 and the TRF2/Rap1 complex by size exclusion chromatography. Elution profiles of each protein obtained from Superose 6 column are shown in (A) with the peak values pointed with arrows. Fractions corresponding to every 0.9 ml were loaded on to a 10% SDSPAGE and stained with either Coomassie Orange (hTRF2 and hRap1) or Coomassie Blue stain (the TRF2/Rap1 complex) (B). The oligomeric state, estimated mass of the each protein with the corresponding elution volumes and the expected size of each protein (C).

The TRF2/Rap1 complex recognizes 3- and 4- way junctions and binds to DNA in as a 4:4

complex

hRap1 and hTRF2 are present at human telomeres and have been identified in cell extracts in complex with each other. However, little is known about the binding preference of this complex on different templates or whether the affinity of the complex is different from its components. To examine the binding preference of the TRF2/Rap1 complex we used the same model junction templates described above and found that a molar ratio of 7:1 TRF2/Rap1 complex to DNA provided optimal binding for EM studies **(Figure 2.4 and Figure 2.5)**. The TRF2/Rap1 complex bound to the Gap25 and telomeric HJ DNAs in a manner similar to hRap1 and hTRF2 alone **(Figure 2.4A and B)**, and localized specifically to the junction of gap25 and telomeric HJ DNAs **(Figure 2.4C)**. No significant difference in binding or preference was observed on the telomeric HJ versus nontelomeric HJs **(Figure 2.4D)**. These results suggest that the 4:4 complex of TRF2/Rap1 can also recognize 3- and 4- way junctions independent of sequences.

When mass analysis was carried out on tungsten shadow cast proteins bound to telomeric HJ DNA **(Table 2.2)** (Corresponding Gaussian distributions are in **Supplementary Figure 2.5**), as described above, the masses obtained were 276 kDa for hRap1 and 476 kDa for the TRF2/Rap1 complex, which were consistent with the negative staining and gel filtration values. The only exception was that the estimated mass for hTRF2 was 315 kDa arguing that it is a tetramer when it is bound to DNA, consistent with the observation of Fouche et al (19). Therefore, hTRF2 and hRap1 bind to telomeric HJ as tetramers while the TRF2/Rap1 complex binds to telomeric HJs as a 4:4 complex of each protein.
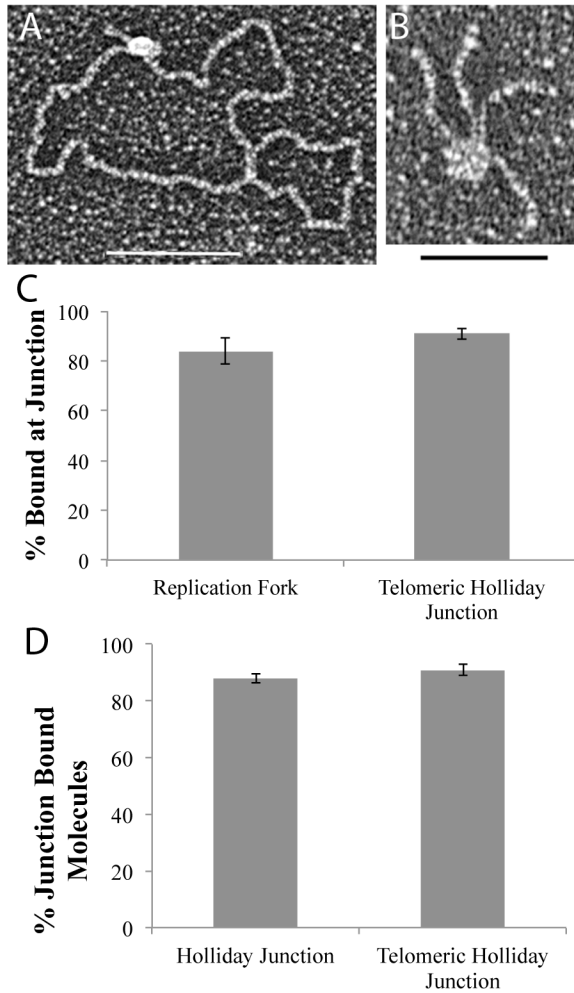
Figure 2.4: The hTRF2/Rap1 complex recognizes replication forks and Holliday junctions. The TRF2/Rap1 complex localizes to the junction site of the replication fork (A) and the telomeric HJ DNA (B). Junction preference of the protein bound molecules on the DNA templates is in (C), while (D) represents the sequence preference of the hTRF2/Rap1 complex on the HJ DNAs. Each EM binding reaction was done in triplicate and 100 molecules each were counted. * is p<0.05, ** is p<0.01 and p<0.001 is represented by ***.

A. Ferritin

B. TRF2

C. Ferritin

D. Rap1
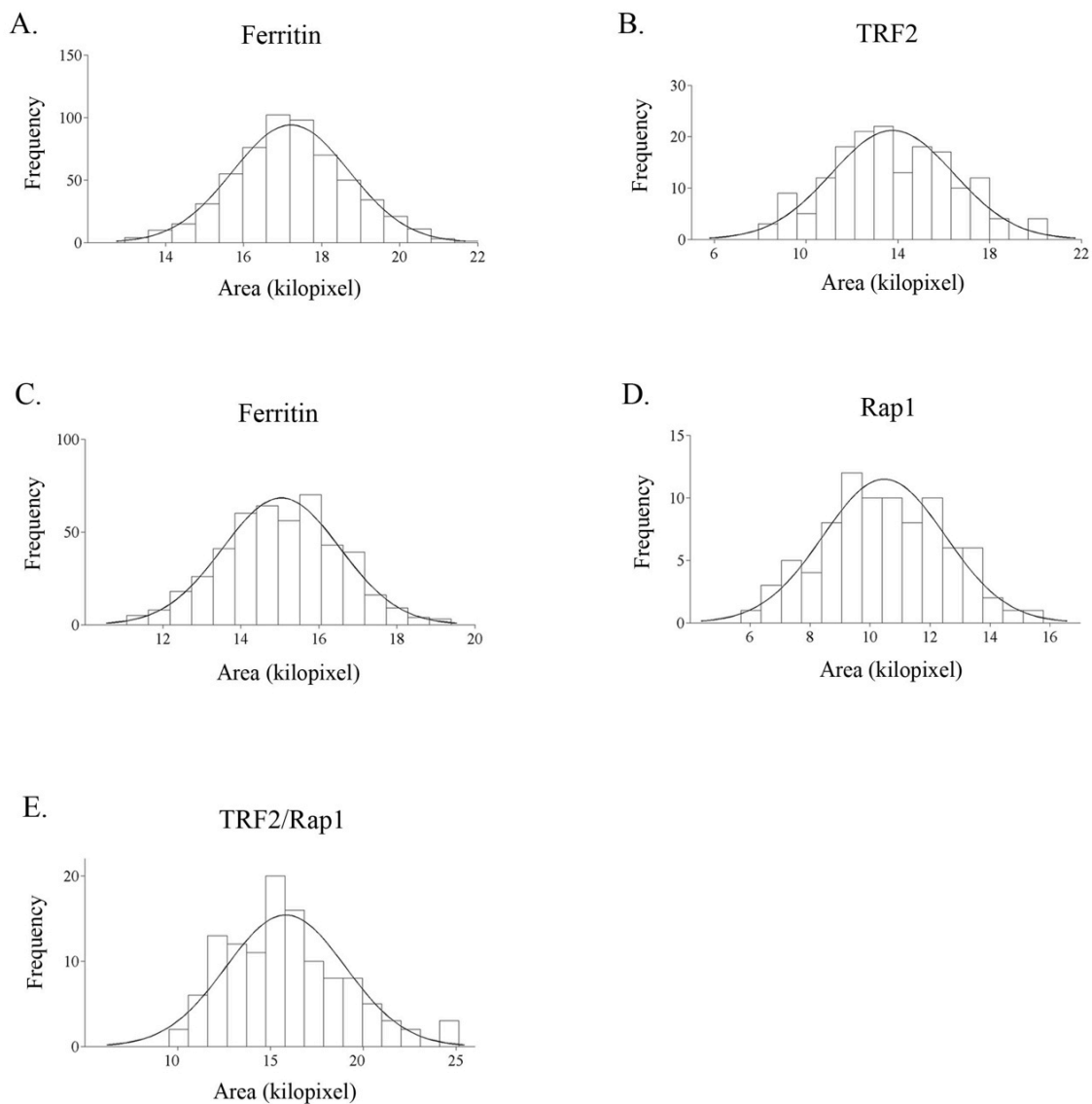
E. TRF2/Rap1

F.

	Ferritin	TRF2	Ferritin	Rap1	TRF2/Rap1
Mean (Area)	17211	13771	15038	10472	15823
Standard error	62	205	70	218	294
95% Confidence Interval	121	402	136	427	576

Supplementary Figure 2.5: The area distributions of hTRF2, hRap1 and the TRF2/Rap1 complex on telomeric Holliday Junction DNA. Gaussian fits on the 2D area projections, and the frequency distributions of each protein are shown in (A-E). Ferritin in (A) is the size standard for the hTRF2 mass analysis while Ferritin in (C) is the size standard for the mass analysis of hRap1 (D) and the TRF2/Rap1 complex (E). Table (F) summarizes the mean area, standard error and the 95% confidence interval for each protein.

	Sample	Number of molecules counted	Mean Area (pixel)	Calculated Mass (kDa)	95% Confidence Interval +/- (kDa)	Number of oligomers
	TRF2	168	13771	315	1.6	Tetramer
Bound to Telomeric Holliday Junction	Rap1	87	10472	256	2.1	Tetramer
	TRF2/Rap1 Complex	119	15823	475	3.3	Tetramer of TRF2 and Rap1

Table 2.2: Mass and oligomeric state of hRap1, hTRF2 and the TRF2/Rap1 complex on telomeric Holliday junctions. Mass, mean area and the oligomeric state of each protein were obtained from the 2D projection analysis of the tungsten shadowcast images as in Figures 2.1 and 2.4). Masses were calculated based on using ferritin as a size standard. Area distributions with Gaussian fits are shown in Supplementary Figure 2.5.

The TRF2/Rap1 complex has higher specificity for telomeric DNA and junction structures

than hTRF2 or hRap1 alone

Incubation of the TRF2/Rap1 complex with the minichromosome template revealed binding along the internal duplex telomeric region **(Figure 2.5A)**, at the ends with the ds-ss junction **(Figure 2.5B)**, or both **(Figure 2.5C)**. Since the duplex telomeric DNA constitutes 1/12 of the total DNA length from each end of the minichromosome molecule, we were able to verify binding at the ds telomeric regions by measuring the length of the DNA from its end to the protein-binding site. When protein molecules were examined in detail, from 60 examples, it appeared that the ds DNA passed through the TRF2/Rap1 complex as contrasted to binding on one side **(Figure 2.5E)**. On the minichromosomes, the major preference for binding was to the ds-ss junction site **(Figure 2.5F)**. However, ~ 1/3 of the bound complexes were localized to the internal duplex telomeric DNA, which is a significantly higher value than the percentage of hRap1 or hTRF2 molecules alone bound to the ds telomeric DNA on the minichromosome where the binding was at background levels for both proteins **(Figure 2.5G)**. These data suggest that in addition to binding to the DNA junction, the TRF2/Rap1 complex has a second binding site available to it on the duplex telomeric DNA; the complex binds more readily to this region than hTRF2 and hRap1 alone.

To examine the influence of sequence on TRF2/Rap1 complex binding we used the linear nontelomeric template with a 3' overhang used for hRap1 binding reactions (described above). Under the same reaction conditions, the amount of the TRF2/Rap1 complex bound to the nontelomeric template was 2.5-fold less than to the minichromosomes **(Figure 2.5H)**. On the minichromosome, the majority of the TRF2/Rap1 complexes were observed at the ds-ss junction site while on the nontelomeric linear template, no preferential binding was observed

(data not shown). These observations reveal that the TRF2/Rap1 complex has high specificity for telomeric sequences with a major preference being the ds-ss junction site at the end of the telomere.

The TRF2/Rap1 complex formed a new structure with the minichromosome template in which both ends of the DNA were joined by the protein complex into a circle: 9% of the DNA was observed in this form **(Figure 2.5D)**. Neither hRap1 nor hTRF2 alone exhibited this property (data not shown). Intermolecular bridges between two or more DNA molecules were present but at very low levels, likely reflecting the low DNA concentration used in the binding reaction. As the TRF2/Rap1 complex concentration was increased, the amount of circles, which might be a different form of t-loops, increased. When the TRF2/Rap1 complex concentration was titrated down to a lower level where hTRF2 no longer makes t-loops, the percentage of DNAs arranged into t-loops by the TRF2/Rap1 complex was 9% **(Figure 2.5I)**. This finding suggests that the TRF2/Rap1 complex has higher capacity to re-model telomeric DNA than TRF2 or Rap1 alone.
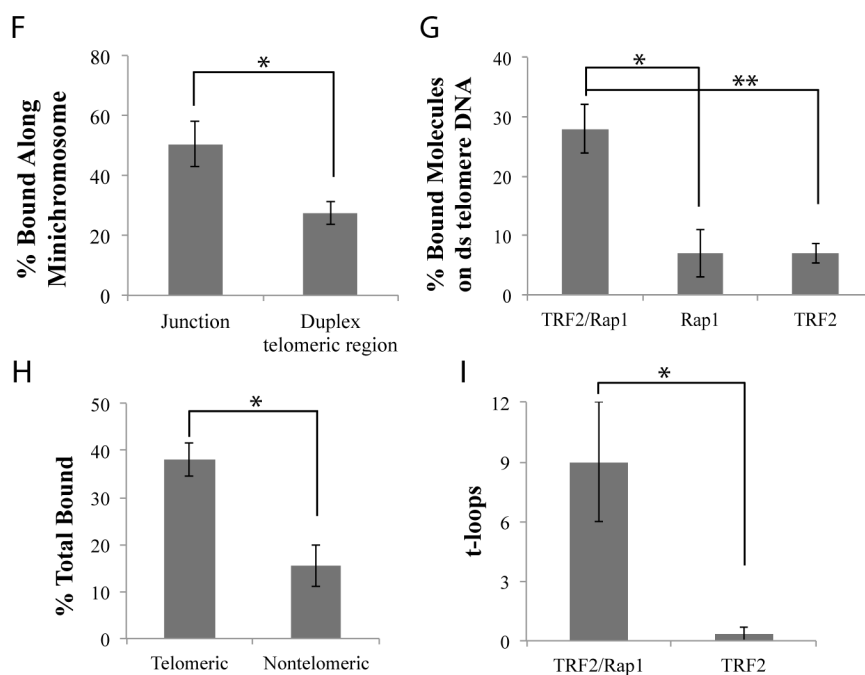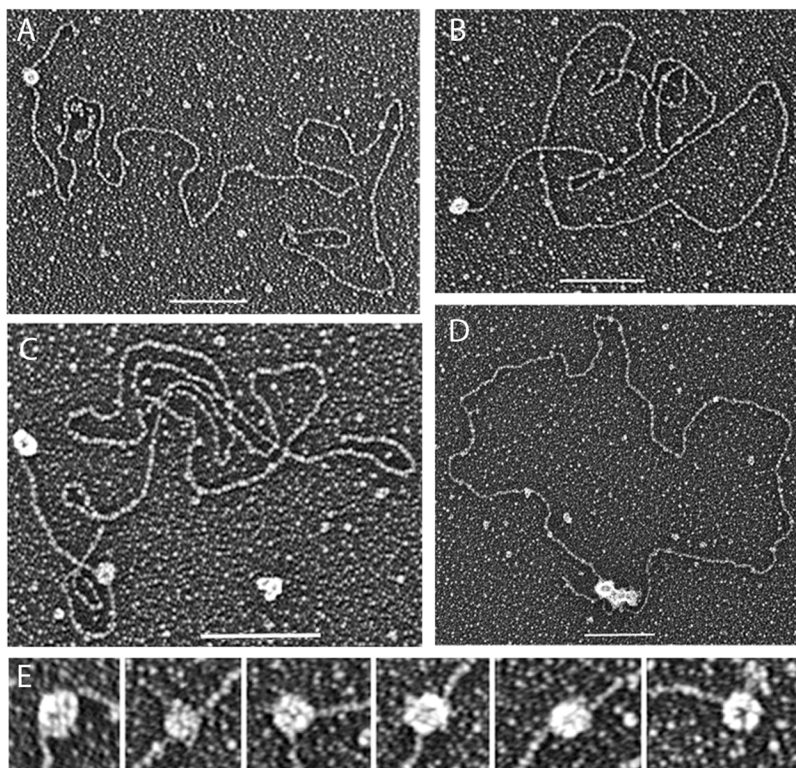
Figure 2.5: The binding properties of the TRF2/Rap1 complex along the minichromosome. Tungsten shadowcast images of the TRF2/Rap1 complex on the minichromosome illustrate the protein bound to the ds-ss junction and/or to the ds telomeric DNA (A-C). An example of the TRF2/Rap1 complex bringing the ends of the minichromosome together to form a circle is depicted in (D). Bars in (A-D) are equivalent to 100 nm. In high magnification images, the ds telomeric DNA appears to pass through the TRF2/Rap1 complex (E). Analysis of the binding preference of the TRF2/Rap1 complex along the minichromosome DNA is in (F), showing significant binding to the internal duplex telomeric sequences. (G) further compares the binding percentages of hRap1, hTRF2 and the complex to the ds telomeric DNA segments of the minichromosomes. The effect of DNA sequence on the binding of the TRF2/Rap1 is shown in (H). (I) demonstrates the comparison of the t-loops formed by the TRF2/Rap1 complex and hTRF2 at 10 nM protein concentration. Corresponding p values are shown by the different number of stars on each graph as in Figure 2.

Kd values for DNA binding confirm the EM observations

The higher specificity of the TRF2/Rap1 complex binding to telomeric sequences, observed by EM, could be due to a change in the substrate affinity when TRF2 and Rap1 form the 4:4 complex. To further explore this, we determined the *Kd* values of the two individual proteins and the complex on telomeric and nontelomeric DNA templates. A telomeric duplex DNA template of 154 bp with or without a 3' 54 mer G rich overhang and a nontelomeric template with a 57 bp ds region that has a 33 nt 3' overhang were prepared, incubated with TRF2, Rap1 or the TRF2/Rap1 complex in similar binding conditions to the EM assays and quantified as described in Experimental Procedures. **Table 2.3** summarizes the *Kd* values obtained for each template and protein, and the Hill coefficients determined for the telomeric 3' overhang DNA. Typical EMSA gels are shown in **Supplementary Figure 2.6** and graphs are shown in **Supplementary Figure 2.7**. Hill coefficients on the model telomere template are 2.4 ± 0.4 for hRap1, 5.3 ± 1.3 for hTRF2, and 3.3 ± 0.6 for the complex. All values are above 1.5 showing that each protein has positive cooperativity of binding. hRap1 has the lowest affinity for ds telomeric DNA (*Kd* of 265 ± 57 nM), and has very similar affinity for telomeric and nontelomeric 3' overhang structures (119 ± 12 nM and 161 ± 25.3 nM). This data is consistent with the EM observations and confirms that hRap1 prefers ds-ss junction sites irrespective of sequence. hTRF2 binds 3 times tighter to ds telomeric regions than hRap1 and both hRap1 and hTRF2 have similar affinities for nontelomeric junctions (161 ± 25.3 and 133 ±16.36 nM). The *Kd* of hTRF2 on ds telomeric DNA is lower than the previously published value of 180 nM (40). The major reason for the tighter binding we observed could be the higher pH of the binding buffer, which inhibited TRF2 aggregation in solution. When telomeric sequences are present, the affinity of hTRF2

for junction sites is nearly 2 times that of hRap1. Consistent with EM, hTRF2 prefers telomeric junction sites over nontelomeric while hRap1 has a slight preference for telomeric junction sites.

When both proteins were present in a 4:4 complex, we observed a great increase in the affinity for telomeric DNA. The binding affinity increased a minimum of 2-fold on ds telomeric DNA and ~10 fold on the model telomere structure containing a 3' overhang. The TRF2/Rap1 complex has a 5 fold higher affinity for 3' ds-ss junctions than ds telomeric regions. The TRF2/Rap1 complex cannot bind to the nontelomeric 3' overhang structures even if the molarity of protein used in the reaction was up to 3 times the molarity in other binding reactions. The significant increase in affinity of the TRF2/Rap1 complex for internal duplex telomeric regions and for telomere 3' overhang structures explains the dual binding of the TRF2/Rap1 complex along the minichromosome.

	Hill coefficient	K_D (nM)		
		Telomeric 3' overhang	Duplex Telomere Template	Nontelomeric 3' overhang
hRap1	2.4 ± 0.4	118.9 ± 12.1	265.0 ± 57.2	161.3 ± 25.3
hTRF2	5.3 ± 1.3	64.7 ± 3.4	85.5 ± 11.6	133.0 ± 16.4
hTRF2/hRap1	3.3 ± 0.6	5.8 ± 0.3	37.0 ± 9.1	No Binding

Table 2.3: Affinity of hRap1, hTRF2 and the TRF2/Rap1 complex on telomeric and nontelomeric DNA templates. The dissociation constants of hRap1, hTRF2 and the TRF2/Rap1 complex on the telomeric and nontelomeric DNA templates with a 3' overhang or on the duplex telomere DNA are presented derived from electrophoretic experiments shown in the supplementary figures 6 and 7. The Hill coefficient values of each protein on the model telomeric DNA is shown with the error values.
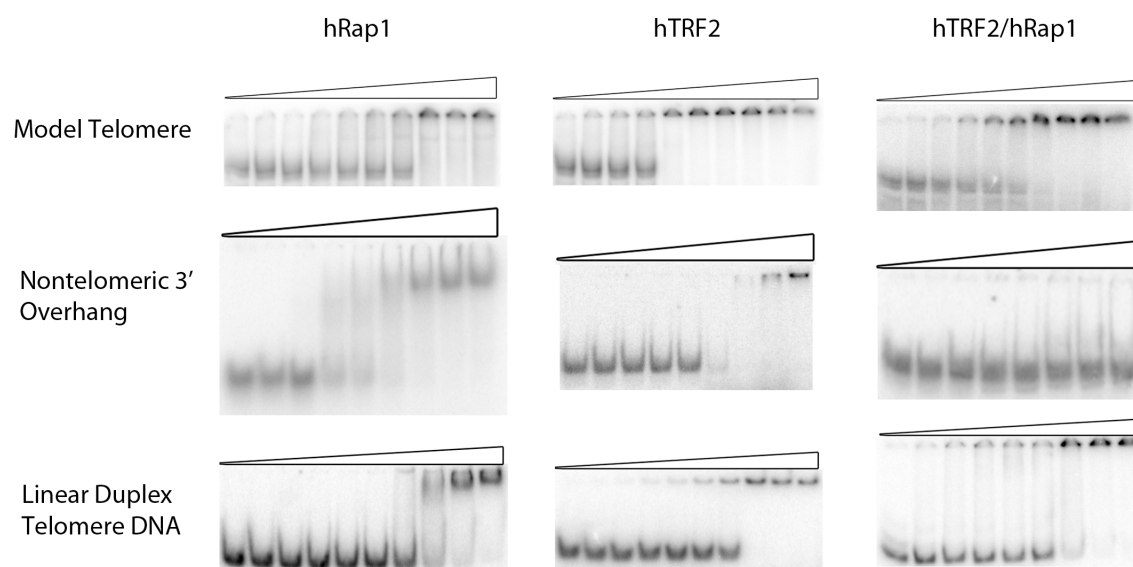
Supplementary Figure 2.6: EMSA analysis of hTRF2, hRap1 and the TRF2/Rap1 complex on radiolabeled templates. Linear duplex telomeric DNA and the nontelomeric 3' overhang template were labeled from their 5' ends with ^{32}P while the model telomeric DNA with a 3' overhang was labeled from its 3'overhang (Experimental Procedures). The binding mixtures were performed in parallel and electrophoresed on 4% native polyacrylamide gels. The results are shown over different ranges of protein concentration due to large differences in the binding affinity. The concentrations of hRap1 in the DNA binding reactions were up to 458 nM (model telomere DNA), 3.2 µM (nontelomeric 3' overhang DNA), and 3.7 µM (linear duplex telomere DNA) while the concentrations of hTRF2 in the reactions were up to 1.5 µM for all DNA templates and the concentrations of the TRF2/Rap1 complex were up to 15 nM (model telomere DNA) and 62 nM (nontelomeric 3' overhang DNA and linear duplex telomere DNA).
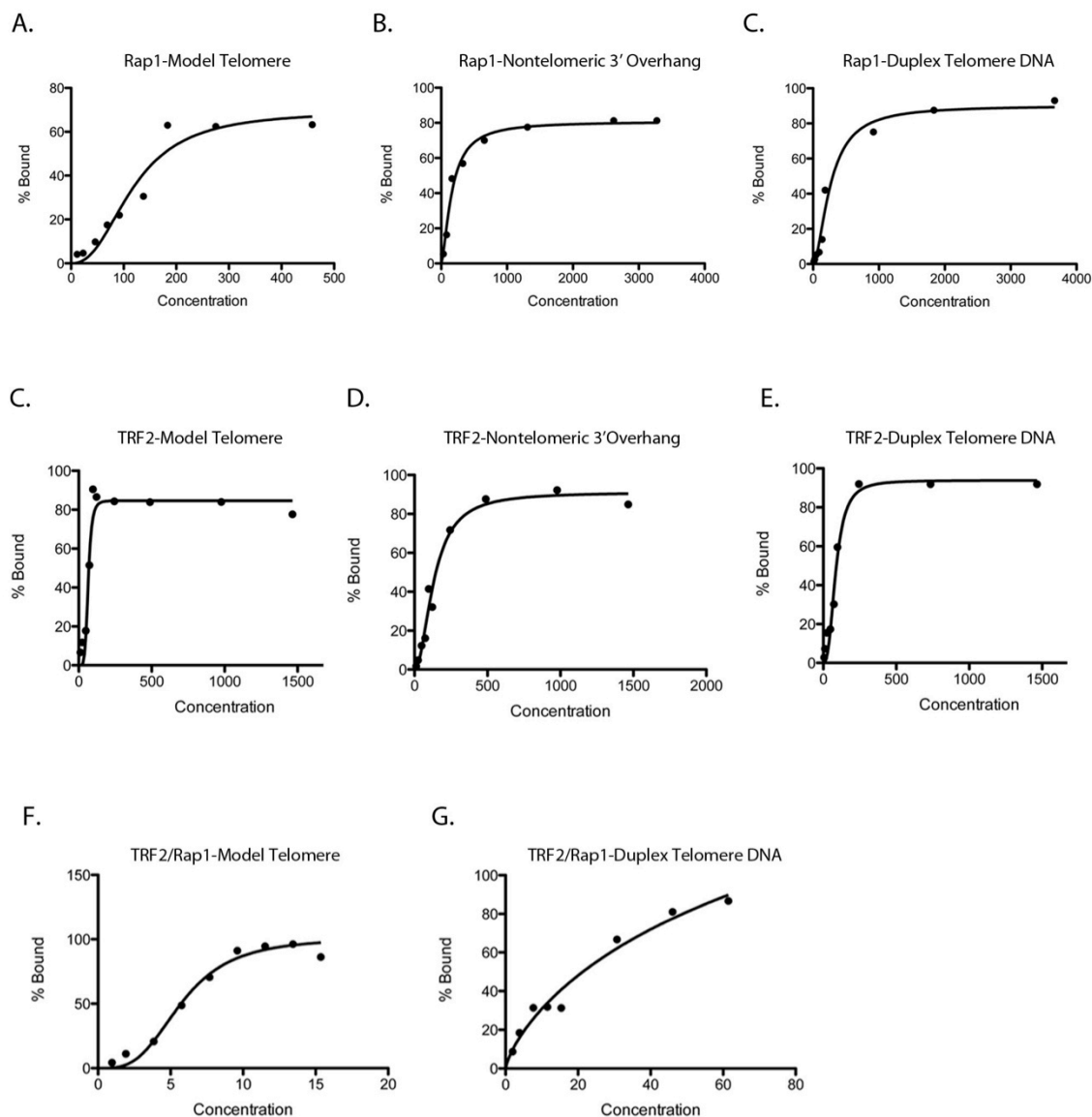
Supplementary Figure 2.7: Binding affinity of hRap1, hTRF2 and the complex on different DNA templates. Curves shown were obtained by fitting data points obtained from EMSAs with one site specific Hill plot (Graphpad prism software) except that the TRF2/Rap1 complex on duplex telomere DNA data points were fit with one site specific binding without Hill plot. Concentrations are shown in nM.

DISCUSSION

To understand the role of hRap1 at telomeres, we studied the DNA binding characteristics of hRap1 and the TRF2/Rap1 complex on different DNA templates that mimic telomeric structures. In this study, we demonstrated that hRap1 directly interacts with ds-ss DNA junctions in the absence of hTRF2. The specificity of hRap1 for junction sites is structure specific rather than sequence specific. The dissociation rates of hRap1 and hTRF2 at nontelomeric 3' ds-ss junction sites are similar in value but in the presence of adjoining duplex telomeric DNA, hTRF2 shows tighter binding. Each protein binds to DNA as a tetramer and with positive cooperativity. The TRF2/Rap1 complex is ~ 500 kDa and consists of 4 molecules each of TRF2 and Rap1. When hRap1 and hTRF2 form a complex, the binding specificity for telomeric DNA increases significantly. As the affinity for ds telomeric DNA increases by more than 2-fold, the affinity for the 3' telomeric ds-ss junction sites increases by more than 10-fold. Interestingly, the affinity of the TRF2/Rap1 complex for linear nontelomeric DNA with a 3' overhang is lost, indicating that upon complex formation, there is a significant change in substrate specificity.

In this work we found that hRap1 directly binds to DNA and has a preference for ds-ss junction sites. However, it was previously reported that hRap1 does not bind to DNA on its own but rather needs hTRF2 for binding (10). The reason for this contradiction is most likely the different sets of DNA templates employed. Li *et al* tested the DNA binding ability of hRap1 using a 72 bp DNA template with ds human telomeric repeats or ds and ss yeast telomeric DNA templates and did not observe binding upon addition of hRap1. They suggested that the lack of binding was due to the neutral character of the myb domain. We used much longer DNA templates and ones with ds-ss junction structures or with 3- and 4-

way junctions. We observed that hRap1 does not bind ds telomeric DNA but has high preference for ds-ss junctions. Moreover, hRap1 has similar *Kd* values for telomeric and nontelomeric ds-ss junctions demonstrating that its myb domain does not directly bind to telomeric duplex DNA, and NMR studies revealed that the myb domain does not contain charged residues on the surface (41) in contrast to the myb domain of hTRF2, which is positively charged (40). On the other hand, the affinity of hTRF2 for ds-ss ends increases in the presence of telomeric sequences, which could reflect the presence of its positively charged myb domain. Thus, hRap1 appears to recognize ds-ss junction structures in the absence of hTRF2 and without a sequence preference.

Previously, Li *et al* failed to detect a change in the dissociation rate of hTRF2 when TRF2 is in complex with hRap1. With a similar template we observed a 2-fold increase in binding affinity for ds telomeric DNA and 10-fold increase in binding affinity for 3' telomeric ds-ss ends. The reason of the difference may lie in the assay conditions. In their assay, hTRF2 and hRap1 were added to the ds telomeric DNA sequentially and DNA was already bound by hTRF2 when hRap1 was added. However, in our assays, the TRF2/Rap1 complex was formed first. Consistent with their finding (10), we did not observe any positive cooperativity of binding on a duplex telomere template. In contrast, the TRF2/Rap1 complex shows positive cooperativity of binding in the presence of a 3' overhang structure and the affinity of binding increases by 10-fold compared to TRF2 alone.

We observed that both hRap1 and hTRF2 have the ability to bind to nontelomeric junction sites with similar affinities, raising the question what domain is responsible for the DNA binding ability of hRap1. We previously showed that the N terminal basic domain facilitates the binding of hTRF2 to junction structures independent of sequence through a 16

amino acid stretch with 8 positive charges (19). Similarly, an 18 amino acid stretch with 5 positive charges is present at the N terminal BRCT domain of hRap1 (10). This is consistent with the finding of a single BRCT domain involved in the DNA binding of TopBP1 (42), XRCC1 (43) and Replication factor C (44) to ds-ss junctions. Thus, the BRCT domain of hRap1 may account for its binding to these ds-ss junctions. Alteration of telomere length upon deletion of the BRCT domain in Rap1 underlines the importance of this domain for its function (11,45). In the future, further analysis with deletion mutants should help elucidate the role of the BRCT domain of hRap1, including how the binding of hRap1 to DNA is affected by its interaction with hTRF2, and how the binding properties of hTRF2 to DNA are affected when it is complexed with hRap1. It is always possible that the effect of hRap1 on hTRF2 is dependent on its DNA binding activity and that the increase in its affinity for telomeric DNA may be due to allosteric interaction between hRap1 and hTRF2. Moreover, the DNA binding activity of hRap1 may be important for non-telomeric functions at internal sites on the chromosomes, but at the telomere, requires hTRF2 as a partner.

hRap1 binds to junction sites and also modulates the binding of hTRF2 along linear telomeric DNA by facilitating a positive cooperativity of binding and by increasing its affinity for telomeric sequences. Moreover, we found that the duplex telomeric region is required for binding of the complex since even if the 3' overhang structure was present, the TRF2/Rap1 complex could not bind when duplex telomeric sequences were absent. Indeed the TRF2/Rap1 complex bound to the ds telomeric DNA in the absence of a 3' ds-ss junction with a ~ 2 fold higher affinity than the individual components. The observation that one fourths of the TRF2/Rap1 complexes localized within the ds telomeric region and three fourth of the TRF2/Rap1 complexes were at the ds-ss junction site (or both at the junction

site and the ds telomeric DNA) indicate that once the 3' overhang is available, the TRF2/Rap1 complexes can slide and until it reaches the ds-ss junction where it binds tighter. These observations together lead to the following model.

During telomere extension, the newly synthesized DNA will be exposed and needs to be coated with histones and/or shelterin components that bind to the ds telomeric DNA. The 3' overhang is generated by the end processing machinery later in S-phase, and thus should not be available as a binding site (46). We propose that **(Figure 2.6)** TRF2 and Rap1 bind to the ds telomeric DNA as a complex due to the higher affinity of the complex for ds telomeric DNA than either individual component. Once the 3' overhang is formed the TRF2/Rap1 complex slides to the junction site and it remain due to its ~ 6-fold lower *Kd* value for the junction over duplex telomeric DNA, and thus covers the open telomere end. On long telomeres the TRF2/Rap1 complex could initiate t-loop formation and prevent DDR or as the telomere length becomes shorter (to a level where the t-loops cannot be formed) then with its very high affinity for the 3' ds-ss telomeric end, the TRF2/Rap1 complex could suppress the unnecessary DDR by blocking the end and making it less accessible to DNA repair proteins. Indeed, the finding that very short telomeres without t-loops exist stably *in vivo* (47), that the TRF2/Rap1 complex can prevent NHEJ of telomeres with less than 10 repeats and in the absence of Rap1, can only be compensated by significantly high amounts of hTRF2 to facilitate end protection (12,48) provide support to this model. Depending on what parameter one measures, the loss of Rap1 in cells may be considered to be important for end protection or not (12,28). This illustrates importance of further work that will better link end protection in vivo with the physical complexes and structures formed at telomere by the shelterin.
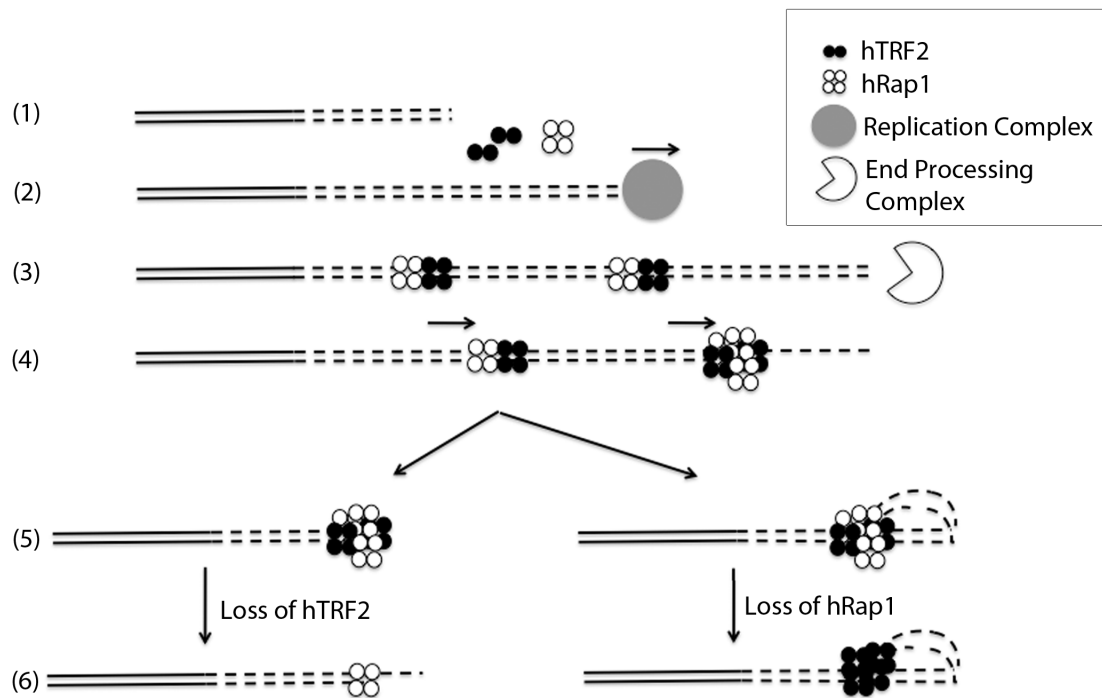
Figure 2.6: Model for the TRF2/Rap1 complex loading onto and protecting the telomere. (1) The nontelomeric and the telomeric DNA are presented in full lines and in dashed lines, respectively. (2) The new telomeric DNA is synthesized in S-phase. (3) hTRF2 and hRap1 bind to the newly synthesized ds telomeric DNA as a complex and then end resection occurs forming the 3' overhang. (4) The TRF2/Rap1 complexes slide along the ds telomeric DNA towards the newly formed 3' overhang. (5) End protection by the TRF2/Rap1 complex occurs through t-loop formation or simply by blocking the exposed end. (6) In the absence of hTRF2, hRap1 binds to the junction site to prevent NHEJ (or HDR in mice) and in the absence of hRap1, hTRF2 prevents DDR by forming t-loops.

REFERENCES

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

2. Liu, D., O'Connor, M. S., Qin, J., and Songyang, Z. (2004) *J Biol Chem* **279**, 51338-51342

3. Palm, W., and de Lange, T. (2008) *Annu Rev Genet* **42**, 301-334

4. Hsu, H. L., Gilley, D., Galande, S. A., Hande, M. P., Allen, B., Kim, S. H., Li, G. C., Campisi, J., Kohwi-Shigematsu, T., and Chen, D. J. (2000) *Genes Dev* **14**, 2807-2812

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

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

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

8. Bianchi, A., Stansel, R. M., Fairall, L., Griffith, J. D., Rhodes, D., and de Lange, T. (1999) *EMBO J* **18**, 5735-5744

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

10. Li, B., Oestreich, S., and de Lange, T. (2000) *Cell* **101**, 471-483

11. O'Connor, M. S., Safari, A., Liu, D., Qin, J., and Songyang, Z. (2004) *J Biol Chem* **279**, 28585-28591

12. Bae, N. S., and Baumann, P. (2007) *Mol Cell* **26**, 323-334

13. Bombarde, O., Boby, C., Gomez, D., Frit, P., Giraud-Panis, M. J., Gilson, E., Salles, B., and Calsou, P. *EMBO J* **29**, 1573-1584

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

15. Konishi, A., and de Lange, T. (2008) *Genes Dev* **22**, 1221-1230

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

17. Denchi, E. L., and de Lange, T. (2007) *Nature* **448**, 1068-1071

18.   d'Adda di Fagagna, F., Reaper, P. M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N. P., and Jackson, S. P. (2003) *Nature* **426**, 194-198

19.   Fouche, N., Cesare, A. J., Willcox, S., Ozgur, S., Compton, S. A., and Griffith, J. D. (2006) *J Biol Chem* **281**, 37486-37495

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

21.   Court, R., Chapman, L., Fairall, L., and Rhodes, D. (2005) *EMBO Rep* **6**, 39-45

22.   Mao, Z., Seluanov, A., Jiang, Y., and Gorbunova, V. (2007) *Proc Natl Acad Sci U S A* **104**, 13068-13073

23.   Sarthy, J., Bae, N. S., Scrafford, J., and Baumann, P. (2009) *EMBO J* **28**, 3390-3399

24.   Takai, K. K., Hooper, S., Blackwood, S., Gandhi, R., and de Lange, T. *J Biol Chem* **285**, 1457-1467

25.   Martinez, P., and Blasco, M. A. *Nat Rev Cancer* **11**, 161-176

26.   Yang, D., Xiong, Y., Kim, H., He, Q., Li, Y., Chen, R., and Songyang, Z. *Cell Res* **21**, 1013-1027

27.   Chen, Y., Rai, R., Zhou, Z. R., Kanoh, J., Ribeyre, C., Yang, Y., Zheng, H., Damay, P., Wang, F., Tsujii, H., Hiraoka, Y., Shore, D., Hu, H. Y., Chang, S., and Lei, M. *Nat Struct Mol Biol* **18**, 213-221

28.   Sfeir, A., Kabir, S., van Overbeek, M., Celli, G. B., and de Lange, T. *Science* **327**, 1657-1661

29.   Subramanian, D., and Griffith, J. D. (2005) *The Journal of biological chemistry* **280**, 42568-42572

30.   Compton, S. A., Ozgur, S., and Griffith, J. D. (2010) *The Journal of biological chemistry* **285**, 13349-13356

31.   Compton, S. A., Tolun, G., Kamath-Loeb, A. S., Loeb, L. A., and Griffith, J. D. (2008) *The Journal of biological chemistry* **283**, 24478-24483

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

33.   Kim, S. H., Kaminker, P., and Campisi, J. (1999) *Nature genetics* **23**, 405-412

34.   Griffith, J. D., and Christiansen, G. (1978) *Annual review of biophysics and bioengineering* **7**, 19-35

35.   Griffith, J. D., Makhov, A., Zawel, L., and Reinberg, D. (1995) *Journal of molecular biology* **246**, 576-584

36.   Lee, S., Cavallo, L., and Griffith, J. (1997) *The Journal of biological chemistry* **272**, 7532-7539

37.   Ye, J. Z., Donigian, J. R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A. N., Chait, B. T., and de Lange, T. (2004) *The Journal of biological chemistry* **279**, 47264-47271

38.   Oganesian, L., and Karlseder, J. (2011) *Molecular cell* **42**, 224-236

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

40.   Hanaoka, S., Nagadoi, A., and Nishimura, Y. (2005) *Protein science : a publication of the Protein Society* **14**, 119-130

41.   Hanaoka, S., Nagadoi, A., Yoshimura, S., Aimoto, S., Li, B., de Lange, T., and Nishimura, Y. (2001) *Journal of molecular biology* **312**, 167-175

42.   Yamane, K., and Tsuruo, T. (1999) *Oncogene* **18**, 5194-5203

43.   Yamane, K., Katayama, E., and Tsuruo, T. (2000) *Biochemical and biophysical research communications* **279**, 678-684

44.   Kobayashi, M., Figaroa, F., Meeuwenoord, N., Jansen, L. E., and Siegal, G. (2006) *The Journal of biological chemistry* **281**, 4308-4317

45.   Li, B., and de Lange, T. (2003) *Molecular biology of the cell* **14**, 5060-5068

46.   Chow, T. T., Zhao, Y., Mak, S. S., Shay, J. W., and Wright, W. E. (2012) *Genes & development* **26**, 1167-1178

47.   Capper, R., Britt-Compton, B., Tankimanova, M., Rowson, J., Letsolo, B., Man, S., Haughton, M., and Baird, D. M. (2007) *Genes & development* **21**, 2495-2508

48.   Xu, L., and Blackburn, E. H. (2007) *Molecular cell* **28**, 315-327

CHAPTER 3: STRUCTURAL CHARACTERIZATION OF LONG G-RICH TELOMERIC SINGLE STRANDED DNA AND FILAMENTS FORMED BY HUMAN POT1 AND THE HUMAN POT1-TPP1 COMPLEX

SUMMARY

G-quadruplexes have various functions in the cells, ranging from regulation of telomerase to controlling expression of onco-genes such as c-Myc, c-Kit, KRAS, pRb and Bcl-2. Understanding the structure of the G-quadruplexes is key to understanding their function. Short oligonucleotides are of limited use in modeling the G-quadruplexes formed during human telomere extension or telomeric transcription, which are of much greater size. Therefore we have utilized an *in vitro* telomere extension system to generate high molecular weight ssDNA capable of forming G-quadruplexes. Electron Microscopy was used to determine the structure of this long human G-rich DNA and its interaction with hPot1 and the Pot1/TPP1 complex. Here we provide the first evidence of G-quadruplex motifs as uniform beads, and with each bead being composed of ~487 nucleotides. These are the longest quadruplex structures yet identified. Moreover, we show that hPot1 forms filaments along the single stranded DNA as it opens up the G-quadruplexes. Similarly, the hPot1/TPP1 complex also form filaments and these filaments are more rigid than hPot1 filaments indicating that the DNA binding properties of hPot1 are changed upon complex formation with hTPP1. At the 3' telomeric overhang hPot1 wraps the G-rich DNA around itself. These findings together can explain how hPot1 relieves replicative stress of C-strand replication and how the Pot1-

TPP1 complex can act as telomerase processivity factor while hPot1 can protect telomeric end from exonuclease action.

INTRODUCTION

Runs of Guanines in DNA and RNA molecules tend to form higher order structures, called G-quadruplexes (1). G-quadruplexes arise from G-tetrads (G4) which are four guanines assembled by Hoogsteen base pairing (2). Computational analysis revealed that there are >375,000 G-quadruplex motifs in the human genome and these motifs cluster in specific regions frequently at oncogene promoters, 5' untranslated regions (UTRs), mitotic/meiotic DSB sites, ribosomal DNA, minisatellites and immunoglobulin heavy chain switch region and telomeres (3-5). The presence of G-quadruplexes at distinct sites suggests that G-quadruplexes may be important regulators of biological functions including gene expression, recombination and telomere extension. Therefore, these higher order structures are a significant target for drug design.

Human telomeres end with a G-rich single stranded 3' overhang of 100-200nt in length (6). Oligonucleotides with telomeric sequences show banding patterns in non-denaturing polyacrylamide gels suggesting that telomeric overhangs can form G4 *in vitro* (7). Quadruplexes formed within the telomeric overhang inhibit the activity of telomerase, which is overexpressed in ~85% of cancers (8). The telomere binding protein Pot1 binds to the telomeric overhang (9), prevents unnecessary activation of ATR dependent double strand break repair (DSBR) at chromosome ends (10) and regulates G-quadruplex formation at the telomeric ends (11). Telomere uncapping and telomere length phenotypes that arise in the absence of hPot1 reveal the importance of hPot1 in telomere end protection (12,13). Despite

hPot1 having two oligonucleotide/oligosaccharide (OB) folds, TPP1 is required for the correct localization of hPot1 to telomeres (14) increasing its affinity for telomeric DNA by 10-fold *in vitro* (15,16). Moreover, TPP1 recruits telomerase *in vivo* (16) while the hPot1-TPP1 complex increases telomerase processivity *in vitro* (15) suggesting an important role of the hPot1-TPP1 complex during telomere length maintenance.

The opportunity for Hoogsten base-pairing frequency increases during DNA replication, transcription and recombination (17) as duplex DNA is melted and the simple strands are exposed. This may be particularly the case during DNA replication where gaps transiently occur on the lagging strand that would increase the chance of G4 formation. G4 DNA cannot act as a template for the replication machinery and this may cause replication fork stalling, which in turn results in replicative stress (18). Telomeric DNA has an especially high risk of G4 formation during telomere replication because of its G-rich repetitive nature and propensity of the fork to slip backwards (19). Therefore, mammalian telomeres resemble fragile sites and additional factors like TRF1 (20) and/or RecQ helicases are required for their efficient replication (21,22). In WRN (a RecQ helicase) deficient cells, hPot1 is required for efficient telomere C-rich strand replication (23). However, the exact mechanism how hPot1 compensates for WRN loss remains unknown. To understand the structure of G quadruplexes formed at telomeres during replication and telomeric transcription, we analyzed the structure of long single stranded human telomeric DNA and asked how hPot1 and hPot1-TPP1 complexes regulate these higher ordered structures.

EXPERIMENTAL PROCEDURES

DNA Templates- The model telomere template was synthesized as previously described (24). Briefly, a plasmid (pOST6) was linearized with BsmBI and NotI restriction enzymes (New England Biolabs Inc., Ipswich, MA) and different sized oligonucleotides (see below) were ligated to one end to generate a duplex DNA with either a 54, 90 or 120 nt 3' terminated single stranded tail (IDT, Coralville, IA) (24).

To create the 5'-Flap DNA substrate the plasmid pRST5Nick (19) containing a C-less cassette was nicked using Nb.BbvCI (New England Biolabs Inc.). A 575 nt ssDNA 5' Flap was displaced using Klenow fragment (exo-) (New England Biolabs Inc.) in a strand displacement reaction in the presence of 0.5 mM of dTTP, dATP and dGTP (19). The template was cleaned using Zymo clean and concentrator spin columns (Zymo Research, Irvine, CA).

Telomeric ssDNA circles were synthesized using oligonucleotides with either G-rich or C-rich sequences (see below) containing 20 telomeric repeats (IDT, Coralville, IA). Oligos of 124 or 120 nt were ligated using CircLigase™ and the remnant linear oligos were digested with ExoI and Exo III (New England Biolabs Inc.) according to the manufacturers instructions (Epicentre Biotechnologies, Madison, WI). Circular products were confirmed by visualization on a 7% denaturing PAGE.

To synthesize the long single stranded G-rich DNA template, a model telomere template with a 3' G-rich overhang was annealed to 120mer C-rich circles in 1:10 molar excess and extended by Φ29 polymerase for 16 hours as described by the manufacturer (New England Biolabs Inc.). For EM analysis, the newly synthesized DNA was deproteinized with 20 μg proteinase K (Roche, Indianapolis, Indiana), 0.02M EDTA and 0.4% SDS in a 50 μl volume

and passed over a 2ml column of 2% agarose beads (Agarose Bead Technologies, Spain). Long C-rich DNA was synthesized using 124mer G-rich circles and 120mer C-rich complementary oligonucleotide similar to as described above.

Oligonucleotide sequences-

58mer: 5'-phosAGGG(TTAGGG)$_9$-3',

94 mer: 5'-phosAGGG(TTAGGG)$_{15}$-3',

124mer: 5'-phosAGGG(TTAGGG)$_{20}$-3',

120mer: 5'-phos(AATCCC)$_{20}$-3'

Purification of human Pot1- Baculovirus for GST-hPot1v1 was kindly provided by Ming Lei (25). For protein expression, 500ml of SF9 insect cells in Sf-900 II SFM media (Invitrogen GIBCO, Grand Island, NY). 1x10^6 cells per ml were infected with a multiplicity of infection (MOI) of 1. After 48hrs cells were pelleted and stored at -80°C. Cells were lysed in 50ml of Buffer A (50mM Tris 7.6, 400mM NaCl, 0.1% NP40 and 10% glycerol) for 30 minutes on ice and the lysate was sonicated and cleared at 15000 rpm for 1 hour. The cleared lysate was incubated with 2ml of glutathione beads (Clontech Laboratories Inc., Mountain View, CA) pre-equilibrated with buffer A for 1 hour at 4°C followed by extensive washing with Buffer B (50mM Tris pH7.6, 150mM NaCl, 10% glycerol). Untagged hPot1 was cleaved from the glutathione beads by adding protease 3°C (10 units) to the wash buffer and incubated overnight at 4°C with rotation. To remove the protease from hPot1, the sample was subsequently incubated with Talon™ metal affinity resin (Clontech, Palo Alto, CA) to

78

capture the histidine-tagged protease 3°C. Human Pot1 was stored in 50mM Tris pH7.6, 150mM NaCl, 5mM DTT, 20% glycerol and stored at -80°C.

Purification of TPP1- the His-Sumo-TPP1(87-545) construct was a gift of Dr. Ming Lei. His-Sumo-TPP1 was expressed in *Escherichia coli* BL21 DE3 cells with 1 mM 1-thio-β-D-galactopyranoside (IPTG) (Roche, Indianapolis, Indiana) for 4 hours at 37°C, purified as described in Sowd et al (26) and stored in 50 mM Tris pH 8.0, 150 mM NaCl, 20% glycerol, 250 mM Imidazole.

The hTPP1-Pot1 complex was formed by incubating equal masses of hPot1 and hTPP1 proteins on ice for 45 min. The complex was used for the DNA binding reactions.

Circular dichroism Spectral Analysis- For CD spectral analysis, synthesized DNA molecules were extracted in phenol/chloroform followed by one chloroform/Isoamyl alcohol wash and ethanol precipitation. G-rich ss DNA was dissolved in double distilled water and diluted to 416 ng/μl in the presence of either 150 mM KCl with 5 mM KPO_4, 150 mM NaCl with $NaPO_4$ or no salt. Similarly, C-rich DNA was diluted to 165 ng/μl in the presence or absence of 150 mM KCl and 5 mM KPO_4. DNA samples containing salt were melted at 95°C for 5 minutes and gradually cooled down to 4°C with 1°C decrease per minute. CD analysis of G-rich and C-rich ss telomeric DNA was performed at the Macromolecular Interactions Facility (University of North Carolina at Chapel Hill, NC) using Chirascan Plus fitted with a Peltier temperature control unit (Applied Photophysics Ltd, Leatherhead, UK). CD values were recorded at 25°C from 220 nm to 320 nm with 0.5 nm step sizes in a cuvette with a 1 mm

79

path length. Baselines of buffers were subtracted from the CD of DNA molecules and smoothed with Chirascan Prodata software.

DNA-protein binding reactions- To determine the mass of protein that is bound to the model telomere overhang (50 or 90 nucleotides), 234 nM Pot1 or 76 nM Pot1-TPP1 complex was incubated with 2.9 nM DNA template for 15 minutes at 37°C. To form filaments on long ss telomeric DNA, 140 nM of hPot1 or hPot1-TPP1 complex was incubated with 1500 ng DNA in 50 μl of 10 mM Tris-HCl (pH 7.5) and 0.1 mM EDTA. DNA-protein complexes were diluted to 150 μl without gluteraldehyde fixation and 20 μl of each sample were prepared for EM as described below.

Electron Microscopy- DNA and DNA-protein complexes were adsorbed onto copper EM supports covered with thin carbon foils for 3 minutes in the presence of a buffer containing 2.5 nM or 0.3 nM spermidine. Samples were washed in water followed by a series of ethanol dehydration steps, air-dried, and rotary shadowcast with tungsten at 1×10^6 torr (27).Samples were visualized using a Tecnai 12 TEM (FEI Inc., Hillsboro, OR) at 40kV and images were collected on a Gatan Orius CCD camera (Gatan, Pleasanton, CA) with supporting software (Gatan).

Image Analysis- The lengths and widths of hPot1, hPot1/TPP1 filaments were measured from digital micrographs using an Orius CCD camera with supporting software (Gatan). Image J software (NIH, Bethesda, MD) was used to analyze the 2D projections of the G-beads and the proteins bound at the telomeric overhangs. The 2D projections of the proteins were analyzed by ImageJ (NIH) software and converted to mass in kDa as described earlier (28)

using ferritin and streptavidin as standards. The kink angles and the length of DNA-protein filaments were also analyzed with ImageJ (NIH). For the statistical analysis Graphpad Prism 5 Student's t-test was used (Graphpad Software, Inc., La Jolla, CA). Images for publication were arranged and contrast optimized using Adobe Photoshop CS5 (Adobe Systems, San Jose, CA).

RESULTS

G-rich telomeric DNA exists as a "beads on a string" conformation

Analyzing the structure of G quadruplexes formed at telomeres is important for understanding the mechanism of telomere replication/transcription and how telomerase is inhibited by the quadruplexes. It is not clear yet if long telomeric DNA has distinct structural features or shares similar characteristics with shorter oligonucleotides. To mimic the structure of quadruplexes formed during replication/transcription at telomeres, single stranded telomeric DNAs up to 16 kbs long were synthesized using a model telomeric DNA and G- or C-rich nanocircles as a template for extension by Phi29 polymerase. The length was hard to measure because we do not know how either the C-rich or G-rich DNA is compacted. If it has the same extension as ds DNA, then the molecules would range from 465 to 16000 nucleotides and the mean length would be ~3500 nts **(Supplementary Figure 3.1A)**. However, it is likely that the G-rich ss DNA is more compact, making them even longer. Double stranded DNA after tungsten coating is ~30 Å wide (Data not shown). However, the G-rich ss DNA appear much thicker and beaded **(Figure 3.1C)**. Its width varies from 83 Å to 117 Å, with the diameter of the beads being 67 to 95 Å. In contrast to G-rich ss DNA **(Figure 3.1B)**, C-rich ss DNA does not form defined structures and its width is 18 ± 6 Å **(Figure 3.1A)**. Higher magnification images show that ss G-rich DNA is composed

of bead units attached by thin ss linker DNA **(Figure 3.1C)**. This data suggests that G-rich

DNA exists in a beads on a string like structure even in the absence of monovalent ions.
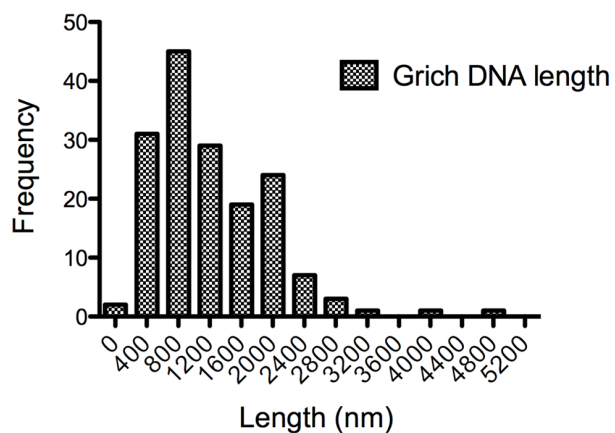
Supplementary Figure 3.1: Frequency distribution of G rich ss DNA length (nm) determined by Gatan software.
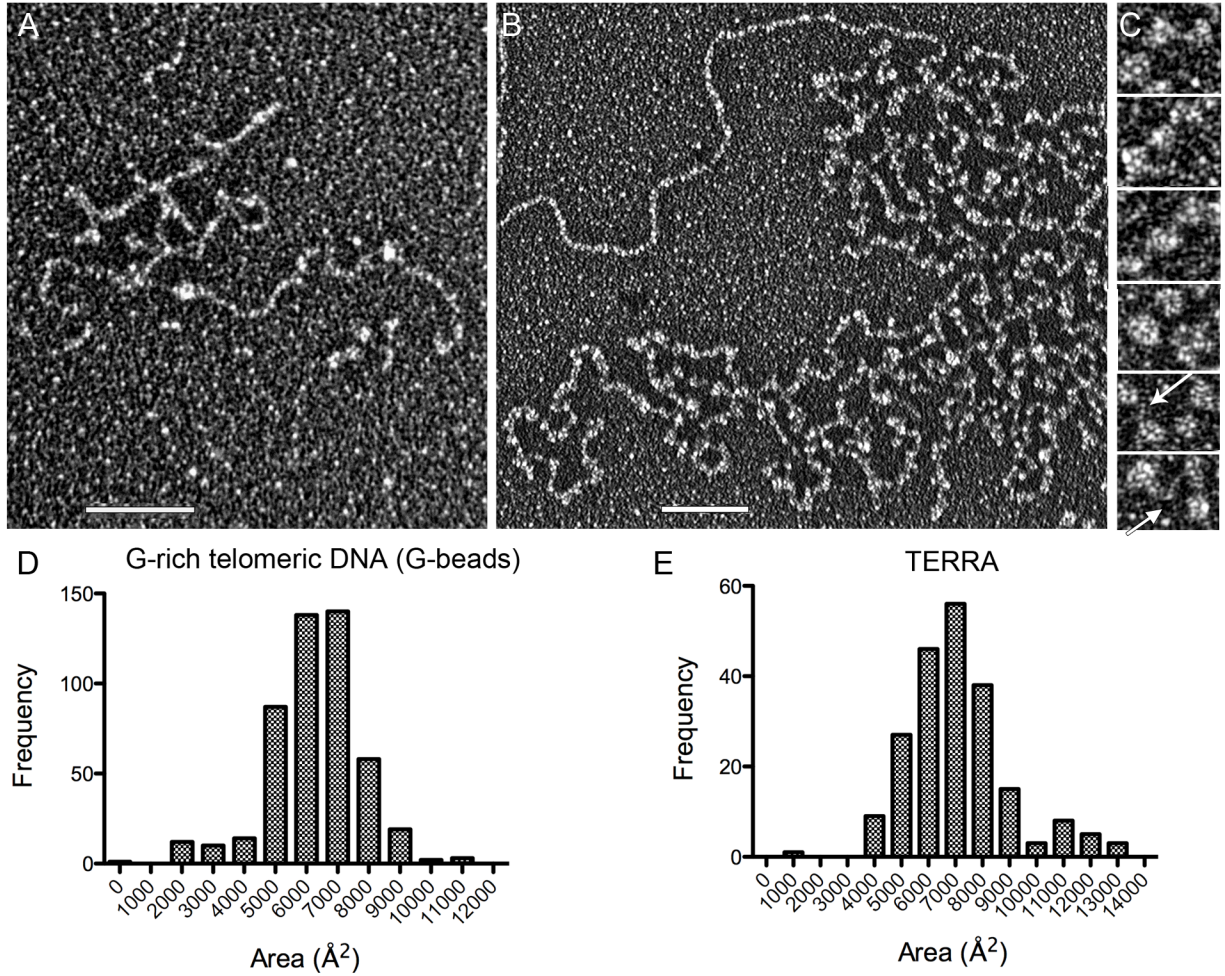
Figure 3.1: G-rich and C-rich telomeric DNA present distinct structures by EM. Tungsten shadowcast images of long C-rich telomeric DNA does not show a defined structure **(A)** while G-rich ss telomeric DNA shows uniform units **(B)**. Higher magnification images of G-rich ss DNA show bead motifs and arrows point the linker DNA between each bead motif **(C)**. Images are shown in reverse contrast and each size bar is 50 nm. The 2D area projections of G-beads show a narrow distribution with a mean of 6278 Å^2 **(D)** while the mean of TERRA's area distribution is 7092 Å^2 **(E)**.

Each G quadruplex bead motif (G-bead) is composed of ~490 nts

To determine the mass of each bead unit, the 2D area projections of G-beads were determined. The 2D area projection analysis of the bead units presents a narrow Gaussian distribution indicating that G-beads are uniformly sized structures with a mean of 6278 ± 67 Å^2 **(Figure 3.1D)**. A G-rich RNA (TERRA) of ~575 nts was prepared as in Randall et al (29). When prepared for EM, TERRA looks like small single beads. (**Supplementary Figure 3.2**). Comparison of the 2D projections provides a rough estimate for G-beads **(Figure 3.1E)** (28). Conversion of 2D area projections to mass illustrated that the size of G-beads is ~160 kDa **(Table 3.1)**. Regarding that 1 kb of ss DNA is 330 kDa, each bead could be composed of ~490 nts **(Table 3.1)**. This data suggests that G-rich telomeric DNA is composed of uniform ~490 nts long bead motifs.
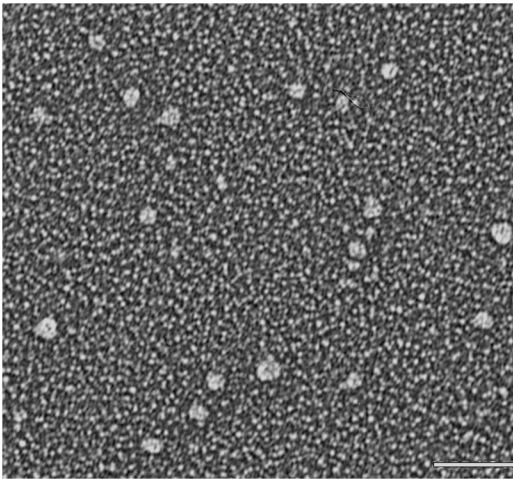
Supplementary Figure 3.2: TERRA forms compact structures. Tungsten shadowcasted images of TERRA present higher order structure formation. Image is in reverse contrast and size bar equals to 50 nm.

	# Particles Analyzed	Mean Area (Å²)	Standard Error (Å²)	Mass (kDa)	Length (nts)
G-Beads	484	6278	67	161 ± 0.2	487 ± 5
TERRA	212	7092	127	194 ± 0.2	575

Table 3.1: Size analysis of G-beads. The estimated mass values obtained from the statistical analysis of the area distributions derived from tungsten shadowcast images as in Figure 1 and the corresponding ss DNA length. Area distributions with Gaussian fits are shown in Supplementary Figure 2.

Long telomeric G-rich DNA contain both parallel and antiparallel G quadruplexes

G quadruplexes can exist in 3 different forms: parallel, antiparallel or hybrid structures. To gain structural information about the type of the G-quadruplexes formed along the long stretches of G-rich telomeric DNA, CD spectrophotometry was utilized in the presence or absence of monovalent ions. Introduction of Na^+ favors antiparallel G quadruplexes as K^+ favors both parallel and hybrid type G quadruplexes. In the absence of Na^+ or K^+, C-rich ss DNA has a single peak at 276 nm indicating no secondary structure formation while G-rich ss DNA presents a minimum at 247 nm and a maximum at 287 nm, which is a common feature of hybrid type G quadruplexes **(Figure 3.2A)**. This data demonstrates that G-quadruplexes can form during telomere extension and are stable structures even in the absence of monovalent ions. To mimic physiological conditions, 150 mM KCl was added to the ss G- or C-rich DNA and gradually cooled down. In the presence of KCl, C-rich ss DNA did not show a change in structure **(Figure 3.2B)**. Surprisingly, the peak values of G-rich ss DNA remained the same upon addition of K^+ or Na^+ions **(Figure 3.2C)**. G-quadruplexes formed by long DNA molecules are the highly stabile and the structure does not change with physiological levels of Na^+ or K^+.
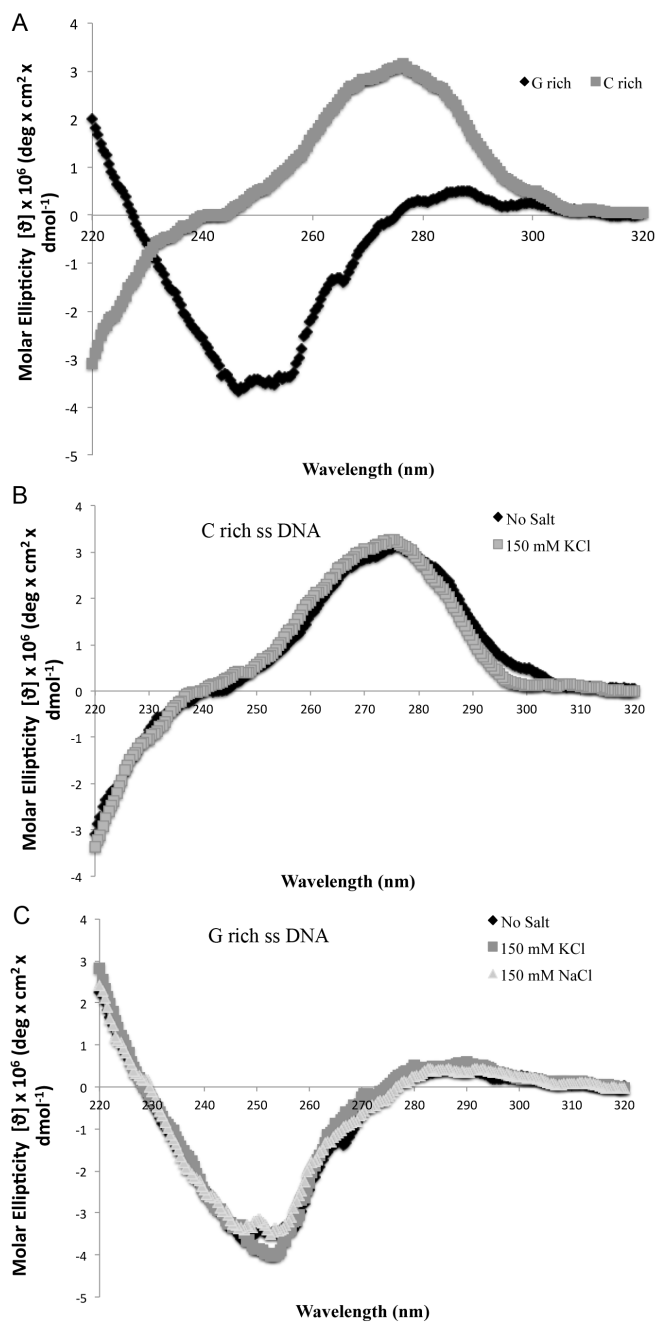
Figure 3.2: G-rich ss telomeric DNA is composed of hybrid type G-quadruplexes. CD spectral analysis of C-rich and G-rich ss telomeric DNA in the absence of salt ions is shown in **(A)**. **(B)** shows the CD spectra of C-rich ss DNA in the presence of 150 mM Na^+ or no salt while **(C)** shows the CD spectra of G-rich ss telomeric DNA in the presence of 150 mM Na^+ and K^+ ions or no salt.

hPot1 and hPot1-TPP1 complex form filaments along G-rich DNA and open up the G-beads

G quadruplexes should be unfolded *in vivo* so that the transcription or replication complexes can progress. It was previously demonstrated that hPot1 and the hPot1-TPP1 complex could unfold G quadruplexes when its composed of four telomeric repeats. However, four repeats do not represent the actual telomere structure and the DNA binding properties of hPot1 and the hPot1-TPP1 complex on very long DNA templates has not been analyzed before. Moreover, we demonstrated that G quadruplexes are much bigger structures and each G-bead can be composed of ~ 490 nucleotides. Even though hPot1 and the hPot1/TPP1 complex can unfold G quadruplexes of 4 repeats, resolution of G-beads may need additional cellular factors. Therefore to determine if hPot1 and the hPot1/TPP1 complex can open up G-beads on, hPot1 or hPot1-TPP1 complexes were bound to G-rich ss telomeric DNAs with the preexisting G quadruplexes. Secondary structures induced by G-rich ss telomeric DNA **(Figure 3.3A)** presented more open structures upon protein binding. In the presence of hPot1 and the hPot1-TPP1 complex, EM revealed that protein bound DNA molecules present open and uniformly coated structures indicating that hPot1 **(Figure 3.3B)** and the hPot1-TPP1 complex **(Figure 3.3C)** form filaments on G-rich DNA. To determine if hPot1 and the hPot1/TPP1 complex open up the G quadruplexes, the thickness of the filaments from various locations were measured and compared to the thickness of G-quadruplex. Thickness distribution of G DNA (**Supplementary Figure 3.3A**), and the filaments formed by hPot1 **(Supplementary Figure 3.3B)** and hPot1-TPP1 complex **(Supplementary Figure 3.3C)** fit a Gaussian distribution.
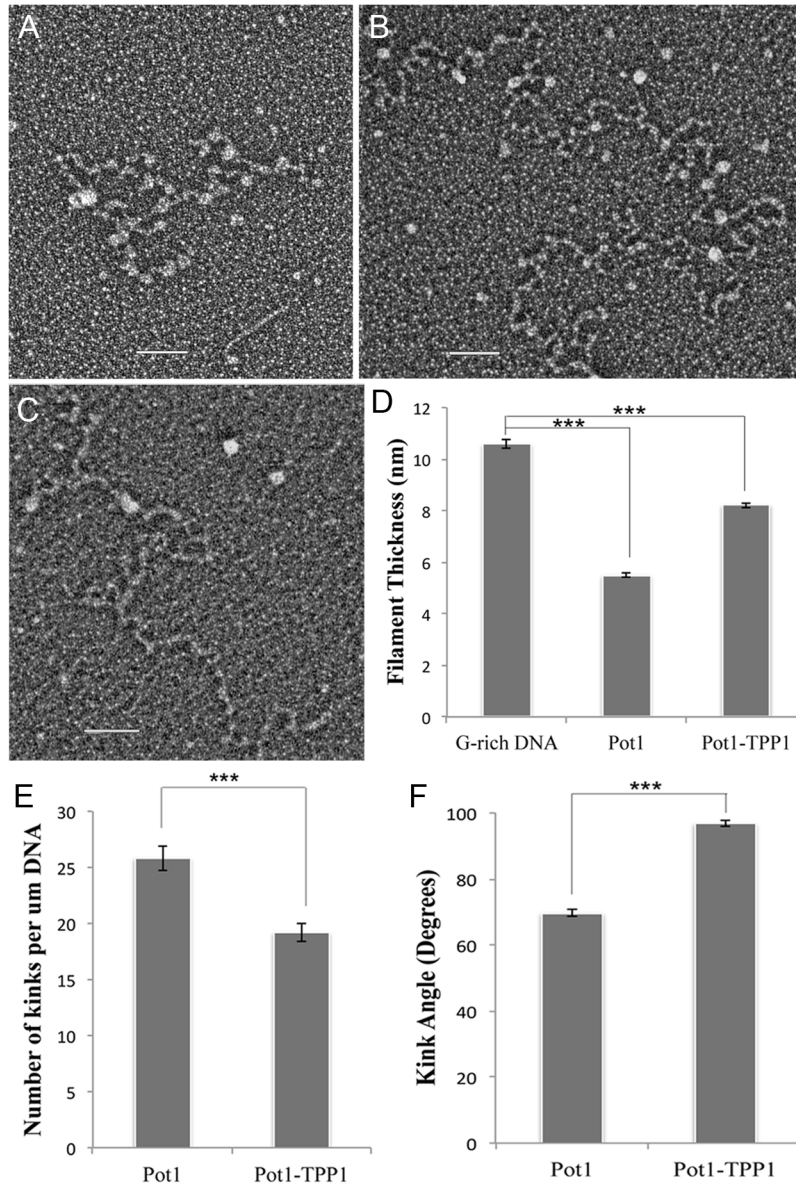
Figure 3.3: hPot1 and hPot1-Tpp1 open G-beads as they form filaments. Tungsten shadowcast images of G-rich ss DNA **(A)**, hPot1 bound to G-rich ss DNA **(B)**, hPot1-TPP1 bound to G-rich ss DNA **(C)**. Images are shown in reverse contrast and each size bar corresponds to 50 nm. Chart **(D)** represents thickness of G-rich ss DNA, hPot1 filaments, and hPot1-TPP1 filaments. hPot1-TPP1 filaments have less number of kinks **(E)** and higher number of nucleotides between each kink **(F)** than hPot1 filaments. Chart in **(G)** shows the kink angle difference between hPot1 and hPot1-TPP1 filaments. * is p<0.05, ** is p<0.01 and p<0.001 is represented by *** or ****.

A

G- rich ss telomeric DNA

B

Pot1 Filament

C
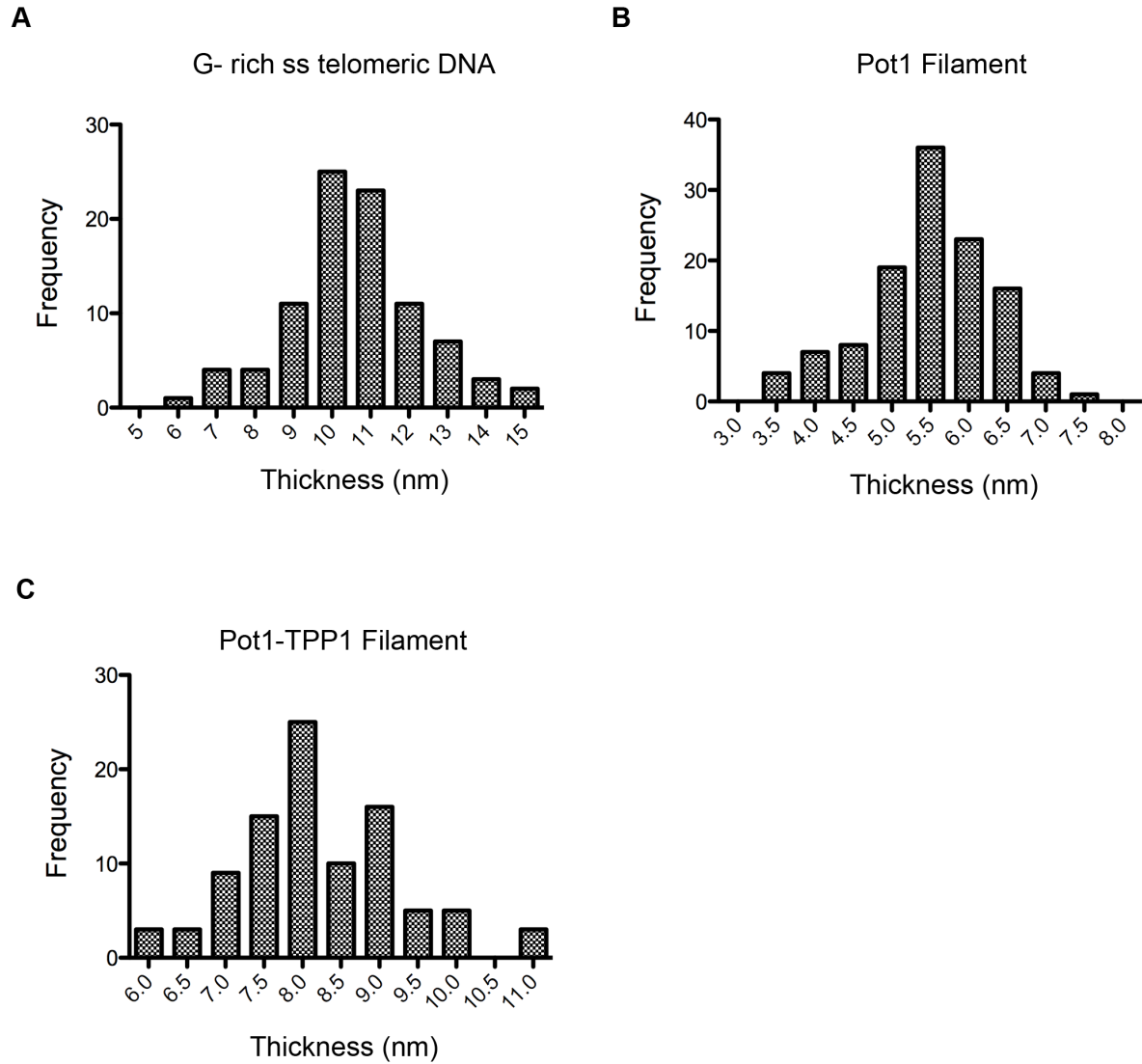
Pot1-TPP1 Filament

Supplementary Figure 3.3: Histograms of filament thicknesses. (A) represents the width distribution of G-rich single stranded telomeric DNA while (B), and (C) present the width distribution of filaments formed by hPot1 and the hPot1-TPP1 complex when they are bound to the G-rich ss telomeric DNA respectively.

The average thickness of G-rich telomeric DNA was found as 11 nm, while the thickness of the hPot1 filament was 5.5 nm and the hPot1-TPP1 filament was 8.2 nm **(Table 3.2)**. In each case, the filaments formed by the proteins were significantly thinner than G-rich DNA alone **(Figure 3.3D)** suggesting that hPot1and hPot1-TPP1 complex has the ability to open up the preexisting G quadruplexes by forming filaments along them.

DNA binding properties of hPot1 changes upon complex formation with TPP1 such that the affinity of the complex increases by 10 fold (16). Also hPot1 in complex can slide along G-rich DNA while hPot1 alone cannot (11). Therefore, the filaments formed by the two proteins can differ in their stiffness. The stiffness of the filaments can be implied from the kink angles and the number of kinks along a certain DNA length. Stiffer filaments tend to have bigger kink angles due to lower twist flexibility and lower number of kinks per unit DNA length. To determine the number of kinks per unit DNA length, the total number of kinks were counted and divided by the total length of the DNA. There is significant difference between the average kink number of the hPot1 filaments and the Pot1-TPP1 filaments **(Figure 3.3E)**. While hPot1 filament has ~25 kinks per um of DNA **(Supplementary Figure 3.4A)**, hPot1-TPP1 filament has ~19 kinks per um DNA **(Supplementary Figure 3.4B)**. To determine the kink angle of each filament, all the kink angles along a single filament were measured and their mean values were calculated **(Supplementary Figure 3.5 A and 3.5B)**. The mean kink angle of all hPot1 filaments presented a significantly smaller angle (69°) than filaments formed by hPot1-TPP1 (98°) **(Figure 3.3F)**. These data together suggest that as a result of filament formation, DNA is stiffer and as hPot1 forms more flexible filaments, hPot1-TPP1 forms more rigid filaments.

	# Particles Analyzed	Average Thickness (nm)	Standard Error (nm)
G-rich telomeric DNA	34	10.58	0.2
hPot1	122	5.5	0.1
hPot1-TPP1	97	8.2	0.1

Table 3.2: Thickness comparison of G-rich ss DNA, hPot1 filaments and hPot1-TPP1 complex filaments. The statistical analysis of the filaments widths, which were determined by Gatan software are presented in nm. Histograms of filament widths are shown in Supplementary Figure 3.3.
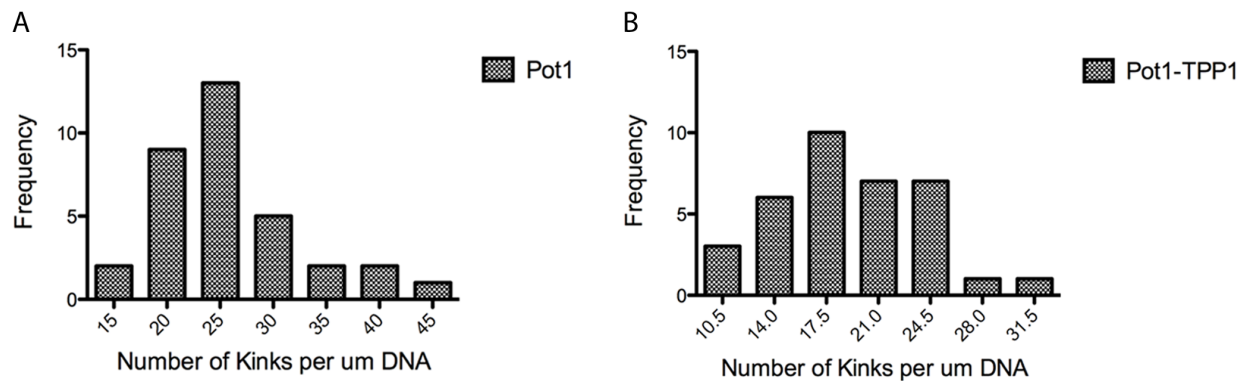
Supplementary Figure 3.4: Frequency distribution of number of kinks per unit DNA length. (A) and (B) show the filaments of Pot1 and Pot1-TPP1 complex respectively. Unit DNA length is um. Statistical analysis is done by Student's t-test (Graphpad Prism 5).
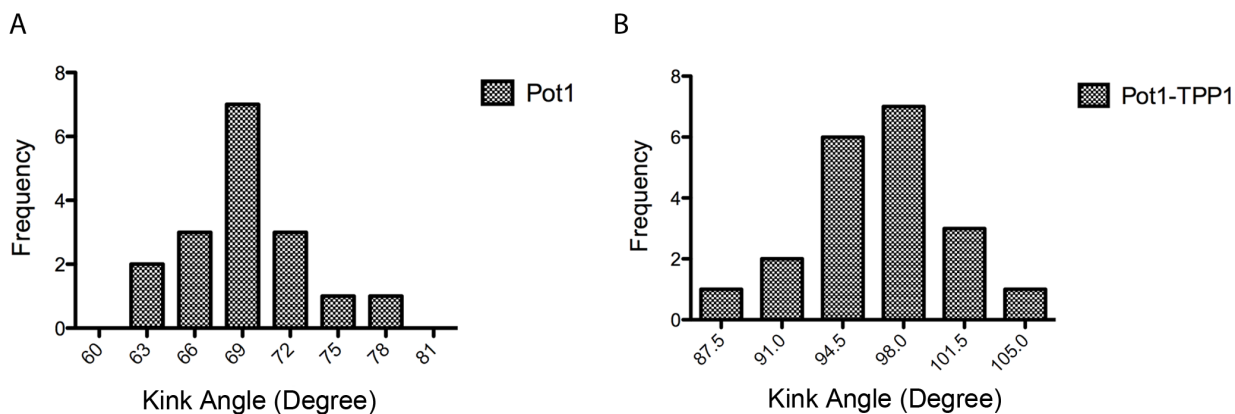
Supplementary Figure 3.5: Histogram of kink angles. Each angle was measured by ImageJ angle tool and Graphpad Prism5 was used for the statistical analysis.

hPot1 present different modes of binding and can compact ss DNA by ~ 2 fold

We showed that both hPot1 and the hPot1-TPP1 complex relaxes G-quadruplexes by forming filaments on long G-rich DNA however, they could employ different DNA binding mechanisms to protect the chromosome ends. To determine the differences between the DNA binding properties of hPot1 and the hPot1/TPP1 complex at chromosome ends in more detail, we used a model telomere and telomeric 3-way junction templates to mimic the human chromosome ends and telomeric fork structures. Mass of each protein bound to the DNA template was determined by measuring its 2D projection and then the mean value was converted in to mass in kDa by using Ferritin as a size standard (Materials and Methods). When hPot1 or the hPot1-TPP1 complex was incubated with model telomeres at saturating levels (~ 95 ± 2 % total binding), DNA bound protein mass differed depending on the length of the overhang and the protein used **(Table 3.3)**. When the overhang length was 50 nts, the mass of hPot1 was 163 ± 1 kDa and the mass of the hPot1-TPP1 complex was 132 ± 1.5 kDa. When the overhang length was increased to 90 nts, the mass of hPot1 increased to 314 ± 1.2 kDa but the hPot1-TPP1 complex mass stayed similar (139 ± 3.8 kDa) suggesting that the DNA binding ability of the hPot1-TPP1 complex was independent of the DNA length. Mass of TPP1 is ~ 60.7 kDa (14) and hPot1 and hTPP1 is suggested to be in 1:1 ratio (30). Therefore, Pot1-TPP1 complex can bind to model telomere overhangs as a monomer of each protein.

In solution, hPot1 exists as a monomer with a mass of 75 kDa **(Supplementary Figure 3.6)**, which is consistent with previous publications (~71 kDa) (31). While hPot1 binds to 50 nts overhangs as two monomers, it binds to 90 nts overhangs as four monomers **(Table 3.3)**. If only the two OB folds were responsible for the DNA binding ability of hPot1 as suggested

previously (25), then each monomer of hPot1 should bind to ~9-10 nts and the mass of hPot1

bound to 50 nts and 90 nts overhangs should be 350 kDa (pentamer) and 630 kDa (nonamer).

However, only half of the expected oligomer numbers were observed at the overhangs and in

each case Pot1.

	DNA length	# Particles Analyzed	Observed Mass (kDa)	St error (kDa)	Monomer Mass (kDa)	Oligomers
hPot1	50 nts overhang	66	163	1	71.4	2.3
	90 nts overhang	168	314	1.24	71.4	4.4
	124 nts circle	91	711	9	71.4	10
hPot1-TPP1	50 nts overhang	37	132	1.5	61+71.4	1:1
	90 nts overhang	23	139	3.8	61+71.4	1:1

Table 3.3: Mass analysis of hPot1and hPot1-TPP1 at model telomere overhangs. 2D area projections of the proteins were determined by ImageJ. The mass and the oligomeric states of each protein were determined by using Ferritin as a mass standard.

A

B

	Projected Area (Å²)	Standard Deviation of Projected Area	# Particles Measured	Estimated Mass	Predicted Oligomerization State
hPot1	1983	753	102	75	monomer

C
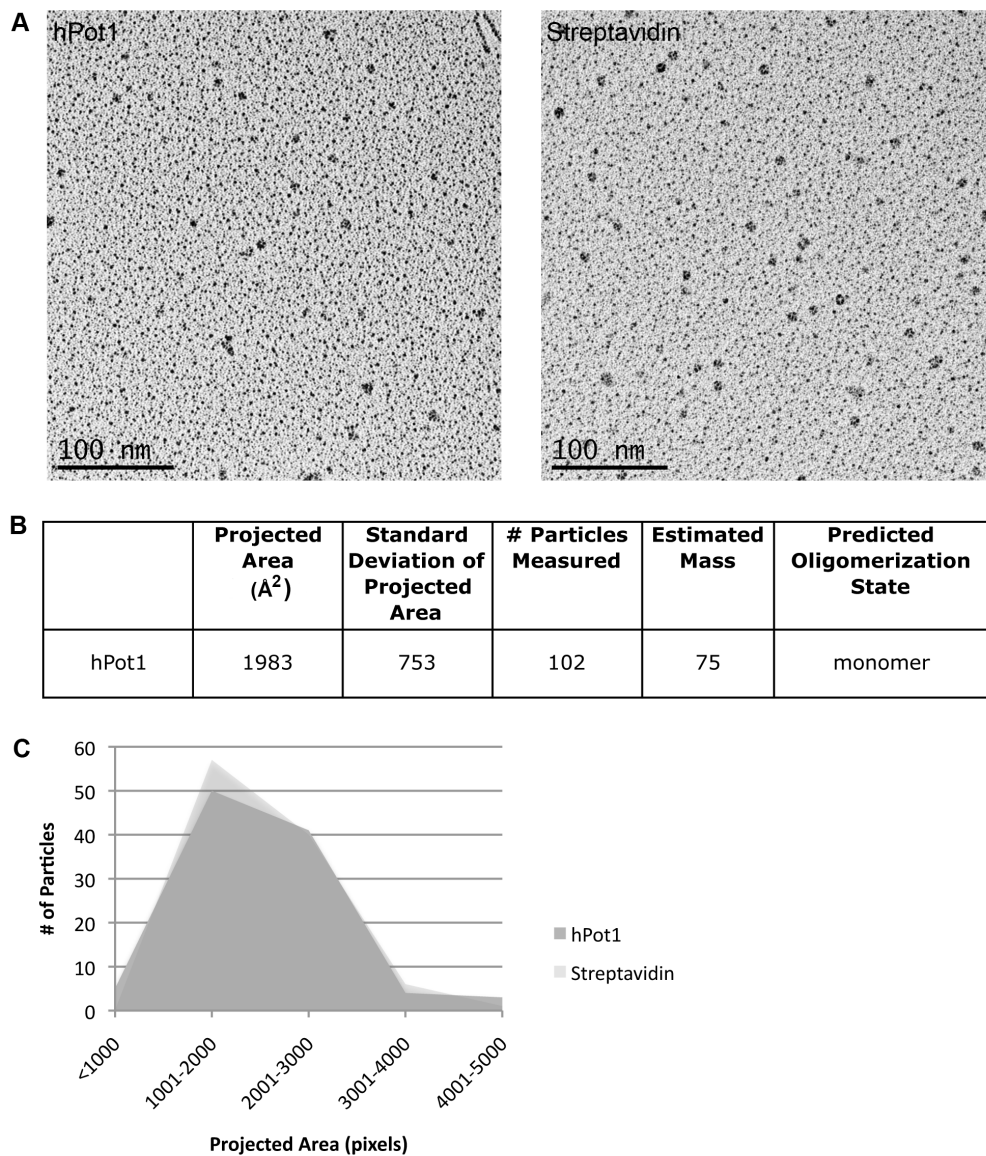
Supplementary Figure 3.6: Mass analysis of hPot1 in solution. (A) hPot1 and Streptavidin were prepared side-by-side for tungsten shadowcasting for direct size comparison. (B) The projected areas of hPot1 bound to the model telomere and free streptavidin were compared and used to estimate the mass of hPot1 bound to the telomere substrate. (C) the distribution of projected areas (Pixels) of hPot1 and streptavidin were compared.

monomer binds to ~20 nts. This data demonstrated that the full-length hPot1, which has 3 OB folds, might bind to more nts compared to the truncated hPot1 with two OB folds.

Some single strand DNA binding proteins with OB folds, such as RPA represent different modes of binding and thus it can bind to different number of nucleotides, either 8-10, 13-14 or 30 nts. To test if hPot1 has differential DNA binding modes, we determined the mass of hPot1 on a rigid DNA template, G-circles and compared to the mass values on model telomere **(Table 3.3, Supplementary Figure 3.7)**. G-circles are composed of 124 nts of pure telomeric repeats and thus up to 12 hPot1 monomers can bind. The mass of hPot1 on G-circles was 711 kDa (10 hPot1 monomers), indicating that hPot1 can bind to ~12 nts on rigid DNA substrate. This finding is consistent with the previously suggested DNA binding motif for hPot1 however, it shows that nearly twice as much hPot1 molecules bind to the rigid DNA structures than flexible overhangs. Therefore, depending on the DNA template the full-length hPot1 can bind to different number of nts. Another example of a single strand DNA binding protein with different modes of binding is E. coli single strand binding protein (SSB). With its different modes of DNA binding, SSB can wrap the DNA around itself and compact the DNA by ~3-folds. To test if hPot1 has the DNA compaction ability, we used a DNA template with 3-way junction that has ~570 nts long ss G-rich telomeric DNA (19). The length of SSB coated displaced strand was 26 ± 0.7 nm while the length of hPot1 coated displaced strand was 41 ± 1 nm suggesting that hPot1 compacts ss overhang structures by ~ 2 fold **(Figure 3.4A and B)**. This data suggests that hPot1 has different modes of DNA binding and when it binds to ~20 nts, it can compact the G-rich ss DNA by ~2-fold and illustrates why hPot1 cannot slide while the complex can.
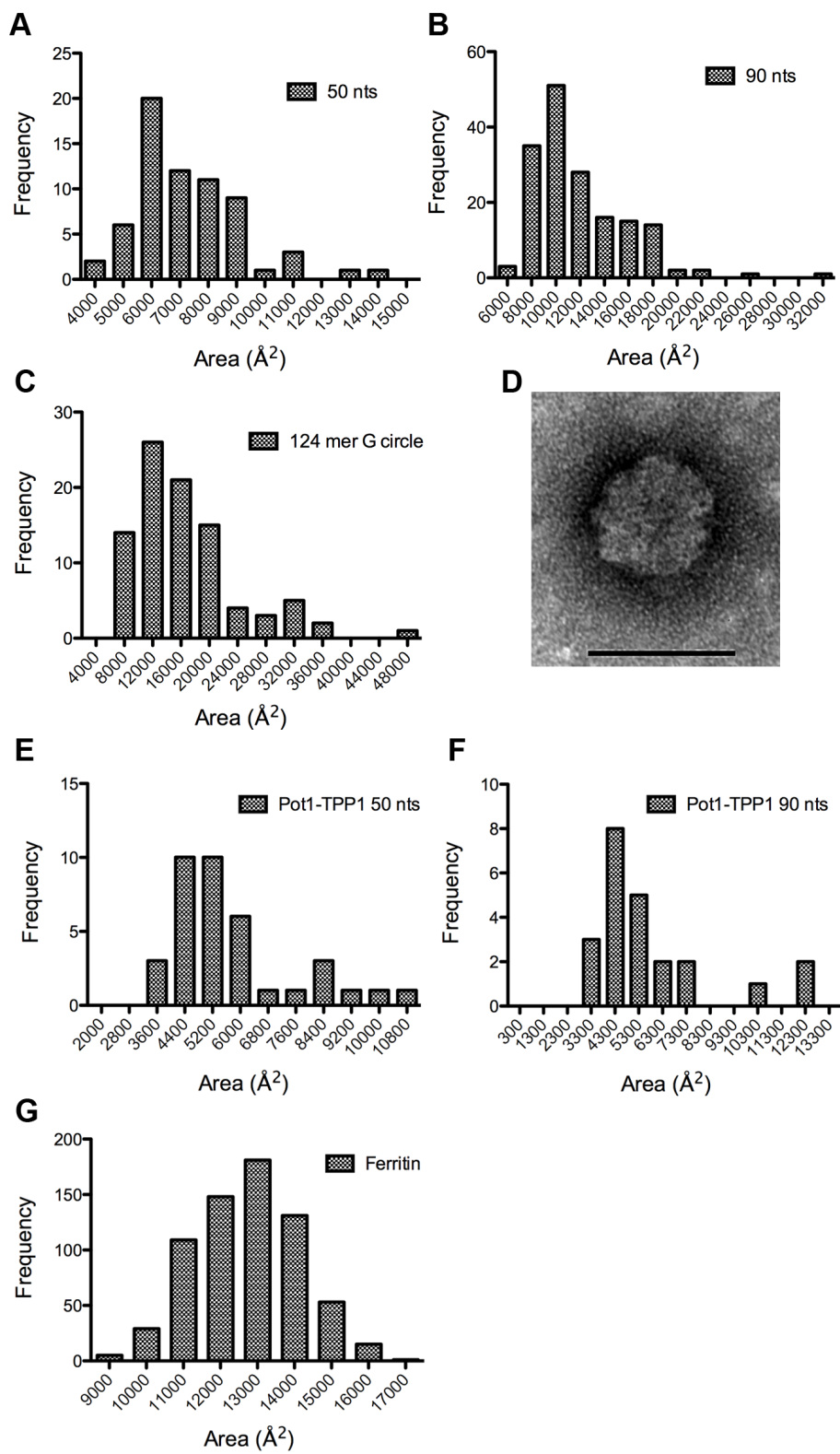
Supplementary Figure 3.7: Mass analysis of hPot1 and the Pot1-TPP1 complex when bound to different DNA templates. (A) and (B) present the 2D area projection of hPot1 when its bound to model telomeres with 50 nts and 90 nts overhangs. The frequency distribution of the 2D area projection of hPot1 when it is bound to 124mer G-circles (C) and a representative EM image (D) are shown. Histograms of the Pot1-TPP1 complex at 50 nts (E) and 90 nts (F) overhangs of model telomere have similar means. Ferritin was used as a size standard for the mass conversion of the 2D projection values of each protein (G).
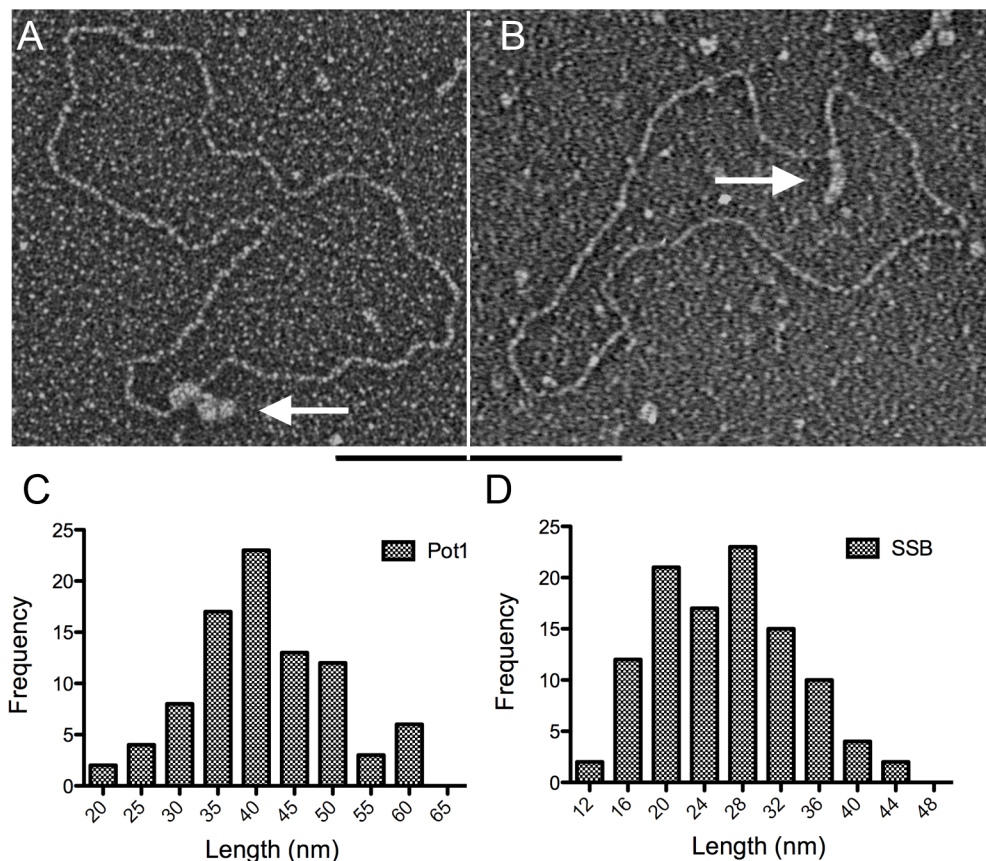
Figure 3.4: hPot1 compacts ss G-rich telomeric DNA by ~ 2 fold. hPot1 and SSB binds to the ss G-rich strand of the 3-way junction template in **(A)** and **(B)** respectively. Images are shown in reverse contrast and the size bar corresponds to 1μm. Arrows indicate the DNA-protein complexes. The length analysis of hPot1 **(C)** and SSB **(D)** bound at the displaced strand, shows the mean value of the DNA-Pot1 complex as ~40 nm and the mean length of the SSB-DNA complex as ~26 nm.

DISCUSSION

To understand the structures formed during telomere extension or transcription, we have determined the structure of telomeric G quadruplexes and DNA binding characteristics of hPot1 and its binding partner TPP1 along these higher order structures and telomeric overhangs. In this study we demonstrated that G quadruplexes compact into uniformly sized bead like structures and these beads are connected through linker DNA. In contrast to G-rich ss telomeric DNA, C-rich ss telomeric DNA lacked to form any defined secondary structure. We also showed that each bead is composed of ~490 nts and composed of both parallel and antiparallel G-quartets. G-beads did not show a significant change in structure upon addition of K^+ or Na^+ ions indicating their high stability. hPot1 and the hPot1-TPP1 complex can both open up G-beads by forming filaments along the G-rich DNA and filaments present differences in their rigidity and thickness. While hPot1 filaments have smaller kink angles, hPot1-TPP1 filaments are more rigid and have lesser number of kinks as well as bigger kink angles. At the 3' overhang of model telomeres, the mass of hPot1-TPP1 complex was ~ 130 kDa, which could be 1 hPot1 and 1 TPP1. While the mass of the complex was independent of the overhang length, hPot1 binds ~ 20 nts at model telomere ends and ~12 nts on G–circles. Moreover, hPot1 compacted the displaced strand of pRST5 by 2 fold indicating that hPot1 and the hPot1-TPP1 complex have distinct DNA binding abilities along G-rich telomeric DNA and hPot1 can present different modes of binding.

In this work we showed that G-rich telomeric DNA is composed of both parallel and antiparallel G quadruplexes. This data is consistent with (32). Long G-rich DNA was suggested to exist in a beads-on-a-string conformation (33). EM analysis confirmed that

102

model. Furthermore we showed that each G-bead is composed of ~ 490 nts and these are the biggest uniform G quadruplex structures that are identified so far. Yu et al also suggested a beads-on-a-string conformation for the G-rich DNA but the bead sizes were smaller than what we have identified (33). Moreover, we did not observe a change in the G-quadruplex structure in the presence or absence of Na^+ or K^+ in contrast to Yu et al (33). The major difference between the two studies can be the length of DNA used. They used oligos up to 120 nts long while the DNA molecules used in our studies were kbs long suggesting that as the DNA length increases, the stability of the G quadruplexes can increase. The size of G-beads and their stability can explain why complexes can stall the replication/transcription complexes and result in replicative stress.

Recently, Hwang et al demonstrated that hPot1 unfolds the telomeric G quadruplex in a stepwise manner and in the presence of TPP1, the hPot1-TPP1 complex can slide along the quadruplex DNA (11). Consistently, we observed that both hPot1 and the hPot1-TPP1 complex could unfold preexisting G quadruplexes and we further showed that both hPot1 and the hPot1-TPP1 complex form filaments along the long telomeric DNA as they open the G-quadruplexes. Moreover, we demonstrated that filaments formed by hPot1 can be more flexible and have higher kink angles while filaments formed by the hPot1-TPP1 complex can be more rigid and continuous, which is consistent with the sliding ability of the hPot1-TPP1 complex as indicated in Hwang et al (11). Moreover, the mass of the hPot1-TPP1 complex did not change depending on the overhang length of the model telomere. In this case, the sliding motion does not completely explain why telomeric end does not get saturated with the Pot1-TPP1 complex. One would expect to see more protein binding to the overhang as the length increases even if the complex slides from 3' end towards the ds-ss junction site.

Alternatively, once TPP1-Pot1 complex binds to the far 3' end of the overhang, the complex can pull the overhang end with it as it slides towards the ds-ss junction site. A loop- or a hairpin-like structure would form and as the complex moves G-quadruplexes would form end resolve. This mechanism would enable end protection as well as the mobility ability to the TPP1-Pot1 complex.

We observed that more hPot1 monomers bound as the overhang length increased and the number of nucleotides that each Pot1 monomer bound was ~20 nts in each case. We have also showed that hPot1 compacts the G-rich telomeric DNA by ~ 2 fold. Compaction ability has not been identified before and the identified DNA binding motif of hPot1 was ~9-10 nt. However according to our findings this motif is only valid when the DNA substrate is a circle. When DNA is a flexible structure such as an overhang, then hPot1 can bind to ~20 nts and compacts DNA by ~ 2 fold. The reason why we observed the unique DNA binding ability of hPot1 could be that the DNA substrates in our experiments were very long, not immobilized and we used full-length hPot1 rather than a splice variant (11,25).

Based on our data we suggest the model in which strand separation during transcription and the discontinuous synthesis of the lagging strand would cause G quadruplex formation **(Figure 3.5)**. The resulting G quadruplexes can be very big and they can stall telomerase and replication/ transcription complexes. Telomeric ss DNA binding protein hPot1 can bind to the G-beads to relieve the replicative stress by opening up quadruplexes and forming filaments along G-rich DNA. This finding can explain, why hPot1 is required for efficient telomere C-rich strand replication in the absence of WRN. In the presence of TPP1, the hPot1-TPP1 complex would slide along the G-rich DNA (11), unfold the preexisting G quadruplexes and form continuous filaments. These filaments would prevent activation of

ATR and refolding of the G quadruplexes. Moreover, the filaments would create a bigger surface area for telomerase to be loaded and increase the processivity of telomerase. Once the replication/transcription finishes hPot1 can bind to the 3' overhang and compact the single stranded DNA by ~2 fold either by wrapping or binding between each G-quadruplex motifs. In the presence of TPP1, the complex could bind to the far end of the overhang, pull the ss DNA as it slides and by this way it can prevent nucleolytic attack, activation of ATR and the access of telomerase to the telomeric end.
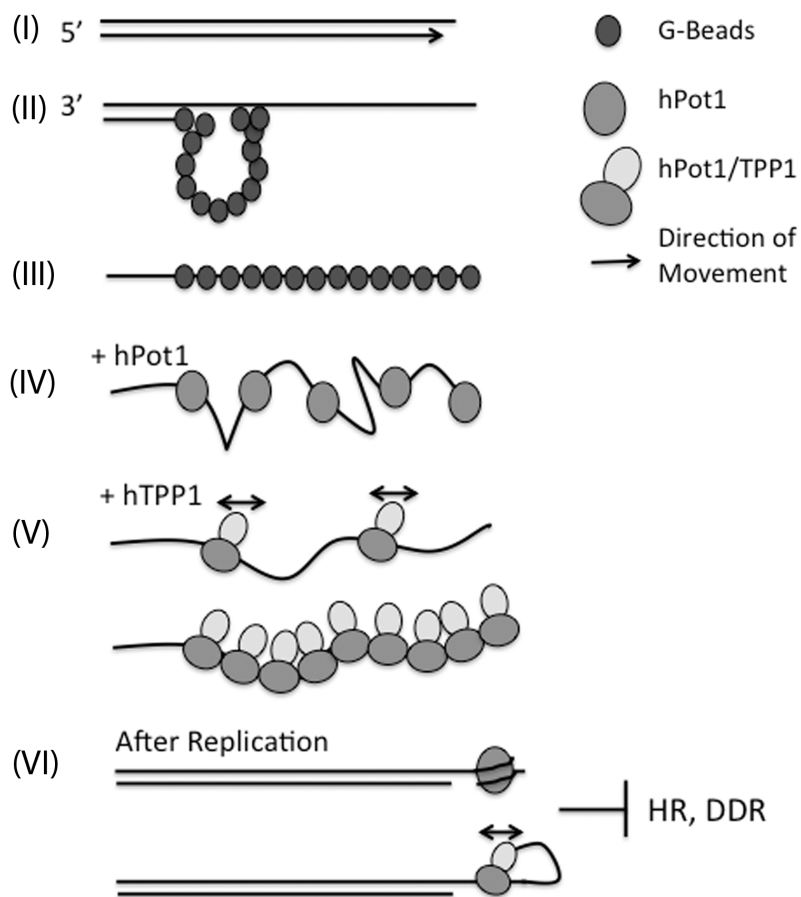
Figure 3.5: Model of structures formed by G-rich telomeric DNA and their regulation. Leading strand is synthesized continuously **(I)** in contrast to lagging strand synthesis. G-quadruplexes form on the lagging strand or when the two strands melt for transcription **(II)**. G-quadruplexes exist as a beads on a string pattern **(III)**. hPot1 opens up the G-quadruplexes and forms flexible filaments **(IV)**. In the presence of TPP1, Pot1 slides and the Pot1-TPP1 complex make smoother filaments **(V)**. At the end of replication or transcription, hPot1 compacts the single stranded G-rich telomeric DNA and TPP1/Pot1 complex holds the ss overhang end and pulls it and prevents nucleolytic attack **(VI)**.

REFERENCES

1. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) *Proc Natl Acad Sci U S A* **48**, 2013-2018

2. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) *Biochemistry* **30**, 4460-4472

3. Huppert, J. L., and Balasubramanian, S. (2005) *Nucleic Acids Res* **33**, 2908-2916

4. Nakken, S., Rognes, T., and Hovig, E. (2009) *Nucleic Acids Res* **37**, 5749-5756

5. Eddy, J., and Maizels, N. (2006) *Nucleic Acids Res* **34**, 3887-3896

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

7. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., and Blackburn, E. H. (1987) *Cell* **51**, 899-908

8. Healy, K. C. (1995) *Oncol Res* **7**, 121-130

9. Baumann, P., and Cech, T. R. (2001) *Science* **292**, 1171-1175

10. Denchi, E. L., and de Lange, T. (2007) *Nature* **448**, 1068-1071

11. Hwang, H., Buncher, N., Opresko, P. L., and Myong, S. (2012) *Structure* **20**, 1872-1880

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

13. Veldman, T., Etheridge, K. T., and Counter, C. M. (2004) *Curr Biol* **14**, 2264-2270

14. Liu, D., Safari, A., O'Connor, M. S., Chan, D. W., Laegeler, A., Qin, J., and Songyang, Z. (2004) *Nat Cell Biol* **6**, 673-680

15. Wang, F., Podell, E. R., Zaug, A. J., Yang, Y., Baciu, P., Cech, T. R., and Lei, M. (2007) *Nature* **445**, 506-510

16. Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., O'Connor, M. S., and Songyang, Z. (2007) *Nature* **445**, 559-562

17. Belotserkovskii, B. P., Liu, R., Tornaletti, S., Krasilnikova, M. M., Mirkin, S. M., and Hanawalt, P. C. (2010) *Proc Natl Acad Sci U S A* **107**, 12816-12821

18.   Rizzo, A., Salvati, E., Porru, M., D'Angelo, C., Stevens, M. F., D'Incalci, M., Leonetti, C., Gilson, E., Zupi, G., and Biroccio, A. (2009) *Nucleic Acids Res* **37**, 5353-5364

19.   Fouche, N., Cesare, A. J., Willcox, S., Ozgur, S., Compton, S. A., and Griffith, J. D. (2006) *J Biol Chem* **281**, 37486-37495

20.   Sfeir, A., Kosiyatrakul, S. T., Hockemeyer, D., MacRae, S. L., Karlseder, J., Schildkraut, C. L., and de Lange, T. (2009) *Cell* **138**, 90-103

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

22.   Laud, P. R., Multani, A. S., Bailey, S. M., Wu, L., Ma, J., Kingsley, C., Lebel, M., Pathak, S., DePinho, R. A., and Chang, S. (2005) *Genes Dev* **19**, 2560-2570

23.   Arnoult, N., Saintome, C., Ourliac-Garnier, I., Riou, J. F., and Londono-Vallejo, A. (2009) *Genes Dev* **23**, 2915-2924

24.   Arat, N. O., and Griffith, J. D. (2012) *J Biol Chem* **287**, 41583-41594

25.   Lei, M., Podell, E. R., and Cech, T. R. (2004) *Nat Struct Mol Biol* **11**, 1223-1229

26.   Sowd, G., Wang, H., Pretto, D., Chazin, W. J., and Opresko, P. L. (2009) *J Biol Chem* **284**, 34682-34691

27.   Griffith, J. D., and Christiansen, G. (1978) *Annual review of biophysics and bioengineering* **7**, 19-35

28.   Compton, S. A., Tolun, G., Kamath-Loeb, A. S., Loeb, L. A., and Griffith, J. D. (2008) *J Biol Chem* **283**, 24478-24483

29.   Randall, A., and Griffith, J. D. (2009) *J Biol Chem* **284**, 13980-13986

30.   Takai, K. K., Hooper, S., Blackwood, S., Gandhi, R., and de Lange, T. *J Biol Chem* **285**, 1457-1467

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

32.   Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R. A., and Yang, D. (2006) *Nucleic Acids Res* **34**, 2723-2735

33.   Yu, H., Gu, X., Nakano, S., Miyoshi, D., and Sugimoto, N. (2012) *J Am Chem Soc* **134**, 20060-20069

CHAPTER 4: SINGLE STRANDED TELOMERIC CIRCLES COMPLEMENTARY TO THE NATURAL TELOMERE OVERHANG CAN SERVE AS TEMPLATES FOR TELOMERASE-INDEPENDENT TELOMERE EXTENSION

SUMMARY

The Alternative Lengthening of Telomeres (ALT) pathway is a recombination-based telomere length maintenance mechanism observed in ~15% of human cancers. Single stranded and double stranded telomeric circles (t-circles) are markers for ALT in human cells. How they participate in this pathway is poorly understood but has been suggested to involve rolling circle replication (RCR) or recombinational mechanisms. To investigate how t-circles may facilitate telomere extension in the absence of telomerase we employed highly purified phage T4 replication proteins, a model telomere DNA with a 3'overhang of the G-rich strand, and 120 nt single stranded t-circles formed from G-rich or C-rich telomeric repeats. We show that C-rich t-circles can anneal to the 3'overhang of the telomere and act as a template for the synthesis of new telomeric DNA. We observed both RCR and coupled synthesis of the leading and lagging strands. The newly synthesized DNA was composed of telomeric repeats and electron microscopic analysis showed that the replication products can be up to 32 kb. In contrast to C-rich t-circles, G-rich t-circles did not initiate replication. These results show the central role of C-rich t-circles in telomere maintenance in ALT cells and suggest how they can generate rapid telomere length changes.

INTRODUCTION

Telomeres are the nucleoprotein structures that maintain genome integrity by distinguishing natural chromosome ends from double stranded (ds) DNA breaks and prevent inappropriate DNA damage response, end-to-end fusions and nucleolytic degradation (1). Human telomeres consist of 5 to 15 kb tracts of TTAGGG repeats and end in a 3' single stranded (ss) DNA overhang of ~150 nt on the G-rich strand (2,3). The very end of the telomeric DNA can loop back and invade the upstream ds region, forming a lariat like structure (t-loop), (4) which is regulated by both telomere-specific proteins (shelterin) and a larger number of general DNA repair factors acting in concert to protect the telomere ends from unnecessary DNA repair activity (5).

For telomeres to function, they must maintain their length beyond a certain critical value. In human somatic cells, progressive shortening of telomeres with each cell division due to the end replication problem eventually results in genomic instability, apoptosis, or senescence (6-9). Germline cells and a majority of cancer cells upregulate the reverse transcriptase enzyme telomerase which restores telomeric repeats to the DNA ends and compensates for telomere shortening (10). However roughly 10-15% of human cancers maintain their telomeres beyond the critical length in the absence of telomerase activity. These cells, termed ALT for Alternative Lengthening of Telomeres, appear to employ homologous recombination pathways to maintain telomere length (11). Characteristics of ALT cells are telomeres that are heterogeneous in size, show rapid length changes (12), a high frequency of telomere sister chromatid exchange (SCE) events without an increase in SCE elsewhere in the genome (13), and promyelocytic leukaemia (PML) bodies (14). In addition, extrachromosomal DNA consisting of telomeric repeats has been noted in ALT

110

cells (15-17). We (15) showed that ALT cells contain abundant heterogeneous ss and ds DNA circles and that these circles (t-circles) consist of telomeric repeats. Thus t-circles are a molecular marker of ALT.

T-circles were first discovered as extrachromosomal DNA in the mitochondria of the yeast *Candida parapsilosis* by two-dimentional agarose gels and electron microscopy (EM) (18). In this yeast, the mitochondrilal DNA is linear yet no telomerase activity has been detected in the mitochondrion suggesting that t-circles provide a primary substrate for telomere maintenance. Other yeast species with linear mitochondrial DNA including *C. salmanticensis*, C. *metapsilosis*, and *Pichia philodendri* (19) also contain mitochondrial t-circles. T-circles have now been detected in species as divergent as *Kluyveromyces lactis* (20,21), *Caenorhabditis elegans* (1), *Arabidopsis thaliana* (22), and *Xenopus laevis* (23), in addition to mammalian ALT cells (15), (17). T-circles also arise in telomerase positive cells when certain telomerase components are overexpressed or mutated. Increased t-circle formation was observed in human somatic cells containing a Ku86 deletion (24) and in *A. thaliana* with Ku70 deletion (22). Deletion of the basic terminus of human TRF2 (TRF2ΔB) results in the appearance of t-circles and this mutation also interferes with recruitment of XRCC3 (25), ORC2 (26) and WRN helicase (27) to telomeres. The size of these circles ranges from 100 bases to more than 30 kb (28). Studies in *K. lactis* showed that t- circles have a role in telomere maintenance and that introduction of artificial t-circles into these yeast cells could lead to telomere elongation (29). This wide prevalance, presence in normal *C. elegans* cells (1) and yeast mitochondria (18), and their appearance when a number of telomere-related proteins are mutated suggests that t-circles are an important player in

telomere maintenance. Tomaska *et al* (30) have suggested that they may preceed telomerase in evolution as a means of maintaining telomere length.

It is currently unclear how, at a molecular level, t-circles arise or are maintained. Possible origins include the trimming of t-loops or intrachromosomal recombination within telomeric arrays (30,31). In ALT cells, the reduction of XRCC3 or NBS1 (32) was found to reduce the level of t-circles greatly although the cells remained phenotypically ALT. The way in which t-circles are utilized to extend telomeres is also debated. One model suggests that priming occurs extrachromosomally with the t-circles generating rolling circle tails that produce long duplex telomeric DNA segments that can then integrate back into the telomere. In another model, telomere elongation and rolling-circle replication occurs together with the 3' overhang of the telomere annealing to a ss t-circle of the complementary sequence which then creates a rolling circle template. The latter mechanism is illustrated in **Figure 4.1**. The goal of this study was to test the latter model in which telomere elongation occurs via t-circle priming from the telomeric 3' end. To do this we used model telomere templates and a well-defined replication system.
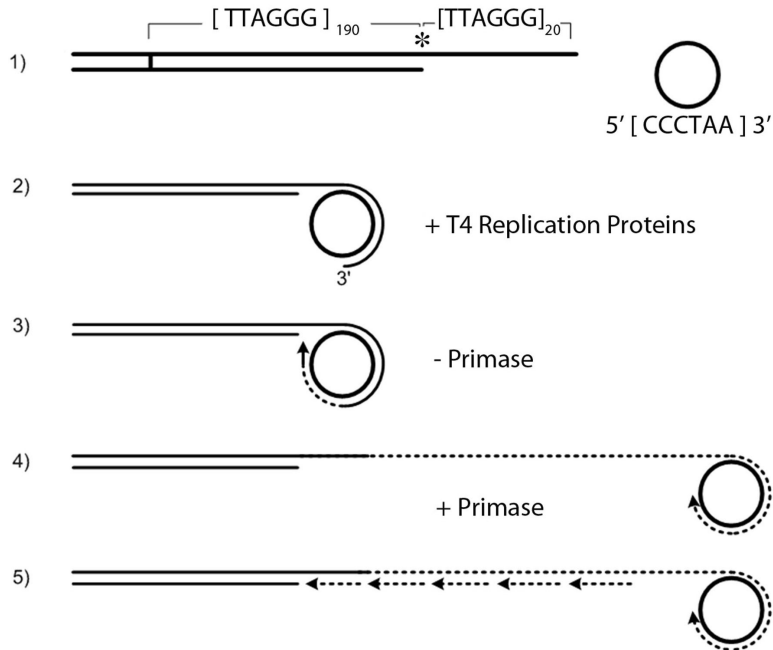
Figure 4.1: Model of t-circle mediated telomere extension primed from the 3' end of the telomere. 1) The model telomere consists of nontelomeric sequences followed by 190 TTAGGG repeats ending in a single stranded tail of 20 TTAGGG repeats. Asteric shows the site of radioactive labeling. 2) C-rich t-circle annealed to the complementary sequences in the 3' overhang. 3) Annealed circle is used as a template for the new synthesis of DNA by the T4 replisome. Priming occurs from the 3' end of the model telomere. 4) RCR through the C-circle with T4 replisome. 5) Lagging strand synthesis in the presence of primase

EXPERIMENTAL PROCEDURES

Plasmids and model templates:

A model linear telomere containing a 3' ss overhang was generated from a linearized plasmid DNA (pRST5) as described previously (33). A 124 nt oligo 5'-AGGG[TTAGGG]$_{20}$-3' was synthesized and PAGE purified by Integrated DNA Technologies (Iowa City, IA) and 50 pm of the oligo was labeled with 50 pmol of γ ^{32}P-ATP in 70 mM Tris-HCl pH 7.6, 10 mM MgCl$_2$, 5 mM dithiothreitol (DTT), with 20 units of T4 polynucleotide kinase (NEB Beverly, MA) at 37°C for 30 min. A ten fold molar excess of the oligo over the linear template (using 12.5 pmol of oligo) was incubated in 50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP with 200 units of T4 DNA ligase in a 200 μl volume at room temperature for 20 min. The free ss oligos were removed with an ultra 0.5 Centrifugal filter as directed by manufacturer (Amicon Bioseparations, Raleigh, NC).

A circular plasmid containing a ~400 nt displaced 5' terminated ss tail (pGLGAP) was prepared as described previously (34). The 5- terminal phosphate was removed with 3 units of Antarctic phosphatase (NEB Beverly, MA) (1 pmol template) in a 20 μl volume of Antarctic phosphatase reaction buffer for 15 min at 37°C followed by heat inactivation for 5 min at 65°C. The 5' terminus was labeled with γ ^{32}P-ATP using 20 units of T4 polynucleotide kinase (NEB Beverly, MA) as above.

Generation of telomeric circles:

C circles: Oligonucleotides 5'-(phosphate)[CCCTAA]$_{20}$-3' were synthesized and PAGE purified by Midland Certified Reagent Company (Midland, TX). Lyophilized oligos were resuspended in 20 μl of circularization buffer (33 mM Tris acetate pH 7.8, 66 mM KCl, 0.5

mM DTT 2.5 mM MnCl$_2$, 1 M betaine), and CircLigase II (Epicentre Biotechnologies, Madison, WI) (100 units) ss DNA ligase was added for 16 h at 60°C followed by heat inactivation of the enzyme for 10 min at 80°C. Exonuclease I (NEB Beverly, MA) (20 units) and exonuclease III (NEB Beverly, MA) (100 units) were then added for 3 h at 37°C to remove any non-circularized ss DNA substrate, followed by heat inactivation at 80°C for 30 min.

G circles: A 124 nt oligo 5'-(phosphate)AGGG[TTAGGG]$_{20}$-3' was synthesized and PAGE purified by Integrated DNA Technologies (lowa City, IA). The lyophilized oligos were resuspended in 20 μl of circularization buffer and circularized exactly as above including post treatment with exonucleases.

Replication reactions with purified T4 proteins

To generate templates consisting of the linear model telomere annealed to a ss t-circle, or control oligos, the radioactively labeled model telomere DNA (0.05 pmol in a 20 μl volume) was incubated with a 5 fold molar excess of t-circles or linear telomere oligos in 2 mM Hepes pH 7.5 with gradual cooling from 65°C to 37°C over 30 min.

The T4 replication proteins 43 polymerase, 61 primase, 51 helicase loader and 32 single strand binding protein were the generous gift from Dr. Nancy Nossal, and 41 helicase primase, 45 clamp, 44/62 clamp loader were purified by Dr. Sezgin Ozgur in this laboratory (to be described elsewhere). Full double strand replication reactions were carried out using 30 nM 43 polymerase, 64 nM 61 primase, 95 nM 59 helicase loader, 242 nM 44/62 clamp loader, 162 nM 45 clamp, 2 μM gene 32 protein and 328 nM 41 helicase at 37°C in the

presence of 25 mM K-Hepes (pH 7.6), 60 mM K acetate, 6 mM Mg acetate, 2 mM ATP, 0.25 mM CTP, 0.25 mM GTP, 0.25 mM UTP, 0.25 mM dATP, 0.25 mM dGTP, 0.25 mM dTTP, 0.25 mM dCTP, 20 ng/µl BSA, 10 mM beta mercaptoethanol in a 50 µl volume. The reactions (10 µl) were quenched at 0, 1, 2, 4, 8, 20 min with 3 µl of 90 mm EDTA, 6% SDS, 30 % glycerol, 0.25 % bromophenol blue, 0.25 % xylene blue, incubated at 55°C for 20 min and loaded on to 0.5% agarose gel. The dried gels were imaged using a Typhoon 9400 phosphoimager (Amersham Biosciences, Piscataway, NJ). For ss extensions, all the conditions were same as the ds extensions except that gene 61 primase was excluded.

Electron microscopic analysis of the replicated DNA

Unlabeled pGLGAP and the pRST5 model telomere annealed to C-rich t-circles were incubated with the T4 replication proteins as described above for 4 min. The reactions were stopped with 50 mM EDTA, deproteinized with 30 mM SDS and 4 µg proteinase K (Roche Applied Science, Indianapolis, IN) at 55°C for 1 h. The DNA was then purified by chromatography over agarose beads (ABT Inc., Burgos, Spain). To visualize the products of the ds extension reactions, the sample was prepared for EM by the cytochrome C the drop spreading method (35). DNA templates replicated in the absence of gene 61 primase to generate only long ss tails were purified as above. To visualize this DNA, *E. coli* SSB protein (USB Inc., Cleveland, OH) (50 ng) was added to 100 µl of the agarose bead eluate and incubated at room temperature for 20 min, followed by fixation with glutaldehyde (0.6% final concentration) for 5 min. The sample was then adsorbed to thin glow discharge treated carbon foils followed by rotary shadowcasting with tungsten at 1 x 10^{-6} Torr as described previously (36). An FEI Technai 12 instrument (Hillsboro, OR) at 40 kV was used to capture

the images using a Gatan Ultrascan US4000SP digital camera (Gatan, Pleasanton, CA). Images for publication were arranged and contrast optimized using Adobe Photoshop (Adobe Systems, San Jose, CA).

RESULTS

Generation of telomeric circles

ALT cells contain extrachromosomal circular DNA composed of 5'-TTAGGG-3' repeats (15) and can be found in ss, nicked, or ds form. In this study we constructed ss t-circles consisting of pure mammalian telomeric repeats as templates for telomerase independent replication extension reactions. A ssDNA specific ligase, CircLigase II (See Materials and Methods) was used to generate the t-circles from linear ss oligos **(Figure 4.2A and B lane I)**, yielding 120 nt t-circles of 5'-CCCTAA-3' and 5'-TTAGGG-3' repeats **(Figure 4.2A and B lanes II-III)**. Analysis of the circularization products on denaturing DNA gels showed that the circles were of a single size, and free of higher molecular weight circles **(Figure 4.2A and B lane II)**. The circularization efficiency was ~40% for C-rich t- circles and ~60% for G-rich t- circles **(Figure 4.2A and B lane II)**. *E. coli* exonucleases Exo I and Exo III digest linear but not circular ssDNA. Any remaining linear oligos were thus removed by treatment with these two enzymes (See Materials and Methods). The G-rich oligo was more resistant to the exonucleases likely due to the ability of the DNA to form higher order G-quadruplexes. Gel analysis however confirmed that ~80% of the C-rich products and ~60% of G-rich products remaining after exonuclease treatment were the desired t-circles **(Figure 4.2A and B lane III)**.
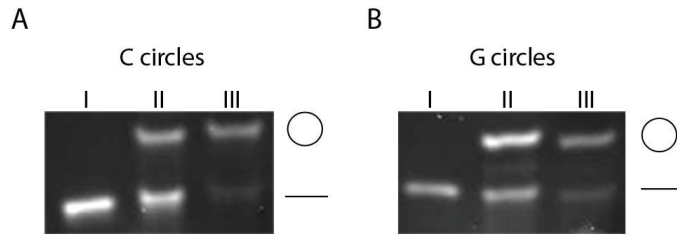
Figure 4.2: Synthesis of single-stranded t-circles. Oligos 124 nt long consisting of the C-rich telomeric repeat (A) or the G-rich telomeric repeat (B) were treated with CircLigase II (See Materials and Methods) and electrophoresed on 7 M urea gels and stained with Sybergold. Lanes I show the unligated oligos, lanes II show the ligated oligos, and Lanes III show the ligated oligos after treatment with ExoI and ExoIII.

C-rich t-circles mediate telomerase independent replication through a rolling circle

mechanism

Human chromosomes end in duplex 5'-TTAGGG-3' repeats with a 3' terminal ss overhang (3). This is modeled in the pRST5 plasmid (33) composed of non-telomeric plasmid sequences followed by a ~550 bp duplex tract of telomeric repeats. Linearization within the telomeric repeats with BsmB1 generates a 4 base overhang, used to anneal a 124 nt long G-rich telomeric oligo to generate a 120 nt 3' overhang **(Figure 4.1)**. At least 90% of the templates contained the overhang as assessed by the binding of *E. coli* SSB protein at one end of the DNA as seen by EM (See Materials and Methods) (data not shown).

To examine the extension of the telomeric repeats in this model telomere by a non-telomerase based mechanism employing t-circles, the model DNA alone, or mixed with G-rich t-circles, C-rich t-circles, G-rich linear oligos, or C-rich linear oligos was incubated with the T4 replication enzymes. The model template was labeled with ^{32}P on the 5' phosphate of the 124 nt oligo used to generate the 3' overhang to facilitate following the course of the

extension reaction by gel electrophoresis and autoradiography. A circular plasmid containing a single replication fork at a known location (34), pGLGAP, was used as a positive control for the replication assays. Because the T4 replication system is highly processive and replication proceeds rapidly, very high molecular weight DNA molecules were observed starting from 8 min of incubation of the forked pGLGAP template with the T4 proteins **(Figure 4.3A lane III)**. As longer DNA molecules were generated, they became too large to pass through the gel and remained in the well **(Figure 4.3A lanes III and IV)**. By 15 min a significant amount of the pGLGAP molecules had undergone replication **(Figure 4.3A lane IV)**.

When the model telomere template was annealed with the C-rich t-circles and used as a template for replication by the T4 proteins, a gradual increase in higher molecular weight products accumulating in the wells was observed after 4 min of incubation **(Figure 4.3B lane IV)**. At 20 min nearly all of the model telomere DNA had replicated as evidenced by its shift to slower moving species **(Figure 4.3B lane VI)**. As a control, incubation of the T4 replication proteins with the model template annealed to the G-rich t-circles failed to generate higher molecular weight products **(Figure 4.3D)**. Further, no high molecular weight products were detected when the model template was mixed with the linear G-rich or C-rich oligos, or incubated by itself with the T4 factors **(Figure 4.3E, F, and G respectively)**. If extension of the telomeric sequences in the presence of the C-rich t-circles follows a rolling circle mechanism, then leaving out T4 primase should result in long ss tails due to the lack of priming on the newly synthesized ss DNA tail. When this was carried out, the labeled template was again shifted to higher molecular weight species that accumulated in the wells

of the gel **(Figure 4.3C)**. This result supports a rolling circle mechanism of telomere elongation.

Alu I and Hpa II cleave duplex DNA at 5'-AGCT-3' and 5'-CCGG-3' sites respectively. These sequences will be present in most any long DNA segment but not in telomeric DNA which lacks these cut sites. To further examine the nature of the replication products, following 15 min of extension of the telomeric template with the C-rich t-circles (a time when ~50% of the DNA was extended), the products were treated with Alu I and Hpa II for 90 min **(Figure 4.4B)**. As a control, pGLGAP, which contains Alu I and Hpa II sites was replicated and treated in parallel **(Figure 4.4A)**. As expected, complete digestion of the pGLGAP replication products into small sized fragments was observed **(Figure 4.4A lanes III and IV)**. However with the model telomere DNA that had been extended with the C-rich t-circles, two different sized bands on the gel were observed **(Figure 4.4B lanes III and IV)**. The lower molecular weight band corresponds to the unreplicated model telomere. Its size is smaller than the full-length model telomere due to Hpa II and Alu I sites within the plasmid segment. The high molecular weight material including DNA in the well of the gel corresponds to the uncleaved extended telomeric DNA formed by t-circle dependent replication. These results confirm that the replication products are composed of telomeric repeats and that the C-rich t-circles act as substrates in the extension reaction.
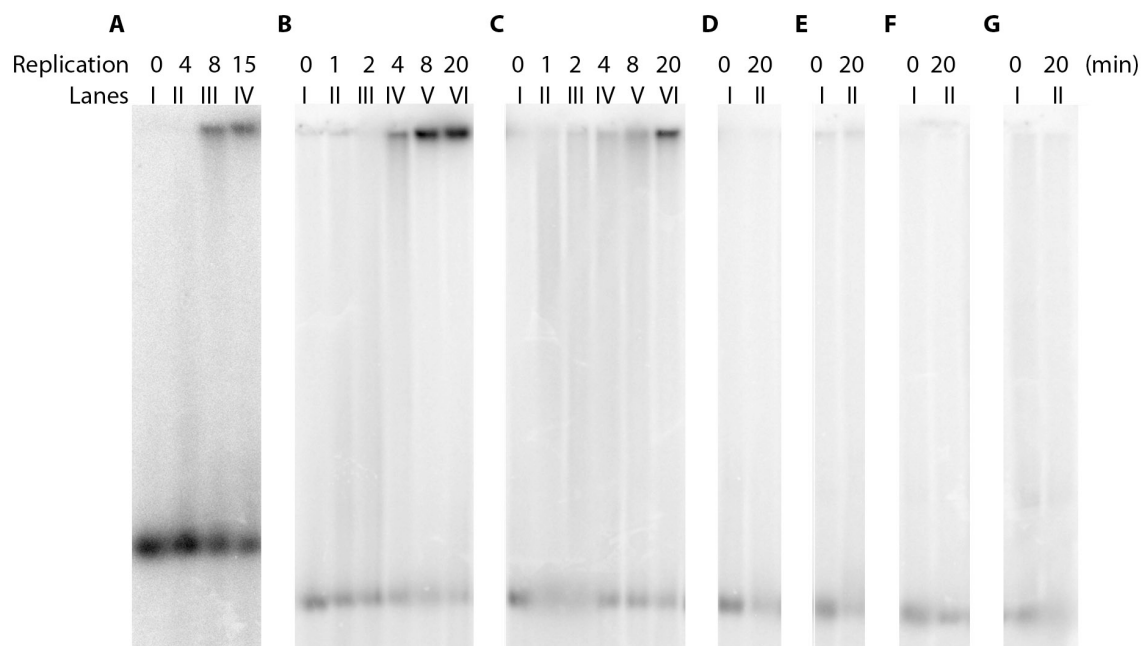
Figure 4.3: T-circle mediated telomere extension. Different templates for t-circle mediated telomere extension by the T4 replication proteins and a control ³²P labeled forked circle were incubated with the purified T4 replication proteins for 0 to 20 min and then electrophoresed on 0.5% agarose gels. (A) pGLGAP, a plasmid circle containing a pre-formed fork as a control; (B) ³²P labeled pRST5 model telomere annealed to C-rich t-circles, (C) C-rich t-circles in the absence of gp 61 primase for single strand extension, G-rich t-circles (D), C-rich linear telomeric oligo (E), G-rich linear telomeric oligo (F), and the model telomere alone (G). Reaction time increases from lanes I to VI in each panel.
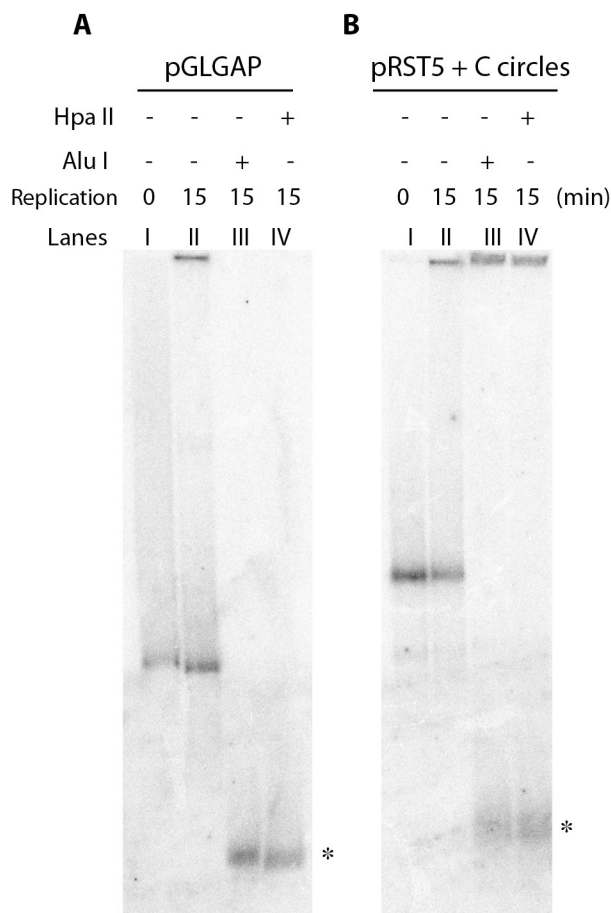
Figure 4.4: Restriction digestion of the replication products confirms the amplification of telomeric repeats. The telomeric sequences extended by t-circle dependent replication should be resistant to cleavage by restriction enzymes while the replicated pGLGAP plasmid would be cleaved. (A) pGLGAP containing a ^{32}P labeled pre-formed fork and replicated with the T4 proteins. (B) pRST5 model telomere ^{32}P labeled at the overhang, annealed to C-rich t-circles and extended for 15 min with T4 replication proteins (lanes II). Both pGLGAP and the model telomere were treated with either Alu I (lanes III) or Hpa II (lanes IV) for 90 min following replication. Asterics correspond to the digested nontelomeric sequences. Repetitive sequence containing replication products remain in the wells in lanes III and IV (B).

Visualization of extension products by electron microscopy

EM was used to visualize the products of the telomere extension reactions. Reactions with the model telomere annealed to C-rich t-circles (4 min) were carried out followed by deproteinization and purification over an agarose bead column. The DNA was then spread on a film of denatured cytochrome-C protein at an air-water interface (See Materials and Methods). This provided a direct means of visualizing very long DNA products. The EM analysis confirmed that the majority of the DNA molecules were extended with products varying in length from 4 to 32 kb **(Figure 4.5A, B and D)**.

The products of the 4 min extension reactions lacking T4 primase (see above) were deproteinized, passed through a sizing column, and then incubated with SSB protein to stain the ssDNA. The resulting complexes were prepared for EM by mounting onto thin carbon supports, air drying and rotary shadowcasting with tungsten (See Materials and Methods). Examination revealed very long ssDNA segments bound by SSB **(Figure 4.5C)**, further supporting a rolling circle mechanism of telomere extension.
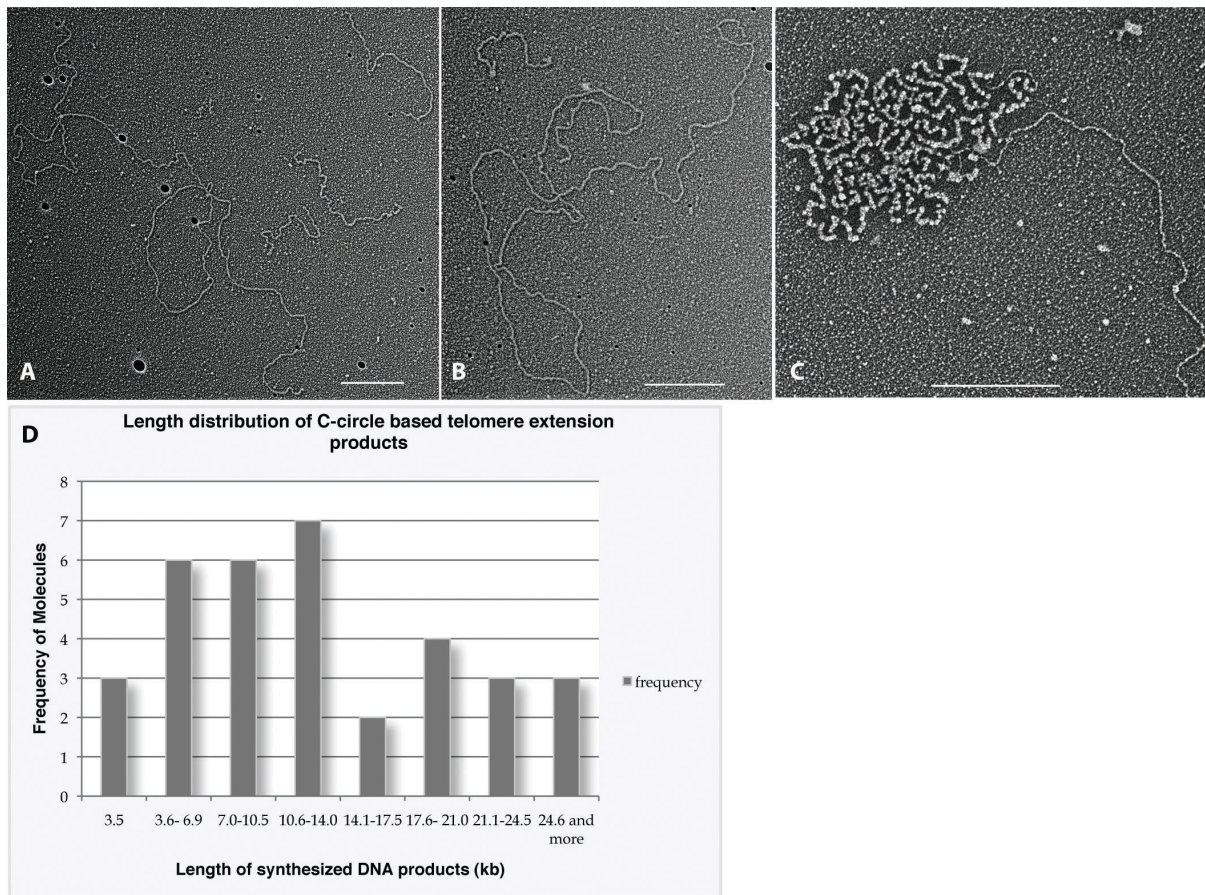
Figure 4.5: EM analysis of the products of t-circle mediated telomere extension. (A, B) The model telomere, pRST5 was annealed to C-rich t-circles and extended by replication for 4 min with the complete set of T4 replication factors. The purified DNA was spread on a denatured film of cytochrome C at an air-water interface and prepared for EM (See Materials and methods). The lengths of the molecules are: 4.6, 6, 6 and 32.7 kb (A), and 3.5 kb, and 31 kb in (B). (C) Single stranded DNA formed from the replication of the model telomere as in A and B but in the absence of pg61 primase was coated with SSB protein and mounted onto thin carbon films and rotary shadowcast with tungsten to reveal that very long single stranded products are produced in the absence of primase. All images are shown in reverse contrast. Size bars correspond to 500 nm in A, B and 200 nm in C. (D) Histogram representing the size distribution of synthesized DNA molecules.

DISCUSSION

The aim of this study was to design an *in vitro* system that would allow us to address the question of whether t-circles can mediate telomere extension in the absence of telomerase. A well-defined replication system, the highly purified T4 replication proteins, was employed along with a model telomere template. Single stranded C-rich t-circles and G-rich t-circles composed of pure telomeric repeats were prepared and these t-circles or the linear oligos of the same sequences were incubated with the model telomere template in the presence of the T4 proteins. The results demonstrate that the telomeric tract in the model telomere can be extended to over 50 times its initial length in a reaction dependent on the replication factors and the C-rich t-circles. G-rich t-circles or the linear oligos could not mediate telomere extension. The products were resistant to four base cutters and were single stranded when T4 primase was omited.

We chose to use the bacteriophage T4 system for this study as the proteins are highly purified, available in our laboratory and are free of nucleases. The mechanism of replication by the T4 proteins is well characterized and provides an excellent model for the complex multienzyme replication systems in eukaryotic organisms (37). Moreover, the T4 replication complex can self-assemble *in vitro* and the parameters of synthesis are similar to those measured *in vivo* (38). The aim in this study was to ask if telomeric circles can initiate telomere extension and the T4 replication proteins have shown that this can occur. Now that this is known, in future studies we will examine the ability of semi-purified extracts from human cells to carry out similar extensions.

The potential role of C-rich t-circles in telomere replication has been discussed previously. Lindström *et al* (39) prepared nanocircles with telomeric repeats and annealed a

125

primer to show that Φ 29 DNA polymerase would extend the primer. Similarly, Henson *et al* (40) used C-rich t-circles with a linker sequence acting as a primer for extension with Φ 29 DNA polymerase in their assay to detect ALT activity and showed that C-rich t-circles can self-prime RCR *in vitro*. Our study extends these reports by showing that C-rich t-circles can anneal to a 3' overhang typical of that found on human telomeres and then use this overhang as a primer for telomere extension. Moreover, we have shown that a well characterized replication system comprised of a DNA polymerase, its sliding clamp, clamp loader, primase, helicase, helicase loader, and a single strand binding protein will use this template in a very efficient manner to extend the 576 bp telomeric tract to over 50 times its starting length. This extension was observed as new ds telomeric sequences or ss telomereic DNA if primase was omitted.

T-circles have been suggested to result from the cleavage of t-loops (15) which resemble homologous recombination (HR) intermediates and thus would be prone to cleavage by proteins that resolve such structures. This suggestion is supported by the finding that the size of t-loops and t-circles are closely correlated (15) , (20) and the observation that the continued presence of t-circles in ALT cells requires normal levels of two recombination factors, XRCC3 and Nbs1 (32). Telomere rapid deletion, which is a Rad 52 dependent shortening of overly long telomeres in *S. cerevisiae,* results in t-circle formation, is also believed to have a t-loop intermediate (41). A similar telomere trimming mechanism and t-circle formation is observed in telomerase positive human cells when their telomeres are overextended by overexpression of telomerase components (42). Deletion of Ku 70/80 in *A. thaliana* (22), induced loss of Ku86 in human somatic cells (24), overexpression of TRF2 lacking its basic domain (25), depletion of ORC2 (26) or WRN helicase (27) all result in the

126

appearance of t-circles. These observations argue that alteration or abrogation of any of a number of genes needed for normal telomere maintenance or telomere-related DNA repair result in changes in the telomere, signaled by the appearance of t-circles. Nonetheless the exact proteins involved in t-circle generation or maintenance in ALT cells remains unclear.

The mechanism of telomere length maintenance in the absence of telomerase is not well understood. Two different general recombination based mechanisms have proposed: unequal telomere sister chromatid exchange (43) and HR dependent DNA replication in which the template can be t-loops, sister chromatids or t-circles (44-46). The mechanism for which there is the most experimental evidence is circle mediated telomere extension, which falls within the latter general mechanism (15,18,25,29). The results of our study described here further extend our understanding of how RCR priming can occur and show that it can be initiated from the end of the telomere. Moreover, our results provide direct clues for why telomeres in ALT cells are able to expand so quickly.

In our model, telomere extension occurs directly from the 3' telomere overhang. In such an expansion, the 3' ss telomere overhang anneals to a C-rich t-circle and acts as a primer for new DNA synthesis by using the circle as a template. Circle-mediated replication would then result in very long arrays of telomeric sequences being rapidly added to the telomere (Figure 1). Double stranded t-circles can also promote circle based telomere extension however, the 3' overhang first needs to invade the duplex circle and then extension can be catalyzed. In contrast, in a model based on RCR occurring extrachromosomally, in that case, telomeric circles produce new telomeric DNA, which is then added to the chromosome end. Here both C-rich t-circles and G-rich t-circles would have similar probabilities of replication. The observations of Henson *et al* (40) provides further support for our model. They showed that

there are a hundred-fold fewer G-rich t-circles than C-rich t-circles in ALT cells (~1000 C-rich t-circles per ALT cell) and that C-rich t-circles are present in blood of ALT-positive osteosarcoma patients (40). Also they showed that C-rich t-circles are ~750 times more abundant in ALT cells than in telomerase positive cells or in non-immortalized cell strains. Finally, ALT activity (extension of the partially double stranded C-rich t-circles with Φ 29 DNA polymerase) was directly correlated with the abundance of C-rich t-circles with the level of C-rich t-circles decreasing rapidly when the ALT activity was inhibited. Their findings together with the work described here illustrates why C-rich t-circles would be specific to ALT cells and how they are able to extend telomeres in the absence of telomerase.

128

REFERENCES

1.   Raices, M., Verdun, R. E., Compton, S. A., Haggblom, C. I., Griffith, J. D., Dillin, A., and Karlseder, J. (2008) *Cell* **132**, 745-757

2.   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

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

4.   Bianchi, A., Stansel, R. M., Fairall, L., Griffith, J. D., Rhodes, D., and de Lange, T. (1999) *EMBO J* **18**, 5735-5744

5.   Palm, W., and de Lange, T. (2008) *Annu Rev Genet* **42**, 301-334

6.   Harley, C. B., Futcher, A. B., and Greider, C. W. (1990) *Nature* **345**, 458-460

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

8.   d'Adda di Fagagna, F., Reaper, P. M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N. P., and Jackson, S. P. (2003) *Nature* **426**, 194-198

9.   Hayflick, L. (1965) *Exp Cell Res* **37**, 614-636

10. Shay, J. W., and Wright, W. E. *FEBS Lett* **584**, 3819-3825

11. Dunham, M. A., Neumann, A. A., Fasching, C. L., and Reddel, R. R. (2000) *Nature genetics* **26**, 447-450

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

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

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

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

16. Tokutake, Y., Matsumoto, T., Watanabe, T., Maeda, S., Tahara, H., Sakamoto, S., Niida, H., Sugimoto, M., Ide, T., and Furuichi, Y. (1998) *Biochemical and biophysical research communications* **247**, 765-772

17.   Nabetani, A., and Ishikawa, F. (2009) *Mol Cell Biol* **29**, 703-713

18.   Tomaska, L., Nosek, J., Makhov, A. M., Pastorakova, A., and Griffith, J. D. (2000) *Nucleic Acids Res* **28**, 4479-4487

19.   Nosek, J., Rycovska, A., Makhov, A. M., Griffith, J. D., and Tomaska, L. (2005) *J Biol Chem* **280**, 10840-10845

20.   Cesare, A. J., Groff-Vindman, C., Compton, S. A., McEachern, M. J., and Griffith, J. D. (2008) *Mol Cell Biol* **28**, 20-29

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

22.   Zellinger, B., Akimcheva, S., Puizina, J., Schirato, M., and Riha, K. (2007) *Mol Cell* **27**, 163-169

23.   Cohen, S., and Mechali, M. (2002) *EMBO Rep* **3**, 1168-1174

24.   Wang, Y., Ghosh, G., and Hendrickson, E. A. (2009) *Proc Natl Acad Sci U S A* **106**, 12430-12435

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

26.   Deng, Z., Dheekollu, J., Broccoli, D., Dutta, A., and Lieberman, P. M. (2007) *Curr Biol* **17**, 1989-1995

27.   Li, B., Jog, S. P., Reddy, S., and Comai, L. (2008) *Mol Cell Biol* **28**, 1892-1904

28.   Basenko, E. Y., Cesare, A. J., Iyer, S., Griffith, J. D., and McEachern, M. J. *Nucleic Acids Res* **38**, 182-189

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

30.   Tomaska, L., Nosek, J., Kramara, J., and Griffith, J. D. (2009) *Nat Struct Mol Biol* **16**, 1010-1015

31.   Lustig, A. J. (2003) *Nat Rev Genet* **4**, 916-923

32.   Compton, S. A., Choi, J. H., Cesare, A. J., Ozgur, S., and Griffith, J. D. (2007) *Cancer research* **67**, 1513-1519

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

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

35.   Thresher, R., and Griffith, J. (1992) *Methods Enzymol* **211**, 481-490

36.   Griffith, J. D., and Christiansen, G. (1978) *Annual review of biophysics and bioengineering* **7**, 19-35

37.   Nossal, N. G. (1992) *FASEB J* **6**, 871-878

38.   Mueser, T. C., Hinerman, J. M., Devos, J. M., Boyer, R. A., and Williams, K. J. *Virol J* **7**, 359

39.   Lindstrom, U. M., Chandrasekaran, R. A., Orbai, L., Helquist, S. A., Miller, G. P., Oroudjev, E., Hansma, H. G., and Kool, E. T. (2002) *Proc Natl Acad Sci U S A* **99**, 15953-15958

40.   Henson, J. D., Cao, Y., Huschtscha, L. I., Chang, A. C., Au, A. Y., Pickett, H. A., and Reddel, R. R. (2009) *Nature biotechnology* **27**, 1181-1185

41.   Lin, C. Y., Chang, H. H., Wu, K. J., Tseng, S. F., Lin, C. C., Lin, C. P., and Teng, S. C. (2005) *Eukaryot Cell* **4**, 327-336

42.   Pickett, H. A., Cesare, A. J., Johnston, R. L., Neumann, A. A., and Reddel, R. R. (2009) *The EMBO journal* **28**, 799-809

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

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

45.   Muntoni, A., Neumann, A. A., Hills, M., and Reddel, R. R. (2009) *Hum Mol Genet* **18**, 1017-1027

46.   Cesare, A. J., and Reddel, R. R. (2010) *Nat Rev Genet* **11**, 319-330

CHAPTER 5: CONCLUSIONS AND FINAL THOUGHTS

The repetitive character of the G-rich telomeric DNA strand enables the formation of unique secondary structures such as t-loops and G-quadruplexes at chromosome ends (1,2). While both t-loops and G-quadruplexes prevent exonucleolytic attack to the telomeric overhang (3), they need to be resolved in S-phase for replication to occur. Failure to do so results in replicative stress and check point activation (4). Moreover, t-loops are suggested to be the origin of t-circles (3), which are the markers of ALT (5). Therefore, regulation of t-loop and G-quadruplex structures could be important for normal cellular processes and for prevention of ALT. ALT is a recombination based telomere extension mechanism, found in ~15% of cancer cases, though the exact molecular mechanism of ALT remains to be solved (6). Rap1, TRF2 and Pot1 are known to suppress recombination and thus might have role in the repression of ALT (7). The aim of this study was to understand the functional interactions of shelterin components hRap1, hPot1 and their associated complexes TRF2/Rap1 and hPot1-TPP1 with telomeric DNA. This study also addressed how shslterin components regulate telomere topology, and if t-circles can act as a substrate for ALT. Findings in this work suggest that prevention of ALT in healthy cells could be a two-step process. While the DNA binding ability of Rap1 induces the first level of prevention, Pot1 brings the second level of control through its G-quadruplex resolution function.

In the second chapter of this study, I demonstrated that hRap1 directly interacts with DNA in the absence of hTRF2 and prefers to bind to the junction structures without a

sequence preference. Moreover, hRap1 forms a tetrameric complex with hTRF2 and the complex has a ~2 fold higher affinity for ds telomeric regions and ~10 fold higher affinity for telomeric ds-ss junction sites when compared to hTRF2 alone. Even though hRap1 cannot initiate t-loop formation on its own, the TRF2/Rap1 complex facilitates a higher percentage of t-loops than TRF2 alone indicating that hRap1 modulates t-loop dynamics by affecting hTRF2. This finding is important for telomere end protection and the prevention of t-circle formation.

In the third chapter, I illustrated that G-rich ss telomeric DNA forms uniform higher order bead structures that are connected through linker DNA. Each bead is composed of ~480 nucleotides and is a mixture of both parallel and antiparallel type G-quadruplexes. hPot1 and the hPot1-TPP1 complex can open up the preexisting G-quadruplexes on their own and form filaments along the G-rich DNA as they melt the quadruplexes. The hPot1-TPP1 complex binds independent of the telomere overhang length, while hPot1 alone saturates the telomeric overhang as monomers by binding to ~20 nucleotides at a time and compacting the DNA length two fold. This finding indicates that hPot1 alone and the hPot1-TPP1 complex present different DNA binding properties while modulating G-quadruplex structures and which would prevent recombination between G quadruplexes.

In the fourth chapter, I demonstrated that C-rich t-circles act as a substrate for the telomere extension reaction while G-rich t-circles or linear G- or C-rich DNA cannot. Modulation of the primase in the multi component *in vitro* replication system with the T4 replication complex illustrated that telomere extension with C-circles is both rolling circle replication and coupled replication. The length of the DNA products presented a heterogeneous distribution and could extend over 50 kbs in 4 minutes, meaning that the rate

133

of extension around the C-circle is ~100 rounds/minute. This finding could explain the heterogeneous character and rapid changes in ALT lengthened telomeres.

The mechanisms of ALT are suggested to be recombination mediated (8) and deletion of hTRF2, hPot1 and hRap1 result in T-SCEs indicating that they can suppress ALT (7,9,10). The finding that there is no change in the total levels of hRap1 in ALT cells (11) indicates that ALT initiation is not dependent on Rap1's expression levels but rather dependent on its localization on the DNA and/or the post-translational modifications. Therefore, understanding the DNA binding ability of hRap1 is a key to understand how hRap1 exerts these functions. The work presented here suggests three different ways how hRap1 can suppress ALT initiation. First, hRap1 has a high preference for ds-ss junctions, which could suppress HR directly **(Model 1A)**. Second, hRap1 can stabilize telomeric ends and prevent t-circle formation by interacting with hTRF2 and increasing hTRF2's affinity for junctions and ability to form/stabilize t-loops **(Model 1B)**. Finally, the TRF2/Rap1 complex presents higher affinity for ds telomeric sequences which would facilitate its binding to subtelomeric sites **(Model 1C)**. Since there is no evidence suggesting one model over the other, we believe that all three mechanisms could be occurring simultaneously. In support of Model 1C, yRap1 is engaged in subtelomeric silencing (12). There is also evidence suggesting that hRap1 and mRap1 are enriched at subtelomeric regions (12-15). Therefore we believe that hRap1 could induce subtelomeric silencing as yRap1 does.

In our model we believe that hRap1 could recruit chromatin modifiers such as ATRX to induce silencing and to decrease TERRA expression. ATRX/DAXX localize to heterochromatin and PML bodies (16). Loss off or mutations in ATRX, DAXX and H3.3 results in an increase in TERRA levels and the presence of ALT-like long telomeres in

pancreatic neuroendocrine cancers, pediatric glioblastomas, other tumors and 22 ALT cell lines (17-22). These findings suggest that ALT is epigenetically regulated and that TERRA transcription can have a role in ALT initiation and maintenance, which is consistent with our model. It would be interesting to find out whether hRap1 interacts with chromatin modifiers directly and whether it induces heterochromatin formation at telomeres. This might affect the transcriptional levels of TERRA, and thus have a direct effect on ALT formation.

Post-translational modifications could also alter the interaction of hRap1 with telomeric DNA or with its binding partner hTRF2 and thus could also lead to ALT phenotypes. In fact, sumoylation of hRap1, hTRF2 and hTRF1 by SMC5/6 complex facilitates APB formation as well as targeting of telomeric DNA to PML bodies and HR (23). Moreover, inhbition of SMC5/6 complex blocks HR, shortening of telomeres and causes senescence in ALT cells (23). Therefore sumoylation is critical for removing Rap1 from telomeres and APB formation. This supports our model in which hRap1's interaction with telomeric DNA is necessary to inhibit ALT.

The exact mechanism of telomeric recombination in APBs is not known and the two suggested mechanisms are unequal TSCEs and HR-dependent DNA replication (24). However, neither of these models explains why TERRA levels are high in various ALT cell lines and cancers. Recent data suggests that ALT is epigenetically controlled and we believe that hRap1 could be controlling this mechanism by binding to subtelomeric sites and silencing TERRA transcription. A similar mechanism that initiates recombination through transcription occurs at immunoglobulin switch regions (S-regions) (25). S-regions are composed of repetitive G-rich DNA and recombination is mediated through G-quadruplexes (25). Recombination only occurs when S-region is transcribed (26). Transcription causes

loop formation and cleavage of the loops results in circular side products (27). Consistently, Rap1 is enriched at S-regions and deletion of mRap1 results in increase S-region transcription (13) suggesting the importance of hRap1's silencing function in recombination repression.

When protection by Rap1 fails, the second level of protection against ALT can arise from the filament formation ability of hPot1 and the hPot1-TPP1 complex **(Model 2)**. As hPot1 and the hPot1-TPP1 complex form filaments, they would open up the G-quadruplexes and prevent the interactions between G-quadruplexes. If Pot1 and hPot1-TPP1 complex are lost, then the G quadruplexes can initiate recombination between each other without any regulation and that could lead to interstrand recombination and TSCEs. Indeed, deletion of mPot1 results in ~10% TSCEs (28). Moreover, it has been demonstrated that the G-quadruplex structures formed at MSH43 minisatellite mediate recombination *in vitro* potentially by forming bimolecular quadruplexes or octaplexes between interstrand G-quadruplexes (29). This data is consistent with our model and it is possible that G-quadruplexes formed at telomeric sequences could initiate recombination once unregulated.

If G-quadruplex formation is transcription coupled as in the case of switch recombination, t-circles can arise. These circles would be single stranded G-circles and double stranded t-circles, in which case the circle would be RNA/DNA hybrids. Double stranded t-circles would have an intact C-rich telomeric DNA strand and TERRA strand. Separation of TERRA or exonucleolytic digestion of RNA strand would result in C-circles, which in turn can act as a substrate for RCR. When TSCEs occur, the recombined sister chromatids could be extended as suggested in HR-dependent DNA replication model. After extension, the strands could be resolved by cellular factors that recognize G-quadruplexes

136

such as MRN complex. The MRN complex is the first protein to be identified as a component for ALT-mediated telomere maintenance (30,31). The MRN complex's ability to recognize G-quadruplexes and how the conserved nuclease Mre11 cleaves G-quadruplex DNA has been reported (32). These two findings also support our model and suggest that recombination could be G-quadruplex mediated. It is always a possibility that many other cellular factors may be involved in this process and future research is needed to illuminate their function.

The model suggested here is consistent with the previously suggested mechanisms of ALT. Furthermore, it incorporates the recent findings about the epigenetic control and the findings of this study to elucidate how DNA binding properties of hRap1 and hPot1 can prevent ALT. The origin of t-circles was suggested to be t-loops but conclusive experimental proof is lacking. This model suggests that t-circles can arise in an alternative way as well. Finally, according to this model G-quadruplexes are important for recombination in ALT. Therefore G-quadruplex stabilizers could prevent recombination by stabilizing the interactions within the G-beads and be used in the treatment of ALT. Recent initiatives for G-quadruplex stabilizers as ALT therapeutics is giving promising results (33,34). So far stabilizers were designed by using short oligonucleotides as a template however, this work suggested that G-beads are very big and complex motifs. Therefore further structural analysis of the G-beads would help in designing new therapeutics with better pharmacokinetic properties and higher target specificity.
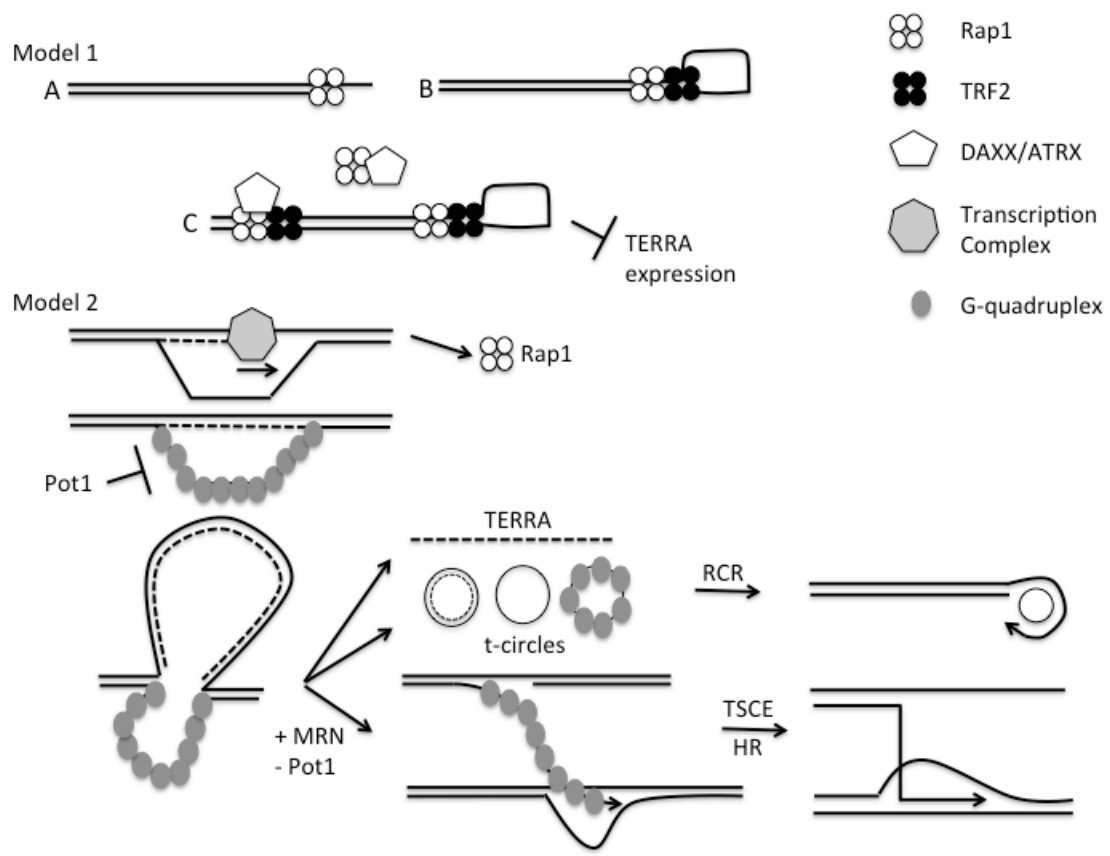
Figure 5.1: Models of hRap1 and hPot1 mediated ALT inhibition. hRap1 could suppress ALT by directly binding to the junction sites (Model 1A), by facilitating t-loop formation (Model 1B) or by recruiting chromatin modifiers such as DAXX/ATRX to silence the telomeres and this would prevent TERRA expression (Model 1C). Removal of hTRF2 and hRap1 from telomeres by sumoylation and/or mutations or loss of DAXX/ATRX would remove the silencing effect and telomeres would be prone to TERRA expression (Model 2). For the recombination to occur G-quadruplexes should form on the untranscribed strand. hPot1 and the hPot1-TPP1 complex could prevent ALT by forming filaments, which would open up the G quadruplexes and also cover them and prevent the interactions between G qaudruplexes. Loss of Pot1 and the Pot1-TPP1 complex would lead to recombination and TSCEs. After the recombined telomeres are extended, MRN complex would cleave the recombined G-quadruplexes to release the strands. In the meantime, if the recombination occurs intrachromosomally, loop structures could form. Cleavage of loops by cellular factors would result in the formation of ds t-circles and G-circles. The t-circles could have nicks on them and exonuclease action would cause C-circles. C-circles would then act as a substrate for the telomere extension by RCR.

REFERENCES

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

2.   Neidle, S., and Parkinson, G. N. (2003) *Curr Opin Struct Biol* **13**, 275-283

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

4.   Rizzo, A., Salvati, E., Porru, M., D'Angelo, C., Stevens, M. F., D'Incalci, M., Leonetti, C., Gilson, E., Zupi, G., and Biroccio, A. (2009) *Nucleic Acids Res* **37**, 5353-5364

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

6.   Shay, J. W., and Bacchetti, S. (1997) *European journal of cancer* **33**, 787-791

7.   Jiang, W. Q., Zhong, Z. H., Henson, J. D., and Reddel, R. R. (2007) *Oncogene* **26**, 4635-4647

8.   Dunham, M. A., Neumann, A. A., Fasching, C. L., and Reddel, R. R. (2000) *Nature genetics* **26**, 447-450

9.   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

10.   Sfeir, A., Kabir, S., van Overbeek, M., Celli, G. B., and de Lange, T. (2010) *Science* **327**, 1657-1661

11.   Zhang, Y., Cai, L., Wei, R. X., Hu, H., Jin, W., and Zhu, X. B. (2011) *Oncol Lett* **2**, 1327-1332

12.   Iglesias, N., Redon, S., Pfeiffer, V., Dees, M., Lingner, J., and Luke, B. (2011) *EMBO Rep* **12**, 587-593

13.   Martinez, P., and Blasco, M. A. *Nat Rev Cancer* **11**, 161-176

14.   Yang, D., Xiong, Y., Kim, H., He, Q., Li, Y., Chen, R., and Songyang, Z. (2011) *Cell research* **21**, 1013-1027

15.   Arat, N. O., and Griffith, J. D. (2012) *J Biol Chem* **287**, 41583-41594

16.   McDowell, T. L., Gibbons, R. J., Sutherland, H., O'Rourke, D. M., Bickmore, W. A., Pombo, A., Turley, H., Gatter, K., Picketts, D. J., Buckle, V. J., Chapman, L., Rhodes, D., and Higgs, D. R. (1999) *Proc Natl Acad Sci U S A* **96**, 13983-13988

17. Lovejoy, C. A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S., Petrini, J. H., Sung, P. A., Jasin, M., Rosenbluh, J., Zwang, Y., Weir, B. A., Hatton, C., Ivanova, E., Macconaill, L., Hanna, M., Hahn, W. C., Lue, N. F., Reddel, R. R., Jiao, Y., Kinzler, K., Vogelstein, B., Papadopoulos, N., and Meeker, A. K. (2012) *PLoS Genet* **8**, e1002772

18. Jiao, Y., Shi, C., Edil, B. H., de Wilde, R. F., Klimstra, D. S., Maitra, A., Schulick, R. D., Tang, L. H., Wolfgang, C. L., Choti, M. A., Velculescu, V. E., Diaz, L. A., Jr., Vogelstein, B., Kinzler, K. W., Hruban, R. H., and Papadopoulos, N. (2011) *Science* **331**, 1199-1203

19. Heaphy, C. M., de Wilde, R. F., Jiao, Y., Klein, A. P., Edil, B. H., Shi, C., Bettegowda, C., Rodriguez, F. J., Eberhart, C. G., Hebbar, S., Offerhaus, G. J., McLendon, R., Rasheed, B. A., He, Y., Yan, H., Bigner, D. D., Oba-Shinjo, S. M., Marie, S. K., Riggins, G. J., Kinzler, K. W., Vogelstein, B., Hruban, R. H., Maitra, A., Papadopoulos, N., and Meeker, A. K. (2011) *Science* **333**, 425

20. Schwartzentruber, J., Korshunov, A., Liu, X. Y., Jones, D. T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A. M., Quang, D. A., Tonjes, M., Hovestadt, V., Albrecht, S., Kool, M., Nantel, A., Konermann, C., Lindroth, A., Jager, N., Rausch, T., Ryzhova, M., Korbel, J. O., Hielscher, T., Hauser, P., Garami, M., Klekner, A., Bognar, L., Ebinger, M., Schuhmann, M. U., Scheurlen, W., Pekrun, A., Fruhwald, M. C., Roggendorf, W., Kramm, C., Durken, M., Atkinson, J., Lepage, P., Montpetit, A., Zakrzewska, M., Zakrzewski, K., Liberski, P. P., Dong, Z., Siegel, P., Kulozik, A. E., Zapatka, M., Guha, A., Malkin, D., Felsberg, J., Reifenberger, G., von Deimling, A., Ichimura, K., Collins, V. P., Witt, H., Milde, T., Witt, O., Zhang, C., Castelo-Branco, P., Lichter, P., Faury, D., Tabori, U., Plass, C., Majewski, J., Pfister, S. M., and Jabado, N. (2012) *Nature* **482**, 226-231

21. Wong, L. H., McGhie, J. D., Sim, M., Anderson, M. A., Ahn, S., Hannan, R. D., George, A. J., Morgan, K. A., Mann, J. R., and Choo, K. H. (2010) *Genome Res* **20**, 351-360

22. Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., and Allis, C. D. (2010) *Proc Natl Acad Sci U S A* **107**, 14075-14080

23. Potts, P. R., and Yu, H. (2007) *Nat Struct Mol Biol* **14**, 581-590

24. Cesare, A. J., and Reddel, R. R. (2010) *Nat Rev Genet* **11**, 319-330

25. Yu, K., and Lieber, M. R. (2003) *DNA Repair (Amst)* **2**, 1163-1174

26. Daniels, G. A., and Lieber, M. R. (1995) *Proc Natl Acad Sci U S A* **92**, 5625-5629

27. Reaban, M. E., Lebowitz, J., and Griffin, J. A. (1994) *J Biol Chem* **269**, 21850-21857

28. Palm, W., Hockemeyer, D., Kibe, T., and de Lange, T. (2009) *Mol Cell Biol* **29**, 471-482

29. Boan, F., and Gomez-Marquez, J. (2010) *Chembiochem* **11**, 331-334

30. Jiang, W. Q., Zhong, Z. H., Henson, J. D., Neumann, A. A., Chang, A. C., and Reddel, R. R. (2005) *Mol Cell Biol* **25**, 2708-2721

31. Zhong, Z. H., Jiang, W. Q., Cesare, A. J., Neumann, A. A., Wadhwa, R., and Reddel, R. R. (2007) *J Biol Chem* **282**, 29314-29322

32. Ghosal, G., and Muniyappa, K. (2007) *J Mol Biol* **372**, 864-882

33. Temime-Smaali, N., Guittat, L., Sidibe, A., Shin-ya, K., Trentesaux, C., and Riou, J. F. (2009) *PLoS One* **4**, e6919

34. Huang, F. C., Chang, C. C., Wang, J. M., Chang, T. C., and Lin, J. J. (2012) *Br J Pharmacol* **167**, 393-406