MUTAGENIC POTENTIAL OF TELOMERIC REPEATS AND THE ROLE OF WERNER SYNDROME HELICASE PROTEIN IN FACILITATING TELOMERIC DNA REPLICATION

by

Rama Rao Damerla

B.Sc Genetics, Microbiology and Chemistry, Osmania University, India, 2003

M.Sc Genetics, Osmania University, India, 2005

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Rama Rao Damerla

It was defended on

April 18th, 2011

And approved by

Dissertation Advisor: Patricia Opresko Ph.D.
Assistant Professor
Department of Environmental and Occupational Health
Graduate School of Public Health, University of Pittsburgh

Committee Chairperson: Susanne Gollin, Ph.D.
Associate Professor
Department of Human Genetics
Graduate School of Public Health, University of Pittsburgh

Committee Member: M. Michael Barmada, Ph.D.
Associate Professor
Department of Human Genetics
Graduate School of Public Health, University of Pittsburgh

Committee Member: Laura Niedernhofer, M.D., Ph.D.
Associate Professor
Department of Molecular Genetics and Biochemistry
School of Medicine, University of Pittsburgh

Committee Member: Robert Sobol, Ph.D.
Assistant Professor
Departments of Pharmacology & Chemical Biology and Human Genetics
School of Medicine, University of Pittsburgh

MUTAGENIC POTENTIAL OF TELOMERIC REPEATS AND THE ROLE OF THE WERNER SYNDROME HELICASE PROTEIN IN FACILITATING TELOMERIC DNA REPLICATION

Rama Rao Damerla, Ph.D.

University of Pittsburgh, 2011

Chromosome termini form nucleoprotein structures called telomeres that consist of tandem repeats of TTAGGG DNA sequences (mammals) and telomeric proteins. Telomeres play a critical role in cell survival and genomic stability. Biochemical studies showed that the G-rich strand of telomeres can fold into secondary DNA structures called G-quadruplexes (G4-DNA), which are thought to impact telomere length regulation and telomeric DNA stability. G4 DNA structures are capable of interfering with DNA synthesis by blocking DNA polymerases *in vitro* and are proposed to hinder replication *in vivo*. We cloned telomeric repeats into reporter cassettes on shuttle vectors and replicated them in normal human somatic cells to determine if telomeric repeats induce mutations and deletions due to their ability to fold into G4 DNA structures. We demonstrated for the first time that G-rich telomeric repeats, in spite of their G4 DNA forming ability are stable upon replication in normal human cells. In contrast, ciliate telomeric sequences that form more stable G4 DNA than human telomeric sequences, induce more mutations.

Stochastic telomere loss is seen in the premature aging disorder Werner Syndrome, which is caused by loss of the RecQ helicase protein WRN. We hypothesized that WRN deficiency leads to replication fork stalling and collapse due to G4 DNA formed by telomeric repeats

resulting in deletions of DNA sequence. Shuttle vectors with a telomeric or control sequence were replicated in U2OS cells deficient or proficient for WRN. Replication of shuttle vectors in normal cells did not influence shuttle vector mutant frequencies, while WRN depleted cells exhibited elevated mutant frequencies for both telomeric and control vectors but the increase was significantly higher for the telomeric vector. We demonstrated that WRN is involved in suppressing mutagenesis in shuttle vectors with telomeric sequences. We are also testing DNA synthesis in plasmids through regions of single stranded DNA containing telomere repeats in WRN proficient and deficient cells. Public health significance: Shortened telomeres are associated with age related diseases such as heart disease, cancer and premature aging disorders. These assays will help us investigate factors that cause accelerated telomere loss with the goal of preventing or delaying disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

AFM	Atomic Force Microscopy
ATM	Ataxia Telangiectasia Mutated
ATR	ATM and Rad3 related
bp	base pair(s)
D-loop	Displacement loop
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
G4 DNA	G-Quadruplex DNA
GD	Gapped Duplex
HR	Homologous Recombination
HSV-tk	Herpes Simplex Virus Thymidine Kinase
kb	Kilobase pair
NHEJ	Non-Homologous End Joining
nt	nucleotide(s)
POT1	Protection of Telomeres 1
RAP1	Transcriptional Repressor/Activator Protein 1
RPA	Replication Protein A
ssDNA	single-stranded DNA
SV	Shuttle Vector
Terc	Telomerase RNA Component
TERRA	Telomeric Repeat Containing RNA

Tert	Telomerase Reverse Transcriptase
TIN2	TRF1 Interacting Nuclear Factor 2
t-loop	Telomere Loop
TPP1	TIN2 and POT1 interacting Protein 1
TRF1/2	Telomere Repeat binding Factor 1/2
UV	Ultra Violet light
WRN	Werner helicase/exonuclease
WS	Werner Syndrome

PREFACE

A famous song in Indian classical music composed by Tyagaraja starts with the line –

“endarō mahānubhāvulu andariki vandanamulu”

(Meaning: So many great souls and salutations to them all)

I am really fortunate to have found myself in the midst of such wonderful people and would like to acknowledge some of them for their encouragement and support without which this thesis would not be possible. Firstly, I owe my deepest gratitude to my advisor Dr. Patricia Opresko for giving me the opportunity and funding to pursue my PhD under her guidance in her laboratory. Dr. Opresko is a thorough professional with an excellent work ethic and I am grateful to have matured professionally working under her. I would also like to extend my gratitude to the distinguished members of my thesis committee Dr. Michael Barmada, Dr. Susanne Gollin, Dr. Laura Niedernhofer and Dr.Robert Sobol for their guidance. The Opresko lab has been my second home for the past five years. I had the pleasure of working with a number of fine individuals in the Opresko lab and would like to acknowledge the support I received especially from Greg Sowd, Ashley Glumac, Fujun Liu (Frank), Kevin Luong, Jerry Nora, Kelly Knickelbein and Noah Buncher. I really appreciate the help from the faculty, staff, students and the administration in the Dept. of Human Genetics and at the Dept. of

Environmental and Occupational Health over the past five years. I am grateful to Dr. Nikhil Bhagwat and Dr. Sangita Suresh, who provided valuable advice throughout my PhD.

My stay in Pittsburgh since the fall of 2005 would not have been more comfortable without the support from my family and friends. I would particularly like to thank Dr. Nagarjun Konduru, Smt. Seshu Kolluri and Sri. Venkataramarao Kolluri for taking good care of me during my initial years in Pittsburgh. My friends Ajaei, Divyasheel, Samvith, Pushkar, Tarun, Puneet, Naveena and from the Lumberjacks cricket club made my stay in Pittsburgh very exciting. I am also grateful to my relatives in New Jersey Mrs. Laxmi Dasika, Mr. Adi Dasika, Mr. Prakash Kolluri, Mrs. Sangita Kolluri, Aditya, Adarsh and Siddharth for providing a home away from home and making all my holidays memorable.

I am deeply indebted to my parents Smt. Ratna Damerla and Sri. Balachandra Damerla for their undying love and sacrifices. I thank my sister Asha and brother Vivek for their constant encouragement and support. I would also like to thank my late aunt Dr. Sundari without whose help I wouldn't have thought of coming to Pittsburgh for my PhD. Finally, I would like to acknowledge my wife Prafulla for her unconditional support and encouragement throughout my PhD.

1.0 GENERAL INTRODUCTION

1.1 BRIEF HISTORY OF TELOMERE BIOLOGY

The word telomere literally comes from two Greek words “telos” which means end and “mer”, meaning part. As the name suggests, telomeres are located at the ends of chromosomes and form nucleoprotein cap like structures that are critical for cell viability and the stability of the genome. Telomeres have been studied in the scientific world even before the discovery of the double helical structure of DNA by Watson and Crick in 1953. Cytogenetic experiments in the 1930s by Hermann Muller and Barbara McClintock showed that broken ends of chromosomes resulted in chromosomal instability (De Lange et al., 2006). Advances in microscopy in the early twentieth century helped scientists to observe chromosomes and led to the idea that genes were arranged as a “string of beads” all along the chromosomes. Muller used X-rays as a mutagen on fruit flies to identify different chromosomal anomalies. He observed that the broken ends of chromosomes could rejoin and that this phenomenon was only possible for broken chromosomes but not for naturally occurring chromosomal ends, which he called “free ends” and coined the term telomere (Muller, 1938). Barbara McClintock’s experiments with maize showed that broken chromosomal ends had different properties from a normal chromosomal end. These experiments further led her to postulate the “breakage-fusion-bridge” cycle (McClintock, 1941) which explains how broken chromosomal ends fuse to form dicentric chromosomes in anaphase (anaphase bridges) and finally result in a break again after cell division thereby causing genomic instability. The Breakage-fusion-bridge cycle model is now considered a hallmark of genomic instability and a

critical step in malignancy. Discovery of the Watson and Crick's DNA structure raised further questions on the telomeres since linear chromosomes have an "end replication problem" which does not allow complete replication of DNA on the lagging strand (Figure 1-1) (Watson, 1972). Several models were put forth to explain telomere maintenance after DNA replication. These ranged from forming hairpin structures to transient fusions at chromosomal ends to explain complete replication through telomeres (De Lange et al., 2006). It was later discovered in the eighties that a reverse trascriptase enzyme called telomerase (Greider and Blackburn, 1985) is responsible for specifically adding telomeric repeats after DNA replication and is instrumental in maintaining telomere lengths. DNA sequencing and advances in cytogenetics and fluorescent microscopy techniques like fluorescent in situ hybridization (FISH) and new variants of FISH have made it possible to study telomere lengths and telomere dysfunction efficiently. The discovery of mutations in telomerase components resulting in the manifestation of several diseases have renewed interests in studying telomeres and understanding telomere roles in age related diseases and cancer (Armanios, 2009). A wealth of epidemiological studies suggest that short telomeres are associated with a number of age related diseases and cancer. Short telomere length has been found to be associated with risk and/or incidence of, cardiovascular disease, stroke, osteoporosis and obesity and risks for diabetes and certain cancers (Samani and van der Harst, 2008; Zee et al., 2010; Kim et al., 2009; Wentzensen et al., 2011). Finally Elizabeth Blackburn and Carol Greider along with Jack Szostak were awarded the Nobel Prize for physiology and medicine in 2009 for their early efforts in understanding telomeres and the discovery of the enzyme telomerase in ciliated protozoans. They elucidated that telomeres have a definitive structure for protection of chromosomal ends and that the reverse transcriptase enzyme telomerase is responsible for elongation of telomeres after DNA replication.

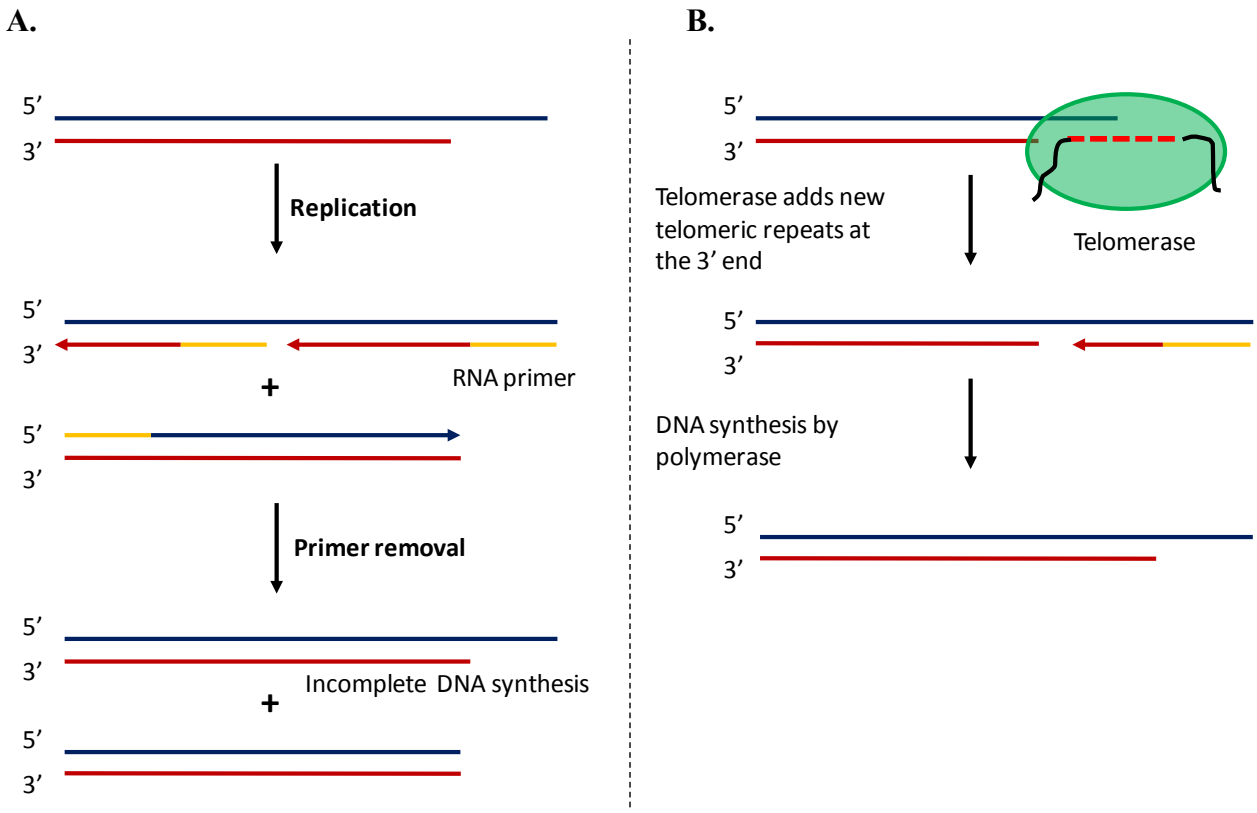


Figure 1-1. End replication problem.

The schematic on the left describes the end replication problem during replication of duplex linear DNA. A) A replication complex where the top strand is the lagging strand and the bottom strand is the leading strand. The replication machinery is not able to able to synthesize new repeats at the end due to the lack of an RNA primer, which results in the shortening of chromosomes with successive cell divisions. B) Telomerase is a reverse transcriptase enzyme that uses an integral RNA template and synthesizes new G-rich telomeric repeats at the 3’ end. A DNA polymerase synthesizes complimentary C- rich repeats leading to telomere maintenance.

1.1.1 Telomere Structure

Telomeres in mammals consist of tandem repeats of TTAGGG sequences (Moyzis et al., 1988) that range from 9 to 15 kilobase pairs (kb) (de Lange, 2005) in humans while mice have much longer telomeres ranging from 40 to 80kb (Chang, 2005). The G rich strand extends into a 3' overhang (50 to 300 nucleotides (nt)) that is formed by an unknown nuclease and the C rich strand always terminates at the 5' end in the sequence CCCAATC-5' (Sfeir et al., 2005) indicating a tightly regulated nucleolytic processing (Figure 1-2). Telomeres are associated with shelterin, a protein complex consisting of TTAGGG- repeat binding factor 1 (TRF1), TRF2, protection of telomeres 1 (POT1), transcriptional repressor/activator protein (RAP1), TRF1 interacting protein 2 (TIN2) and POT1 and TIN2 organizing protein (TPP1) (Baumann and Cech, 2001; Kim et al., 1999; Li et al., 2000; O'Connor et al., 2006). Shelterin helps in remodeling the telomeric repeats to perform a capping function. Electron microscopy studies in mouse and human cells revealed that the G rich single stranded overhang can invade homologous double stranded telomeric repeats resulting in a lasso-like structure called the telomeric loop or the t-loop (Griffith et al., 1999) (Figure 1-2). The displacement of double stranded repeats by the invading single stranded G rich sequence pairing with the C rich strand forms a displacement loop or the D-loop (Figure 1-2). This configuration of DNA is further stabilized by the interactions with the shelterin protein complex. Three of the shelterin components TRF1, TRF2 and POT1 directly recognize TTAGGG repeats. TRF1 and TRF2 bind to the duplex repeats while POT1 binds to the single stranded TTAGGG repeats present at the 3' overhang and in the D loop (Figure 1-2). TRF1 and TRF2 recruit TIN2 and RAP1 respectively while POT1 and TPP1 are bound together at the single stranded DNA (ssDNA) regions of the telomere. TIN2 interacts with TPP1 to bridge the double stranded regions with the single stranded regions in the

telomere (de Lange, 2005). Numerous non shelterin proteins also interact with telomeres making important contributions to the maintenance and protection of telomeres. However these proteins are transiently associated with telomeres unlike shelterin, which is present at the telomeres throughout the cell cycle (de Lange, 2005). A majority of these accessory proteins are involved in DNA transactions like replication, DNA damage signaling, recognition and repair (Palm and de Lange, 2008). The interactions of these accessory proteins and their roles in modulating telomere function are discussed in the following section.

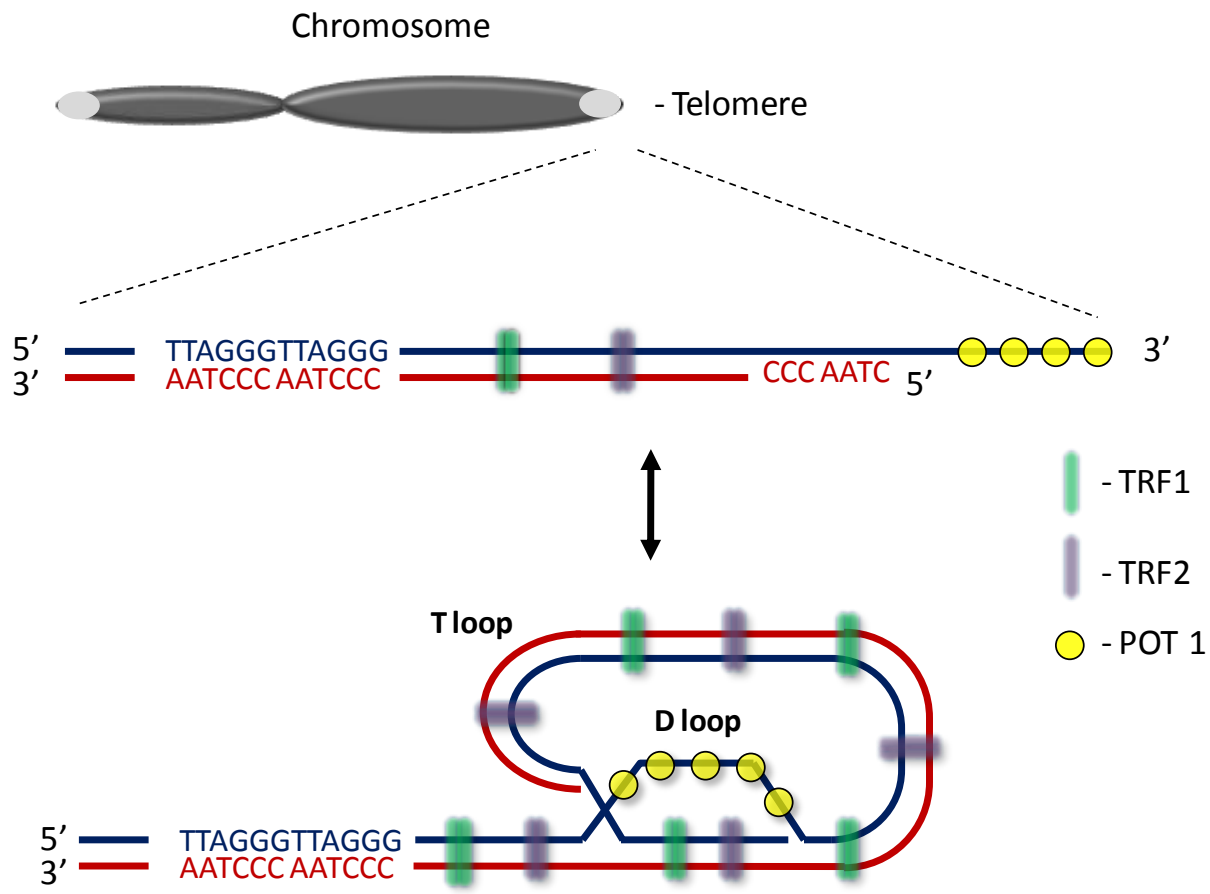


Figure 1-2. Structure of telomere.

Telomeric DNA consists of 9-15 kb duplex TTAGGG repeats in human cells with a 50-300 nt single stranded (ss) TTAGGG overhang on the 3’ strand. Shelterin proteins specifically bind to both the ds and ss portion of the telomere. TRF1 and TRF2 are dimers that specifically recognize and bind to ds telomeric DNA. POT1 binds to ss telomeric DNA. chromosome ends require processing in order to acquire a long 3′ overhang. And the 5′ end always has the sequence CCAATC-5′. Shelterin proteins have the ability to bend the telomeric DNA to form a loop like structure called the “t – loop”. The 3′ overhang is strand invades into the preceding duplex telomeric repeats, forming a D-loop.

1.1.2 Telomere Function

Telomeres have two main functions that are critical for cell survival and genomic stability. Firstly, chromosome ends being linear double stranded (ds) DNA molecules could be recognized as double strand breaks (DSBs) by the DNA damage signaling proteins and processed by the DNA repair machinery either through homologous recombination (HR) or non homologous end joining (NHEJ) pathways. Telomeres have a capping function at chromosomal ends that shield them from being recognized as DSBs (Fagagna et al., 2003). Secondly, the end replication problem of linear dsDNA molecules eventually leads to chromosomal shortening which is approximately 100 bp per population doubling (Levy et al., 1992). The coordinated actions of telomerase, a telomere-specific reverse transcriptase and the DNA replication machinery aid in function of maintaining telomere lengths. Telomerase consists of an enzymatic subunit with reverse transcriptase activity (hTERT), RNA component (hTERC) and another protein component dyskerin (DKC 1) that helps in stabilizing the RNA component (Chen and Greider, 2004). Telomerase activity is robust in embryonic development, stem cells and the germline. However the enzyme telomerase is inactive or expressed in meager quantities in adult tissues and hence telomere lengths shorten during every cell division (Bodnar et al., 1998; Blasco, 2005).

The length of telomeric DNA is an important determinant of function (Bailey and Murnane, 2006). Telomere lengths shorten for every cell division in primary cells (Harley, 1990). When telomere lengths reach a critical length, telomere dysfunction occurs which leads to chromosome instability, end-to-end fusions, and checkpoint-mediated cell cycle arrest (replicative senescence) and/or apoptosis (Celli and de Lange, 2005; Karlseder et al., 1999). Normal cells, in the presence of functional Retinoblastoma protein (pRB) and protein 53 (p53) tumor suppressor genes respond to short telomeres (as low as 13 repeats) (Capper et al., 2007) by

inducing a senescence response (Campisi, 2001; Harley et al., 1990). Cellular evidence shows that critically short telomeres elicit p53-mediated senescence or apoptosis based on the cell type (Karlseder et al., 1999). Short telomeres induce p53 dependent apoptosis in highly proliferative tissues (Hemann et al 2001a). The absence of normal p53 and pRB function leads to genomic instability characterized by chromosomal fusions followed by random breakage and fusion cycles (Deng et al., 2008). These cells could then stabilize telomeres by expressing telomerase predisposing them to neoplastic transformation (Counter et al., 1998). Some telomerase negative tumors adopt an HR based mechanism to maintain telomere lengths referred to as the alternate lengthening of telomeres (ALT) pathway (Bryan et al., 1997).

Shelterin proteins play a major role in regulating telomere lengths thereby preserving telomere function. Immunofluoroscence studies in immortalized cell lines using fluorescent tagged TRF1 revealed higher amounts of TRF1 at longer telomeres and the overexpression of TRF1 gradually shortened telomere lengths until they reached a specific length (van Steensel and de Lange, 1997). Expression of a dominant negative allele of TRF1 results in telomere elongation (van Steensel and de Lange, 1997). TRF1 is therefore presumed to influence telomere length by sequestering the action of telomerase at telomeres. So, accumulation of TRF1 at long telomeres would block the action of telomerase while short telomeres would have fewer TRF1 molecules and therefore, would have a greater chance of being elongated by telomerase. This mechanism of negative length regulation occurs at each individual telomere, and would ultimately contribute to homogenous telomere lengths among chromosomes. It has been shown that TRF2 protects critically short telomeres from chromosomal end fusions and has the ability to delay senescence (Karlseder et al., 2001). Loss of TRF2 is accompanied by activation of checkpoints which is usually dependent on Ataxia Telangiectasia Mutated protein (ATM) kinase

in mammalian cells (Takai et al., 2003). Overexpression of TRF2 in human fibroblasts results in increased rates of telomere shortening (Karlseder et al., 2002) but also delayed senescence and allowed for cells to enter senescence with much shorter telomeres when compared to control cells. Expression of a dominant negative mutant of TRF2 ($TRF2^{\Delta B \Delta M}$) causes displacement of TRF2 from the telomeres and results in loss of 3' overhangs followed by chromosomal end-to-end fusions containing telomeric DNA at the sites of fusion (van Steensel., 1998). These phenotypes occur due to the loss of the 3' overhang at telomeres which is critical for the formation of t-loops and is also a substrate for telomerase. Therefore, TRF2 plays an important role in preserving telomere function by restricting chromosomal end-to-end fusions which could lead to genomic instability, by maintaining proper telomere structure. Unprotected telomeres resemble double strand breaks which are bound by the DNA repair protein complex Ku70/80. Ku 70/80 complex binds to DNA double stranded breaks (DSB) and recruits other DNA repair proteins to facilitate end to end fusions through a nonhomologous end joining pathway (NHEJ). The t-loop structure is proposed to block the Ku70/80 complex from loading on the telomere resulting in prevention of NHEJ (Celli et al., 2006). Both TRF1 and TRF2 therefore act as negative regulators of telomere length but by different mechanisms. Telomeres are uncapped as a result of deletion of TRF2 which results in activation of the DNA damage signaling pathway that is dependent on Ataxia Telangiectasia Mutated protein (ATM) (Wu and de Lange, 2008). ATM phosphorylates the histone variant H2AX within the telomeric regions (Zhang et al., 2006). When H2AX is phosphorylated on serine 139 it is referred to as γH2AX. γH2AX is formed as a response to DNA double-strand breaks (DSB) (Burma et al., 2001). The same signaling also occurs at dysfunctional telomeres providing further evidence that dysfunctional telomeres are recognized as DSBs. This phenomenon has been used widely to study telomere dysfunction and

the localization of γH2AX at telomeres form distinct foci in interphase nuclei that are referred to as telomere dysfunction induced foci (TIFs) (Takai et al., 2003). TIFs are not formed upon inactivation of ATM in the absence of TRF2 which proves that TRF2 is required for shielding chromosomal ends from being recognized as DSBs by the ATM-dependent DNA damage signaling pathways (Denchi and de Lange, 2007). Proteins that participate in DSB repair pathways also interact with shelterin proteins (Celli et al., 2006) further strengthening the result that chromosomal ends need a capping function to escape from being recognized as double strand breaks. Chromosome oriented FISH or Co-FISH, an assay which enables one to differentiate the replication products from lagging strand and leading strand DNA synthesis, revealed that TRF2 deletion could result in telomere sister chromatid exchanges (T-SCEs) only in the absence of Ku70 complex (Celli et al., 2006). This suggests a strong interplay between shelterin proteins and the DNA repair machinery to protect telomere structure, thereby maintaining telomere length and function. POT1 is a telomere specific ssDNA binding protein that plays an important role in maintaining telomere structure and function. *In vitro* experiments with fission yeast and human POT1 proteins showed that POT1 also inhibits extension of a telomeric primer by telomerase and further supported a capping function (Baumann and Price, 2010). Moreover, POT1 depletion in human cells results in telomere elongation (Veldman et al., 2004). POT1 specifically binds to ssDNA in the telomeric 3’ overhang (Veldman et al., 2004) and is responsible for the maintenance of the CCAATC-5’ sequence which becomes randomized upon POT1 deletion (Hockemeyer et al., 2005). POT1 depletion results in activating the DNA damage signaling pathway that is dependent on ataxia telangiectasia and Rad3-related protein (ATR), which is responsible for signaling a DNA damage checkpoint in response to single strand breaks and stalled replication forks (Denchi and de Lange, 2007).

Since telomeric repeats are G-rich sequences they have the potential to fold into secondary DNA structures called G-quadruplex structures (G4 DNA) (Maizels, 2006). Numerous *in vitro* studies have shown that the G-rich telomeric repeats from various species adopt such conformations. Though one study in ciliates allowed for the visualization of G4 DNA *in vivo* (Schaffitzel et al., 2001), not much is known about how often telomeres adopt these structures and how they might modulate telomere function. More about G4-DNA and the implications in telomere maintenance are discussed in section 1.4.

Telomeres were considered to be transcriptionally silent until a recent finding demonstrated that telomere DNA is transcribed to give rise to non coding RNA called telomeric repeat-containing RNA or TERRA in mammalian cells (Azzalin et al., 2007). The exact functions of TERRA are being investigated but TERRA is thought to play an important role in regulating telomere length.

1.2 TELOMERE REPLICATION

As mentioned previously the end replication problem results in the loss of telomeric sequences from every round of cell division. There is no known origin of replication within the telomeres (Verdun and Karlseder, 2007). The closest origin is in the subtelomeric region and therefore telomeres are replicated in only one direction. Since telomeric DNA is always replicated in the same direction, the G-rich strand always acts as the template for discontinuous lagging strand synthesis and the C-rich strand acts as the template for the continuous leading strand synthesis. Telomeric chromatin, due to its repetitive nature has the ability to form various unusual structures that cause potential problems to replication fork progression. Alternate DNA structures

such as G4 DNA, t-loops and heterochromatin-like structures have been detected *in vivo* that could result in stalled replication forks (Gilson and Geli, 2007). Stalled replication forks within telomeric repeats could collapse to give rise to DSBs or form folded back regressed structures that are prone to endonuclease action and aberrant recombination (Gilson and Geli, 2007). Efficient telomere replication involves interplay between shelterin complex proteins, replication machinery and a number of accessory proteins that tightly regulate telomeric DNA replication. It has been shown that Taz1, a homologue of TRF1 and TRF2 in *S.pombe*, is essential for telomere replication (Miller et al., 2006). Inhibition of Taz1 causes stalled replication forks both at telomeres as well as at interstitial telomeric repeats suggesting that this phenotype is sequence specific rather than position or structure specific (Miller et al., 2006). In cells expressing telomerase, the replication machinery and telomerase act in a coordinated fashion and POT1 plays an important role in coordinating these processes (Smogorzewska and de Lange, 2004). Studies in *S. cerevisiae* showed that telomerase cannot extend the 3' single strand overhang if the replicative DNA polymerase delta and primase polymerase alpha are inactive, strengthening the idea that telomerase and the replication machinery act in a coordinated manner (Diede and Gottschling, 1999). Okazaki fragment processing during lagging strand replication allows for transient ssDNA regions to accumulate at the telomeres that provide opportunities for the G-rich repeats to fold into G4 DNA (Duquette et al., 2004). Though G4 DNA has been shown to block DNA synthesis *in vitro*, (Zahler et al., 1991) there is no *in vivo* evidence to show that G4 DNA structures play a role in replication fork stalling at telomeres. Therefore understanding mechanisms of telomere replication could unlock several clues pertaining to telomere maintenance and function.

1.3 G-QUADRUPLEX DNA

The G rich strand of the telomeres possesses the ability to form secondary DNA structures called G4 DNA. Each G4 DNA structure is comprised of planar arrays of four guanines stabilized by Hoogsteen bonds known as quartets (Figure 1-3). Quartets are stacked on top of each other to form a quadruplex (Figure 1-3) (Maizels, 2006). The number of quartets in each quadruplex varies based on the number of guanines in each guanine rich repeat. For example quadruplexes formed from human telomeric TTAGGG repeats have 3 quartets in each quadruplex whereas ciliates that have TTGGGG or TTTTGGGG as their telomeric repeats possess four quartets in each quadruplex. This renders greater stability to the quadruplex when compared to human quadruplex DNA (Lee et al., 2008). Telomeric DNA has been shown to form G4 DNA spontaneously both *in vivo* and *in vitro* (Schaffitzel et al., 2001; Luu et al., 2006). The possibility of G4 DNA formation in the telomeres occurs either in the 3' overhang, displaced DNA in the D-loop, or in the G-rich strand that is replicated by the lagging strand DNA replication machinery. Cells deficient in WRN, POT1 or flap endonuclease 1 (FEN1) showed loss of telomeres replicated from the G-rich lagging strand. The loss of G-rich sequences is attributed to the absence of these proteins that prevent formation of secondary DNA structures like quadruplexes formed during telomere replication (Crabbe et al., 2004; He et al., 2006; Saharia et al., 2008). G4 DNA folded telomeric DNA is not accessible to the action of the enzyme telomerase *in vitro* (Zahler et al., 1991). It is possible that G4 DNA formation could have a role in regulating telomere length. G4 DNA structures have also been shown to block replicative DNA polymerases *in vitro* (Kamath-Loeb et al., 2001). It is possible that telomeric repeats can form alternate DNA structures that block replication and ultimately lead to loss of the telomeric DNA. Though G4 DNA structures have been shown to form at telomeres, their roles in modulating

telomere replication and telomere lengths are yet to be established. Recent evidence showed that transcripts from human telomeres, TERRA RNA, forms parallel G4 DNA structures localized to telomeres in cells, providing more *in vivo* evidence for G4 DNA formation (Xu et al., 2010).

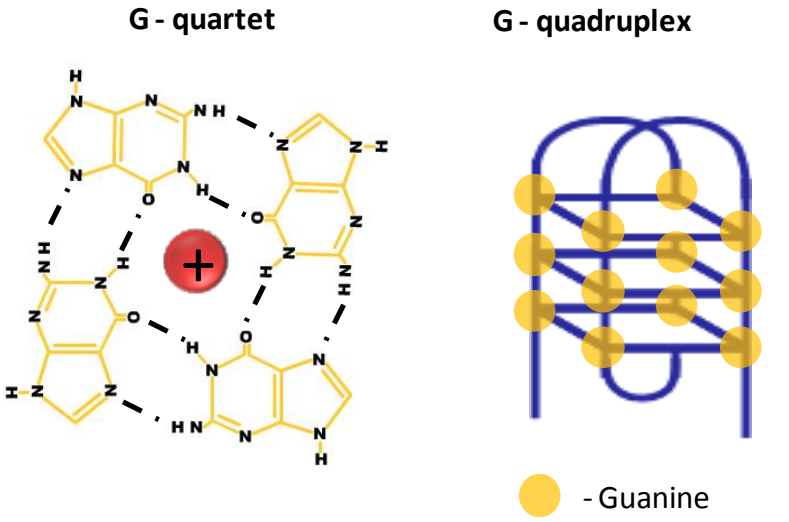


Figure Error! No text of specified style in document.**-1. G4 DNA structure.**

G4 DNA structures, represented on the right are formed by planar arrangements of G-quartets. Hoogsteen base pairing enables 4 guanine molecules, each coming from different G-rich repeats of telomeric DNA to form square planar structures called G-quartets which are stabilized by metal ions.

1.1 TELOMERES AND HUMAN DISEASE

A vast spectrum of genetic disorders with genomic instability, premature aging and cancer predisposition are associated with defects in telomere maintenance (Blasco, 2005). Defects in telomerase and its components are responsible for some cases of dyskeratosis congenita, aplastic anemia, Hoyeraal–Hreidarsson syndrome, idiopathic pulmonary fibrosis and liver disease

(Armanios et al., 2007; Dokal, 2000; Rocha et al., 1998; Vulliamy et al., 2001; Yamaguchi et al., 2005). These disorders exhibit accelerated telomere shortening. A number of diseases that are caused by defects in proteins that play accessory roles in preserving telomere function have been found. RecQ helicases are a conserved family of DNA helicases that have been shown by numerous studies to perform important functions in telomere replication. Bloom syndrome (BS) and Werner syndrome (WS) are caused by loss of RecQ helicase proteins BLM and WRN respectively that are implicated in telomere replication and repair (Opresko et al., 2004). Defects in DNA repair signaling proteins cause premature aging disorders with telomere dysfunction such as ataxia telangiectasia (AT) (Metcalfe et al., 1996), seckel syndrome (Pennarun et al., 2010) and Ataxia telangiectasia like disorder (ATLD) (D'Amours and Jackson, 2002). Premature onset of age related symptoms and predisposition to cancer are hallmarks of these disorders (Blasco, 2005).

Somatic cells in culture can only replicate for a limited number of generations. This phenomenon was discovered in 1961 by Hayflick and Moorehead (Hayflick and Moorhead, 1961). The rates of cell divisions also gradually decrease and the cells ultimately enter replicative senescence. Bodnar et al showed that this phenomenon was due to telomere shortening and could be rescued by exogenous expression of telomerase (Bodnar et al, 1998). In most somatic cells telomere lengths shorten with each cell division due to down regulation or absence of telomerase (Blasco, 2005). This ultimately results in telomere lengths reaching a critical length that triggers activation of a p53 or the retinoblastoma protein pRB dependent tumor suppressor response that induces replicative senescence (Karlseder et al., 2002). Cells with defective p53 and pRB pathways can avoid this barrier and eventually these cells undergo chromosome end to end fusions promoting genomic instability and apoptosis leading to a crisis

state when the majority of the cell die (Karlseder et al., 1999). Activation of telomerase or the ALT pathway can rescue some of these cells and allow further cell divisions which are not regulated, ultimately resulting in carcinogenesis. Therefore replicative senescence leads to age related pathologies that are a common feature of premature aging disorders accompanied by accelerated telomere loss. Bypass of replicative senescence and the crisis stage due to defects in checkpoint proteins results in carcinogenesis (Kim et al., 1994; Chin et al., 1999).

1.5 TELOMERE DYSFUNCTION AND WERNER SYNDROME

1.5.1 Werner Syndrome

Werner Syndrome (WS) was first described by Dr. Otto Werner in a sibship that presented with symptoms of aging that included cataracts and scleroderma like skin (Epstein et al., 1966). It is an autosomal recessive adult onset segmental progeroid syndrome caused by mutations in the RecQ helicase gene WRN (Yu et al., 1996). Except for a patient with two missense mutations, all the other mutations identified so far in the WRN gene have been due to premature termination of the protein that fails to localize to the nucleus (Huang et al., 2006). The premature aging symptoms usually manifest in the third decade of life. Initial symptoms include lack of a growth spurt that occurs at puberty, short stature, skin atrophy, loss and graying of hair and osteoporosis. This is followed by development of symptoms of age-related disorders that include cataracts (preva0lence of 100%), osteoporosis (91%), hypogonadism (80%), diabetes mellitus (71%). Median age of death is around 54 years of age with the common cause being malignancy and myocardial infarction (Huang et al., 2006). Though WS patients show a wide range of age

related disorders, the incidence of Alzheimer's disease is comparable to the general population (Huang et al., 2006). WS syndrome patients present with a high incidence of sarcomas of mesenchymal origin. The Japanese population has a high prevalence of the disease. Out of the 1200 patients reported from 1904 to 1996, 845 of them were from Japan with an estimated 23 new cases every year. This frequency is attributed to a founder mutation in the Japanese population (Yu et al., 1996). One of the many differences between WS and other premature aging disorders such as Hutchinson-Gilford progeria syndrome (HGPS) and Rothmund Thomson syndrome (RTS) is the late onset of disease symptoms seen in WS (Puzianowska-Kuznicka and Kuznicki, 2005). Since patients present with age related phenotypes early in life, WS is considered as a good model for studying normal aging and age-related disorders.

1.5.2 Werner Syndrome Helicase Exonuclease - WRN

Positional cloning helped scientist to map the *WRN* gene on the short arm of chromosome 8 (8p11.1 – 21.1) (Yu et al., 1996) and spans more than 250 kb which includes 35 exons out of which 34 are coding exons (Goddard et al., 1996). WRN is a DNA dependent ATPase with 3' to 5' helicase, 3' to 5' exonuclease activity and ssDNA annealing activity (Gray et al., 1997; Huang et al., 1998; Machwe et al., 2005). WRN is a RecQ helicase that belongs to the sf2 super family of helicases (Opresko et al., 2003). The RecQ family of helicases is a highly conserved group of proteins with important roles in various DNA metabolism pathways including DNA replication, repair, recombination and DNA damage signaling. Patients with RecQ helicase disorders are predisposed to cancer and premature aging (Bohr, 2008). Numerous cellular and biochemical studies also reveal that deficiencies in RecQ helicases cause genomic instability which is an important step in malignancy. Bacteria and yeast have only one RecQ homolog while humans

have 5 RecQ helicases WRN, BLM, RECQ4, RECQ1 and RECQ5 and mutations in 3 of them are known to cause rare genetic disorders involving genomic instability (Figure 1-4). Werner syndrome (WS) is associated with defects in WRN, Bloom syndrome (BS) is associated with defects in BLM and Rothmund Thomson syndrome (RTS) is associated with defects in RECQ4 (Mohaghegh and Hickson, 2002). Premature aging phenotypes are a hallmark of WS and RTS but less pronounced in BS. Not much research has been done on RECQ4 but WRN and BLM share many common cellular interactions, some of which are discussed in the following section. However WRN differs from the rest of the RecQ helicase proteins due to the presence of the additional 3' to 5' exonuclease activity that is absent in other RecQ helicase members in humans (Figure 1-4) (Huang et al., 1998). WRN is also ubiquitously expressed while BLM is found more predominantly in tissues that have a high turnover rate (Chang, 2005).

Figure 1-4. RecQ helicase family and associated diseases.

Human RecQ helicase family consists 5 members. All five RecQ proteins possess a conserved RecQ helicase domain. Mutations in the human WRN, BLM and RECQ4 genes lead to the genetic instability/ cancer predisposition syndromes Werner syndrome, Bloom syndrome and Rothmund-Thomson syndrome respectively. WRN is the only human RecQ helicase that also has exonuclease activity in addition to helicase activity.

1.5.3 Mouse model of WS

It is surprising to note that Wrn null mice do not exhibit signs of premature aging similar to WS patients (Lombard et al., 2000). Murine telomeres range from 40 to 80 kb compared to 9 -15 kb in humans (Chang, 2005). Moreover human and mouse tissues show discrepancies in telomerase expression. Telomerase activity is present in mouse somatic cells while it is severely reduced in human somatic cells (Chang, 2005). Therefore the likelihood of telomere based replicative senescence due to WRN depletion may not be possible in Wrn null mice and hence the absence of age related pathologies. When the telomerase deficient mouse *mTerc*$^{-/-}$ was crossed into a *Wrn*$^{-/-}$ background, late generation *mTerc*$^{-/-}$ *Wrn*$^{-/-}$ mice with shortened telomeres presented with symptoms associated with WS patients that included wound healing defects, osteoporosis, cataracts, hypogonadism, type II diabetes and premature death (Chang et al., 2004). These symptoms were not observed in *mTerc*$^{-/-}$*Wrn*$^{+/+}$ age matched controls suggesting that telomere dysfunction and shortened telomeres are required to see the effects of WRN depletion.

1.5.4 Biological and Cellular Functions of WRN

A wealth of research has shown that WRN plays an important role in DNA replication. Firstly, WS cells show extended S-phase and undergo premature replicative senescence (Poot et al., 1999). Cells with WRN deficiency or mutations in WRN show hypersensitivity to exogenous agents that block DNA replication revealing that WRN has roles in progression of replication forks. WS lymphoblasts are sensitive to the topoisomerase I inhibitor campothecin which induces apoptosis in S-phase (Poot et al., 1999). WS cells are also hypersensitive to 4-Nitroquinoline 1-oxide (4NQO) which produces bulky base damage like ultra violet radiation

(UV) induced cyclobutane pyrimidine dimers (CPDs) (Ogburn et al., 1997). WRN interacts with Replication protein A (RPA), an important single stranded DNA binding protein that is localized to replication forks during DNA replication. *In vitro* studies show that WRN unwinds short forks (25 bp) but in the presence RPA, it can unwind up to 1000 bp (Brosh et al., 1999). WRN also co localizes with RPA upon treatment with Hydroxyurea (HU) which is known to cause replication fork arrest by reducing the purine dNTP pools (Constantinou et al., 2000). WRN is actively involved in resolving different DNA substrates that are intermediates in replication and repair processes such as forks, displacement loops (D-loops), Holliday junctions, regressed forks (chicken foot structures) and G4 DNAes (G4) (Rossi et al., 2010). WRN interacts physically with a number of replication proteins including RPA, Proliferating Cell Nuclear Antigen (PCNA) and DNA polymerase δ (Kamath-Loeb et al., 2000). WRN's interaction with polymerase δ helps in enhancing the latter's DNA synthesis property even through G-rich sequences that form G4 DNA structures (Kamath-Loeb et al., 2001). WRN also interacts with another replication enzyme (FEN1), which functions in resolving replication intermediates particularly at Okazaki fragments on lagging strand synthesis (Brosh et al., 2001). These studies reveal that WRN plays a significant role in DNA replication and may participate in resolving alternate DNA structures that block the progression of replication forks.

In addition to its role in replication, WRN has different roles in transcription, DNA damage signaling, DNA repair pathways and telomere maintenance (Rossi et al., 2010). WRN is localized in the nucleolus and relocates to the nucleoplasm upon DNA damage (Gray et al., 1998). A number of studies show that WRN is involved in base excision repair (BER). Firstly, WRN deficient cells are hypersensitive to alkylating agents, methyl methanesulfonate (MMS) and temozolomide (Sidorova et al., 2008). Oxidative stress also produces more oxidative damage

as a result of less repair in WRN deficient cells compared to control cells (Kusumoto et al., 2008). WRN also interacts physically and functionally with a number of BER proteins including pol β (Harrigan et al., 2006), Nei endonuclease VIII-like 1 (NEIL1) and poly ADP-ribose polymerase 1(PARP1) (Rossi et al., 2010).

Numerous studies have reported WRN's roles in DSB repair pathways. WRN interacts with several proteins that are responsible for carrying out either NHEJ or HR which are the two main pathways to repair DSBs. NHEJ is mediated by DNA-dependent protein kinase catalytic subunit (DNA-PKcs) and Ku70/80 heterodimer. These proteins together form the DNA-PK complex which plays critical roles in the repair of DNA double-strand breaks. WRN interacts with both DNA-PKcs and Ku. Ku recruits WRN to broken ends of DNA and also stimulates its exonuclease activity (Li and Comai, 2001). However, WS cells are only slightly sensitive to ionizing radiation and therefore WRN may not play a significant role in NHEJ (Oshima et al., 2002). WRN is also involved in HR by its interaction with the Mre11-Rad50-NBS1 (MRN) complex through NBS1. The MRN complex stimulates the DNA unwinding activity of WRN (Cheng et al., 2004). WRN also interacts with breast cancer type 1 susceptibility protein, BRCA1 which is an important protein aiding in the removal of interstrand DNA cross links in the HR pathway for repairing DSBs. BRCA1 stimulates WRN's helicase activity (Cheng et al., 2006). These interactions suggest that WRN may be involved in resolving recombination intermediates which is important for successful HR and repair of DSBs.

1.5.5 Role for WRN at telomeres

WRN plays a significant role in telomere maintenance. WS fibroblasts enter premature senescence and this phenomenon can be rescued by exogenous expression of telomerase suggesting telomere dysfunction causes premature senescence in WS cells (Wyllie et al., 2000). WRN interacts with shelterin proteins POT1, TRF1 and TRF2 (Opresko et al., 2005; Opresko et al., 2004; Opresko et al., 2002). Colocalization studies in ALT cells with fluorescent-tagged proteins TRF1 (marker for telomeric DNA) and WRN showed that they localize with RPA and Rad51 foci (represent sites of DNA repair through HR) (Yeager et al., 1999; Sakamoto et al., 2001). This suggests that WRN is present at telomeres during telomeric DNA repair and replication (Opresko et al., 2003). *In vitro* biochemical experiments further reveal that WRN can unwind telomeric D-loops which have to be dissociated to allow telomere replication and that TRF1 and TRF2 regulate this process (Opresko et al., 2004). POT1 has also been found to enhance the helicase activity of WRN in resolving telomeric D-loops (Opresko et al., 2005). This suggests that telomeric proteins may function in recruiting WRN to resolve alternate DNA structures formed by telomeric repeats.

Crabbe et al showed that HeLa cells expressing a dominant negative helicase dead WRN exhibits an increase in sister telomere loss (STL; loss of the telomeric signal from one of two sister chromatids) which represents telomere dysfunction during replication (Crabbe et al., 2004). A similar phenotype but with more severity was observed in HeLa cells treated with a telomerase inhibitor and expressing the dominant negative helicase dead WRN. Experiments using the chromosome oriented FISH (CO-FISH) technique allowed them to differentiate between lagging and leading strand DNA synthesis and they found that cells in which WRN was inhibited preferentially showed problems in lagging strand synthesis (Crabbe et al., 2004). These data

indicate that WRN is required for efficient replication of the G-rich strand of telomeres, which as previously discussed is fraught with potential problems during DNA replication leading to telomere loss. WRN deficient cells that are defective in check point proteins p53 and p16 further revealed that telomere dysfunction leads to genomic instability in these cells (Crabbe et al., 2007). These phenotypes, which include chromosome fusions, lacking telomere sequences at the fusion sites could be rescued by complimenting these cells with WRN or telomerase. These data indicated that WRN defects lead to telomere loss and genomic instability. Moreover, as discussed in the previous section 1.6.3, late generation *Wrn* and *Tert* double knockout mice show phenotypes similar to WS patients indicating that short telomeres are a prerequisite for the onset of premature aging symptoms. These studies together reveal that WRN has important roles at telomeres especially in telomere replication and repair.

1.6 STATEMENT OF PROBLEM

Telomeric DNA is made up of tandem repeats of G-rich sequences that are bound by shelterin proteins in a sequence specific manner. Although the composition of telomeric sequences varies among different species, the G-rich nature and the association with telomeric proteins and other repair proteins to protect chromosomal integrity remains conserved. DNA can adopt more than 10 different types of non-B DNA structures which usually form in tracts of repeat sequences (Zhao et al., 2010). Mutagenesis studies on non-B DNA forming sequences such as Z-DNA (left handed double helix) and H-DNA (triplex) forming sequences show elevated mutagenicity and a significant increase in deletions in mutation spectra of Z DNA forming sequences (Wang et al., 2006; Wang and Vasquez, 2004). Studies involving replication micro satellite repeats on shuttle

vectors have shown that the mutagenic potential is sequence specific (Hile et al., 2000). Telomeric sequences can also form alternate G4 DNA structures that can block the action of telomerase and DNA polymerases during replication (Figure 1-5) (Kamath-Loeb et al., 2001) which could eventually lead to telomere loss. The first goal was to determine if telomeric sequences possess high mutagenic potential due to their ability to form G4 DNA structures that can block replication fork progression and lead to deletions.

A number of proteins play important roles in telomere replication and repair and in maintaining telomere homeostasis. WRN, which belongs to the RecQ helicase family of proteins, is one such protein that has been shown to have important roles in telomere maintenance. An increase in stochastic loss is seen in WS cells which can be prevented by introducing either WRN or telomerase. The loss of telomeric repeats in WS cells almost always occurs from the G-rich strand during replication and telomere dysfunction is required to manifest WS phenotypes. This led to the models that WRN may have roles in resolving alternate DNA structures such as G4 DNA at telomeres to facilitate DNA replication. My next goal was to determine if WRN facilitates replication of telomeric repeats by preventing spontaneously induced mutations and deletions.

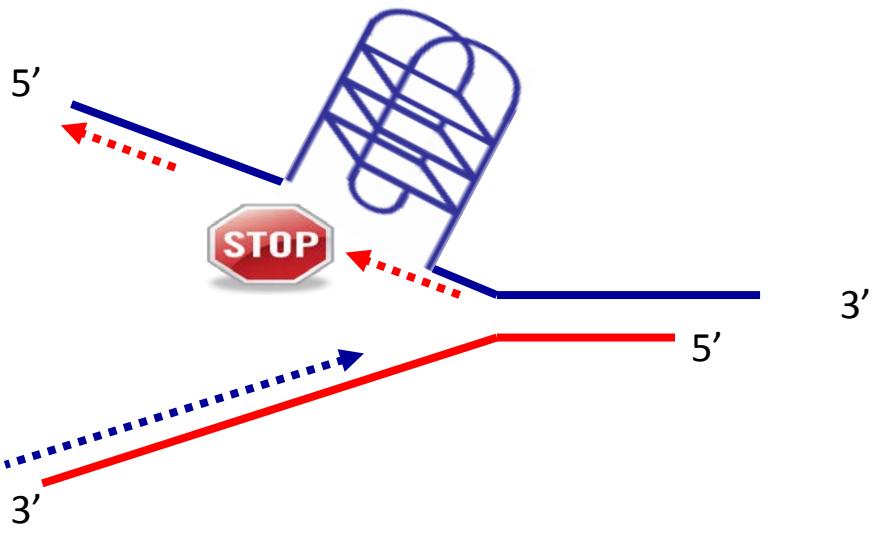


Figure 1-5. Model of G4 DNA blocking replication fork progression.

Transient ssDNA regions during discontinuous lagging strand DNA synthesis of G-rich telomeric sequences (solid blue lines) can fold into G4 DNA structures that block the progression of replication fork.

1.7 PUBLIC HEALTH SIGNIFICANCE

According the National Institute of Aging (NIA) the earliest of the baby boomers in the USA will be in their late 60s by the end of 2011. Shortened telomeres are associated with age related disorders such as heart disease, atherosclerosis, cancer and premature ageing disorders (Blasco, 2005). We are investigating factors that lead to accelerated telomere loss with the goal of preventing or delaying disease. Studying telomere replication and factors affecting this process could provide important insights into understanding the mechanisms of telomere loss that is a hallmark of age related disease and cancer. Premature disorders like WS are useful models to study aging and cancer since they present with age related symptoms at an earlier stage in life. Understanding these mechanisms of telomere loss and the factors affecting this process could

open new avenues to design innovative therapeutic targets to preserve telomeres that would ultimately lead to preventing cancer or delaying the onset of age related pathologies.

2.0 TELOMERE REPEAT MUTAGENECITY IN HUMAN SOMATIC CELLS IS MODULATED BY REPEAT ORIENTATION AND G4 DNA STABILITY

This chapter was published as a research article in DNA Repair:
Rama Rao Damerla, Kelly E. Knickelbein, Devin Kepchia, Abbe Jackson, Bruce A. Armitage, Kristin A. Eckert, Patricia L. Opresko, Telomeric repeat mutagenicity in human somatic cells is modulated by repeat orientation and G-quadruplex stability, DNA Repair, Volume 9, Issue 11, 10 November 2010, Pages 1119-1129, ISSN 1568-7864, DOI: 10.1016/j.dnarep.2010.07.014.

2.1 INTRODUCTION

Telomeres are nucleoprotein structures at chromosome ends that critically impact lifespan and health, as well as cell viability and genome stability (Armanios, 2009; Blasco, 2005; Palm and de Lange, 2008). Progress in recent years indicates that the inability to completely replicate chromosome ends is not the only source of telomere attrition, and that inappropriate processing by DNA repair enzymes or failures in telomere replication can cause rapid telomere loss (reviewed in (de Lange, 2005)). Telomeres consist of an array of repeat sequences that interact with specific proteins to prevent the chromosome ends from being recognized as double strand breaks (d'Adda di Fagagna et al., 2003; Takai et al., 2003). Mammalian telomeres comprise of

TTAGGG repeats, and human telomere lengths vary from 5 -15 kb and terminate in a 3' ssDNA tail that is 50 -500 nt long (Riethman, 2008). The 3' tails can invade preceding telomeric repeats to form a lariat like t-loop/D-loop structure that is further stabilized by the shelterin protein complex (Griffith et al., 1999; Stansel et al., 2001). Shelterin proteins TRF2 and TRF1 bind duplex telomeric DNA and POT1 binds to single strand TTAGGG repeats (Broccoli et al., 1997; Lei et al., 2004), and together they recruit the remaining shelterin proteins TIN2, RAP1, and TPP1(de Lange, 2005). How these proteins influence the fundamental processes of DNA repair and replication in telomeric repeats has yet to be fully realized.

Cellular evidence indicates that telomeres are fraught with potential obstacles to DNA replication and require specific proteins to prevent stalling. In *S. cerevisiae* DNA replication fork stalling is greatly increased at telomeres in the absence of the Rrm3p helicase (Ivessa and Zakian, 2002). In S. *pombe* and humans the telomeric proteins Taz1 and TRF1, respectively, are required to prevent replication fork stalling at telomeres (Miller et al., 2006; Sfeir et al., 2009). The precise mechanism is not known, but some evidence suggests that TRF1 recruits helicases BLM and RTEL to dissociate alternate DNA structures (Crabbe et al., 2004). The consequences of fork stalling in the telomeres can be loss of telomeric DNA or aberrant telomere structures including doublets that resemble broken telomeres (Crabbe et al., 2004; Martinez et al., 2009; Sfeir et al., 2009). Telomere doublets are induced by aphidicolin treatment which stalls replication forks and induces breaks at fragile sites (Sfeir et al., 2009). The mechanistic models of mutagenesis in repetitive sequences involve stalling and/or dissociation of the DNA replication fork due to road blocks (Hile and Eckert, 2004). Studies in yeast and bacteria demonstrate that sites of stalled replication forks are susceptible to chromosomal breakage (Cha

and Kleckner, 2002; Courcelle et al., 2003; Ivessa and Zakian, 2002). Thus, replication-mediated breaks in telomeres may represent an important source of telomeric loss.

Possible sources of replication fork stalling at telomeres include oxidative DNA damage which preferentially occurs at G runs (Petersen et al., 1998), or alternate DNA structures including the t-loop/D-loop or G-quadruplex (G4) DNA which can form in ssDNA with tandem guanines. Telomeric DNA forms G4 structures spontaneously *in vitro* and *in vivo* (Duquette et al., 2004; Luu et al., 2006; Paeschke et al., 2008; Paeschke et al., 2005; Yang et al., 2009; Zahler et al., 1991), that block DNA polymerase progression *in vitro* (Kamath-Loeb et al., 2001). G4 structures consist of planar arrays of quartets, and each quartet is formed by four guanines interacting through Hoogsteen base pairing (Maizels, 2006) (Figure 2-3A). The number of quartets in a quadruplex influences the stability of the structure and depends on the number of guanine residues (Lee et al., 2008). The potential for G4 formation in the telomeres exists either in the 3'overhang, displaced DNA in the D-loop, or in the G-rich sequences present on the lagging strand. Okazaki fragment processing during lagging strand DNA synthesis is expected to produce transient regions of ssDNA, and G4 DNA folds in ssDNA regions (Duquette et al., 2004; Neaves et al., 2009). Cells deficient in the Werner syndrome protein (WRN), POT1 or FEN1 exhibit preferential loss of telomeres replicated from the G-rich lagging strand (Crabbe et al., 2004; Saharia et al., 2008; Wu et al., 2006), suggesting these proteins may function in preventing and/or dissociating G4 structures. Furthermore, an agent that stabilizes G4 DNA induces defects in telomere replication and causes telomeric aberrations (Rizzo et al., 2009). Whether G4 structures can interfere with telomere replication in normal cells has yet to be established.

Previous work indicates that sequences with the ability to form various alternate structures exhibit increased mutagenic potential (reviewed in (Wang and Vasquez, 2006)). In these studies shuttle vectors with mutation reporter genes have been invaluable. The insertion of sequences with the potential to form H-DNA and Z-DNA adjacent to a reporter gene induced breaks and large deletions in the shuttle vector after transfection into normal mammalian cells (Wang et al., 2006a; Wang and Vasquez, 2004). The impact of G4 DNA on shuttle vector stability is unknown, but studies in yeast and worms suggest that G4 structures can be mutagenic. Loss of DOG-1 helicase in *C. elegans* leads to deletions in genes containing G-runs (Cheung et al., 2002), and loss of Pif1 helicase in *S. cerevisiea* promotes instability in an artificial human G-rich minisatellite in the yeast genome (Ribeyre et al., 2009). However, the fidelity of telomeric repeat replication and the impact of G4 potential on the mutagenicity of telomeric repeats in human cells are largely unexamined.

Studies of ciliated protozoa provide evidence for G4 formation at telomeres and G4 resolution during replication. Ciliates contain a macronucleus consisting of up to 10^8 small DNA molecules that are terminated by telomeres consisting of about 20 bp of duplex DNA and a 16 nucleotide 3' G-rich ssDNA tail (reviewed in (Lipps and Rhodes, 2009)). This high concentration of telomeres allowed for the detection of G4 DNA by immuno-staining with antibodies raised against G4 structures (Schaffitzel et al., 2001). DNA replication occurs exclusively in a distinct replication band (Prescott, 1994) in which G4 DNA is not detected (Schaffitzel et al., 2001). G4 formation is regulated by telomere binding proteins TEBP-α and TEBP-β (Paeschke et al., 2008; Paeschke et al., 2005). These studies suggest that G4 DNA is resolved during telomere replication in ciliates. In this study our goal was to test the mutagenic potential of telomeric repeat sequences and their ability to induce breaks and deletions upon

replication in normal human cells, using a well established shuttle vector mutagenesis assay. We hypothesized that the mutagenicity of telomeric repeats correlates with G4 forming potential and thermal stability. To test this we examined various telomeric repeats that differ in G-quartet numbers and compared repeats with the G-rich sequence on the lagging strand versus the leading strand. We show that the ciliate repeats from *T. thermophila* (TTGGGG) and *O. nova* (TTTTGGGG) form more stable G4 DNA than human repeats (TTAGGG) *in vitro*. We demonstrate that while all of the vectors with various telomeric repeats exhibited low mutant rates after replication in human cells, the orientation of the human telomeric repeats (G-rich lagging versus leading strand) and the stability of the potential G4 structures significantly affected the vector mutant rates. We also observed an increase in mutagenic events in the ciliate telomeric repeats compared to the human repeats. However, in contrast to H-DNA and Z-DNA forming sequences, our data indicate that normal human cells possess the ability to effectively manage G4 forming sequences, particularly human telomeric repeats, during replication.

2.2 MATERIALS AND METHODS

2.2.1 Reagents

Oligonucleotides containing telomeric repeat sequences and primers used in sequencing reactions were ordered from Integrated DNA technologies Inc. (Coralville, IA) (Table 2-1). Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). 5-fluoro-2'-deoxyuridine (FUdR) and chloramphenicol (chlor) were purchased from Sigma Chemical Co. (St.Louis, MO). Hygromycin and gentamycin were purchased from EMD Chemicals Inc

(Gibbstown, NJ) and Fisher BioReagents respectively. Proteinase K and cell culture reagents RPMI-1640 and FBS were purchased from Invitrogen Corporation (Carlsbad, CA).

2.2.2 UV melting curves

All DNAs used in the UV melting curve experiments were purified by gel filtration chromatography, except for 5-GT-$(TTAGGG)_{10}$-TC-3', which was purified by denaturing polyacrylamide gel electrophoresis. DNA stock solutions were prepared in pure water and concentrations were determined by UV absorbance at 260 nm at 85°C on a Varian Cary 3 Bio spectrophotometer. At high temperature the bases are presumably unstacked, and the extinction coefficient can be calculated as the sum of the individual bases. The DNA base extinction coefficients were obtained from the literature (Dawson, 1986). Solutions containing 2.5 μM DNA in 10 mM Tris-HCl (pH 7), 100 mM KCl and 1.0 mM EDTA were prepared in 1 cm pathlength quartz cuvettes. Samples were placed in a Varian Cary 3 Bio spectrophotometer equipped with a thermoelectrically controlled multicell holder. The solutions were heated to 90°C and equilibrated for 5 min. Then a cooling gradient was applied at a rate of 1°C/min down to 25°C, when samples were equilibrated for 5 min before starting a heating gradient at the same rate up to 90°C. The absorbance at 295 nm was recorded as a function of temperature every 0.5°C. Melting temperatures were estimated by calculating the first derivative of the melting curve then determining the maximum value. Each melting curve was normalized by dividing the entire curve by the minimum absorbance values at 295 nm.

2.2.3 Cell culture

The non-tumorigenic human lymphoblastoid cell line LCL-721 was used in all experiments. These cells are EBV-transformed and were established from a clinically normal female donor. They were cultured in RPMI-1640 supplemented with 10% FBS and 50 µg/ml gentamycin as described (Hile et al., 2000).

2.2.4 Construction of shuttle vectors containing telomeric repeat sequences

Vectors containing various telomeric repeat sequences within the 5' coding region of the HSV-*tk* were constructed as previously described (Eckert et al., 2002). Briefly, oligonucleotides containing telomeric repeat sequences were annealed and cloned into the *BsiW*I and *Mlu*I sites of HSV-*tk* gene on the pGTK4 plasmid (Eckert et al., 1997) (Table 2-1). Telomeric repeats from human $[TTAGGG]_6$, $[TTAGGG]_{10}$, $[CCCTAA]_6$, $[CCCTAA]_{10}$ and ciliates $[TTGGGG]_{10}$, $[CCCCAA]_{10}$, $[GGGGTTTT]_5$ and $[CCCCAAAA]_5$ were inserted in-frame into the HSV-*tk* gene between bases 111 and 112 (Figure 2-4). Only the sequence of the HSV-*tk* antisense strand will be referred to throughout the manuscript. To avoid introducing a rare codon that would lead to insufficient production of the HSV-*tk* reporter gene product, one of the telomere repeats of the $[TTAGGG]_6$, $[TTAGGG]_{10}$, and $[TTGGGG]_{10}$ vectors was interrupted and the $[GGGGTTTT]_5$ repeats began with a run of Gs rather than Ts (Table 2-1; Figure 2-8). Vector names remained the same for simplicity. The HSV-*tk* gene was then subcloned into pND123 shuttle vector to generate the pJY parent control vector (Drinkwater and Klinedinst, 1986; Hile et al., 2000) and the various telomeric repeat containing shuttle vectors. All vectors were introduced and

amplified in the DH5α bacterial strain and the telomeric repeat sequence in each vector was confirmed by DNA sequence analysis.

Table 2-1.Oligonucleotides used for construction of telomeric vectors

Vector Name	Sequence (5' to 3') for the antisense strand of the plasmid
[TTAGGG]6[a]	caacgccgtacgacggaagctatggcctcgagaGTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGG acgcgcagcctggtcgaacgcagacgcgtcaa
[CCCTAA]6	caacgccgtacgacggaagctatggcctcgagaCCCTAACCCTAACCCTAACC CTAACCCTAACCCTAA acgcgcagcctggtcgaacgcagacgcgtcaa
[TTAGGG]10[a]	caacgccgtacgacggaagctatggcctcgagaGTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGac gcgcagcctggtcgaacgcagacgcgtcaa
[CCCTAA]10	caacgccgtacgacggaagctatggcctcgagaCCCTAACCCTAACCCTAACC CTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAAac gcgcagcctggtcgaacgcagacgcgtcaa
[TTGGGG]10[a]	caacgccgtacgacggaagctatggcctcgagaTGGGGTTGGGGTTGGGGTTG GGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTac gcgcagcctggtcgaacgcagacgcgtcaa
[CCCCAA]10	caacgccgtacgacggaagctatggcctcgagaCCCCAACCCCAACCCCAACC CCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAAac gcgcagcctggtcgaacgcagacgcgtcaa
[GGGGTTTT]5[b]	caacgccgtacgacggaagctatggcctcgagaCGGGGTTTTGGGGTTTTGGG GTTTTGGGGTTTTGGGGTTTTGacgcgcagcctggtcgaacgcagacgcgtc aa
[CCCCAAAA]5	caacgccgtacgacggaagctatggcctcgagaCCCCAAAACCCCAAAACCCC AAAACCCCAAAACCCCAAAACCacgcgcagcctggtcgaacgcagacgcgt caa

Insert fragments were constructed by annealing to the complimentary strand for each (not shown).

[a]The interruption of one telomeric repeat was necessary to avoid codons that code for rare tRNAs in *E. coli* that cause a significant reduction in HSV-*tk* gene product.

[b]The insert began with a run of Gs, rather than Cs, to avoid codons that code for rare tRNAs in *E. coli*.

2.2.5 HSV-tk mutational analyses of telomeric repeat shuttle vectors

Vectors (10 μg) were electroporated into 10^7 LCL721 cells and plasmid-bearing cells were selected and cloned as previously described (Eckert et al., 2002). To test for the frequency of pre-existing mutations in the vectors generated spontaneously during vector propagation and selection in *E. coli*, background HSV-*tk* mutant frequencies of all vectors were calculated as previously described (Eckert et al., 1997). To avoid the selection and propagation of LCL721 clones that received a vector with a pre-exisiting mutation, cells were cloned at densities of 1-20 cells/well that were less than 0.1 x (1/μ), where μ is the background HSV-*tk* mutant frequency in *E. coli* (Eckert et al., 2002). Each clonal population was propagated for 26 – 29 generations in media containing 150 μg/ml hygromycin, after which an alkaline extraction method (Hile et al., 2000) was used to isolate shuttle vector DNA from $2 – 3 \times 10^8$ cells.

To determine HSV-*tk* mutant frequencies, the isolated shuttle vectors were transformed into FT334 bacteria (r*ecA*113, *upp, tdk*) followed by selective plating on Vogel Bonner Minimal Salts media (VBA) supplemented with 50 μg/ml chlor with and without 40 μM FUdR (Eckert and Drinkwater, 1987). Chlor selects for plasmid-bearing bacteria and FUdR selects for bacteria containing shuttle vectors with a mutation that inactivates the HSV-*tk* gene product. The HSV-*tk* mutant frequency was determined as the number of FudR resistant colonies divided by the total number of plasmid-bearing colonies. HSV-*tk* mutant rates were calculated as the mutant frequency divided by the number of cell generations the clone was propagated to at the time of DNA isolation. Median mutant rates were determined for each telomeric shuttle vector and were analyzed by the non-parametric Mann Whitney test for pair-wise comparisons. Values determined to be outliers by the Grubb's statistical test for outliers were excluded.

2.2.6 Generation of HSV-tk telomeric shuttle vector mutation spectra

Restriction enzyme digests and DNA sequencing were used to determine the types and locations of mutations in the mutant shuttle vectors isolated after replication in the human cells. For this, plasmids from 5-10 independent HSV-*tk* mutants from at least 3 clones for each shuttle vector were isolated and analyzed as described previously (Eckert et al., 2002). After electroporation of replicated shuttle vectors into FT334 cells, the bacteria were placed on ice and aliquoted into multiple tubes containing 1 ml VBA broth. After the 2h recovery period at 37^{o}C, each culture was plated on selective media, and one FUdR-resistant mutant was isolated for plasmid purification and sequencing. This ensures that any mutational hotspots were not due to division of bacteria harboring HSV-*tk* mutant vectors during the 2 h recovery. To identify mutants with large deletions or rearrangements the plasmids were digested with *Ava*I and *Bgl*II restriction enzymes (Figure 2-4). For this analysis each mutant obtained from a single clone was considered independent since the low resolution of this assay cannot distinguish between potential siblings and independent events. Sequence changes and mutations within the promoter and coding region of the HSV-*tk* gene, as well as the telomeric inserts, were determined by dideoxy DNA sequencing at ACGT Inc. (Wheeling, IL). DNA sequence analysis was done using Align-X software of Vector NTI Advance (Invitrogen Corporation). Although rare, some mutants from the same clone exhibited the identical mutation due to either that mutation occurring independently in different plasmids within a clone, or due to an early mutagenic event that was replicated multiple times producing mutant siblings. To maintain rigor and consistency, mutants from the same clone that exhibited the identical mutation were considered siblings and that mutation was scored once. The same mutations occurring in different clones were considered independent.

2.3 RESULTS

2.3.1 Ciliate telomeric repeats form more stable G4 structures compared to human

Previous biophysical studies showed that *O. nova* telomeric $(GGGGTTTT)_3GGGG$ substrates formed significantly more stable G4 structures compared to human telomeric $(GGGTTA)_3GGG$ substrates (Lee et al., 2008). We directly compared the G4 structure stability of the human, *O.nova* and *T.thermophila* telomeric repeats in the context of the shuttle vector flanking sequence in one orientation. The various telomeric repeats were inserted in-frame between positions 111 and 112 of the HSV-*tk* reporter gene cassette on the shuttle vector, as previously described for other repeat sequences (Eckert et al., 2002; Hile et al., 2000) (Figure 2-4). G4 formation and relative thermal stability of the inserted sequences was measured by standard UV melting experiments. G4 structures exhibit hypochromic transitions as a function of temperature at 295 nm absorbance, which can serve as a signature for G4 formation (Mergny et al., 1998; Roy et al., 2007) (Figure 2-1). Melting curves for oligonucleotides GT-$(TTAGGG)_4$-TC, GT-$(TTAGGG)_6$-TC, GT-$(TTGGGG)_4$-TC and GT-$(TTTTGGGG)_4$-TC in 100 mM KCl (Figure 2-3) yielded melting temperatures (Tm) of 52.5, 53.5, 83.0 and 81.0^{o}C, respectively. Similar results were obtained with 100 mM NaCl, which also promotes G4 folding, although the G4 stabilities were decreased relative to 100 mM KCl (Table 2-2). Although the *O. nova* G4 DNA exhibited large hysteresis between the heating and cooling curves, the melting curves and Tm values were not dependent on substrate concentration (Figure 2-2). This indicates that an intra-molecular G4 structure was formed but that the re-folding rate was slow relative to the cooling rate for the experiment. These data confirm that the telomeric repeats

with flanking sequence can form intra-molecular G4 DNA, and that the ciliate G4 units with four quartet's exhibit increased thermal stability compared to human G4 units with three quartets.

Next we asked whether sequences that can form two G4 units (at least 8 G-runs) fold into more stable structures compared to sequences that form a single G4 unit (< 8 G-runs). We compared the thermal stabilities of TC-$(TTAGGG)_6$-GT to TC-$(TTAGGG)_{10}$-GT, which represent the repeat number inserted into the shuttle vector (Figs. 2-3A and B). The Tm value for ten repeats at 42.8°C in KCl was 9.3-10.3°C lower than the Tm for the shorter oligonucleotides containing 4 or 6 repeats of the human telomeric sequence. Thus, the potential to form two G4 structures for TC-$(TTAGGG)_{10}$-GT did not increase G4 stability. Importantly, these studies confirm that telomeric repeats of lengths that were inserted into the shuttle vector (6 and 10) can form uni-molecular G4 structures.

Table 2-2. Melting temperatures oC determined for G-4 oligonucleotides in 100 mM NaCl and 100 mM KCl.

	$(T_2AG_3)_4$	$(T_2AG_3)_6$	$(T_2G_4)_4$	$(T_4G_4)_4$
NaCl	39.0	41.0	52.1	52.6
KCl	52.5	53.5	83.0	81.0

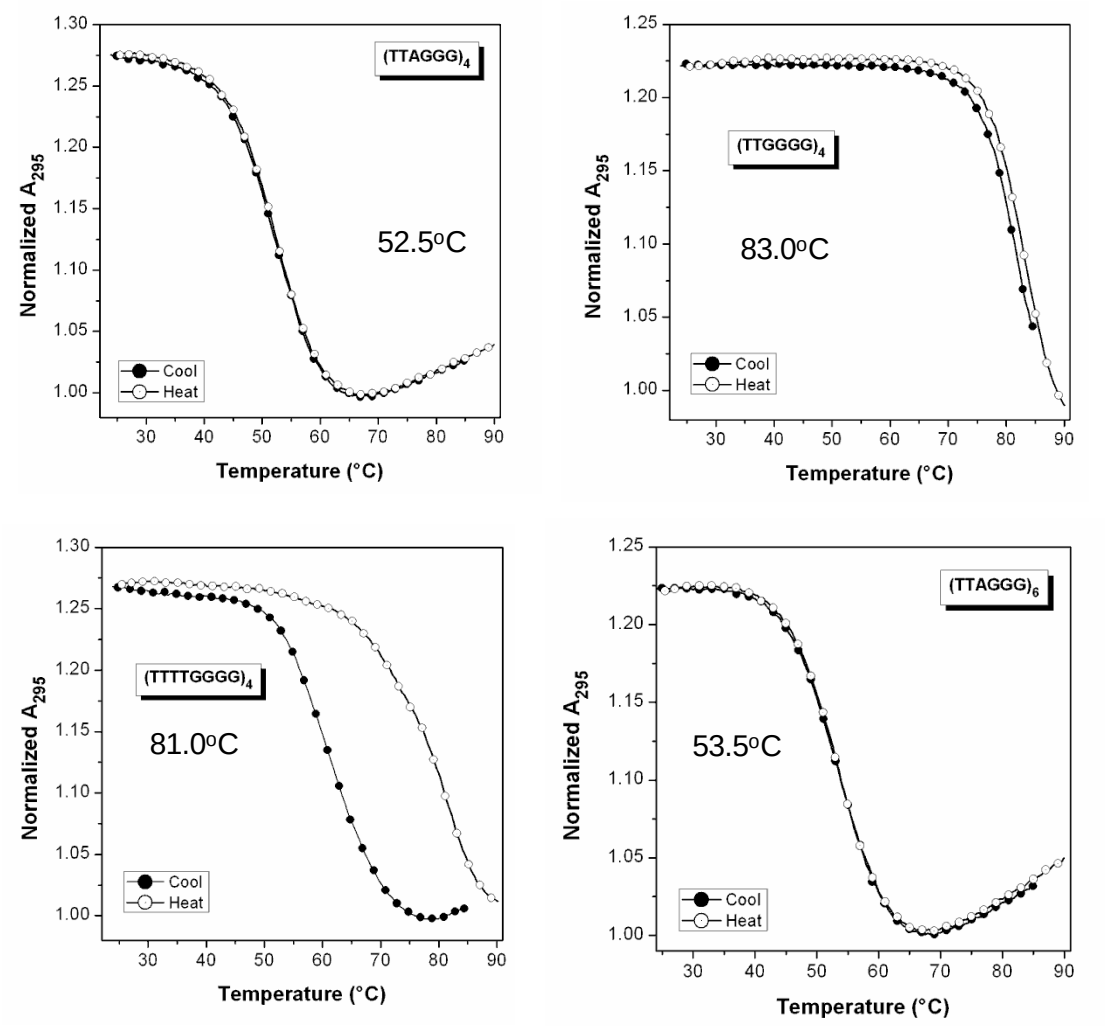


Figure 2-1. UV melting curves for ciliate and human telomeric repeats.

Melting curves for oligonucleotides GT-$(TTAGGG)_4$-TC, GT-$(TTAGGG)_6$-TC, GT-$(TTGGGG)_4$-TC and GT-$(TTTTGGGG)_4$-TC were recorded at 295 nm in solutions containing 100 mM KCl. The overlap of heating (open circle) and cooling (closed circle) curves show reversible transitions for each, except oligonucleotide GT-$(TTTTGGGG)_4$-TC which shows a hysteresis.

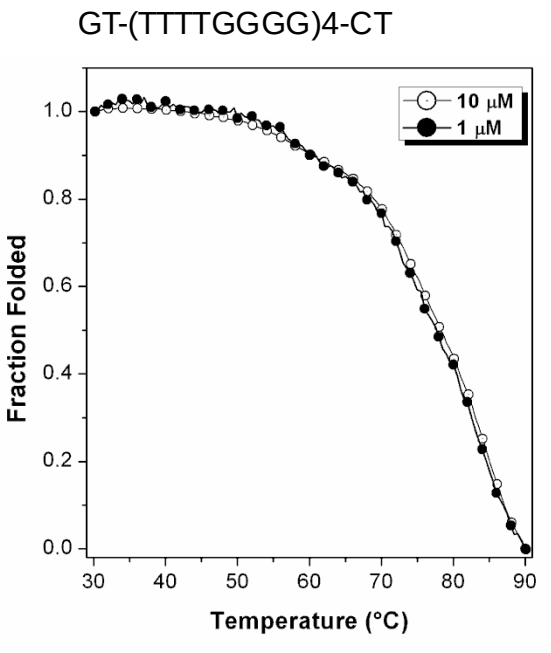


Figure 2-2. UV melting curves for O. nova telomeric repeats is not dependent on oligonucleotide concentration.

Melting curves for oligonucleotide GT-$(TTTTGGGG)_4$-TC recorded at 295 nm in solutions containing 100 mM KCl. The oligonucleotide concentrations were 10 μM (open circle) and 1 μM (closed circle).

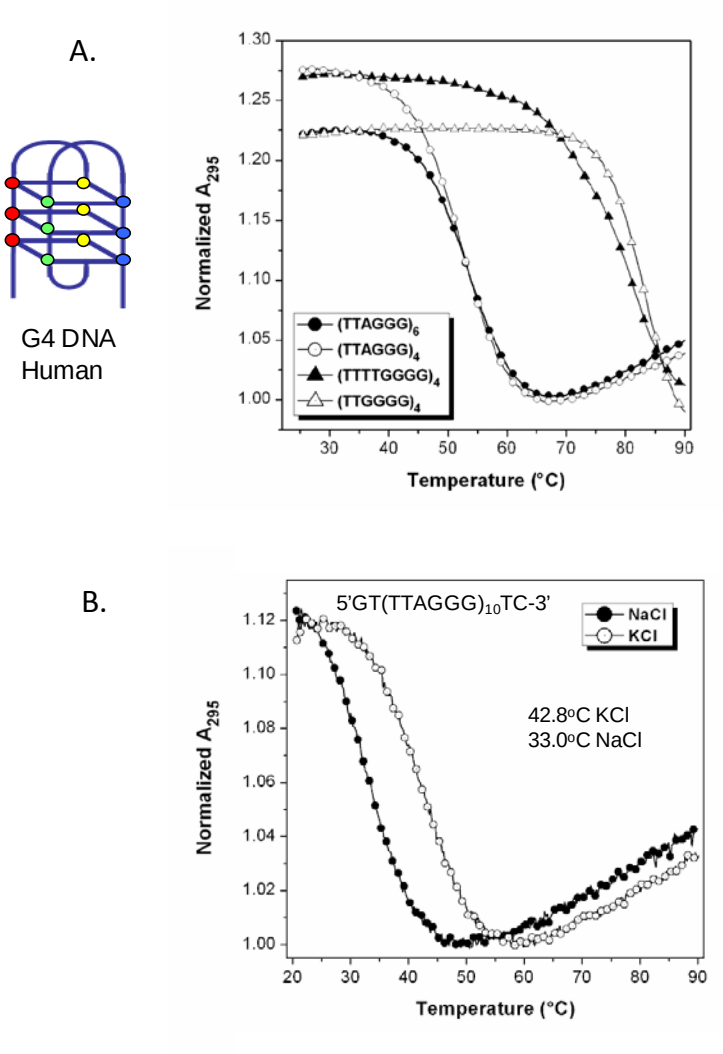


Figure 2-3. Ciliate telomeric repeats form G4 DNA structures of greater thermal stability compared to human telomeric repeats.

(A) UV Melting curves showing intra-molecular G-quadruplex formation in telomeric ssDNA with flanking sequence. Melting curves for oligonucleotides GT-$(TTAGGG)_4$-TC, GT-$(TTAGGG)_6$-TC, GT-$(TTGGGG)_4$-TC and GT-$(TTTTGGGG)_4$-TC recorded at 295 nm in solutions containing 100 mM KCl. A schematic of one possible G4 conformation for human telomere repeats is shown. Grey balls indicates Gs, and grey lines indicate a quartet formed by base pairing between four Gs. (B) UV melting curves for the olignucleotide GT-$(TTTAGGG)_{10}$-TC in two different salt solutions, 100 mM KCl and 100 mM NaCl.

2.3.2 Experimental system for analysis of telomeric DNA mutagenesis

Deletions leading to telomere loss even in cells lacking the WRN helicase are rare but can still critically impact cell function (Bai and Murnane, 2003; Crabbe et al., 2004). The shortest telomere, rather than average telomere length, determines cell survival and genome stability (Hemann et al., 2001b). Therefore, we required a highly sensitive assay to detect the types of mutagenic events that impact telomere structure and function, namely deletions and rearrangements. The *in vitro/ex vivo* shuttle vector HSV-*tk* mutagenesis assay was chosen for its proven ability to detect rare spontaneous mutations (frequencies as low as $1x10^{-5}$) (Hile et al., 2000), and for the multiple unique advantages it offers for analyzing telomeric repeat replication. First, episomal vectors avoid complications of random insertion at sites of endogenous telomeres. Second, loss of telomeric DNA in the vector will not impact cell survival, unlike loss of endogenous telomeres. Third, *ori*P episomes are used to study human chromosome replication because they replicate once in S-phase, form chromatin structure, and segregate with sister chromatids (Sears et al., 2004; Sugden, 2002). Lastly, the vectors have a defined EBV *oriP* replication origin and are replicated by the host proteins and the *oriP* binding protein EBNA-1 which lacks enzymatic activity (Dhar et al., 2001; Yates et al., 1985).

Replication of the shuttle vectors initiates at the *oriP* DS element and terminates at the 20 tandem FR repeats, which causes primarily unidirectional replication (Aiyar et al., 2009; Dhar and Schildkraut, 1991; Ermakova et al., 1996; Gahn and Schildkraut, 1989; Lindner and Sugden, 2007; Platt et al., 1993; Wang et al., 2006c). This affords the unique opportunity to examine the mutagenic potential of the TTAGGG (G-rich) sequence when replicated by lagging strand versus leading strand DNA synthesis. Single strand gaps during lagging strand replication are thought to permit G4 folding. The correct orientation on human chromosomes is the 5'-TTAGGG-3'

sequence on the lagging strand, therefore, we named all the vectors according to the lagging strand sequence. Since the replication fork starts at the 3' end of the HSV-*tk* gene, the lagging strand is the antisense strand. For example, the $[TTAGGG]_{10}$ vector has the correct repeat orientation with the G-rich sequence on the lagging strand (Figure 2-4), and the $[CCCTAA]_{10}$ vector has the reverse orientation with the G-rich sequence on the leading strand. The assay will detect any mutation that inactivates the HSV-*tk* gene product, thus, mutation frequencies and specificities in the coding region serve as an internal control. Importantly, we can detect potential deletions or rearrangements induced by the presence of the telomeric repeats because flanking HSV-*tk* sequence is also affected. The assay is designed to detect events that are most likely to impact the telomere integrity, rather than minor alterations in repeat number.

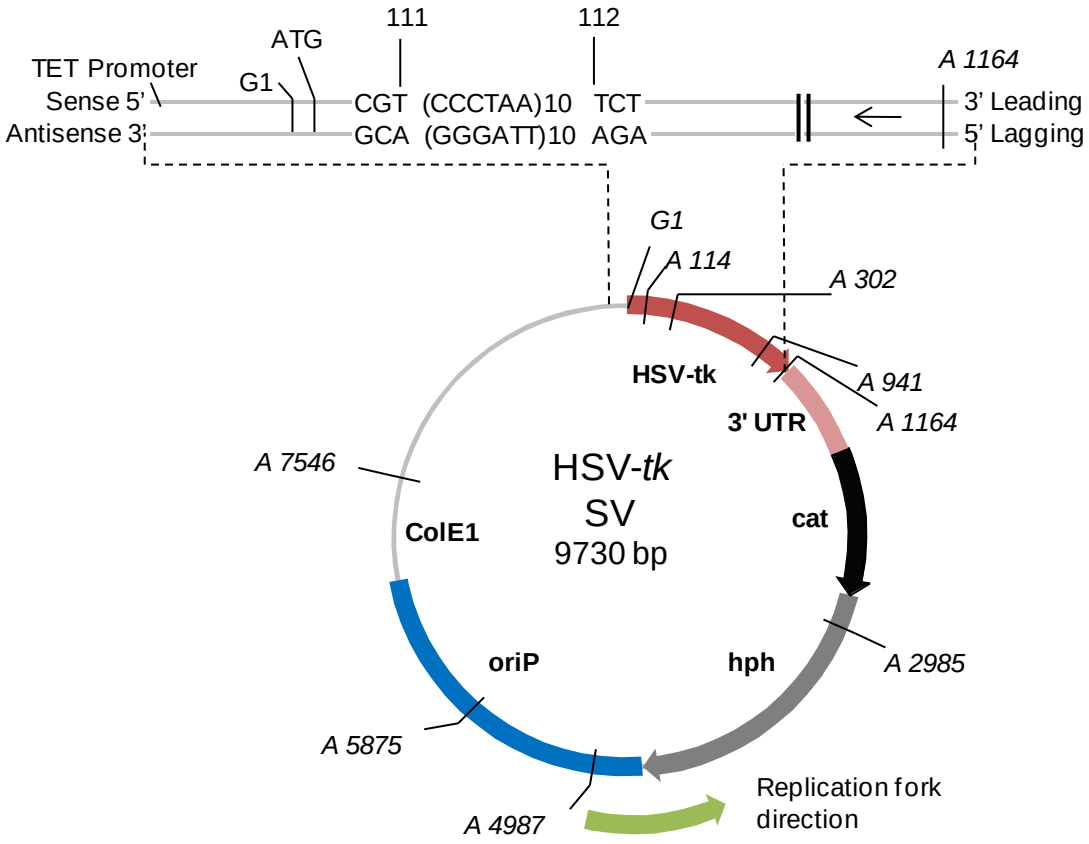


Figure 2-4. pJY shuttle vector indicating the site of telomeric repeat insertion within the HSV-tk gene.

The positions of the mutagenic reporter HSV-*tk* gene including the 3'UTR, EBV (*oriP*) origin of replication, ColE1 bacterial origin of replication, choramphenicol acetyltransferase gene (*cat*) for selection in bacteria, and hygromycin phosphoryltransferase gene (*hph*) for selection in human cells are indicated. The region from the Tet promoter to the 3'end of the HSV-*tk* gene is amplified to indicate insertion of the telomeric repeats in-frame between positions 111 and 112. Human repeats in the correct orientation are shown as an example. The arrow indicates the primary direction of replication in human cells and bacteria. The positions of restrictions sites used to screen for large deletions or rearrangements are shown; A = *Ava*I, and G = *Bgl*II.

2.3.3 Stability of telomeric vectors in bacteria

We first examined the stability of the telomeric repeat vectors in *E.coli* by determining the mutant frequency for each compared to the control vector which lacks telomeric repeats. The shuttle vectors have the bacterial replication origin derived from pBR322 which is also unidirectional (Figure 2-4 arrow) (Viguera et al., 1996) in the same direction as *ori*P. The vectors containing the human telomeric repeats were stable in bacteria and yielded HSV-*tk* mutant frequencies that were similar to the control vector (Figure 2-5, bars 1-5). Increasing the repeat number from six to ten did not significantly impact the vector mutant frequency (Figure 2-5 compare bars 2 and 4; 3 and 5). Surprisingly, the $[TTAGGG]_{10}$ vector with the G-rich lagging strand orientation was highly stable and yielded an average mutant frequency of $0.77x10^{-5}$ that was 4-fold lower than the mutant frequency obtained for G-rich leading strand $[CCCTAA]_{10}$ vector ($3.2x10^{-5}$). The difference was statistically significant whether there were six repeats (p-value = 0.0007) or ten repeats in the vectors (p-value= 0.0002). Therefore, the orientation of the human telomeric repeats influences the stability of the vectors in *E. coli*.

Next we examined the stability of the vectors with ciliate telomeric repeats. In contrast to the human repeats, the mutant frequencies obtained for the ciliate telomeric vectors were statistically similar regardless of repeat orientation (G-rich lagging vs. G-rich leading strand) (Figure 2-5, compare bars 6 to 7 and 8 to 9). However, the *T.thermophila* $[TTGGGG]_{10}$ vector with the G-rich lagging strand sequence yielded a mutant frequency of $5.0\ x10^{-5}$ that was significantly higher than the human $[TTAGGG]_{10}$ vector with the G-rich lagging strand ($0.77x10^{-5}$) (Figure 2-5 compared bars 4 and 6, p-value = 0.0007). These repeats differ by one base. Interestingly, when the G-rich sequence was on the leading strand, the mutant frequencies for the ciliate $[CCCCAA]_{10}$ and human $[CCCTAA]_{10}$vectors were not statistically different

(Figure 2-5). Thus, *T.thermophila* telomeric vectors were less stable than human telomeric vectors only when the repeats were in the G-rich lagging strand orientation.

The vectors containing the *O.nova* $[GGGGTTTT]_5$ and $[CCCCAAAA]_5$ telomeric inserts showed a >100-fold increase in mutant frequency over the control vector (Figure 2-5). Since the *O. nova* repeats are multiples of 8, changes in repeat number alter the reading frame and inactivate the HSV-*tk* gene. To determine if this was the reason for the dramatic increase in mutant frequency we sequenced the mutants. Most (14/15) of the $[GGGGTTTT]_5$ vector mutants had alterations in the telomeric repeat number (+1 repeat, 2 events; - 1 repeat, 12 events). Therefore, the *O.nova* vector affords us the opportunity to examine alterations in repeat number in telomeric repeats that share similar properties as human repeats (*i.e.* tandem G-runs).

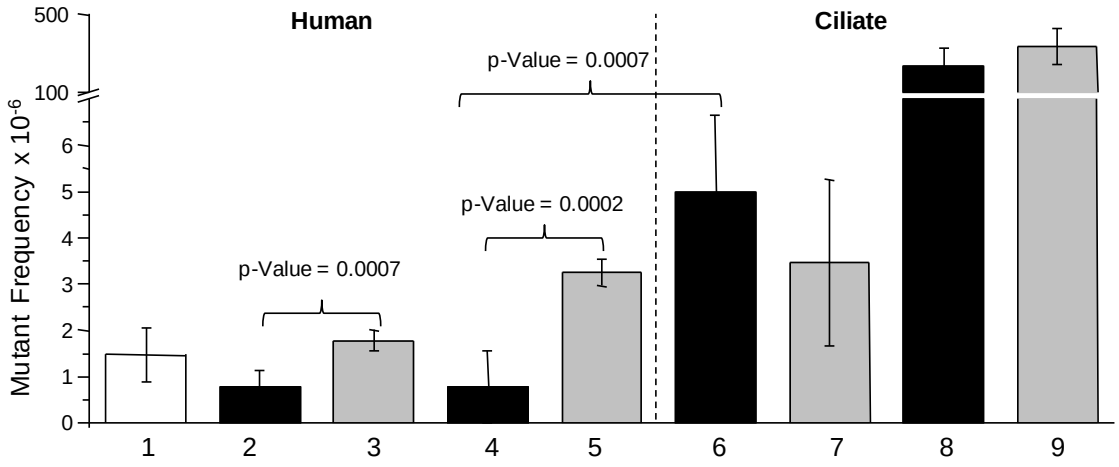


Figure 2-5. HSV-tk mutant frequencies of telomeric shuttle vectors in recA- E. coli.

Mutant frequencies are based on $1.3x10^4$ to $1.7x10^6$ chloramphenicol-resistant bacteria colonies. The sequence of the telomeric repeats on the antisense (lagging) strand is indicated. Black bars indicate that the G-rich strand of telomeric repeats in the vector is replicated as the lagging strand (correct orientation), and the grey bars indicate replication as the leading strand (reverse oriention). Values represent the mean and standard deviations from at least 5 independent experiments.

2.3.4 Replication of telomeric vectors in human cells

The vectors with telomeric inserts were introduced into LCL721 human lymphoblastoid cells by electroporation, and plasmid bearing clones were expanded for 25 to 29 population doublings (Eckert et al., 2002). Mutant rates were calculated from 4 – 7 clones from each independent experiment. Shuttle vector DNA was extracted and transformed into FT334 bacteria and mutant frequencies were calculated and converted to mutant rates (see Materials and Methods). The median mutant rate for the control vector was calculated as 3.3×10^{-6} from a total of 12 clones isolated from two independent experiments. Importantly, the median mutant rates from each experiment were very similar (4.4×10^{-6} and 3.4×10^{-6}, Table 2-3) and were consistent with previous studies (3.6×10^{-6} (Eckert et al., 2002)).

Table 2-3. Mutant rates of shuttle vectors with human telomeric repeats in LCL721 clones

	Mutant frequencies per cell generation x 10^{-6}[a]						
Clone	**No insert**		**$(TTAGGG)_{10}$**		**$(CCCTAA)_{10}$**		**$(CCCTAA)_6$**
Experiment	1	2	1	2	1	2	1
A	21	3.2	1.4	1.8	2.2	3.4	2.4
B	4.1	27	2.0	2.3	3.1	2.8	3.1
C	4.8	3.3	3.8	2.1	4.4	71	4.5
D	3.0	2.7	1.6	5.8	11	3.4	5.0
E	5.6	2.4	3.0	2.7	14	3.7	
F	1.2	3.4	4.4		6.3	3.8	
G					4.8		
Median	3.3		2.3		3.8		3.8
N	12		11		13		4

[a]HSV-*tk* mutant frequency was determined by isolating shuttle vectors from LCL721 clones after 24-35 cell generations, followed by electroporation in reporter *E. coli* and selective plating. Clones A-G represent independent clones, and experiments 1 and 2 represent independent experiments.

2.3.4.1 Stability of human telomeric repeat vectors in human cells

Next we examined the mutagenic potential of vectors with human telomeric repeats after replication in human LCL721 cells. The human telomeric repeat shuttle vectors were highly stable and yielded median mutant rates that were not statistically different from the control vector (Figure 2-6 and Tables 2-3 and 2-5). The data are presented in Figure 2-6 as a box plot in which each box represents the interquartile range (25^{th} to 75^{th} percentiles), horizontal bars in the box represent the medians, and the vertical bars represent the maximal and minimal values. When the G-rich sequence was on the lagging strand the vector median mutant rate was $2.3x10^{-6}$ ($[TTAGGG]_{10}$ vector), which was significantly lower than the mutant rate obtained for the G-rich leading strand vector ($[CCCTAA]_{10}$ vector = $3.8x10^{-6}$)(p-value = 0.0178, Mann Whitney) (Figure 2-6 and Table 2-5). Importantly, the G-rich lagging strand $[TTAGGG]_{10}$ vector has the repeats in the correct orientation as they occur at chromosomal telomeres in humans. Similar to results in bacteria, the human telomeric repeat sequences are highly stable in human LCL721 cells, and the repeat orientation significantly impacts the overall median mutant rates of the vectors.

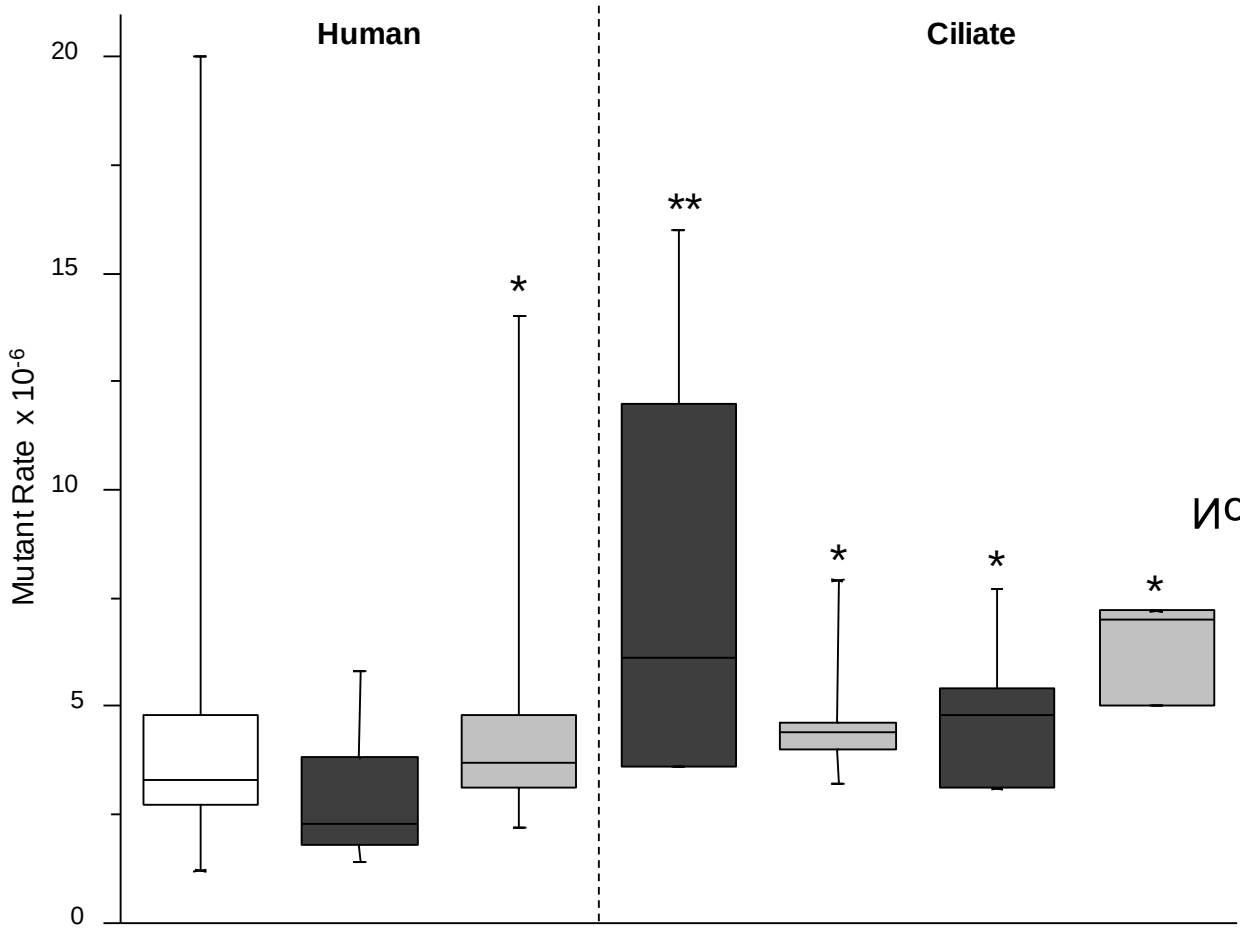


Figure 2-6. Median mutant rates of telomeric shuttle vectors in LCL721 human cells.

Clonal populations of human cells bearing various shuttle vectors were expanded for 26 to 29 generations. Shuttle vector DNA was isolated from each clone and the HSV-*tk* mutant frequency was determined as described in Materials and Methods. The mutant rate was calculated by dividing the mutant frequency by the number of generations the clone was expanded to at the time of shuttle vector isolation (see also Tables 2-3 and 2-4). Each box represents the interquartile range (25^{th} to 75^{th} percentiles). The horizontal bars in the boxes represent the medians. The vertical bars represent the range of values included in the statistics, from the

maximal to the minimal value. The following outliers were excluded: 27 [no insert control]; 71 $[CCCTAA]_{10}$; and 38 $[CCCCAAAA]_5$. * indicates $p \leq 0.05$ relative to $[TTAGGG]_{10}$ (Mann-Whitney rank sum test (two-sided)).

Table 2-4. Mutant rates of shuttle vectors with ciliate telomeric repeats in LCL721 clones.

	Mutant frequencies per cell generation x 10^{-6}[a]				
Clone	**$(TTGGGG)_{10}$**		**$(CCCCAA)_{10}$**	**$(GGGGTTTT)_5$**	**$(CCCCAAAA)_5$**
Electroporation	1	2	1	1	1
A	3.6	4.4	4.6	5.4	5.0
B	9.3	13	4.5	7.7	7.0
C	16	6.1	3.2	3.1	38
D	12	3.6	4.0	4.8	7.2
E			7.9		
F			4.4		
Median	7.7		4.4	5.1	7.1
N	8		6	4	4

[a]HSV-*tk* mutant frequency was determined by isolating shuttle vectors from LCL721 clones after 24-35 cell generations, followed by electroporation in reporter *E. coli* and selective plating. Clones A-F represent independent clones, and experiments 1 and 2 represent independent experiments.

Table 2-5. Statistical comparison of telomeric shuttle vectors mutant rates in LCL721 human cells.

	No Insert	[TTAGGG]10	[CCCTAA]10	[TTGGGG]10	[CCCCAA]10	[GGGGTTTT]5	[CCCCAAAA]5
No Insert	-	0.1309	0.3557	**0.0203**	0.2277	0.2398	0.0604
[TTAGGG]10		-	**0.0178**	**0.003**	**0.0159**	**0.0264**	**0.011**
[CCCTAA]10			-	0.0896	0.512	0.5045	0.1363
[TTGGGG]10				-	0.2451	0.3677	0.9212
[CCCCAA]10					-	0.6095	0.1667
[GGGGTTTT]5						-	0.6286
[CCCCAAAA]5							-

[a]p values are derived from pair-wise comparisons of mutant rates reported in Figure 2-6 and Tables 2-3, 2-4 for each shuttle vector using the Mann-Whitney rank sum test (two-sided).

2.3.4.2 Stability of vectors with human telomere repeats compared to ciliate repeats

Based on the model that G4 DNA in telomeric sequences may lead to replication fork demise, we predicted that telomeric sequences that form more stable G4 DNA would have a greater destabilizing effect on the vectors, compared to the human repeats. Ciliate telomeric repeats (*T.thermophila* and *O.nova*) form G4 structures of greater thermal stability than human sequences (Figure 2-3). We compared *T.thermophila* repeats to human repeats since both are multiples of six. Vectors with the correct G-rich lagging strand orientation yielded a median mutant rate for *T.thermophila* $[TTGGGG]_{10}$ vector that was 3.3-fold higher than the mutant rate obtained for the human $[TTAGGG]_{10}$ vector; $7.7x10^{-6}$ and $2.3x10^{-6}$, respectively (Figure 2-6).

This difference was statistically significant (p-value = 0.003) (Tables 2-4, 2-5). In contrast, the vectors with the G-rich leading strand yielded very similar median mutant rates; $4.4x10^{-6}$ for the *T.thermopila* $[CCCCTT]_{10}$ vector, compared to $3.8x10^{-6}$ for the human $[CCCTAA]_{10}$ vector (Figure 2-6, Tables 2-3 – 2-5). G4 units are expected to fold more frequently in the lagging strand, compared to leading strand, and these data suggest that the *T.thermopila* G4 units may impede replication more than human G4 units. Consistent with this, we observed that the vector mutant rate was nearly 2-fold higher when the *T.thermopila* G-rich sequence was on the lagging strand ($[TTGGGG]_{10}$ vector), compared to its presence on the leading strand ($[CCCCAA]_{10}$ vector) (Figure 2-6, Table 2-4). However this difference was not statistically significant, perhaps due to the strikingly large spread of values obtained for the $[TTGGGG]_{10}$ vector (Figure2-6). In summary, vectors with the correct human repeat orientation (G-rich lagging) were more stable than vectors with *T.thermopila* repeats. The orientation of the telomeric repeats for both human and *T.thermopila* influences the vector mutant rate, but in opposite directions.

2.3.4.3 Vectors with O.nova repeats are more stable in human cells compared to bacteria

To determine whether tandem telomeric repeats are susceptible to the same types of repeat alterations that occur in other simple tandem repeat (STR) sequences (Hile et al., 2000), we took advantage of the *O.nova* (GGGGTTTT) sequence. Alterations in *O.nova* repeat number inactivate the HSV-*tk* gene (Figure 2-4). These repeats form stable G4 structures (Figure 2-3), and resemble human repeats in that they consist of alternating pyrimine and purine runs and the G-rich sequence is present on the lagging strand (Gottschling and Zakian, 1986). Thus, we asked whether these characteristics of telomeric repeats make them susceptible to alterations in repeat number. Replication of the G-rich lagging strand $[GGGGTTTT]_5$ vector yielded a mediate mutant rate of $5.1x10^{-6}$ that was slightly lower than the rate for the vector with the G-

rich leading strand ($[CCCCAAAA]_5$ vector = $7.1x10^{-6}$) (Figure 2-6, Tables 2-4, 2-5). However, this difference was not significant, and neither vector yielded a mutant rate that was statistically different from the control (Table 2-5). This is in stark contrast to the 100- and 300-fold increase in mutant frequencies observed for these vectors in bacteria, compared to controls (Figure 2-5). Thus, while vectors containing *O.nova* repeats are highly unstable in *E. coli*, they exhibit similar stabilities as a vector lacking these repeats after replication in LCL721 human cells.

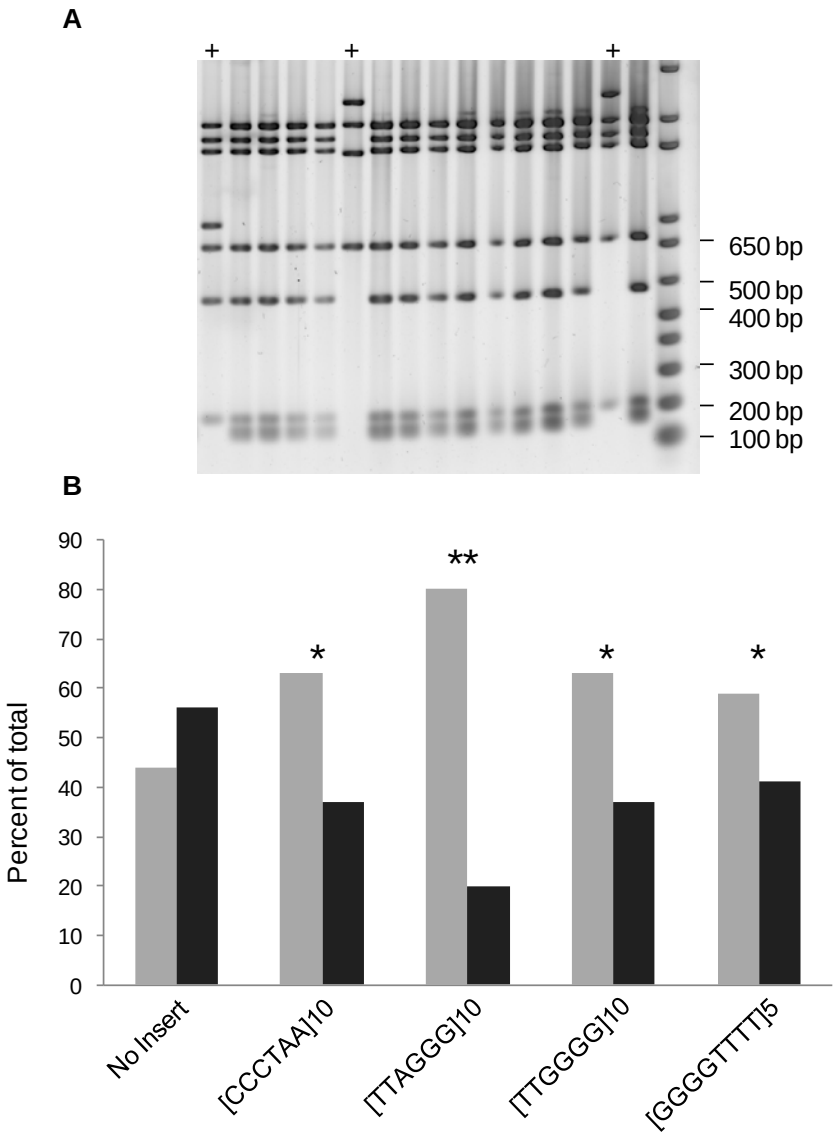


Figure 2-7. Proportion of mutants with large deletions and alterations throughout the shuttle vector. Shuttle vector DNA from 5 to 10 mutants isolated from approximately 3 clones for each shuttle vector was digested with *Ava*I and *Bgl*II restriction enzymes for 2 hr at 37°C. Reactions were run on an 0.8% agarose gel. A representative image is shown in panel *A*. The + indicates an abnormal digest pattern. *B*, the percent of mutant DNA exhibiting a wild type digest pattern (grey bars) or an abnormal digest pattern (black bars) was calculated as a function of total mutant vectors that were digested. All mutants isolated from the same clone were considered independent (see Materials and Methods). p-values calculated from Fisher's exact test.

2.3.5 Insertion of telomeric repeats decreases the frequencies of vector alterations

Next we asked whether the mutation types and frequencies arising within the shuttle vectors were influenced by the various telomeric repeats. We isolated 5 to 10 independent mutants from about 3 clones for each shuttle vector for mutational analysis (see Tables 2-3, 2-4 for clones). Previous reports indicate that H-DNA and Z-DNA forming sequences increase the frequency of large deletions and rearrangements within a shuttle vector (Wang et al., 2006, Wang and Vasquez, 2004). To test whether this was true for the telomeric G4 forming sequences, the mutant vectors were digested with specific restriction enzymes to test the integrity of the shuttle vector as described previously (Hile et al., 2000) (Figure 2-4). The proportion of mutant vectors harboring the human G-rich lagging ($[TTAGGG]_{10}$) or leading ($[CCCTAA]_{10}$) strand repeats that exhibited abnormal digest patterns was significantly lower at 20% or 37%, respectively, compared to the control vector (56%) (p-values were <0.0001 or 0.01, respectively, Fisher's exact test) (Figure 2-7). Similar results were observed for the vector mutants bearing ciliate telomeric repeats $[TTGGGG]_{10}$ and $[GGGGTTTTT]_5$ (p-values = 0.01 and 0.047, respectively; Fisher's exact test). Since the orientation of the ciliate repeats did not significantly alter the overall mutant rates (Figure 2-6), we only analyzed mutants of the vectors with the correct G-rich lagging strand orientation. In summary, the insertion of telomeric repeats decreased the frequency of shuttle vector alterations (i.e. large deletions and rearrangements induced throughout the vectors), and the human repeats in the correct G-rich lagging strand orientation yielded the most significant decrease.

Deletions that span the telomeric repeats could have arisen from breaks induced by the G4 forming sequences. To determine the frequencies of defined deletions and rearrangements that involved loss of the telomeric repeats we sequenced the entire 1110-bp HSV-*tk* gene and

upstream promoter region. The median mutation rate for defined deletions/rearrangements occurring within the HSV-*tk* gene was calculated by multiplying the proportion of these events by the total median mutant rate (Figure 2-6 and Table 2-6). The deletion/rearrangement mutation rates for human telomeric vectors were not greatly influenced by repeat orientation and were decreased approximately 2- to 3- fold ($5.6x10^{-7}$ and $7.6x10^{-7}$), compared to the control and *T.thermophila* vectors ($13x10^{-7}$ and $18x10^{-7}$, respectively) (Table 2-6). Vectors with the *O.nova* repeats exhibited the lowest rates at $1.5x10^{-7}$. However, there was no significant bias for deletions/rearrangements involving loss of the telomeric repeats for any of the vectors (Table 2-6). The proportions of events involving loss of the telomeric repeats were similar to the proportion of events that spanned the repeat insertion site (111-112 bps) for the control vector (Fisher's exact test) (Table 2-6). This indicates that the presence of telomeric repeats did not destabilize the insertion region in the HSV-*tk* gene (Figure 2-4).

Table 2-6. Sequenced deletions and rearrangements arising in the HSV-tk gene that involve the telomeric repeat region after replication in human cells.

Mutation Class	**Telomeric Inserts**[a]				
	No insert	**$[CCCTAA]_{10}$**	**$[TTAGGG]_{10}$**	**$[TTGGGG]_{10}$**	**$[GGGGTTTT]_5$**
Deletions and Rearrangements					
Median mutation rate	$13x10^{-7}$	$5.6x10^{-7}$	$7.6x10^{-7}$	$18x10^{-7}$	$1.5x10^{-7}$
Proportion of total mutations	0.39	0.15	0.33	0.24	0.030
Number	13	6	10	9	1
Loss of repeats (proportion)[b]	6 (0.46)[c]	3 (0.50)	5 (0.50)	4 (0.44)	0
Retain repeats (proportion)[b]	7 (0.54)	3 (0.50)	5 (0.50)	5 (0.56)	1

Values were derived from sequenced HSV-*tk* mutants in Supplemental Tables 2-8 – 2-12. Small deletions ($\leq$ 12 bp) were excluded. Identical mutants from a single clone were counted once to avoid including potential siblings.

[a]DNA sequence of the lagging strand of replication

[b]The proportion is based on the number of deletions and rearrangements from line 3.

[c]Deletions or rearrangements that involved loss of base pairs 111 and 112. This is the region in which telomeric repeats were inserted for the other vectors (see Figure 2-4).

2.3.6 Ciliate telomeric repeats exhibit higher mutation frequencies than human repeats

Next we compared the rates of mutations occurring within the various telomeric repeats. Mutational events that affected only the HSV-*tk* coding sequence served as an internal control, and included base substitutions, frameshifts and small deletions (Table 2-7). The proportions of HSV-*tk* coding sequence mutations were similar among the various vectors (Table 2-7). Furthermore, the spectra of mutations in the HSV-*tk* coding sequence were all dominated by base substitutions, and the distribution of mutations was not altered by the telomeric repeats in an obvious manner (Tables 2-8 to 2-12). However, we observed a significantly higher proportion of mutations within the ciliate $[TTGGGG]_{10}$ and $[GGGGTTTT]_{10}$ repeats compared to the human $[TTAGGG]_{10}$ repeats (p-value = 0.0309 and 0.0309, respectively, Fisher's exact test). This translated to a near 10-fold increase in median mutation rate for the *T.thermophila* repeats ($14x10^{-7}$) compared to human repeats in the same orientation ($1.5x10^{-7}$) (Table 2-7). The occurrence of telomeric mutations in the human repeats was rare and was not significantly influenced by repeat orientation. The mutations detected were alterations in repeat number that were accompanied by mutations elsewhere in the HSV-*tk* gene, and a frameshift within one telomere repeat (Figure 2-8 and Tables 2-8, 2-9). The mutations in the *T.thermophila* repeats also included alterations in repeat number, but more strikingly, we observed a total of 8 mutants containing large deletions with an endpoint in the $[TTGGGG]_{10}$ repeats (Figure 2-8). Since six of these mutants were isolated from the same clone (C) they could potentially be siblings and were scored once (Tables 2-6, 2-7), but it is worth noting that siblings were rare (Tables 2-8 to 2-12). These data indicate that ciliate [TTGGGG] repeats induce deletions likely due to breaks within the repeats.

The *O.nova* $[GGGGTTTT]_5$ vectors exhibited an increase in median telomeric mutation rate ($9.2x10^{-7}$), compared to vectors with human repeats ($1.5x10^{-7}$), due to the potential for alterations in repeat number to shift the HSV-*tk* reading frame (Table 2-7). Consistent with this, all of the sequenced telomeric mutations involved repeat alterations, and the loss of telomeric repeat units occurred more frequently compared to repeat additions (Figure 2-8). Some mutants from the same clone showed the identical mutation and were considered to be siblings (Table 2-7). But we cannot rule out the possibility that these were independent events, especially considering that similar mutations were observed in multiple distinct clones (Figure 2-8). However, none of sequenced *O.nova* mutants exhibited large deletions with endpoints in the $[GGGGTTTT]_5$ region, in contrast to the *T.thermophila* repeats. Thus, the mutation spectra revealed that the *T.thermopila* and *O.nova* repeats yielded higher telomeric mutation rates compared to the human repeats due to distinctly different mechanisms.

Table 2-7. Defined mutational events within the telomeric repeats and HSV-tk coding regions after replication in LCL721 human cells.

Mutation Class	Telomeric Inserts				
	No insert	**$[CCCTAA]_{10}$**	**$[TTAGGG]_{10}$**	**$[TTGGGG]_{10}$**	**$[GGGGTTTT]_5$**
Telomeric Sequence[a]					
Median mutation rate	n/a	$1.0x10^{-7}$	$1.5x10^{-7}$	$14x10^{-7}$	$9.2x10^{-7}$
Number	n/a	1	2	7	6
Proportion of total	n/a	0.025	0.067	0.18	0.18
HSV-*tk* Coding Region[b]					
Median mutation rate	$20x10^{-7}$	$31x10^{-7}$	$15x10^{-7}$	$51x10^{-7}$	$40x10^{-7}$
Number	20	33	19	25	26
Proportion of total	0.61	0.83	0.63	0.66	0.79
Total Mutations	33	41	30	38	33

Values were derived from sequenced HSV-*tk* mutants in Tables 2-8 – 2-12. Identical mutants from a single clone were counted once to avoid including potential siblings

[a]Includes large deletions with a border in the telomeric repeats

[b]Includes base substitutions, frameshifts and small deletions < 12 bp confined to the HSV-*tk* coding sequence.

Replication fork

[TTAGGG]10 human

Clone	no.	3' ctg cga cgc gca GGATTG GGATTG GGATTG GGATTG GGATTG GGATTG GGATTG GGATTG GGATTG GGATTG aga gcg ccg gta 5'
B	1	ctg cga cgc gca GGATT_GGATTG GGATTG GGATT G GGATTG GGATTG GGATTG GGATTG GGATTG GGATTG aga gcg ccg gta
C	1	ctg cga cgc gca _____ aga gcg ccg gta

[TTGGGG]10 *T.thermophila*

Clone	no.	3' ctg cga cgc gca TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT aga gcg ccg gta 5'
A	1	ctg cga cgc gca TGG GGA TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT aga gcg ccg gta
C	1	ctg cga cgc gca TGGGGGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT aga gcg ccg gta
A,B	6,2	ctg cga cgc gca TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT TGG GGT aga
C,D	6,1	ctg cga cgc gca TGG GGT TGG GGT Tga cct gac tcg tct gtc tgg gca cca aaa acc tac cgg acc cgt acc tgg cgt ac a tga ccg cgc tgt gct g (1515 bp deletion)
A	1	ctg cga cgc gca TGG GGT TGG GGT TGG GGT TGG GGT TGG GGg ggg ttt ttg gtg gcg c cgc cta aag acc gcg gcg gcc tgc ttg att tgg (1621 bp deletion)

[GGGGTTTT]5 *O.nova*

Clone	no.	3' ctg cga cgc gca GTT TTG GGG TTT TGG GGT TTT GGG GTT TTG GGG TTT TGG GGC aga gcg ccg gta 5'
A,B	1,2	ctg cga cgc gca G _____ TTT TGG GGT TTT GGG GTT TTG GGG TTT TGG GGC aga gcg ccg gta
A,C	2,1	ctg cga cgc gca G _____T TTT GGG GTT TTG GGG TTT TGG GGC aga gcg ccg gt
A,B	1,1	ctg cga cgc gca GTT TTG GGG TTT TGG GGT TTT GGG GTT TTG GGG TTT TGG GGT TTT GGG GTT TTG GGG C aga gcg ccg gta

Fig. 2.8. Spectra of mutations arising within the telomeric repeats for the various shuttle vectors. The sequences of the various telomere repeats (upper case) with flanking HSV-*tk* coding sequence (lower case) are shown 3′–5′ for the lagging strand of the vector (see Fig. 2-4). Sequence alterations in the mutants are shown in bold. A line indicates deleted sequence. The clone from which the mutant was derived is shown, and the number (no.) next to the clone indicates the number of mutant isolates from the clone that exhibited the indicated mutation (see Supplemental Tables 2-8 – 2-12). Arrow indicates the direction of the replication fork.

Table 2-8. Mutational events in HSV-tk gene of no insert control shuttle vector after replication in LCL721 clones.

5' mutations[a]	Clone[b]	3' mutations[a]	Clone[b]
Deletion 159 bp at -141	**2B**[c]	- G:C 116	D
Deletion 1808 bp at -1	A	T:A to C:G 176	A
G:C to A:T 24	C	C:G to A:T 178	A
Deletion 1565 bp at 52	**B**	C:G to G:C 180	A
-A:T 48	B	G:C to C:G 182	D
-G:C 49	B	G:C to T:A 183	A
T:A to A:T 59	C	Deletion 92 bp at 146	A
Deletion 284 bp upstream to 91	**A**	Deletion 954 bp at 301	2E[c]
C:G to T:A 115	C	G:C to T:A 304	A
Deletion 118 bp at 111	E	C:G to A:T 367	F
Deletion 630 bp at 111	E	C:G to T:A 370	F
Deletion 214 bp upstream to 112	F	-C:G 372-373	F
862 bp deletion upstream to position 501	D	Rearrangement after position 387	A
		+G:C 487- 493	E
		C:G to A:T 543	A
		-C:G 605-610	**B**
		G:C to T:A 716	**A**
		-G:C 846-849	**A**
		Deletion 327 bp at 975	F
		Deletion 115 bp at 991	E

All sequences represent the sense strand.

[a]5' mutations represent mutational events upstream of position 112 in the HSV-tk gene and 3' mutations represent mutational events downstream of position 112.

[b]Clones A-F represent the clones in table 2-3 from which the mutant was isolated and sequenced. Multiple independent mutants were characterized for each clone. Bolded letters denotes that the clone was independent experiment #2 (see Table 2-3).

[c]Multiple isolates from a single clone. Indicates the number of mutants from the designated clone that exhibited the mutation.

Table 2-9. Mutational events in HSV-tk gene of (TTAGGG)10 containing shuttle vector after replication in LCL721 clones.

5' and telomeric mutations[a]	Clone[b]	3' mutations[a]	Clone[b]
Deletion 1401 bp at -81	A	Multiple base substitutions 178 – 183	C
Deletion 193 bp upstream to 1[d]	**C**		
C:G to T:A 37	**C**	CGCCGG:GCGGCC to AGGCCT:TCCGGA	
Deletion of all 10 telomeric repeats[d]	**C**	T:A to G:C 218	**9C**[c]
-C from (AATCCC) in the first telomeric repeat	B	Small deletion 12 bp at 280	E
Deletion 1099 bp upstream to 161	A	G:C to A:T 298	C
Deletion 1188 bp upstream to 247	B	T:A to C:G 299	C
Deletion 1178 bp upstream to 362	B	+G:C 301-303	D
		G:C to A:T 320	D
		+TT:AA between 372 and 373	B
		C:G to A:T 448	C
		-T:A 450-451	A
		-C:G 461-462	7A[c]
		Small deletion 8 bp at 470 (TTCTGGCT)	C
		+G:C 487-493	**C**
		Deletion 719 bp at 502	A
		G:C to A:T 541	**C**
		T:A to G:C 638	C
		G:C to T:A 674	**B**
		Deletion 63 bp at 696	**B**
		Deletion 268 bp at 722	D
		-G:C 824-826	**C**
		T:A to C:G 920	D
		Deletion 551 bp at 1035	A

All sequences represent the sense strand

[a]5' mutations represent mutational events upstream of and including the (TTAGGG)10 region at position 112 in the HSV-tk gene and 3' mutations represent mutational events downstream of the repeats and position 112.

[b]Clones A-F represent the clones in Table 2-3 from which the mutant was isolated and sequenced. Multiple independent mutants were characterized for each clone. Bolded letters denotes that the clone was independent experiment #2 (see Table 2-3).

[c]Multiple isolates from a single clone. Indicates the number of mutants from the designated clone that exhibited the mutation.

[d]Occurred together in the identical mutant from clone **C**.

Table 2-10. Mutational events in HSV-tk gene of (CCCTAA)10 containing shuttle vector after replication in LCL721 clones.

5‘ and telomeric mutations[a]	Clone[b]	3' mutations[a]	Clone[b]
C:G to G:C 4	F	G:C to T:A 172	B
C:G to A:T 69	B	C:G to A:T 202[d]	F
C:G to G:C 89[d]	F	Deletion 47 bp at 206	G
+(TTAGGG)3 in telomeric region[d]	F	G:C to C:G 209	A
Deletion 184 bp upstream to 118	**A**	G:C to T:A 220	**A**
Rearrangement upstream to 437	**D**	G:C to C:G 223	**F**
Deletion 810 bp upstream to 462	**D**	Deletion 65 bp at 266	B
		T:A to A:T 277	F
		C:G to T:A 280	F
		G:C to A:T 281	F
		-C:G 282	F
		G:C to T:A 309	**A**
		T:A to A:T 316	**A**
		-C:G 461-462	**F**
		-G:C 487-493	**D**
		+G:C 487-493	A
		C:G to T:A 610	**4C**[c]
		C:G to A:T 632	B
		A:T to T:A 637	B
		+A:T between 638 and 639	7
		C:G to G:C 694	2F[c]
		G:C to A:T 739	**D**
		A:T to G:C 740	F
		C:G to T:A 763	E
		G:C to C:G 822	F
		Deletion 266 bp at 837	B
		-C:G 877-880	G
		T:A to A:T 972	B
		T:A to C:G 1042	F
		G:C to C:G 1045	F
		G:C to A:T 1134	F
		G:C to A:T 1150	F
		G:C to A:T 1178	F

All sequences represent the sense strand

[a]5’ mutations represent mutational events upstream of and including the (CCCTAA)10 region at position 112 in the HSV-tk gene and 3’ mutations represent mutational events downstream of the repeats and position 112.

[b]Clones A-G represent the clones in Table 2-3 from which the mutant was isolated and sequenced. Multiple independent mutants were characterized for each clone. Bolded letters denotes that the clone was independent experiment #2 (see Table 2-3).

[c]Multiple isolates from a single clone. Indicates the number of mutants from the designated clone that exhibited the mutation.

[d]Occurred together in the identical mutant from clone F.

Table 2-11. Mutational events in HSV-tk gene of (TTGGGG)10 containing shuttle vector after replication in LCL721 clones.

5' and telomeric mutations[a]	Clone[b]	3' mutations[a]	Clone[b]
Deletion 1993 bp -151	**B**	C:G to T:A 130	A
C:G to A:T 109	A	C:G to A:T 245	D
A:C to T:A first telomeric repeat	A	CGC:GCG to AA:TT 280 to 282	B
+(CCCCAA)2	6A, 2B[c]	C:G to A:T 344	B,C,D
Deletion of 8 telomeric repeats followed by 1515 bp deletion from 112	6C,1D[c]	Deletion 1091 bp at 414	D
+C:G in the first telomeric repeat	C	C:G to T:A 448	**A**
Deletion of 6 telomeric repeats followed by 1621 bp deletion from 112	A	-C:G 478-479	**C**
		+G:C 487-493	C
		A:T to G:C 494	C
		T:A to G:C 510	C,D
		G:C to A:T 541	C
		A:T to T:A 664	D
		G:C to T:A 674	C
		A:T to G:C 731	**D**
		Small deletion 4bp 778	**8A[c]**
		T:A to A:T 783	**8A[c]**
		Deletion 93 bp 803	C,D
		G:C to T:A 821	D
		-G:C 824-826	C
		-C:G 938-942	C
		Deletion 64 bp 991	C
		T:A to A:T 1051	D
		Deletion 643 bp 1107	**C**
		T:A to C:G 1130	C,D

All sequences represent the sense strand

[a]5' mutations represent mutational events upstream of and including the (TTGGGG)10 region at position 112 in the HSV-tk gene and 3' mutations represent mutational events downstream of the repeats and position 112.

[b]Clones A-D represent the clones in Table 2-4 from which the mutant was isolated and sequenced. Multiple independent mutants were characterized for each clone. Bolded letters denotes that the clone was independent experiment #2 (see Table 2-4). Some mutations occurred in multiple clones as indicated by distinct letters.

[c]Multiple isolates from a single clone. Indicates the number of mutants from the designated clone that exhibited the mutation.

Table 2-12. Mutational events in HSV-tk gene of (GGGGTTTT)5 containing shuttle vector after replication in LCL721 clones.

5' and telomeric mutations[a]	Clone[b]	3' mutations[a]	Clone[b]
G:C to A:T 49	A, B	C:G to A:T 133	B
Deletion of 1 telomeric repeat	1A,2B[c]	G:C to A:T 166	B
Addition of 1 telomeric repeat	2A,1C[c]	G:C to C:G 189	B
Deletion of 2 telomeric repeats	A,B	G:C to A:T 220	B
Deletion 26 bp at 56	C	G:C to T:A 233	A
		T:A to A:T 236	B
		A:T to T:A 241	C
		C:G to A:T 245	B
		T:A to A:T 316	C
		C:G to A:T 344	C
		+G:C 400-403	C
		G:C to A:T 402	C
		A:T to G:C 494	C
		C:G to G:C 497	C
		T:A to G:C 498	C
		+G:C 499-501	A
		+GGGG:CCCC 499-501	C
		+G:C 511	C
		+G:C 522-523	C
		+G:C 567	C
		T:A to A:T 571	B
		-G:C 655-656	B
		+T:A 1159-1161	C
		+G:C 1173-1177	C

All sequences represent the sense strand.

[a]5' mutations represent mutational events upstream of and including the (TTTTGGGG)5 region at position 112 in the HSV-tk gene and 3' mutations represent mutational events downstream of the repeats and position 112.

[b]Clones A-D represent the clones in Table 2-4 from which the mutant was isolated and sequenced. Multiple independent mutants were characterized for each clone. Bolded letters denotes that the clone was independent experiment #2 (see Table 2-4). Some mutations occurred in multiple clones as indicated by distinct letters.

[c]Multiple isolates from a single clone. Indicates the number of mutants from the designated clone that exhibited the mutation.

2.4 DISCUSSION

In this study we measured the mutagenic potential of various telomeric repeats in *E. coli* and clonal populations of human lymphoblastoid cells as a function of repeat orientation and G4 DNA thermal stability. To our knowledge this is the first report of spontaneous mutation rates of telomeric repeat sequences using a shuttle vector mutagenesis assay. This highly sensitive assay allowed us to quantitate rare and independent mutagenic events that are expected to impact telomere integrity and function, namely large deletions and rearrangements. Importantly, the unidirectional replication origins afforded a unique opportunity to compare leading versus lagging strand replication of the G-rich sequences which possess quadruplex forming potential. As expected we observed that molecules with telomeric repeats from ciliates formed G4 structures with higher thermal stability compared to human repeats *in vitro*. Consistent with this, shuttle vectors with *T.thermophila* $[TTGGGG]_{10}$ repeats exhibited greater mutant frequencies compared to vectors with human $[TTAGGG]_{10}$ repeats in both bacteria and human cells. This translated to a near 10-fold higher rate of mutations within the *T.thermophila* repeats compared to human repeats (Table 2-7). Interestingly, the vectors with the human repeats in the correct orientation (lagging G-rich strand) were highly stable, and exhibited significantly lower mutant frequencies compared to vectors with the reverse repeat orientation (leading G-rich strand) for both bacteria and human cells. In general, we observed that the mutagenic potential of various telomeric repeats is relatively low in human cells, but is influenced by repeat orientation and thermal stability of the folded G4 DNA.

Replication of G-rich sequences as the lagging strand template mimics the process of telomere replication on human chromosomes. The result that the human telomeric repeat vector with the G-rich lagging strand sequence yielded significantly lower mutant frequencies

compared to the vector with G-rich leading strand sequence, in both human cells and bacteria, was unexpected (Figs. 2-4, 2-5). Transient ssDNA in Okazaki fragment processing is predicted to permit G4 folding, and is thought to explain the preferential loss of telomeres replicated from the G-rich lagging strand in cells defective for WRN helicase, FEN1, and POT1 proteins (Crabbe et al., 2004; Saharia et al., 2008; Vallur and Maizels; Wu et al., 2006). G4 impediments to replication were also attributed to the deletions and rearrangements in G-rich repeat sequences in *S. cerevisiea* lacking Pif1 helicase (Ribeyre et al., 2009) and *C.elegans* lacking the FANCJ helicase homolog Dog-1(Cheung et al., 2002). One possible explanation for our results is that quadruplexes did not form, or did not fold more frequently when the G-rich runs were present on the lagging strand versus the leading strand. We believe the former is unlikely because TTAGGG repeats were observed to form G4 DNA on plasmids *in vivo* during transcription (Duquette et al., 2004). Another possibility is that regulated G4 DNA folding is favorable for replication of human telomeric repeats rather than detrimental. A recent study reported a positive role for G4 DNA in the maintenance of telomere lengths, and suggested that intra-molecular G4 DNA promotes telomerase translocation in *S. cerevisiea* (Zhang et al.). We propose that G4 DNA may recruit helicases such as the RecQ helicases WRN and BLM or FANCJ in human cells, and RecQ in bacteria, which all exhibit high affinity for, and unwind G4 structures *in vitro* (Mohaghegh et al., 2001; Wu and Maizels, 2001; Wu et al., 2008). These enzymes are known to act in pathways that prevent replication fork demise, restore stalled replication forks, and promote genomic stability (Wu and Hickson, 2006). Consistent with this, the presence of TTAGGG repeats on the lagging strand had the greatest stabilizing effect on the shuttle vectors (Figs. 2-4, 2-5, 2-7), and offered the potential for G4 folding during Okazaki fragment processing.

Our studies indicate that the thermal stability of G4 structures also influences the mutagenic potential of G4 forming sequences. The higher thermal stability of the ciliate repeat G4 DNA, compared to human repeat G4 DNA (Figure 2-3), agrees with previous work and is due to the extra guanine tetrad in the ciliate repeats (Lee et al., 2008; Risitano and Fox, 2003). Even in the presence of a complementary strand the *T.thermophila* $(TTGGGG)_4$ strand predominately formed G4 DNA, whereas the human $(TTAGGG)_4$ strand existed in a mixture of G4 and duplex DNA (Risitano and Fox, 2003). This correlated with a higher mutant frequency for vectors with the TTGGGG lagging strand repeats, compared to vectors with this sequence on the leading strand, and vectors with the human TTAGGG lagging strand repeats in both bacteria and humans (Figs. 2-5, 2-6). Thus, in the case of the *T.thermophila* repeats, the expected higher G4 potential when the G-rich sequence was on the lagging strand (correct orientation), compared to leading strand (reverse orientation), was detrimental rather than beneficial unlike the human repeats. The G4 forming sequence not only determines thermal stability but also topology (*i.e.* parallel vs. anti-parallel, loop size between tetrads, arrangement) (Burge et al., 2006), therefore, it was proposed that various helicases exhibit differing efficiencies of G4 unwinding depending on the G4 structure (Ribeyre et al., 2009). Humans and *E.coli* may possess an abundance of helicases, such as RecQ helicases, that are well suited for disrupting human telomeric G4 DNA, compared to other types of G4 sequence structures. Recruitment of telomeric proteins TRF1 and TRF2 to the human repeats may further enhance their stability within the shuttle vector. In human cells, we observed a higher rate of mutations within the *T.thermophila* TTGGGG repeats compared to human repeats in either orientation, and we observed large deletions with endpoints in the TTGGGG repeats only (Figure 2-8). These deletions suggest the *T.thermophila* repeats may have impeded replication fork progression. Chromatin immuno-precipitation studies

showed that TRF1 and TRF2 proteins bind the TTAGGGTTA site in the *ori*P in human cells (Deng et al., 2003; Eckert et al., 2002), indicating that the human telomeric sequence we inserted into the shuttle vectors should also bind telomeric proteins. This raises the possibility that replication might have initiated within the telomeric repeats, but this is unlikely because only *oriP* can support plasmid establishment and long term replication in clonal expansion (Wang and Sugden, 2008). TRF1 and TRF2 interact with G4 DNA resolving helicases including WRN and BLM (Lillard-Wetherell et al., 2004; Opresko et al., 2002), and may enhance recruitment of these helicases to the shuttle vectors with human telomeric repeats. Recent work indicates that TRF1 is required to prevent replication fork stalling at human telomeric repeats (Sfeir et al., 2009). It is important to note that the structure of the TRF1 and TRF2 proteins with respect to the progressing replication fork will be influenced by the orientation of their TTAGGG binding sequences, and this might also influence replication fidelity.

Since *O.nova* telomeric repeats are multiples of 8, we could directly compare the stability of sequences with telomeric character and G4 DNA forming potential, to Short Tandem Repeat (STR) sequences elsewhere in the genome. Alterations in repeat number shift the HSV-*tk* reading frame. Surprisingly, the vectors with *O.nova* $(GGGGTTTT)_5$ repeats in either orientation were highly unstable in *E. coli* and exhibited a near 100- to 300-fold increase in mutant frequency compared to controls due to changes in repeat number (Figure 2-5 and data not shown). In stark contrast, these repeats were stably replicated in human LCL721 cells (Figure 2-6), and while some mutants also exhibited alterations in repeat number the frequencies were very low (Table 2-7 and Figure 2-8). The human and bacterial replicative polymerases may differ in their tendency to slip in regions with high G-C content. However, our data indicate that in normal human cells the G4 forming potential of repeat sequences is unlikely to promote DNA

polymerase slippage and primer/template misalignments that could lead to alterations in repeat number. Consistent with this, the vectors with *O.nova* repeats in either orientation were more stable and exhibited lower mutant rates ($5.1x10^{-6}$ and $7.1x10^{-6}$) than the STR sequences $[TTTC/AAAG]_9$ and $[TCTA/AGAT]_9$ ($38x10^{-6}$ and $58x10^{-6}$) reported earlier using our system in human cells (Eckert et al., 2002). While the G4 structures from *O.nova* repeats exhibited similar thermal stabilities as the G4 units from *T.thermophila*, large deletions were only observed in the *T.thermophila* repeats (Figure 2-8). This suggests that the *T.thermophila* repeats likely induced replication fork stalling. RecQ helicases have been found to efficiently disrupt *O.nova* telomeric G4 DNA (Mohaghegh et al., 2001), and the larger (TTTT) loop between G-tetrads compared to the (TT) loop for *T.thermophila* influences the G4 topology (Tang et al., 2008). Thus, not all G4 forming sequences exhibit the same mutagenic potential and mutation specificities.

While differences were observed among the various telomeric repeats, all of the telomeric vectors exhibited relatively low mutant rates after replication in human cells. One limitation of the assay is that the vectors contain only 5 to 10 telomeric repeats, whereas upwards of 800 repeats exist in chromosomal ends (Riethman, 2008). However, the shuttle vector with 10 telomeric repeats approaches the average length of an Okazaki fragment during DNA replication (~140 nt) (Cleary et al., 2002), and thus, mimics the span of ssDNA arising during normal telomere replication that could form G4 DNA. Furthermore, previous reports indicate that a greater number of telomeric repeats does not correlate with more G4 DNA units (Vorlickova et al., 2005). In contrast to the G4 forming telomeric repeats, the insertion of sequences of similar length (25-32 bp) that can form alternate non-B structures (H-DNA and Z-DNA), led to a significant increase in deletions (Wang et al., 2006a; Wang and Vasquez, 2004). These data suggest that some H-DNA and Z-DNA structures may impede replication or provoke breakage

(Wang et al., 2006; Wang and Vasquez, 2004), in stark contrast to the telomeric G4 sequences. Our results are consistent with studies that indicate natural interstitial telomeric sequences of perfect repeats in human cells are not prone to breakage (Ruiz-Herrera et al., 2008).

In summary, our experiments indicate that human and ciliate telomeric repeats are stably replicated in normal human somatic non-tumorigenic cells, and that thermal stability and topology of potential G4 structures and repeat orientation influence the mutagenic potential. Our studies suggest that normal human cells have evolved mechanisms, including a plethora of helicases, to facilitate replication through G4 forming telomeric sequences. Consistent with this, the vectors with human repeats in the correct orientation exhibited significantly lower mutant rates than the control and other telomeric vectors (Figure 2-6), and the lowest proportion of vector alterations (Figure 2-7). These results suggest that human cells have evolved mechanisms for optimizing replication specifically of human telomeres. This may partly explain why telomeres in normal cells are stably replicated despite the tracts of repetitive sequences that can form alternate structures. The mutagenesis assay used in this study now provides a highly sensitive approach for elucidating the roles of various proteins in promoting faithful telomere replication and telomere stability, such as various DNA helicases.

2.5 ACKNOWLEDGEMENTS

This work was supported by NIH grants [RO1 ES0515052 (P.L.O.); RO1 CA100060 (K.A.E.)], the Ellison Medical Foundation (P.L.O.), and the Jake Gittlen Cancer Research Foundation (K.A.E.). We thank members of the Opresko and Eckert lab for critical reading of the manuscript, and Gregory Sowd and Suzanne Hile for technical support and assistance.

3.0 INVESTIGATING THE ROLE OF WRN PROTEIN IN PREVENTING REPLICATION INDUCED DELETIONS AND MUTATIONS IN TELOMERE REPEATS

This chapter is currently being prepared as a manuscript by the following authors: Rama Rao Damerla1, †, Kelly E. Knickelbein1, †, Fu-Jun Liu, Hong Wang and Patricia L Opresko.

†These authors contributed equally to the paper. P.L.O. supervised the project, interpreted the data and wrote the manuscript. H.W. performed the AFM experiments.

3.1 INTRODUCTION

Werner syndrome (WS) is a rare autosomal recessive disorder characterized by numerous features of premature aging and a predisposition to cancer (Kudlow et al., 2007) . WS is caused by loss of the WRN helicase/exonuclease, which is a member of the highly conserved RecQ helicase family (Huang et al., 1998a; Yu et al., 1996b). Cells from WS patients undergo premature replicative senescence, exhibit an extended S-phase, and show increased levels of genomic and telomere instability (Crabbe et al., 2004; PooT et al., 1992; Sidorova et al., 2008). Evidence indicates that WRN protein has a critical role in responding to DNA replication stress, and functions in the prevention or restoration of stalled and broken DNA replication forks

(Cheng et al., 2008; Pichierri et al., 2001). WRN deficient cells exhibit significantly reduced rates of replication fork elongation after cellular exposures to agents that induce replication stress by either generating DNA damage or by depleting nucleotide pools (Sidorova et al., 2008). Consistent with this, WRN is required to prevent replication-associated breaks at fragile sites *in vivo*, and to prevent DNA polymerase δ from stalling at fragile site sequences *in vitro* (Shah et al., 2010; Pirzio et al., 2008). Together the data indicate that WRN has roles in facilitating DNA replication, which are critical for genome maintenance and cancer prevention.

Telomeres are protein-DNA structures that protect the ends of linear chromosomes and are critical for genome stability. Telomere dysfunction contributes to aging related pathologies and carcinogenesis (Artandi and DePinho, 2010; Murnane, 2010). Human telomeres comprise of 5-15 kb of duplex TTAGGG repeats followed by a 50-500 bp single stranded G-rich 3' overhang, and are bound by the 6-member shelterin protein complex (Palm and de Lange, 2008). The 3' overhang folds back and invades the double stranded telomeric DNA to form a displacement loop (D-loop) that stabilizes the lasso-like t-loop end structure (Griffith et al., 1999). Telomere shortening results from the inability to completely replicate chromosome ends and from defects in DNA repair or replication at telomeres (Palm and de Lange, 2008). The loss of telomeric repeats is compensated for by telomerase, but most human somatic cells lack telomerase activity (Bodnar et al., 1998; Greider and Blackburn, 1987). Telomere dysfunction results when telomeric repeats reach a critical length or when shelterin proteins are defective (Hemann et al., 2001b; van Steensel et al., 1998). This activates a DNA damage response resulting in apoptosis or senescence, or chromosome end fusions and aberrations if checkpoint proteins such as p53 are defective (Palm and de Lange, 2008).

Accumulating evidence indicates that WRN protein is required for proper telomere replication. WRN defective cells show an increase in stochastic loss of telomeres that were specifically replicated from the G-rich lagging strand template, and showed elevated sister chromatid exchanges in telomeric regions (Anderson et al., 2008; Arnoult et al., 2009; Crabbe et al., 2004; Hagelstrom et al., 2010; Laud et al., 2005). The premature senescence, genomic instability and stochastic telomere loss phenotypes of WS cells can be rescued by expressing either WRN protein or telomerase, indicating that telomerase can compensate for WRN roles at telomeric ends (Anderson et al., 2008). We and others showed that WRN protein is required to prevent telomere loss resulting from endogenous and environmental effectors of DNA replication stress (Anderson et al., 2008; Liu et al., 2010). However, the precise role of WRN in preserving telomere replication and in preventing telomere loss is not well defined.

Recent data indicate telomeres resemble common fragile sites that are prone to breakage during DNA replication stress. Cells lacking telomeric TRF1, ATR DNA replication stress checkpoint protein, or exposed to the DNA polymerase inhibitor aphidicolin exhibit aberrant chromosomal telomeres thought to arise from fragmentation (McNees et al., 2010; Sfeir et al., 2009). Several obstacles to telomere replication have been described which offer various scenarios for replication stalling and WRN roles at telomeres. 1) We showed previously that WRN efficiently disrupts the telomeric end D-loop that may impede replication fork progression (Opresko et al., 2004b; Opresko et al., 2009; Sowd et al., 2009). 2) Single stranded DNA at the telomeric 3' overhang, the displaced strand of the D-loop, and in Okazaki fragments during telomeric lagging strand DNA synthesis are G-rich and can potentially fold into G4 DNA (Maizels, 2006). G4 DNA are non B-DNA structures that block DNA synthesis by polymerases and are unwound by WRN and BLM helicases *in vitro* (Kamath-Loeb et al., 2001; Mohaghegh et

al., 2001). The findings that BLM helicase suppresses the fragile telomere phenotype (Sfeir et al., 2009), while a G4 DNA stabilizing ligand increased telomere fragility, strengthen the model that G4 folds in the G-rich telomeric strand can interfere with telomeric DNA replication (Rizzo et al., 2009b; Tahara et al., 2006). 3) Potential for the misalignment of DNA primers and templates in repetitive sequences may provoke DNA replication fork stalling (Shishkin et al., 2009). 4) Telomeres are transcribed into non-coding RNAs which can impede telomere replication (Luke and Lingner, 2009). Thus, obstacles that are either specific to telomeric chromosome ends or inherent in the telomeric repeats themselves may impede DNA replication, and failed telomere replication can cause telomere loss (Gilson and Geli, 2007).

In this study we asked if WRN modulates replication of telomeric sequence independently from telomeric chromosome end structures and telomerase extension. Our previous study showed that human TTAGGG repeats are not mutagenic in human cells despite their ability to form G4 DNA, unlike other sequences that can form alternate non-B DNA structures (Damerla et al., 2010; Wang et al., 2006b; Wang and Vasquez, 2004). We hypothesized that WRN deficiency would cause increased replication fork stalling and collapse in telomeric repeats due to G4 DNA, leading to deleterious deletions. A vector mutagenesis assay was used to examine the impact of factors inherent to telomeric sequence independently from chromosomal telomeric end factors. Here we show that insertion of human telomeric repeats did not alter the mutation frequency of the vector *supF* reporter gene upon replication in normal human cells. In contrast, WRN depletion elevated the *supF* mutation frequency for both control and telomeric vectors, but the increase was significantly higher for the telomeric vector and was attributed primarily to a dramatic increase in sequence deletion events. Our results establish that WRN is required to accurately replicate human telomeric sequences, and provide a

mechanism to explain the loss of telomeres replicated from the template TTAGGG lagging strand in WRN deficient cells (Crabbe et al., 2004).

3.2 MATERIALS AND METHODS

3.2.1 Cell culture and reagents.

Human U2OS osteosarcoma cells were cultured in Dulbecco's Modified Eagle Media (DMEM) supplemented with 10% fetal bovine serum, and Penicillin/Streptomycin (Invitrogen, Carlsbad, CA) and grown at 5% O_2, 5% CO_2 and 37°C. U2OS cells proficient and deficient for WRN were generated by stably expressing a scrambled shRNA (shCTRL) and shRNA against WRN (shWRN) respectively, as described in (Liu et al., 2009). Cells were cultured in the presence of hygromycin (200 μg/ml) (EMD Chemicals Inc., Gibbstown, NJ) to maintain selective pressure for shRNA expression. All the restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Chloramphenicol (chlor), 5-bromo-4-chloro-3-indolyl β-d-galactoside (X-Gal) and kanamycin (kan) were purchased from Sigma Chemical Co. (St. Louis, MO) while isopropyl β-d-thiogalactoside (IPTG) was obtained from Fisher Bioreagents (Fairlawn, NJ). All oligonucleotides were from Integrated DNA Technologies Inc (Coralville, IA) (Table 3-1).

3.2.2 Construction of *supF* reporter gene shuttle vectors containing telomeric repeats.

The shuttle vector (SV) pSP189 harboring the *supF* mutagenic reporter gene was kindly provided by Dr. Karen Vasquez (U. of Texas) (Parris and Seidman, 1992). To optimize vector selection in bacteria the chloramphenicol acetyltransferase (*cat*) gene was cloned into the *BamH*I site of pSP189. The sequence (TTAGGG)6 was inserted such that the last repeat was present in the *sup*F D stem of the suppressor tRNA molecule (Figure 3-1) to generate the telomeric SV. For this the following mutations were introduced into the *sup*F gene: G100T, G102A, C176T, and C178A. To generate the scrambled control SV the sequence [CTTGCTTACGACTTACCGTTCATCGTTGA] was inserted 3' to the *supF* gene and the following residues of the D-stem were mutated: T100A, G102A, C176T, and C178A. SV construction was as described previously (Damerla et al., 2010). Briefly, oligonucleotides (Table 3-1) containing the telomeric or scrambled control sequences and the *supF* gene were annealed and cloned into the *Xho*I and *Eag*I restriction sites of the pSP189 vector. Ligation reactions were transformed into the MBM7070 indicator *E. coli* strain (provided by Dr. Karen Vasquez, U. of Texas) and plated on LB agar plates supplemented with chlor (50 μg/ml), X-gal (0.12 mg/ml) and IPTG (0.3 mg/ml). Plasmids were isolated from wild type blue colonies and sequenced at (ACGT, Wheeling, IL) to confirm correct construction.

Table 3-1. Oligonucleotides used for construction of telomeric and scrambled shuttle vectors.

Vector Name	Sequence
Telomeric	5'- ccgctcgagc tgtTTAGGGT TAGGGTTAGG GTTAGGGTTA GGGTTAGGGt tcccgagcgg ccaaagggag cagactctaa atctgccgtc atcgacttcg aaggttcgaa tccttcccct aacaccac ggccgtgca -3'
Scrambled control	5'- ccgctcgagc tgtCTTGCTT ACGACTTACC GTTCATCGTT GAGATAGGGt tcccgagcgg ccaaagggag cagactctaa atctgccgtc atcgacttcg aaggttcgaa tccttcccct atcaccacgg ccgtgca -3'

Insert fragments were constructed by annealing to the complimentary strand for each (not shown).

3.2.3 *SupF* mutational analysis of telomeric and control shuttle vectors.

SupF mutant frequencies of the various SVs were determined after replication in human U2OS cells as described previously (Wang et al., 2006). For transfections, 2 μg SV was mixed with $2x10^6$ U2OS cells in 100 μl of nucleofector kit V solutions and electroporated using the Amaxa Nucleofection system (Lonza). After 48 hours of growth in supplemented DMEM media the SVs were isolated using the PureLink Quick Plasmid Miniprep kit (Invitrogen). Vectors were digested with *Dpn*I enzyme to remove vectors that were not replicated in U2OS cells, and were then transfected into MBM7070 bacteria, incubated for 45 min at 37^{o}C, and plated on selective media containing 50 μg/ml chlor, 0.12 mg/ml X-gal and 0.3 mg/ml IPTG to screen for *supF* mutants by blue-white screening (Wang et al., 2006). All mutant white colonies were confirmed by re-plating on selective media. The *supF* mutant frequency was determined as the number of mutant white colonies divided by the total number of chlor resistant colonies. To test for the recovery efficiency of the SVs after replication in human cells, we co-transfected $1x10^6$ U2OS cells with 1μg of SV harboring the chlor resistant gene and 1μg of pEYFP-C1 (Clontech)

harboring the kan resistant gene, which served as an internal standard to control for differences in transfection efficiency. Both vectors contained an SV40 origin of replication. The vectors were electroporated into human cells as described above, isolated 48h after culturing, and transfected into MBM7070 *E. coli*. The recovery efficiency was calculated as the ratio of chlor resistant bacteria colonies to kan resistant colonies.

3.2.4 Generation of *supF* telomeric and scrambled control shuttle vector mutation spectra.

To obtain independent mutants for analysis, after electroporation of replicated shuttle vectors into MBM7070 bacteria, the culture was placed on ice and aliquoted into multiple tubes containing 200 µl media. After the 45 min recovery period at 37^{o}C, each aliquot was plated on selective media, and one white mutant was isolated per aliquot for plasmid purification and DNA sequencing. This ensures that any mutational hotspots were not due to division of bacteria harboring *supF* mutant vectors during the 45 min recovery. To initially screen for mutants with large deletions or rearrangements the plasmids were digested with *Xho*I and *Eag*I restriction enzymes (Figure 3-1). Sequence changes and mutations within the promoter and coding region of the *supF* gene, as well as the telomeric inserts, were determined by dideoxy DNA sequencing (ACGT Inc, Wheeling, IL). DNA sequence analysis was done using Align-X software of Vector NTI Advance (Invitrogen Corporation). The primer used for DNA sequencing annealed at the 5' end of the *cat* gene to prime sequencing through the *supF* gene. Mutant SVs from the same transfection into human cells that exhibited the identical deletion or rearrangement could have resulted from either that mutation occurring independently in different U2OS cells in culture, or from an early mutagenic event that was replicated multiple times during the 48 hr replication

period. To maintain rigor and consistency, these mutants were considered siblings and were scored once. The same mutations occurring in different clones were considered independent.

3.2.5 Statistics

Means and standard deviations of mutant frequencies were calculated and statistical significance was determined by Student's t-test using Microsoft Excel. Fisher's exact test was used to calculate statistical significance between proportions of deletions/rearrangements using Graphpad. The statistically significant level was set as $p< 0.05$.

3.2.6 AFM imaging and analysis

SupF telomeric SV was diluted to a final concentration of 0.5 or 1 μg/ml in a buffer containing 25 mM Hepes pH7.5, 25 mM Na-acetate, and 10 mM $MgCl_2$. Buffers were pre-heated at 65^{o}C for 15-30 min to dissolve small salt particles that may have accumulated during storage. SV molecules were pre-incubated at 37^{o}C for 15 min immediately prior to deposition onto a freshly cleaved mica disk (SPI Supply, West Chester, PA). All samples were washed with MilliQ water and dried under a stream of nitrogen gas. Images were collected using a MultiModeV microscope (Veeco Instruments, Plainview, NY) using E scanners in tapping mode. Pointprobe® plus noncontact/tapping mode silicon probes (PPP-NCL, Agilent) with spring constants of ~50 N/m and resonance frequencies of ~190 kHz were used. Images were captured at a scan size of 2 μm × 2 μm, a scan rate of 3 Hz, a target amplitude of 0.30 to 0.35 V and a resolution of 512 × 512 pixels. Structures on SV molecules with a peak height over 1 nm were considered G4 DNA according to previous work (Wang et al., 2011).

3.3 RESULTS

3.3.1 Development of telomeric *supF* mutagenesis assay.

To determine whether WRN protein is required for replication of telomeric repeat sequences in human cells, we constructed a shuttle vector containing six telomeric repeats upstream to and within the mutation reporter *supF* gene. This approach offers several critical advantages. 1) Deletions of telomeric DNA in the SV will not affect cell survival. However, loss of telomeric DNA at chromosome ends causes apoptosis or senescence (Palm and de Lange, 2008), and short telomeres and defects in repair proteins (i.e. WRN) may synergistically decrease cell survival. 2) Episomal vectors allow for direct comparison in different genetic backgrounds. Since the vector is not integrated in the genome there are no confounding effects of different integration points on mutagenesis. 3) While the vector telomeric DNA may form G4 DNA during replication (Duquette et al., 2004), it will not form complex telomeric t/D-loops and lacks the substrate for telomerase. Thus, the addition of telomerase inhibitors, as done previously (Crabbe et al., 2004), is not required to unmask WRN roles in preserving telomeric sequence. By using this approach we can examine WRN roles in modulating factors inherent in telomeric sequence, independently from potential confounding effects of complex end structures, telomerase activity and telomere transcription. The new *supF* vector design integrates the last repeat of the inserted $[TTAGGG]_6$ sequence into the D stem of the *supF* tRNA, and the telomeric sequence and *supF* gene are flanked by *Xho*I and *Eag*I restriction sites (Figure 3-1). The control vector contains a 36 bp scrambled sequence of identical nucleotide composition as the $[TTAGGG]_6$ sequence (Table 3-1). These vectors exhibit very low background mutant frequencies in *E. coli*. For the scrambled control vector no mutants were detected in 14,694 colonies, yielding an estimated mutant

frequency of $<6.81x10^{-5}$. The telomeric vector provided a background mutant frequency of $3.91X10^{-5}$ (1 mutant/25549 total colonies) indicating that both vectors are highly stable upon replication in the indicator *E. coli* strain MBM7070.

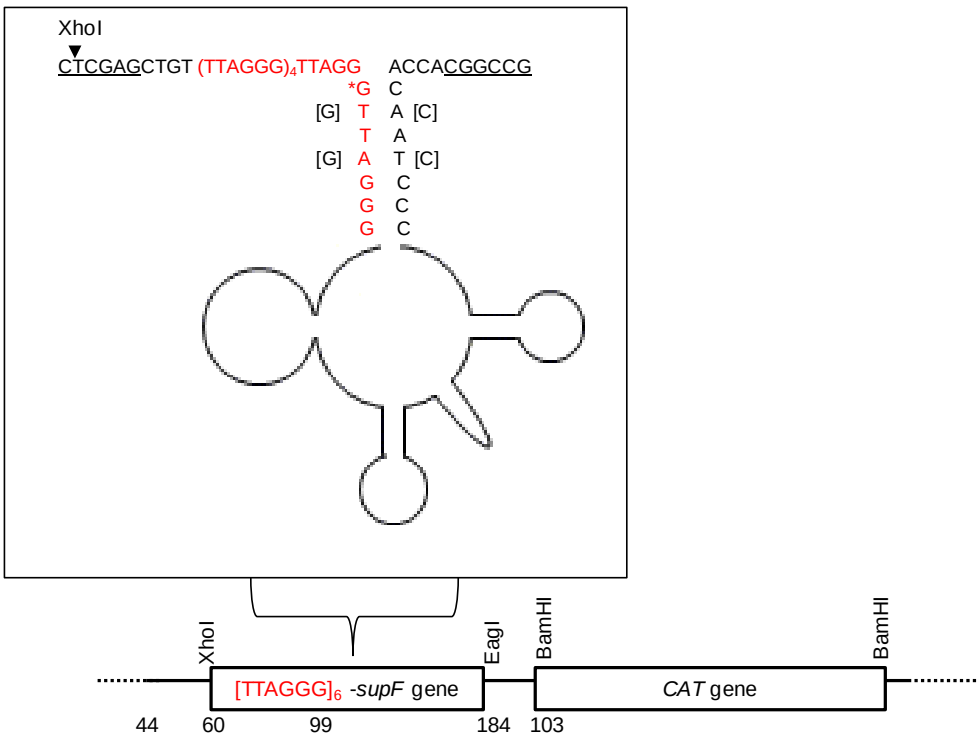


Figure 3-1. Structure of vectors containing telomeric DNA.

Six telomeric repeats were inserted upstream and within the *supF* gene of the pSP189 shuttle vector as shown. Four nucleotides of the *supF* gene were mutated as indicated so the last repeat was located in the stem of the *supF* tRNA hairpin structure. A scrambled telomeric sequence of identical size and nucleotide context was introduced to generate the control shuttle vector (see Table 3-1). The chloramphenicol aceytltransferase gene was introduced at the *BamH*I site. The 5'end of the mature tRNA (*G) is marked as position 99 following the traditional nomenclature.

3.3.2 WRN depletion increases the mutant frequency of the telomeric vector.

The control and telomeric SVs were transfected into human U2OS cells stably expressing either a control shRNA (shCTRL) or an shRNA targeted against WRN (shWRN) (Liu et al., 2009) (Figure 3-2B). WRN expression was decreased to 24% of the control cells (Figure 3-2B). SVs were replicated for 48 hrs, isolated and subjected to *Dpn*I digestion to select for replicated vectors, which were then transformed into the *E.coli* reporter strain and subjected to blue/white screening for *supF* mutants. The mean mutant frequencies for the scrambled and telomeric vectors after replication in shCTRL U2OS cells were very similar at $5.2x10^{-4}$ and $5.6x10^{-4}$, respectively (Figure 3-2). Thus, human telomeric repeats are not mutagenic and are stably replicated in normal human cells in agreement with our previous results from a different shuttle vector mutagenesis system (Damerla et al., 2010). In contrast, replication of the telomeric vector in WRN deficient U2OS cells yielded a significantly higher (4.4-fold) mutant frequency compared to the scrambled vector ($150\ x10^{-4}$ vs. $34\ x10^{-4}$, p-value = 0.047). Thus, vectors with telomeric repeats are more mutagenic in the absence of WRN protein. Consistent with this, WRN depletion increased the control vector mutant frequency 6.5-fold, but more dramatically elevated the telomeric vector mutant frequency 27-fold (p-value = 0.0064), compared to shCTRL cells (Figure 3-2). The vectors with telomeric repeats exhibited an increased dependence on WRN protein to suppress mutagenic events during replication, compared to non-telomeric vectors.

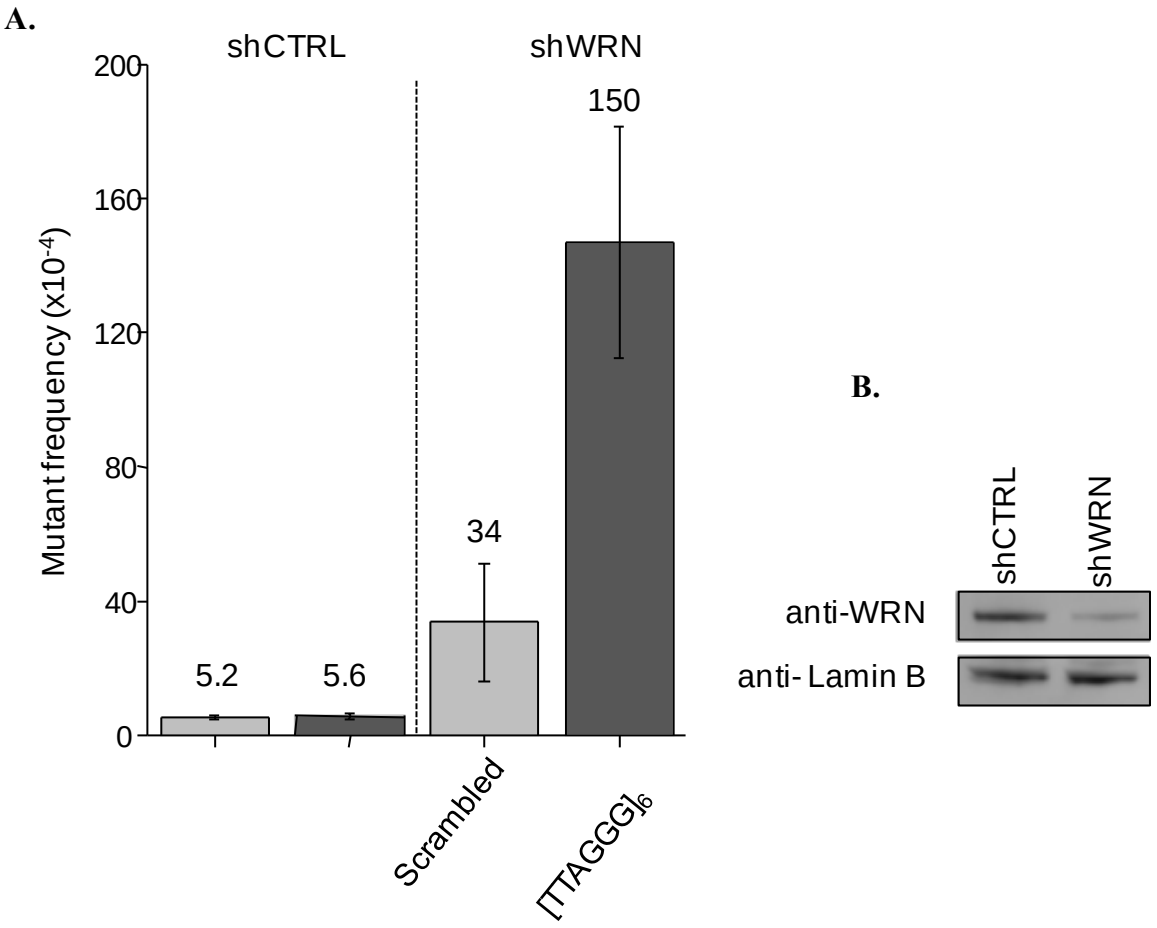


Figure 3-2. WRN depletion significantly increases the supF mutant frequencies for the telomeric vectors.

(A.) *supF* mutant frequencies for scrambled control and telomeric vectors after replication in WRN proficient and deficient cells. Shuttle vectors were replicated in U2OS cells expressing control or WRN shRNAs for 48 h, isolated and transformed into the reporter *E.coli* strain to screen for *supF* mutants. Values are the mean and error bars indicate SEM from at least three independent experiments. (B.) Western Blot shows WRN protein levels in U20S cells stably expressing control or WRN shRNAs. Quantification revealed a 76% knockdown in WRN expression.

3.3.3 Reduced recovery of vectors from WRN depleted U2OS cells.

Next we examined whether insertion of the telomeric repeats altered the efficiency of recovering replicated vectors from the control and WRN depleted human cells. For this the chlor resistant scrambled or telomeric vectors were co-transfected with the kan resistant vector pEYFP-C1, which served as an internal standard, and were isolated 48h after transfection. The mean recovery efficiency (see Materials and Methods for calculation) of the replicated scrambled and telomeric vectors from the shCTRL U2OS cells was 53 for both, indicating the telomeric repeats did not interfere with SV replication or recovery in WRN proficient cells (Figure 3-3). The mean recovery efficiency for the scrambled vector from shWRN U2OS cells was 27, which was slightly higher than for the telomeric vector at 20. WRN depletion resulted in a near 2-fold decrease in recovery for both the scrambled and telomeric vectors compared to WRN proficient cells. This suggests WRN is required for efficient SV replication and/or to prevent large deletions that impact the chlor resistant gene or the mammalian/bacterial origins of replication on the vector.

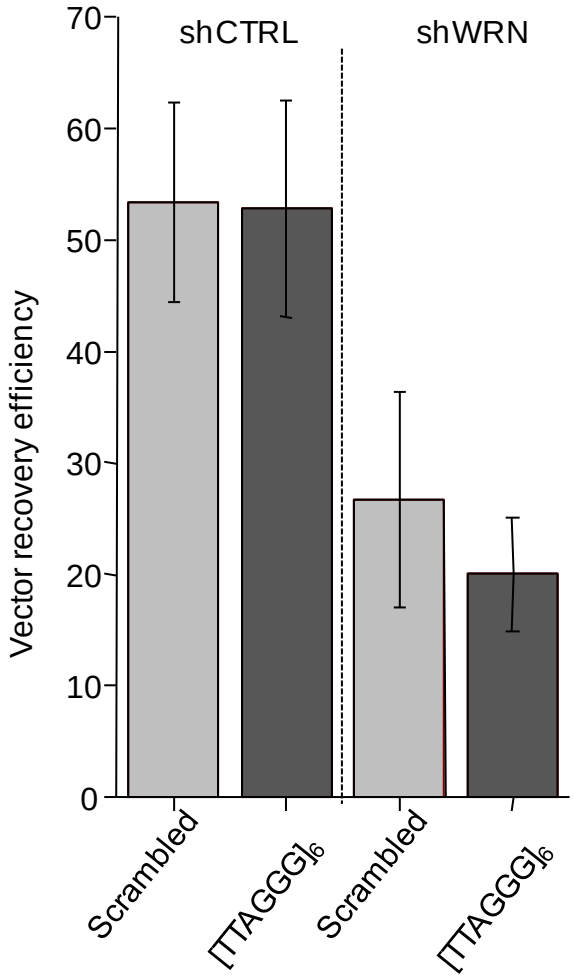


Figure 3-3. The pSP189 vector recovery efficiency is decreased in WRN deficient cells.

The graph shows vector recovery efficiencies for the scrambled control and telomeric vectors after replication in control and shWRN U2OS cells. The kan resistant vector pEYFP-C1 was co-transfected with scrambled or telomeric pSP189 vectors into U2OS cells. Vectors were isolated after 48 h, electroporated into *E. coli* and plated on selective media. The recovery efficiency was calculated as the ratio of chlor to kan resistant colonies. Values are the mean and error bars indicate SEM from two to three independent experiments.

3.3.4 SV with telomeric repeats exhibit a dramatic elevation in deletion events in the absence of WRN.

To determine the mechanisms for the significant increase in *supF* mutant frequency for the telomeric vectors upon WRN depletion, we investigated the mutations arising from vector replication in the human cells. SVs replicated in U2OS cells were transformed into the reporter *E. coli* strain. Plasmids were isolated from independent white mutant bacteria colonies and subjected to double digestion with *Xho*I and *Eag*I to determine whether the telomeric or scrambled sequences and the *supF* gene were present (see Materials and Methods and Figure 3-1). The majority mutant telomeric vectors from the WRN deficient cells exhibited abnormal digest patterns (Figure 3-4) indicating that the *supF* gene was likely deleted. Therefore, vectors were sequenced using a primer that binds to the 5'end of the *cat* gene rather than previously used primers (Wang and Vasquez, 2004), since all mutant colonies were chlor resistant. Mutations sorted into two categories 1) point mutations (PM) within the *supF* gene or 2) deletions and unknown rearrangements (del/rrg). All the PM detected were base substitutions with an obvious bias at G•C base pairs (94%), compare to A•T base pairs (3%) arising primarily at GA•CT dinucleotide site (Figure 3-5), consistent with previous reports for these cells (Bacolla et al., 2011). The telomeric repeats and WRN depletion did not alter this bias. The majority of del/rrg were well defined deletions greater than 500 bp and approximately 30-40% exhibited 1-3 bp of microhomology (Tables 3-3 to 3-6, see after discussion). The telomeric repeats and WRN depletion did not alter the deletion sizes in an obvious way.

Insertion of the telomeric repeats did not significantly alter the overall mutant frequency, but rather altered the types of mutations generated after SV replication in shCTRL U2OS cells (Table 3-2). While the PM and del/rrg frequencies were nearly identical for the scrambled vector

($2.7x10^{-4}$ and $2.5x10^{-4}$, respectively), the PM frequency was 2.2-fold higher than the del/rrg frequency ($3.9x10^{-4}$ and $1.7x10^{-4}$, respectively) for the telomeric vector. Consistent with this, the proportion of del/rrg as a function of total mutations was significantly lower for the telomeric vector (0.70) compared to the scrambled vector (0.52) ($p < 0.01$, Fisher's exact test). This agrees well with our previous result using the HSV-*tk* shuttle vector that the insertion of human telomeric repeats stabilized the vectors and suppressed the occurrence of larger deletions and rearrangements (Damerla et al., 2010).

WRN is required to suppress both del/rrg and PM arising in the SVs upon replication in human cells. WRN depletion increased the PM frequency 2.5-fold for the scrambled SV to $7x10^{-4}$ but yielded a greater increase of 8.5-fold for the telomeric SV to $33x10^{-4}$, compared to the shCTRL cells (Table 3-2). However, WRN deficiency exhibited an even greater effect on del/rrg events for both vectors. The del/rrg frequency for the scrambled vector was moderately increased 11-fold to $27x10^{-4}$, but was dramatically elevated 70-fold to $120x10^{-4}$ for the telomeric SV, compared to shCTRL cells. Next we compared the effect of WRN depletion on the ratio of the proportion of PM to del/rrg. This ratio was similar in shWRN cells for both vectors, indicating WRN depletion shifted the bias toward del/rrg (Table 3-2). However, the shWRN-mediated reduction in the ratio of PM to del/rrg was greater for the telomeric vector (8.5-fold, 2.3 shCTRL vs. 0.27 shWRN), compared to the scrambled vector (4.1-fold, 1.1 shCTRL vs. 0.27 shWRN). Several of the del/rrg events exhibited endpoints within the *supF* gene or telomeric repeats or near the repeats, although endpoints also existed within the scrambled sequence of the control vector (Figure 3-6). In summary, the data indicate that WRN has a greater role in suppressing deletions and rearrangements, compared to point mutations, and that this role is more critical in templates with telomeric repeats.

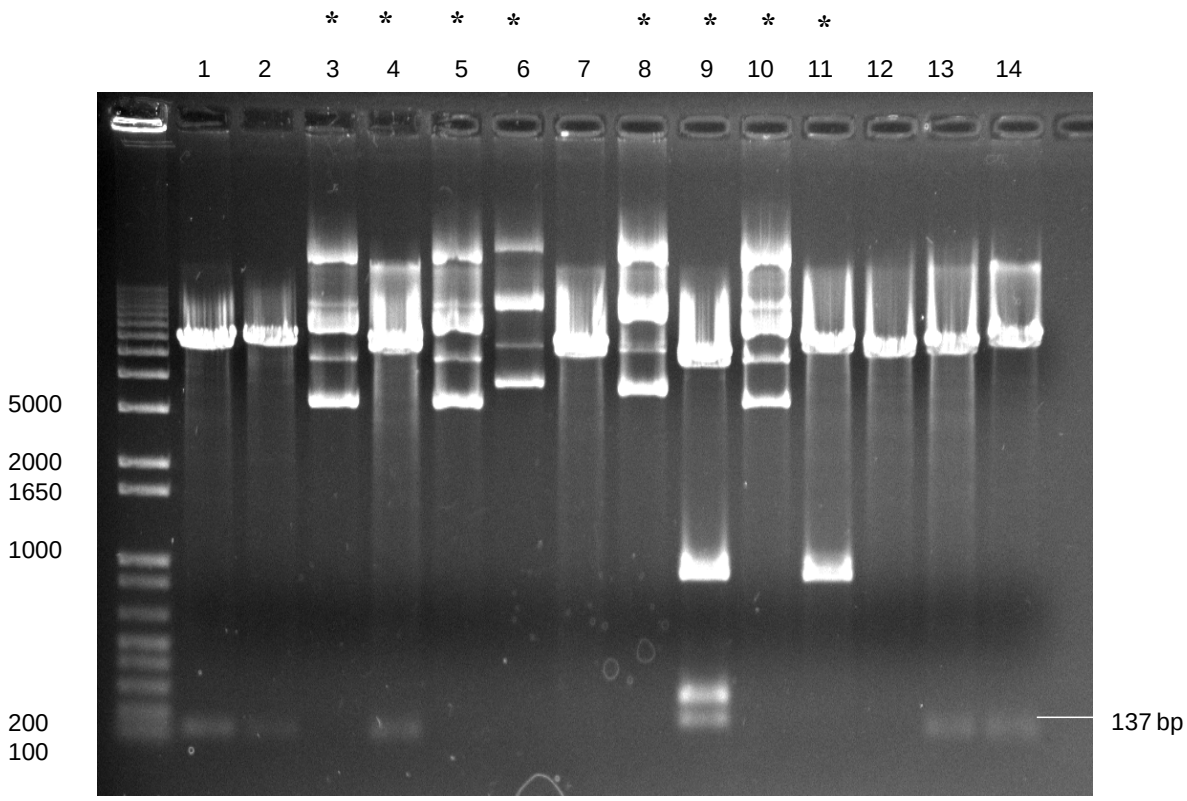


Figure 3-4. WRN depletion induces large deletions within the telomeric SV.

Plasmids isolated from mutant colonies were digested with *Xho*I and *Eag*I and separated on a 1% agarose gel. The 137 bp fragment is the (TTAGGG)6 insert and *supF* gene and the 6000 bp fragment represents the rest of the vector (see lane 1 for example). The telomeric vectors replicated in shWRN U2OS cells yielded a large number of mutants with deletions and rearrangements as indicated by the patterns of fragments shown in the lanes marked with an asterisk.

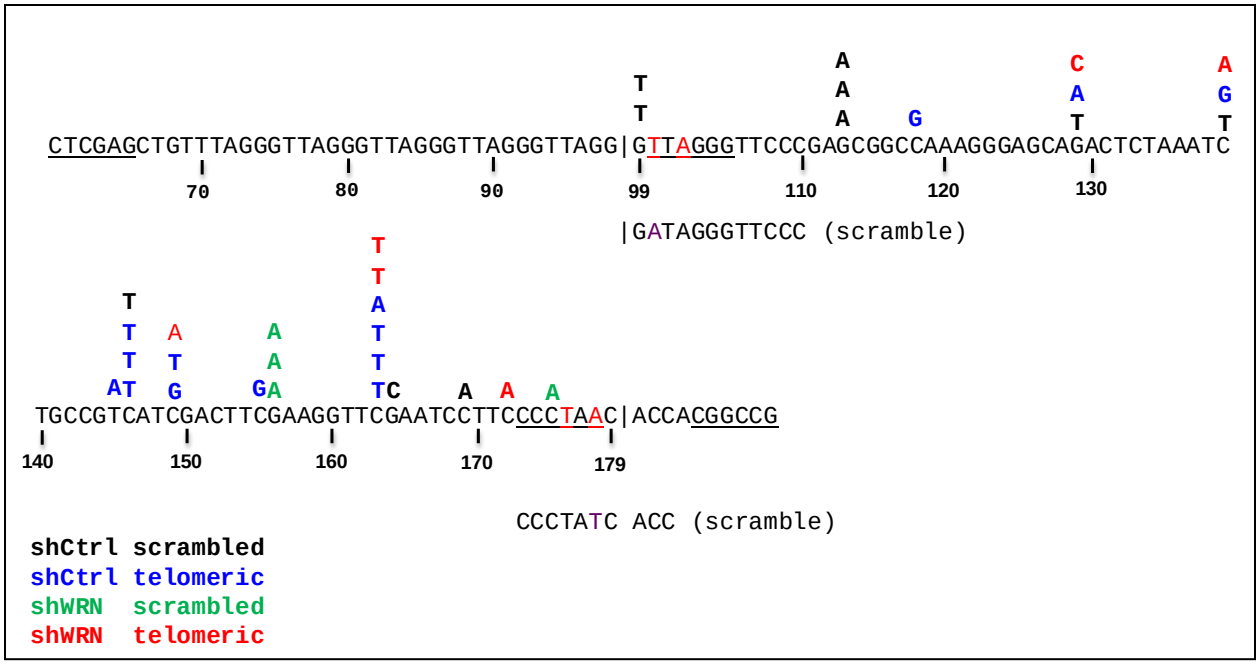


Figure 3-5. SupF inactivating base substitutions.

The *supF* gene containing the telomeric insert with borders marked by "|", is shown 5' to 3'. Regions of the *supF* gene that are different for the scrambled vector are depicted below the corresponding nucleotide positions (scramble). Letters above the sequence indicate base substitutions obtained from all sequenced *supF* vectors from various transfections into U2OS cells (depicted using individual colors).

Table 3-2. Sequenced mutations arising in the supF SV after replication in human cells.

Mutation Class		scrambled shCTRL	$[TTAGGG]_6$ shCTRL	scrambled shWRN	$[TTAGGG]_6$ shWRN
Overall mean mutant frequency		$5.2x10^{-4}$	$5.6x10^{-4}$	$34x10^{-4}$	$150x10^{-4}$
Point Mutations					
	mean mutation frequency	$2.7x10^{-4}$	$3.9x10^{-4}$	$7x10^{-4}$	$33x10^{-4}$
	number	11	16	4	5
	proportion of total	0.52	0.70	0.21	0.21
Deletions and Rearrangements[a]					
	mean mutation frequency	$2.5x10^{-4}$	$1.7x10^{-4}$	$27x10^{-4}$	$120x10^{-4}$
	number of mutations	10	7[b]	15[c]	19[d]
	proportion of total	0.48	0.30	0.79	0.79
Total mutations (mutants)		21 (21)	23 (17)	19 (19)	24 (21)
Ratio proportion of PM: del/rrg		1.1	2.3	0.27	0.27

[a]Identical deletions or rearrangements from the same SV transfection experiment in human cells were scored once to avoid the inclusion of potential SV siblings (progeny from a single mutagenic event).

[b]Includes a mutation in which one (AGGGTT) repeat was inserted in the $[TTAGGG]_6$ sequence

[c]Includes one unknown rearrangement and 14 defined deletions.

[d]Includes four unknown rearrangements, 14 defined deletions, and a mutation in which one (AGGGTT) repeat was deleted from the $[TTAGGG]_6$ sequence.

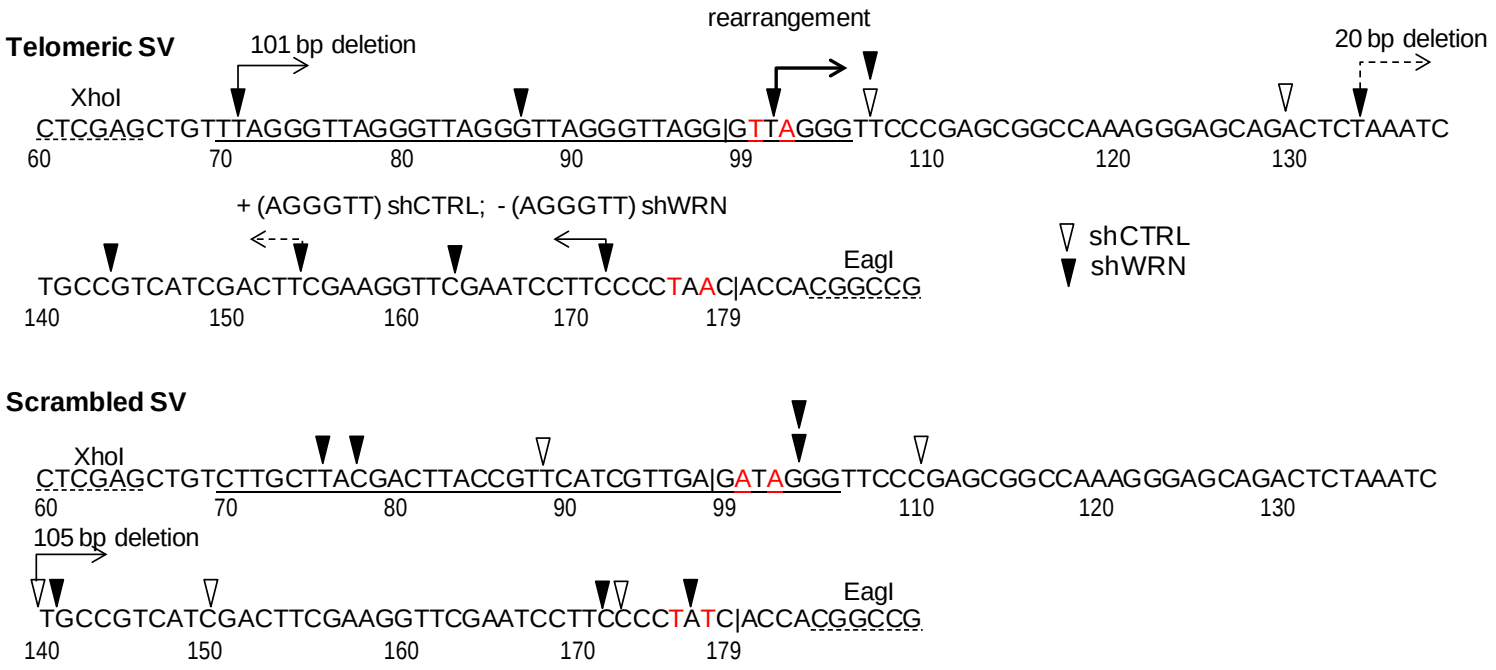


Figure 3-6. Deletions and rearrangements with endpoints in the *supF* gene and telomeric repeats.

The *supF* gene with borders marked by " | ", telomeric or scrambled sequence insert and the *Xho*I and *Eag*I restriction sites are shown 5' to 3'. Bases marked in red are alterations in the tRNA D stem. End points of deletions or rearrangements are marked by an open arrow head (shCTRL) cells or a solid arrow head (shWRN) cells. Deleted or altered sequences are all 3' to arrowheads unless otherwise noted with an arrow. The 5'end of the mature tRNA (*G) is marked as position 99 following the traditional nomenclature.

3.3.5 Duplex supercoiled telomeric SV do not form G4 DNA prior to replication in human cells.

Next we wished to determine whether the deletions induced in the telomeric SV were associated with replication. It is well established that a high-energy state of supercoiled DNA can lead to formations of non-canonical DNA structures, such as cruciforms, Z-DNA, and H-DNA (Mirkin, 2008; Shlyakhtenko et al., 1998). Furthermore, transitions to these structures simultaneously relax the DNA. Preformed alternate structures in supercoiled plasmids can induce deletions independently of DNA replication (Wang et al., 2006). To investigate whether or not telomeric sequences formed G4 structures in the supercoiled SVs prior to transfection into human cells, we visualized telomeric SVs using atomic force microscopy (AFM). Previous work showed AFM can detect the formation of cruciforms and H-DNA in supercoiled plasmids (Lyubchenko, 2004; Shlyakhtenko et al., 1998), and we and others showed G4 DNA generates structures with distinct peaks at heights between 1 and 2 nm in AFM images (Neaves et al., 2009). A representative AFM image of the telomeric SV is shown in Figure 3-7. All circular DNA molecules observed (n=80 molecules) were in a plectonemic shape with formations of large loops between very tightly twisted segments. These images are consistent with previous AFM images of supercoiled DNA under similar sample preparation conditions (Lyubchenko, 2004; Wang et al., 2011). However, no structures with peaks greater than 1 nm in height were observed on the telomeric SVs (n=80 molecules) except at regions where strands of duplex DNA overlapped (Figure 3-7). These regions of overlapping DNA were distinctly different from define peaks formed by a single G4 structure that could fold from six telomeric repeats (Wang et al., 2011). In summary,

contrary to Z-DNA and H-DNA, the data indicate that G4 structures do no form on duplex DNA by introducing supercoiling.

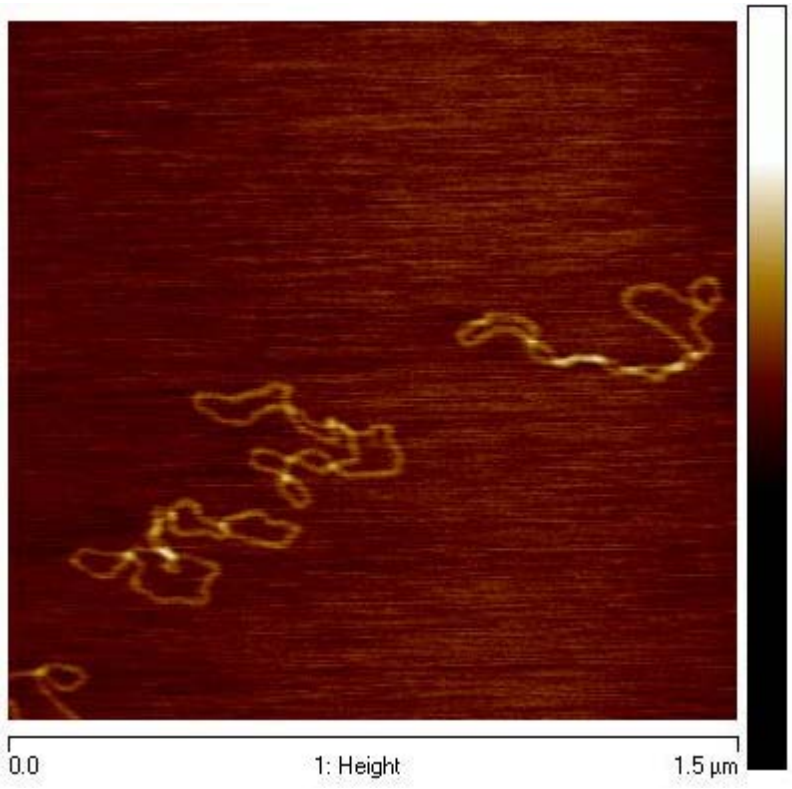


Figure 3-7. Supercoiled telomeric vectors do not form G4 DNA structures.

No G4 structures were formed on supercoiled telomeric SV molecules (0.5 µg/ml) .The image is 1.5 µm × 1.5 µm at 2 nm Z-scale. The light brown regions represent regions where strands of duplex DNA overlap.

3.4 DISCUSSION

Previous studies indicated that WRN protein is required to prevent the loss of chromosomal telomeres replicated from the G-rich lagging strand (Arnoult et al., 2009; Crabbe et al., 2004) which can fold into G4 DNA structures that block DNA synthesis, that are unwound by WRN (Kamath-Loeb et al., 2001a; Mohaghegh et al., 2001). This led to the model that WRN could facilitate telomere replication by resolving G4 DNA on the lagging strand which are capable of blocking replication fork progression. However, telomeres present several challenges to replication that are specific to telomeric chromosome ends or inherent in the telomeric repeats. Here we dissected WRN roles in the replication of telomeric sequences using the *supF* shuttle vector mutagenesis assay to test factors inherent to the repeats, such as G4 DNA, independently from chromosome end factors. Consistent with our previous results (Damerla et al., 2010), we demonstrate that 1) SVs with $[TTAGGG]_6$ sequences are stably replicated in human cells and 2) insertion of the repeats suppresses the frequency of large deletions despite G4 folding potential (Figure 3-2). WRN depletion increased the *supF* mutant frequency for both the telomeric and non-telomeric scrambled SVs, compared to the control cells, but this increase was much greater (27 fold) for the telomeric SVs (Figure 3-2.). The higher *supF* SV mutant frequencies in WRN deficient cells was primarily due to an increase in large sequence deletions and rearrangement events for both vectors. However, WRN depletion caused a more dramatic increase in deletions and rearrangements arising within the telomeric SV (70-fold), compared to the scrambled SV (11-fold). Our results suggest that WRN prevents deleterious mutagenic events and deletions during replication, and that this role is particularly important in templates with telomeric sequence. This provides a mechanism for increased telomere loss in WS cells.

We observed a greater role for WRN in suppressing deletions and rearrangements compared to base substitutions in both telomeric and non telomeric SV (Table 3-2). This increase in deletion frequency is likely an underestimate since the recovery efficiency of SVs from shWRN U2OS cells was ~2 fold lower compared to shCTRL cells (Figure 3-3). There are two possibilities for reduced recovery: 1) WRN is required for SV replication or 2) WRN is needed to prevent large deletions that may compromise the SV chloramphenicol resistant gene, or replication origins. SV lacking these factors would not replicate in human cells or bacteria, or would not confer chloramphenicol resistance to the bacterial colonies in the *supF* mutant screen. We propose the second scenario is more likely for several reasons. First, the result that WRN deficient cells exhibited slightly reduced recovery of telomeric SV compared to scrambled SV suggests that WRN may be required to replicate SVs with G4 forming sequences. However, since there was also less recovery of the scrambled SV from WRN depleted cells compared to the control cells this indicates WRN may have a more general role in SV maintenance or recovery. Second, previous studies of mutation rates at the HPRT chromosomal locus from SV-40 transformed WS fibroblasts revealed an increase in large deletion events (Fukuchi et al., 1989). These studies also confirmed that SV-40 large T-antigen could not compensate for WRN roles in suppressing deletions, consistent with our results using a SV that requires SV-40 large T-antigen for replication.

Our data indicate that the mechanism for SV deletions in WRN depleted cells resulted from replication induced DSBs. Replication fork stalling due to alternate structures or other factors can lead to fork collapse into DSBs (Murnane, 2010). Results from AFM imaging of the super coiled SV did not reveal any evidence for pre-folded alternate structures, including G4 DNA, in the telomeric SV (Figure 3-7). In contrast, the introduction of Z-DNA forming

sequences in the *supF* SV caused the formation of alternate structures prior to transfection into human cells, and induced large deletions that were not dependent on DNA replication presumably due to nuclease cleavage at Z-DNA structures causing DSB formation (Wang et al., 2006). In stark contrast to Z-DNA forming sequence, the introduction of telomeric G4 forming sequences in the *supF* SV actually suppressed the deletion frequency in control cells (Table 3-2). Therefore, we argue that the increase in deletions observed upon WRN depletion, particularly in telomeric SV, resulted from secondary structures or others factors that arose during replication, rather than DSBs arising from nucleolytic processing of preformed secondary DNA structures.

A recent study by Bacolla *et al* showed that WRN depletion resulted in only a 2-fold increase in mutant frequency for the control vector and for the SV with Z-DNA forming sequences (Bacolla et al., 2011). This is vastly different from the 27-fold increase in *supF* mutant frequencies obtained when the telomeric SVs were replicated in WRN deficient cells compared to control cells (Figure 3-2.). This indicates that while WRN can unwind both Z-DNA and G4 DNA, it only suppresses the mutagenicity of the telomeric G4 forming sequences but not the mutagenicity of Z-DNA forming sequences. This may be related to different mechanisms by which the telomeric G4 DNA and Z-DNA forming sequences induce DNA deletions. Z-DNA is pre-folded prior to replication and the deletions are not dependent on replication; unlike the telomeric G4 DNA. The mutation spectrum of the control and Z-DNA sequence SVs from previous work in both WRN proficient and deficient cells consisted of ~78% BS and only 22% deletions (Bacolla et al., 2011). The domination of BS was attributed to oxidative damage which was increased in WRN depleted U2OS cells (Bacolla et al., 2011). In contrast, we observed WRN depletion resulted in a much greater increase in deletions, compared to BS (Table 3-2). This difference may have been attributed to the fact that we cultured the U2OS cells at 5%

oxygen to minimize oxidative damage. In addition, the design of our *supF* telomeric SV allowed for the detection of larger deletions compared to the previous *supF* Z-DNA SV. Recovery of SV and subsequent mutational screening requires intact mammalian and bacterial origins along with an intact drug resistance gene for selection in bacteria. Deletions in the Z-DNA SV were confined to the amplicillin resistance gene 3' to the *supF* gene and the pBR327 replication origin 5' to the *supF* gene. Since we used a chloramphenicol resistant gene cloned at the 5'end of the *supF* gene for selection instead, we could detect deletions into the ampicillian resistance gene up to the SV 40 origin located further upstream. Thus, it was possible for our system of selection to recover a greater number of mutants with large deletions.

The increase in deletion frequencies in WRN depleted cells may have resulted from WRN roles in preventing DSBs during replication by resolving G4 DNA structures, or WRN function in promoting error-free repair of DSBs. While WRN is not required for NHEJ repair of DSBs, WRN is required to minimize nucleotide end resection to prevent large deletions during NHEJ repair (Sallmyr et al., 2008) (Oshima et al., 2002). However, our data showed that the mean deletion sizes in the SVs were similar with or without WRN depletion, and so was the percentage of deletion endpoints that exhibited microhomology at the junction sites. Rather, the shCTRL and shWRN cells only exhibited a significant difference in deletion frequency. Therefore, our data suggest WRN is not suppressing end resection at break points in our SV assay. Rather, we argue that the increase in deletion frequency is a result of WRN's role in preventing replication fork demise, consistent with previous studies that WRN promotes replication fork progression on chromosomes (Sidorova et al., 2008)

We observed in two independent mutagenesis SV assays that the introduction of human telomeric repeats does not raise the SV mutant frequency, despite G4 folding potential, and

suppresses the occurrence of larger deletions after replication in WRN proficient cells (Damerla et al., 2010). We hypothesized that the formation of G4 DNA during telomeric repeat replication may recruit specialized DNA helicases, including WRN and BLM that bind to G4 DNA with high affinity and promote replication fork stability. Therefore, G4 DNA formation at telomeric repeats may be beneficial in WRN proficient cells. However, our data indicate the beneficial effects of telomeric repeat sequences are lost when WRN is depleted. Although the proportion of deletions/rearrangements remained same for telomeric and non-telomeric SV in WRN depleted cells, the increase in deletion frequency was higher for the telomeric SV compared to control vector. This suggests that the absence of WRN causes an increase in fork stalling and collapse at telomeric repeats compared to the scrambled control sequence. WRN may suppress replication fork stalling in telomeric repeats by resolving G4 DNA structures, as observed *in vitro* (Kamath-Loeb et al., 2001). Alternatively, WRN may facilitate stimulate DNA polymerase δ and facilitates DNA replication through repeat sequences (Kamath-Loeb et al., 2001; Shah et al., 2010), and WRN is essential for preventing chromosomal breaks at fragile sites during replication (Pirzio et al., 2008). While we do see a number of deletion endpoints in the telomeric repeats, nearly all of the deletions encompass the full *supF* gene and likely resulted from extensive resection at broken ends prior to rejoining. Given that telomeres resemble fragile site sequences, it is possible that WRN is promoting DNA polymerase δ progression through the repetitive G-rich telomeric sequences through other mechanisms, in addition to G4 DNA resolution. Consistent with this, our work and previous studies showed that although SV-40 large T antigen could resolve G4 DNA it cannot compensate for WRN function in preventing large scale deletions in non-telomeric and telomeric DNA (Fukuchi et al., 1989).

Together, our results suggest that WRN is important for preventing deletions during the replication of telomere repeats. Our data indicate that WRN is required for suppressing deleterious mutagenic events that are induced by factors inherent to telomeric repeat sequences, independently from processes and chromatin structure at chromosomal ends. Our data support the model that WRN acts to resolve alternate structures, such as G4 DNA, that may form at telomeric repeats during replication in normal cells. In WRN deficient cells unresolved G4 DNA or other factors in repetitive sequences may induce fork stalling and collapse, leading to large deletions, which in a cellular context causes stochastic telomere loss, a characteristic of WRN deficient cells (Table 3-3).

3.5 ACKNOWLEDGEMENTS

We would like to acknowledge Dr. Hong Wang for providing us with AFM data, Dr. Fujun Liu for standardizing the *supF* mutagenesis assay in our laboratory and Kelly Knickelbein for performing the *supF* mutagenesis platings. I was involved in writing some parts of the manuscript and analyzing all the mutation data from the *supF* assay. We would also like to thank Dr. Karan Vasquez and Dr. Graham Wang for providing us with MBM7070 bacterial strain and technical support.

Table 3-3. End points of deletions and rearrangements for scrambled vector upon replication in shCTRL U2-OS cells.

Endpoints	n	size of deletion	size of µhomology	size of insertion	position on vector upstream from:
Scrambled vector replicated in shCTRL U2OS					
GCGCCGGGAtctttta--CTGGGGAAATCAATC--ttaaaaaAATTACG	5*	1227	0	15	302
CGCGGCCCTagaaaat--GACCCCTTTAGTTAG--aatttttTTAATGC					
AAGGTACctaaccaa--ttatcTATCACTT	1	1209	0	0	234
TTCCATGgattggtt--aatagATAGTGAA					
AAGGTACctaaccaa--ttatcTATCACTT	4*	1209	0	0	234
TTCCATGgattggtt--aatagATAGTGAA					
CAGTGTgttatc--tcatc**G**ACTTC	1	537	1	0	150
GTCACA**C**aatag--agtagCTGAAG					
GAAGcatttat--gaaatt**C**GGTA	1	162	1	0	195
CTTC**G**taaata--ctttaaGCCAT					
GGTATGgctgatta--aatccttccCCTAAC	1	1971	0	0	174
CCATACcgactaat--ttaggaaggGGATTG					
CCGTTcatcgt--tttaaggg**CA**CCAA	1	183	2	0	89
GGCAA**GT**agca--aaattcccGTGGTT					
TTCCTCTttcagaggtt--gggttcccGAGCGGCC	1	1070	0	0	111
AAGGAGAaagtctccaa--cccaagggCTCGCCGG					
TGGCTgtcttc--tttaaggg**G**CACCAA	1	3882	1	0	271
ACCGA**C**agaag--aaattcccCGTGGTT					
TAAATCtgccgt--ttattcAGGCG	1	105	0	0	245 (other at 140)
ATTTAGacggca--aataagTCCGC					

*potential siblings from replication in human cells

Table 3-4. End points of deletions and rearrangements for [TTAGGG]6 vector upon replication in shCTRL U2-OS cells.

Endpoints	n	size of deletion	size of µhomology	size of insertion	position on vector upstream from:
[TTAGGG]6 vector replicated in shCTRL U2OS					
CCGGctccag--gcaga**CT**CTAAA GGCC**GA**ggtc--cgtctGAGATTT	1	829	2	0	131o
CATTCTgagaat--ccccGCCCTG GTAAGActctta--ggggCGGGAC	1	627	0	0	304o
GTAGATAactacg--cccggatCCGAATTT CATCTATtgatgc--gggcctaCCGTTAAA	1	975	0	0	207o
AAACTTGGtctgac--ttagggttCCCGAG TTTGAACCagactg--aatcccaaGGGGCTC	1	971	0	0	108o
CCATAGttgcctg--tccgaa**TT**TCTGCC GGTATC**AA**cggac--aggcttAAAGACGG	2*	1016	2	0	212o
CCGCCTCAgaaggt--tccgTTATTA GGCGGAGTcttcca--aggcAATAAT	1	1254	0	0	234o

*potential siblings from replication in human cells
o Telomeric sequence is fully or partially deleted

Table 3-5. End points of deletions and rearrangements for scrambled vector upon replication in shWRN U2-OS cells.

Endpoint	n	size of deletion	size of μhomology	size of insertion	position on vector
Scrambled vector replicated in shWRN U2OS					upstream from:
GCGAGacccac--cactcaTCGCAGT	3*	1029	0	0	317
CGCTCtgggtg--gtgagtAGCGTCA					
AAAACctctac--agatagGGTTCCC	1	3877	0	0	104
TTTTGgagatg--tctatcCCAAGGG					
GGGCTTAccatct--cccctaTCACCAC	2*	826	0	0	178
CCCGAATggtaga--ggggatAGTGGTG					
TTCATCcataq--cttacGACTTAC	3*	898	0	0	79
AAGTAGgtatc--gaatgCTGAATG					
GCTGgggag--ccctgCTCGAG	1	1179	0	0	60
CGACcccctc--gggacGAGCTC					
TATTCcaggggaaa--ccttc**C**CCTAT	1	3440	1	0	173
ATAAG**G**tccccttt--ggaaggGGATA					
AAAAAAATgctttattt--gaaattcg**G**TACCCG	1	4111	1	0	197
TTTTTTTA**C**gaaataaa--ctttaagccATGGGC					
TTTAACacatt--gatagGGTTCCC	1	2013	0	0	104
AAATTGtgtaa--ctatcCCAAGGG					
CAGCGatctg--gccaCTCAT	1	1132	0	0	313
GTCGCtagac--cggtGAGTA					
CCTCCATCCagtct--T--cgtaGCAACCAGGC	1	880	0	1	252
GGAGGTAGGtcaga--A--gcatCGTTGGTCCG					
CCTCCATCCagtct--T--cgtaGCAACCAGGC	1	880	0	1	252
GGAGGTAGGtcaga--A--gcatCGTTGGTCCG					
TGAATCaatgcc--aatctGCCGTCA	1	3828	0	0	141
ACTTAGttacgg--ttagaCGGCAGT					
AATCTAaagtata--ttgctt**A**CGACTT	1	973	1	0	77

Table 3-5 Continued.

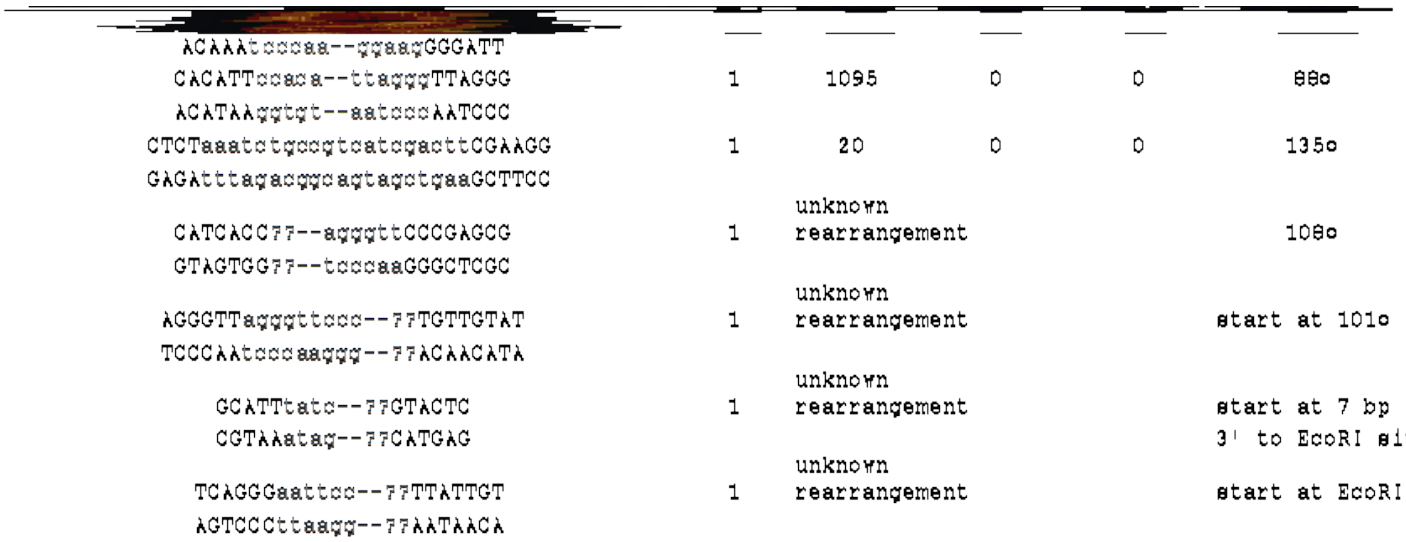
TTAGAT**T**tcatat--aacgaatGCTGAA AATCTAaagtata--ttgctt**A**CGACTT	1	973	1	0	77
TTAGAT**T**tcatat--aacgaatGCTGAA CCGCGAGaccca--??? GGCGCTCtgggt--???	1	unknown rearrangement			starts 714 bp 3' to EcoRI site

*potential siblings from replication in human cells

Table 3-6. End points of deletions and rearrangements for [TTAGGG]6 vector upon replication in shWRN U2-OS cells.

Endpoints	n	size of deletion	size of phomology	size of insertion	position on vector
[TTAGGG]6 vector replicated in WRN deficient U2OS					upstream from:
GGTCCtcggt--gggaattcGAGA	2*	467	0	0	50
CCAGGagcca--cccttaagCTCT					
GGTCCtcggt--gggaattcGAGA	9*	467	0	0	50
CCAGGagcca--cccttaagCTCT					
CAGTTaccaatg--aaggcaCCAATAA	2*	1127	0	0	274o
GTCAAtggttac--ttccgtGGTTATT					
CAGTTaccaatg--aaggcaCCAATAA	3*	1127	0	0	274o
GTCAAtggttac--ttccgtGGTTATT					
TTGCTGcaggca--ATAT--ggttcGAATCC	3*	703	0	4	164o
AACGACgtccgt--TATA--ccaagCTTAGG					
CATTCTgagaat--ccccGCCCTGC	1	588	1	0	304o
GTAAGACtctta--ggggcGGGACG					
GCTGGttctt--tcacttaTTCAGGCG	3*	1247	3	0	243o
CGACCAAGaaa--agtgaataagTCCGC					
TTAATCAGTGaaggcacc--cttattcAGGCGTAGCA	5*	1082	1	0	246o
AATTAGTCACTtccgtgg--gaataagtCCGCATCGT					
TTTCCACAcctggtt--gaatttCTGCCATTCAT	1	1312	1	0	215o
AAAGGTGTGgaccaa--cttaaagACGGTAAGTA					
CCATCCagtctatt--ctgccGTCATCGAC	1	761	0	0	144o
GGTAGGtcagataa--gacggCAGTAGCTG					
TTAAGGgatttgg--ttcatccGCTTATTA	1	4027	1	0	228o
AATTCCCtaaaacc--aagtaggcGAATAAT					

Table 3-6 Continued



ACAAAtcccaa--ggaagGGGATT CACATTcccca--ttagggTTAGGG	1	1095	0	0	88o
ACATAAggtgt--aatcccAATCCC CTCTaaatctgccgtcatcgacttCGAAGG	1	20	0	0	135o
GAGAtttagacggcagtagctgaaGCTTCC CATCACC77--agggttCCCGAGCG	1	unknown rearrangement			108o
GTAGTGG77--tcccaaGGGCTCGC AGGGTTagggttccc--77TGTTGTAT	1	unknown rearrangement			start at 101o
TCCCAAtcccaaggg--77ACAACATA GCATTtatc--77GTACTC	1	unknown rearrangement			start at 7 bp
CGTAAatag--77CATGAG TCAGGGaattcc--77TTATTGT	1	unknown rearrangement			3' to EcoRI siteo start at EcoRI siteo
AGTCCCttaagg--77AATAACA					

*potential siblings from replication in human cells
o Telomeric sequence is fully or partially deleted

4.0 WRN FACILITATES DNA SYNTHESIS *IN VIVO* ACROSS GAPPED TEMPLATES CONTAINING TELOMERIC REPEATS.

I would sincerely like to acknowledge Hong Wang, Ph.D., University of Pittsburgh Cancer Institute, for performing AFM experiments and providing data discussed in sections 4.2.2 and 4.3.2.

4.1 INTRODUCTION

Numerous studies have shown that G-rich telomere repeats can fold into secondary DNA structures called G4 DNA structures in regions of ssDNA. Increasing evidence shows ssDNA regions of telomeric repeats TTAGGG can fold into G4 DNA structures (Duquette et al., 2004; Xu et al., 2009). G4 DNA structures have been proposed to block replication progression and a number of proteins have been shown to resolve these structures *in vitro* (Kamath-Loeb et al., 2001). Our previous study using the *supF* mutagenesis system (chapter 3) revealed that replication of shuttle vectors (SV) with telomeric repeats in the absence of WRN led to a dramatic increase in the frequency of large deletions in the SV compared to telomeric repeats replicated in control cells. One of the reasons that WRN may suppress deletions in telomeric shuttle vectors could be due to WRN roles in resolving G4 DNA structures formed by G-rich telomeric repeats that block replication fork progression. The *supF* and the HSV-*tk* mutagenesis assays were based on the potential of G-rich telomeric repeats to fold into G4 DNA structures

during replication, but no pre-formed G4 DNA was introduced into the cells. In this study we examined the role of WRN in facilitating DNA synthesis through pre-folded G4 DNA structures formed by G-rich telomeric sequences. Since WRN has been shown to resolve G4 DNA structures *in vitro* (Kamath-Loeb et al., 2001), we hypothesized that the absence of WRN would lead to an accumulation of G4 DNA that could block DNA synthesis across G-rich telomeric sequences and lead to deletions. To test our hypothesis, we designed an *in vivo* assay that involves plasmids with regions of single stranded DNA that contain telomeric repeats which can readily fold into G4 DNA structures. This assay was adopted from a previously reported assay that used plasmids with single stranded regions harboring defined DNA lesions to examine and measure translesion synthesis in human cells (Avkin et al., 2002). However, in our assay the "lesion" was a pre-folded G4 DNA in TTAGGG ssDNA rather than damaged DNA bases.

4.2 MATERIALS AND METHODS

4.2.1 Construction of gapped duplex plasmids

The pRStuI and pSStuI phagmid vectors contain the HSV-*tk* reporter and harbor either a functional or non-functional chlor resistance gene (*cat*), respectively (Hile and Eckert, 2008). The sequence (TTAGGG)10 or CCTAA(CCCTAA)9C, was introduced in-frame into the 5' coding region between positions 111 and 112 of the HSV-tk gene as shown in Figure 4-1A and described previously (Damerla et al., 2010 and chapter 2). Plasmids with single stranded regions or gapped duplex (GD) were constructed with either the G-rich telomeric repeats in the ssDNA region (G-rich GD) or the C-rich telomeric sequences (C-rich GD) by hybridizing a linear DNA

fragment from pRStuI to the single stranded circular DNA templates generated from pSStuI plasmid. Single stranded DNA templates were generated and prepared by infecting pSStuI plasmid bearing F' *E. coli* strain DH5αIQ with R408 helper phage. Linear DNA fragments were prepared by digesting pRStuI plasmids with *MluI* and *StuI* and were hybridized to pSStuI ssDNA templates to generate GD molecules containing 157 nt of ssDNA as described previously (Eckert et al., 1997). To eliminate any residual linear fragment that did not hybridize to the circular templates, we introduced a *BsrGI* restriction enzyme site into the *cat* gene of the pRStuI plasmid through site directed mutagenesis. This renders the linear fragment vulnerable to *BsrGI* digestion (Figure 4-1 B), while the GD is resistant since hybridization with the pSStul ssDNA template yields a mismatch at the BsrGI site. The GD was separated from digested linear fragment by 0.8% agarose gel electrophoresis and purified using GeneClean^{TM} (MP Biomedicals, Solon, OH). Purified DNA was concentrated using Ultracel® YM-100 (Millipore, Billerica, MA).

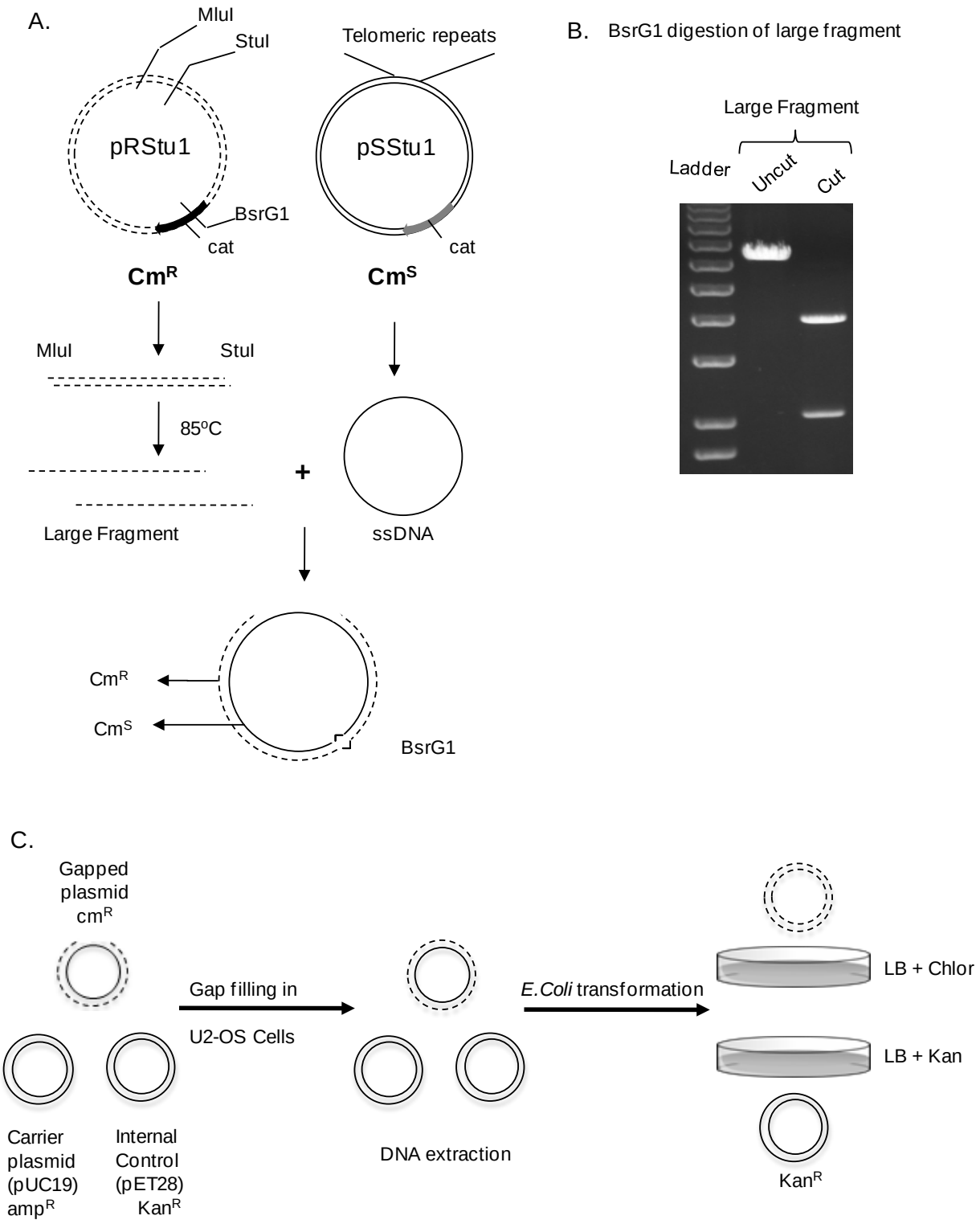


Figure 4-1. Outline of in vivo telomeric gap filling assay.

Figure 4-1. continued

A, A gapped duplex (GD) was constructed with one DNA strand from the plasmid pRStu1 containing a functional *cat* gene (Cm^R) and one strand from pSStu1 plasmid containing a non-functional *cat* gene (Cm^S). pRStu1plasmids are digested with enzymes *MluI* and *StuI*, to yield a linear large fragment (LF). LF is purified, denatured and hybridized with pSStu1 ssDNA circular templates, to manufacture GD. The pSStu1 ssDNA templates harbored an HSV-*tk* gene with either TTAGGG repeats or AATCCC repeats. B, 0.8% agarose gel with a 1 kb plus ladder (Invitrogen, Carlsbad, CA) and *Bsrg1* digestion of the LF is complete and yields two smaller fragments. C, *In vivo* telomeric gap filling assay. A mixture of GD, internal control plasmid (with a kanamycin resistance gene, Kan^R) and carrier plasmid (with an ampicillin resistance gene, Amp^R) is transfected into U2OS cells followed by DNA extraction and transformation into *E.coli* followed by selective plating on chlor and kan. Completely filled GD plasmids are replicated in bacteria. Bacteria harboring plasmids derived from pSStu1 strand are eliminated by selection with chlor while the plasmids derived from pRStu1 strand are retained upon selection with chlor. Internal control plasmid recovery can be calculated by measuring the number of kan resistant colonies.

4.2.2 AFM imaging and analysis

G-rich GD and C-rich GD molecules were diluted to 3 µg/ml in a buffer containing 10 mM Tris-HCl pH7.5 and 10 mM $MgCl_2$ with either 50 mM KCl or 25 mM Na-acetate. GD buffers were pre-heated at 65^{o}C for 15-30 min to dissolve small salt particles that may have accumulated during storage. GD molecules were heated at 65 ^{o}C for 15 min before slow cooling to room temperature to allow for G4 formation, and were pre-incubated at 37^{o}C for 15 min immediately prior to deposition onto a freshly cleaved mica disk (SPI Supply, West Chester, PA). All samples were washed with MilliQ water and dried under a stream of nitrogen gas. Images were collected using a MultiModeV microscope (Veeco Instruments, Plainview, NY) using E scanners in the

tapping mode. Pointprobe® plus noncontact/tapping mode silicon probes (PPP-NCL, Agilent) with spring constants of ~50 N/m and resonance frequencies of ~190 kHz were used. Images were captured at a scan size of 2 μm × 2 μm, a scan rate of 3 Hz, a target amplitude of 0.30 to 0.35 V and a resolution of 512 × 512 pixels. Structures on GD molecules with a peak height over 1 nm were considered G4 DNA based according to previous work (Wang et al., 2011). G4 peak height and G4 length (full-width at half-maximum height) were measured using Nanoscope 7.30 software by using the section analysis function. Duplex DNA regions from the same images as the G4 structures were used as internal standards for the G4 length quantification. The adjusted G4 length was calculated as F = D × R, where F is the adjusted value for G4 length, D is the value from direct measurement, and R is the ratio of the mean cross-section width (10 nm) measured from multiple depositions of the 517 bp long PCR fragments using different imaging probes to the mean value of duplex DNA regions on the GD molecules. Any incompletely formed GD molecules were excluded from data analysis.

4.2.3 *In vivo* telomeric gap filling assay

We followed previous protocols for the *in vivo* quantitative gap filling assay as described previously, with some modification (Figure 4-1C) (Hendel et al., 2008; Shachar et al., 2009). Purified gapped duplex DNA (500 ng) along with an internal control plasmid (pET28, Kan^R) (500 ng) and a carrier plasmid (pUC19, Amp^R) were transfected into 5 million shCTRL and shWRN U2OS cells by the Amaxa nucleofector system (Lonza). After 24 hours of incubation at 5% O_2, 5% CO_2 and 37°C, the plasmids were isolated using the Qiagen Tip-20 kit (Valencia, CA). Then 4ul of the isolated plasmid DNA was transfected into *tk*- FT334 bacteria (Eckert et

al., 1997) using the Bio-Rad Xcell system followed by incubation at 37^{o}C for 1 hour. Transformed bacteria were plated onto LB plates supplemented with chlor to select for bacteria harboring completely filled gapped duplexes and for progeny of the strand that was synthesized in human cells during gap filling. Bacterial cultures were also plated on kan supplemented LB agar plates to select for bacteria harboring the internal control plasmids. The efficiency of gap repair was calculated as the ratio of the number of chlor resistant colonies to the number of kan resistant colonies. Mutations occurring within the gap filled region were determined by DNA sequencing at ACGT Inc. (Wheeling, IL).

4.3 RESULTS

4.3.1 Telomeric gap filling and repair assay

In order to determine whether WRN protein is required for DNA replication past pre-folded G4 DNA, we developed the *in vivo* telomeric gap filling assay which is a modification of a previously established translesion replication (TLR) assay (Avkin et al., 2002). In our case the "lesion" of interest is G4 DNA. Briefly, the assay involves transfection of mammalian cells with a gapped duplex (GD) vector containing a chloramphenicol resistant gene (cm^R) and an unrelated plasmid with a kanamycin resistant gene (kan^R) which serves as an internal standard to control for differences in transfection efficiency. The GD and plasmids lack origins for replication in human cells and are recovered from the cells using an alkaline lysis purification procedure that denatures unfilled GD molecules or nicked plasmids (Avkin et al., 2002). Intact plasmids are introduced into a reporter *E.coli* strain followed by plating on selective media. The cm^R/kan^R transformant ratio represents the efficiency of intact GD recovery or repair (Figure 4-1 C). Recovered GDs are sequenced to determine the repair mechanism. We generated GD molecules containing 157 nt of ssDNA consisting of the sequence $(TTAGGG)_{10}$ (G-rich GD) or $CCTAA(CCCTAA)_9C$ (C-rich GD) flanked by sequences of the HSV-*tk* reporter gene (see Materials and Methods). The G-rich GD represent lagging strand DNA synthesis with G-rich ssDNA regions that can form G4 DNA, whereas the C-rich GD represents leading strand DNA synthesis and should not form secondary structures.

The GDs were generated by hybridizing single stranded pSStu1 template containing telomeric repeats (cm^S) to a linear ssDNA strand generated from the pRStu1 plasmid (cm^R). The hybridization takes place in such a way that only the 60 nt telomeric sequences and the 97 nt flanking regions are in the single stranded region of the GD. Since only the outer pRStu1 strand of the GD has an intact *cat* gene, plasmids DNA derived only from pRStu1 strand of completely filled GD are selected for (Figure 4-1C). To ensure that any contaminating linearized vectors were not present during transfection, GDs were digested with BsrG1 which cleaves any contaminating linear fragments but not the GD (see materials and methods).

4.3.2 Gapped duplex molecules with TTAGGG single stranded regions form G4 structures

To confirm that G4 DNA can form in the G-rich GD vectors with TTAGGG ssDNA we used single molecule AFM imaging. Previous work by others, and our recent study imaging DNA substrates that mimic the 3' ssDNA overhang at telomeres, validated the application of AFM imaging for the direct visualization and quantification of G4 structures (Neaves et al., 2009; Wang et al., 2011). We previously established that structures on TTAGGG ssDNA with an AFM peak height greater than 1 nM represent G4 DNA (Wang et al., 2011). AFM images of the C-rich GD (Figure 4-2A) did not display any structures exhibiting a peak height greater than 1 nm except for one at a site of overlapping duplex DNA molecules (1/18 C-rich GD). The average peak height of structures on C-rich GD is 0.41 (± 0.1) nm (n=23 measurements). On the contrary, 18% (7/39) of the G-rich GD molecules exhibited an individual structure with a peak height greater than 1 nm (Figure 4-2B, white arrow), consistent with G4 DNA measured

previously (Wang et al., 2011). These structures are distinct from regions of overlapping duplex DNA molecules. The average length of the G4 structures on G-rich GD is ~19 nm (see Materials and Methods), which is consistent with two G4 folds in TTAGGG repeat ssDNA (Wang et al., 2011). Not all the G-rich GD molecules displayed G4 structures and increasing the potassium chloride concentration to 100 mM, which promotes G4 folding, did not significantly increase the percent of G-rich GD molecules with G4 folds (data not shown). This is consistent with previous studies suggesting a dynamic equilibrium between unfolded and folded G4 structures (Aad et al., 2010), however, intermediate folds are not detectable by AFM. In summary, single-molecule AFM imaging provided direct evidence that some of the G-rich GD, but not the C-rich GD, transfected into the human cells had pre-folded G4 structures.

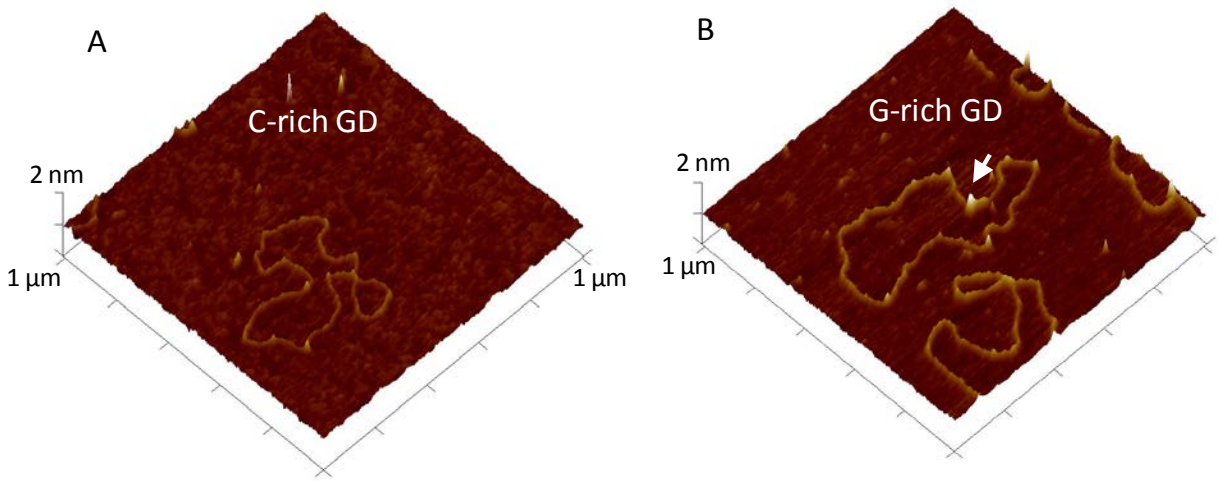


Figure 4-2. Gapped duplex vectors with TTAGGG repeat ssDNA form G4 structures.

A representative AFM image of (3 µg/ml) purified gapped duplex with $(TTAGGG)_{10}$ (A) or $CCTAA(CCCTAA)_9C$ (B) ssDNA. The images are 1 µm × 1 µm at 2 nm Z-scale. The white arrow points to a region forming a G4 structure on a G-rich GD molecule.

4.3.3 G4 DNA does not inhibit GD recovery or repair

In order to examine whether G4 forming ability in the ssDNA region interfered with GD recovery, the G-rich GD and C-rich GD were transfected into U2OS cells expressing a control shRNA. First we had to ensure that GDs were filled by DNA synthesis in human cells, rather than in bacteria. Any unfilled GDs in human cells should be denatured during the alkaline lysis plasmid DNA extraction from human cells (Avkin et al., 2002). To test this, we performed an experiment wherein 500 ng of internal control plasmid along with 500 ng of C-rich GD were treated with alkali and introduced directly into the *E.coli* reporter strain without prior transfection into human cells. We observed that the gap repair efficiency of the C-rich GD upon alkaline lysis

was 0.001 which was approximately 80 fold lower compared to the lowest gap repair efficiency of C-rich GD in human cells (Table 4-1). This suggests that unfilled GD plasmids after passage through human cells are unlikely to survive alkaline lysis during DNA extraction. We then transfected G-rich and C-rich gaps into shCTRL U2OS cells to determine the influence of G4 DNA forming sequences on the repair efficiency. Two independent transfections a and b (Table 4-1) for each GD yielded highly variable repair efficiencies that differed by almost 2 fold. Although the gap repair efficiencies of C-rich and G-rich GD into shCTRL cells were quite variable, we did not observe any difference between the mean gap repair efficiency of C-rich GD (0.11) and G-rich GD (0.12) indicating that G4 DNA does not inhibit GD recovery or repair in U2OS cells.

Table 4-1. Gap repair efficiencies for the *in vivo* telomeric gap filling assay.

Cell Line	GD	Experiment	Cm^R	Kan^R	Gap Repair Efficiency Cm^R/Kan^R
n/a	C-rich + Alkali		4	4230	0.001
shCTRL	C-rich	a	7	50	0.14
		b	5	64	0.080
shCTRL	G- rich	a	14	172	0.080
		b	167	1116	0.15
shWRN	C- rich	a	370	1160	0.32
		b	120	2030	0.060
shWRN	G- rich	a	640	670	0.96
		b	100	1370	0.070

Plasmid mixtures containing the indicated C-rich or G-rich GD (cm^R), internal control plasmid (pET.SUMO.TPP1 – kan^R) and carrier plasmid pUC19 were transfected into the indicated U2OS cell lines followed by DNA extraction

and *E.coli* transformation. Efficiency of gap repair was calculated as the ratio of cm^R colonies to kan^R colonies. The letters a and b in column 3 denote independent experiments.

4.3.4 Highly variable repair efficiencies of gapped duplex plasmids in shWRN U2OS cells

We calculated gap repair efficiencies of C-rich and G-rich GDs in U2OS cells that were either proficient or deficient for WRN from two independent experiments. The gap repair efficiencies were even more variable in shWRN cells for both the C-rich and G-rich GD compared to the shCTL cells. A 5 fold difference between two independent experiments was observed when C-rich GD was transfected into WRN deficient cells, and a 14 fold difference in repair efficiency for the G-rich GD was observed (table 4-1.)

4.3.5 Gap filling accuracies of C-rich and G-rich GD

To determine if gap repair efficiencies correlated with accuracy of gap filling, we sequenced plasmid DNA from randomly chosen cm^R colonies from each independent transfection into U2OS cells. We successfully recovered accurately filled gaps for both G-rich and the C-rich GD in shCTRL and shWRN cells from two independent experiments. We also found that some of the plasmids did not yield any sequencing data (Table 4-2) suggesting the occurrence large deletions. Plasmid DNA from cm^R colonies was also digested with restriction enzymes *BsshII* and *BglII* to test for the presence of the HSV-*tk* gene and any abnormal events that correspond to deletions or rearrangements in the gap filled plasmids. We again observed that the gap filling accuracy varied considerably among all experiments. This variability in gap filling accuracy correlated with the highly variable gap repair efficiencies.

Table 4-2. Gap repair accuracy determined by DNA sequencing and restriction enzyme digest analysis.

Cell Line	GD	Experiment	Cm^R	Accurately filled GD	Accuracy of gap filling (%)	BsshII-BglII Digest	
						Normal	Abnormal
shCTRL	C-rich	a	9	8	89	7	2
		b	4	1	25	2	2
shCTRL	G- rich	a	10	10	100	10	0
shWRN	C- rich	a	10	10	100	10	0
		b	5	2	40	2	3
shWRN	G- rich	a	10	6	60	3	7
		b	5	5	100	5	0

Plasmid mixtures containing the indicated GD were introduced into either shCTRL or shWRN U2OS cells. Plasmids were extracted and transfected into bacteria. Plasmids DNA isolated from cm^R colonies are subjected to DNA sequence analysis and also digested using restriction enzymes *BssHII* and *BglII* to look for abnormalities in the HSV-*tk* gene. a and b denote two independent transfections into U2OS cells. Sequencing and restriction enzymes digest data were acquired from only one independent experiment for the G-rich GD in shCTRL U2OS cells.

4.4 DISCUSSION

We developed an *in vivo* telomeric gap filling assay to determine if WRN has a role in facilitating DNA synthesis across gapped duplex plasmids with telomeric repeats which form prefolded G4 DNA structures. The advantages of using this assay are: (1) DNA synthesis across gapped plasmids with pre-formed G4 DNA structures can be analyzed. (2) These plasmids do not replicate in human cells. Therefore the amount of plasmid DNA recovered from the human cells is a direct measure of gap repair in the human cells. (3) This assay can be used in almost any cell line since the need for episomal replication does not arise.

We analyzed gap repair efficiencies across GD with C-rich or G-rich telomeric sequences in the ssDNA region of the gapped plasmids. Atomic force microscopy was used to confirm the formation of G4 DNA in G-rich GD but not in C-rich GD (Figure 4-2). The gap repair efficiencies for C-rich and G-rich GD were similar in control cells but varied from 0.080 to 0.15 (table 4-1). The gap repair efficiencies for C-rich and G-rich GDs in WRN depleted cells revealed a greater variability (0.060 to 0.32 for C-rich GD and 0.070 to 0.96 for G-rich GD; table 4-1). We observed similar variability upon analyzing the gap filling accuracies for the telomeric GDs in control and WRN depleted cells (Table 4-2). However we could successfully get accurate gap filling of G-rich and C-rich GD occurred in both control and WRN depleted cells.

The *in vivo* telomeric gap filling assay is based on the *in vivo* translesion replication (TLR) assay published earlier by Avkin *et al* in 2002 (Avkin et al., 2002). The TLR assay was highly reproducible and the coefficient of variance for gap repair efficiency (bypass efficiency) in six experiments was only 6.5% (Avkin et al., 2002). We were not able to obtain reproducible results in our assay. The sources for variability in the telomeric gap filling assay could be due to: (a) Larger regions of ssDNA in the gaps may be more susceptible to nicking and breakage during

eletroporation. The ssDNA region of the GD we use in our assay is around 150 nt long compared to only 20 nt in the TLR assay (Avkin et al., 2002). AFM data showed that approximately 18% of the G-rich GD form G4 DNA in one GD preparation that was examined. Potential differences in the fraction of G-rich GDs that form G4 DNA among independent GD preparations could result in variability of repair efficiency. (b) The potential differences in WRN expression levels among various passages of shWRN U2OS cells could also account for the variability. We noticed that WRN protein expression increases with increasing passage of shWRN U2OS cells, possibly due to loss of the shWRN transcription. We also found a two fold variability in gap repair efficiency of GDs in shCTRL cells (Table 4-1), which was lower than the variability of gap repair efficiencies obtained from shWRN U2OS cells. This two-fold variability may represent the inherent variability of the assay.

4.4.1 Future directions

Recovery of accurate gap filling of GD after transfection into control and WRN deficient cells indicates that our assay design was successful. If the inherent variability of the assay is high than performing a greater number of multiple independent experiments will be necessary to discern any statistical differences in gap repair efficiencies among gap types (G-rich vs. C-rich) and cell types (shCTRL vs shWRN). Multiple transfections of the same GD preparation into human cells would reveal if the high variability is due to differences in GD preparations. Larger ssDNA regions in the gaps could also contribute to variability, since ssDNA is susceptible to the action of nucleases in the cell which result in deleterious events or cleavage independently of DNA synthesis across GD. Therefore reducing the size the ssDNA region in the gap molecules might also decrease the variability we observed in the repair efficiency and gap filling accuracy of the

telomeric gapped duplex plasmids. Finally, we can test for variability in WRN protein expression in the shWRN U2OS cells among independent experiments by Western Blot analysis using an anti-WRN antibody. If the variability in the telomeric gap filling assay can be reduced, then this will be a highly valuable assay for determining the consequences of G4 DNA formation during DNA replication, and the factors that ensure accurate DNA synthesis past G4 DNA folds.

4.5 ACKNOWLEDGEMENTS

We would like to acknowledge Dr. Hong Wang, University of Pittsburgh Cancer Institute for performing AFM experiments and providing us with data.

5.0 GENERAL DISCUSSION

Repetitive sequences present throughout the genome can possibly adopt a variety of secondary DNA structures that differ from the normal right handed double helical B DNA (Wang and Vasquez, 2006). Telomeres comprise of repetitive DNA sequences which can form non B DNA structures such as G4 DNA. Since repetitive sequences promote stalled replication forks due to their ability to form alternate DNA structures, it is possible that telomeric sequences may also promote replication fork stalling and collapse leading to DSBs and large deletions. Since telomere lengths are an important determinant of telomere function (Baird, 2008), it is important to understand mechanisms of telomere loss and factors regulating telomere length homeostasis. Our first goal was to understand if the mutagenicity of telomere repeats depended on sequence specificity, i.e. if the G-rich nature of telomeric sequences were prone to cause mutations and deletions during replication. Our next goal was to understand factors affecting telomere stability. Numerous studies revealed that accessory proteins other than the shelterin complex proteins at the telomere to play important functions in telomere replication and repair (de Lange, 2005). WRN is one such enzyme with helicase and exonuclease activity and defects in WRN causes a premature aging syndrome called Werner syndrome which is characterized by increased stochastic telomere loss especially telomeres derived from the G-rich strand (Crabbe et al., 2004). We therefore asked the question if WRN has a role in facilitating DNA replication of telomeric repeat sequences.

We used three independent mutagenesis assays to test our hypothesis. Mutagenesis assays involving shuttle vectors have been powerful tools for studying DNA repair, recombination and replication. More recently a number of studies involving shuttle vectors have been published on genomic instability caused by non-B DNA structure forming sequences. We used two different SV mutagenesis systems. The HSV-*tk* mutagenesis system is based on the Epstein-Barr virus origin of replication and the SVs replicate along with the host cell replication cycles. The *supF* mutagenesis system involves a more transiently replicating shuttle vector which utilizes the Simian Virus 40 replication origin.

The HSV-*tk* mutagenesis system was used to determine the mutagenic potential of (1) G-rich repetitive sequences based on their opportunity to fold into G4 DNA structures (single-stranded DNA during lagging strand replication increases the opportunity for G4 folding) and (2) mammalian and ciliate telomeric repeats which form quadruplex structures with different thermal stabilities *in vitro.* This assay was previously used to study the mutagenesis of microsatellite DNA sequences, which can also adopt alternate DNA structures. Varied mutagenic potential with both losses and gains of sequences were detected based on sequence specificity of the microsatellite repeats (Hile et al., 2000). There are several advantages of the HSV-*tk* shuttle vector mutagenesis assay. In bacteria the mutations that compromise the reporter gene (HSV-*tk*) activity can be readily selected through selective plating. Therefore the burden of screening for mutant colonies does not arise. The mammalian replication origin on the SV further ensures that the SV replicates only once every cell cycle using the host replication machinery and has a unidirectional origin (Sears et al., 2004; Sugden, 2002). We made use of the unidirectional replication origin to asked the question if there was a difference in mutagenic potential when G-rich telomeric sequences were replicated on the lagging or the leading strand. Mutant rates of

shuttle vectors with different telomeric inserts from human and ciliate telomeric repeats after being replicated in human cells were calculated. However, the mutant rates we observed were highly variable and did not fit a normal distribution curve. We used non parametric statistics and hence median was used as a measure of central tendencies among various groups, and we used the Mann-Whitney test to determine differences among various groups. The vectors with human G-rich repeats on the lagging strand (correct orientation of chromosomal telomeres in normal cells) exhibited significantly lower mutant rates than the vectors with G-rich sequences on the leading strand (Figure 2-6). This suggests that human cells have evolved mechanisms to ensure proper replication through telomeres. We also found that G4 DNA with greater stability (as seen in G4 DNA of *T.thermophila* and *O.nova*) were significantly more mutagenic compared to human telomeric repeats. We were unsuccessful in using this assay to elucidate the role of WRN protein in promoting faithful telomere replication and telomere stability. The lymphoblastoid cells LCL-721 used in this assay did not survive WRN depletion by both retroviral or lentiviral transduction methods. Therefore the *supF* mutagenesis assay was employed to test the hypothesis that WRN prevents mutations and deletions during telomere DNA replication.

The *supF* mutagenesis system was previously used to measure the mutagenic potential of non-B DNA forming sequences like H DNA and Z DNA (Wang et al., 2004; Wang et al., 2006). Previous studies showed that transfection of Z DNA containing SV into mammalian cells resulted in large scale deletions in the SV. The advantages of using the *supF* mutagenesis system is that a high copy number of SV can be transfected into U2OS cells which ensures sufficient recovery of replicated shuttle vectors. Also, these SVs have the gene that encodes the SV-40 large T antigen which allows for replication of these SV in almost any cell line. Using the *supF* mutagenesis assay we observed that telomeric SV are replicated stably in normal human U2OS

cells which agreed with the previously reported results using the HSV-*tk* mutagenesis assay. Upon WRN depletion we observed that the mutant frequencies of telomeric SV increased dramatically along with the frequency of deletions compared to replication of the same telomeric SV in control U2OS cells. These data indicate that WRN prevents large deletions during replication at templates containing telomeric sequences.

The *supF* mutant frequencies were not highly variable when compared to some of the mutant rates obtained from the HSV-*tk* shuttle vectors (Table 2-3 and 2-4). The HSV-tk mutagenesis system involves a laborious task of replicating the SV for 20 to 30 generations in human cells. Therefore mutations that occur in the SV at early or late generations could skew the mutant rates and result in high standard error in the measurements. This problem can be circumvented using the *supF* SVs since they are only replicated transiently (48 hours) in the host cell. The HSV-*tk* mutagenesis assay is not conducive for multiple independent experiments. We were only able to accomplish a maximum of two independent experiments of transfecting telomeric SV into human cells. But using the *supF* assay we could easily calculate standard errors while comparing mutant frequencies since each mutant frequency was obtained from one independent experiment of transfecting SV into human cells.

With the help of the shuttle vector mutagenesis assays we were able to show for the first time that telomeric sequences in spite of their G4 DNA forming potential are stably replicated in normal human cells. At the same time, we also showed that telomeric sequences are more prone to deletions upon replication compared to scrambled sequences in a WRN deficient background. This suggests that stable telomere DNA replication in normal cells demands robust regulation by a host of DNA replication, repair and telomere associated proteins. Helicases like WRN play a vital role in facilitating telomeric DNA replication most likely by resolving alternate DNA

structures like G4 DNA formed by telomeric repeats during replication. Through the gap filling assay briefly discussed in chapter 4, we plan to dissect the role of WRN in allowing DNA synthesis across preformed G4 DNA structures formed by telomeric repeats. We have successfully established two highly sensitive mutagenesis assays to determine the stability of telomeric repeat sequences and factors affecting the stability of telomeric repeats upon DNA replication. The high sensitivity of these assays can now be used to determine several factors that influence telomere stability such as:

1. Telomeric DNA associated proteins that help in replication, repair or DNA damage signaling at telomeres such as shelterin proteins or other RecQ helicases
2. Environmental agents that cause genotoxic stress such as heavy metals.
3. Potential chemotherapeutic agents that target telomeres such as G4 DNA binding ligands.

RecQ helicases have important roles in various metabolic pathways and preventing genomic instability in the cell. Patients with defects in RecQ helicase genes are predisposed to age related disorders and cancer. Moreover, numerous studies have shown that cells deficient for RecQ helicases WRN and BLM are sensitive to various DNA damaging agents, some of which are used as chemotherapeutic agents. Stratagies that can increase sensitivity of cancer cells to anti cancer therapies can enhance the efficacy of chemotherapeutic or radiotherapeutic treatment methodologies. Recent advances in RNA interference (RNAi) are proposed to knockdowm various proteins that could sensitize cancer cells to anti cancer treatment. The expression of short hairpin RNA (shRNA) by viral vectors to knockdown specialized helicases like WRN in cells could lead to this increased sensitivity to chemotherapy.

The inability of telomerase and DNA polymerase to synthesize DNA past a G-quadruplex folded telomeric substrate has led to the emergence of G4 DNA stabilizing agents as promising candidates for cancer therapy. A number of studies have shown that cells exposed to G4 DNA stabilizing compounds such as RHPS4 and telomestatin exhibit anti tumor effect, increased growth arrest, apoptosis and telomere dysfunction (Salvati et al., 2007; Tauchi et al., 2006; Tahara et al., 2006). Since WRN plays an important role in telomere maintenance, siRNA mediated inhibition of WRN could be used in combination with G4 DNA stabilizing agents as possible therapeutic avenues to treat cancer. Further research is required in this field to understand the importance of G4 DNA or helicases like WRN in telomere regulation and their potential in preventing or delaying human disease.

BIBLIOGRAPHY

Aad, G., Abbott, B., Abdallah, J., Abdelalim, A.A., Abdesselam, A., Abdinov, O., Abi, B., Abolins, M., Abramowicz, H., Abreu, H., *et al.* (2010). Observation of a centrality-dependent dijet asymmetry in lead-lead collisions at sqrt[S(NN)] =2.76 TeV with the ATLAS detector at the LHC. Phys Rev Lett *105*, 252303.

Aiyar, A., Aras, S., Washington, A., Singh, G., and Luftig, R.B. (2009). Epstein-Barr Nuclear Antigen 1 modulates replication of oriP-plasmids by impeding replication and transcription fork migration through the family of repeats. Virol J *6*, 29.

Anderson, R.A., Wallace, W.H., and Baird, D.T. (2008). Ovarian cryopreservation for fertility preservation: indications and outcomes. Reproduction *136*, 681-689.

Armanios, M. (2009). Syndromes of Telomere Shortening. Annual Review of Genomics and Human Genetics *10*, 45-61.

Armanios, M.Y., Chen, J.J.-L., Cogan, J.D., Alder, J.K., Ingersoll, R.G., Markin, C., Lawson, W.E., Xie, M., Vulto, I., Phillips, J.A., *et al.* (2007). Telomerase Mutations in Families with Idiopathic Pulmonary Fibrosis. New England Journal of Medicine *356*, 1317-1326.

Arnoult, N., Saintome, C., Ourliac-Garnier, I., Riou, J.F., and Londono-Vallejo, A. (2009). Human POT1 is required for efficient telomere C-rich strand replication in the absence of WRN. Genes Dev *23*, 2915-2924.

Artandi, S.E., and DePinho, R.A. (2010). Telomeres and telomerase in cancer. Carcinogenesis *31*, 9-18.

Avkin, S., Adar, S., Blander, G., and Livneh, Z. (2002). Quantitative measurement of translesion replication in human cells: Evidence for bypass of abasic sites by a replicative DNA polymerase. Proceedings of the National Academy of Sciences of the United States of America *99*, 3764-3769.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science *318*, 798-801.

Bacolla, A., Wang, G., Jain, A., Chuzhanova, N.A., Cer, R.Z., Collins, J.R., Cooper, D.N., Bohr, V.A., and Vasquez, K.M. (2011). Non-B DNA-forming sequences and WRN deficiency independently increase the frequency of base substitution in human cells. J Biol Chem.

Bai, Y., and Murnane, J.P. (2003). Telomere instability in a human tumor cell line expressing a dominant-negative WRN protein. Hum Genet *113*, 337-347.

Bailey, S.M., and Murnane, J.P. (2006). Telomeres, chromosome instability and cancer. Nucleic Acids Research *34*, 2408-2417. Nucl. Acids Res. 34(8): 2408-2417.

Baird, D.M. (2008). Mechanisms of telomeric instability. Cytogenetic and Genome Research *122*, 308-314.

Baumann, P., and Cech, T.R. (2001). Pot1, the Putative Telomere End-Binding Protein in Fission Yeast and Humans. Science *292*, 1171-1175.

Baumann, P., and Price, C. (2010). Pot1 and telomere maintenance. FEBS Lett *584*, 3779-3784.

Blasco, M.A. (2005). Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet *6*, 611-622.

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). Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. Science *279*, 349-352.

Bohr, V.A. (2008). Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance. Trends in Biochemical Sciences *33*, 609-620.

Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. (1997). Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. Nat Genet *17*, 231-235.

Brosh, R.M., Jr., Orren, D.K., Nehlin, J.O., Ravn, P.H., Kenny, M.K., Machwe, A., and Bohr, V.A. (1999). Functional and physical interaction between WRN helicase and human replication protein A. The Journal of biological chemistry *274*, 18341-18350.

Brosh, R.M., Jr., von Kobbe, C., Sommers, J.A., Karmakar, P., Opresko, P.L., Piotrowski, J., Dianova, I., Dianov, G.L., and Bohr, V.A. (2001). Werner syndrome protein interacts with human flap endonuclease 1 and stimulates its cleavage activity. The EMBO journal *20*, 5791-5801.

Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res *34*, 5402-5415.

Burma, S., Chen, B.P., Murphy, M., Kurimasa, A., and Chen, D.J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. The Journal of biological chemistry *276*, 42462-42467.

Campisi, J. (2001). From cells to organisms: can we learn about aging from cells in culture? Exp Gerontol *36*, 607-618.

Capper, R., Britt-Compton, B., Tankimanova, M., Rowson, J., Letsolo, B., Man, S., Haughton, M., and Baird, D.M. (2007). The nature of telomere fusion and a definition of the critical telomere length in human cells. Genes & Development *21*, 2495-2508.

Celli, G.B., and de Lange, T. (2005). DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. Nature cell biology *7*, 712-718.

Celli, G.B., Denchi, E.L., and de Lange, T. (2006). Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. Nature cell biology *8*, 885-890.

Cesare, A.J., and Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet *11*, 319-330.

Cha, R.S., and Kleckner, N. (2002). ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. Science *297*, 602-606.

Chang, S. (2005). A mouse model of Werner Syndrome: what can it tell us about aging and cancer? Int J Biochem Cell Biol *37*, 991-999.

Chang, S., Multani, A.S., Cabrera, N.G., Naylor, M.L., Laud, P., Lombard, D., Pathak, S., Guarente, L., and DePinho, R.A. (2004). Essential role of limiting telomeres in the pathogenesis of Werner syndrome. Nat Genet *36*, 877-882.

Chen, J.-L., and Greider, C.W. (2004). Telomerase RNA structure and function: implications for dyskeratosis congenita. Trends in Biochemical Sciences *29*, 183-192.

Cheng, W.-H., Kusumoto, R., Opresko, P.L., Sui, X., Huang, S., Nicolette, M.L., Paull, T.T., Campisi, J., Seidman, M., and Bohr, V.A. Collaboration of Werner syndrome protein and BRCA1 in cellular responses to DNA interstrand cross-links. Nucleic Acids Research *34*, 2751-2760.

Cheng, W.H., Muftic, D., Muftuoglu, M., Dawut, L., Morris, C., Helleday, T., Shiloh, Y., and Bohr, V.A. (2008). WRN is required for ATM activation and the S-phase checkpoint in response to interstrand cross-link-induced DNA double-strand breaks. Mol Biol Cell *19*, 3923-3933.

Cheng, W.H., von Kobbe, C., Opresko, P.L., Arthur, L.M., Komatsu, K., Seidman, M.M., Carney, J.P., and Bohr, V.A. (2004). Linkage between Werner syndrome protein and the Mre11 complex via Nbs1. J Biol Chem *279*, 21169-21176.

Cheung, I., Schertzer, M., Rose, A., and Lansdorp, P.M. (2002). Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA. Nat Genet *31*, 405-409.

Cleary, J.D., Nichol, K., Wang, Y.-H., and Pearson, C.E. (2002). Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat Genet *31*, 37-46.

Chin, L., Artandi, S.E., Shen, Q., Tam, A., Lee, S.-L., Gottlieb, G.J., Greider, C.W., and DePinho, R.A. (1999). p53 Deficiency Rescues the Adverse Effects of Telomere Loss and Cooperates with Telomere Dysfunction to Accelerate Carcinogenesis. Cell 97, 527-538.

Constantinou A, Tarsounas M, Karow JK, Brosh RM, Bohr VA, et al. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. EMBO Rep. 2000;1:80.

Counter, C.M., Hahn, W.C., Wei, W., Caddle, S.D., Beijersbergen, R.L., Lansdorp, P.M., Sedivy, J.M., and Weinberg, R.A. (1998). Dissociation among in vitro telomerase activity, telomere maintenance, and cellular immortalization. Proceedings of the National Academy of Sciences *95*, 14723-14728.

Courcelle, J., Donaldson, J.R., Chow, K.H., and Courcelle, C.T. (2003). DNA damage-induced replication fork regression and processing in Escherichia coli. Science *299*, 1064-1067.

Crabbe, L., Jauch, A., Naeger, C.M., Holtgreve-Grez, H., and Karlseder, J. (2007). Telomere dysfunction as a cause of genomic instability in Werner syndrome. Proceedings of the National Academy of Sciences of the United States of America *104*, 2205-2210.

Crabbe, L., Verdun, R.E., Haggblom, C.I., and Karlseder, J. (2004). Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. Science *306*, 1951-1953.

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). A DNA damage checkpoint response in telomere-initiated senescence. Nature *426*, 194-198.

D'Amours, D., and Jackson, S.P. (2002). The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. Nat Rev Mol Cell Biol *3*, 317-327.

Damerla, R.R., Knickelbein, K.E., Kepchia, D., Jackson, A., Armitage, B.A., Eckert, K.A., and Opresko, P.L. (2010). Telomeric repeat mutagenicity in human somatic cells is modulated by repeat orientation and G-quadruplex stability. DNA Repair (Amst).

Dawson, R.M.C. (1986). Data for biochemical research, 3rd edn (Oxford, Clarendon Press).

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev *19*, 2100-2110.

De Lange, T., Lundblad, V., and Blackburn, E.H. (2006). Telomeres Second Edition.

Denchi, E.L., and de Lange, T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. Nature *448*, 1068-1071.

Deng, Y., Chan, S.S., and Chang, S. (2008). Telomere dysfunction and tumour suppression: the senescence connection. Nat Rev Cancer *8*, 450-458.

Deng, Z., Atanasiu, C., Burg, J.S., Broccoli, D., and Lieberman, P.M. (2003). Telomere Repeat Binding Factors TRF1, TRF2, and hRAP1 Modulate Replication of Epstein-Barr Virus OriP. J Virol *77*, 11992-12001.

Dhar, S.K., Yoshida, K., Machida, Y., Khaira, P., Chaudhuri, B., Wohlschlegel, J.A., Leffak, M., Yates, J., and Dutta, A. (2001). Replication from oriP of Epstein-Barr virus requires human ORC and is inhibited by geminin. Cell *106*, 287-296.

Dhar, V., and Schildkraut, C.L. (1991). Role of EBNA-1 in arresting replication forks at the Epstein-Barr virus oriP family of tandem repeats. Mol Cell Biol *11*, 6268-6278.

Diede, S.J., and Gottschling, D.E. (1999). Telomerase-mediated telomere addition in vivo requires DNA primase and DNA polymerases alpha and delta. Cell *99*, 723-733.

Dokal, I. (2000). Dyskeratosis congenita in all its forms. Br J Haematol *110*, 768-779.

Drinkwater, N.R., and Klinedinst, D.K. (1986). Chemically induced mutagenesis in a shuttle vector with a low-background mutant frequency. Proc Natl Acad Sci U S A *83*, 3402-3406.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev *18*, 1618-1629.

Eckert, K.A., and Drinkwater, N.R. (1987). recA-dependent and recA-independent N-ethyl-N-nitrosourea mutagenesis at a plasmid-encoded herpes simplex virus thymidine kinase gene in Escherichia coli. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *178*, 1-10.

Eckert, K.A., Hile, S.E., and Vargo, P.L. (1997). Development and use of an in vitro HSV-tk forward mutation assay to study eukaryotic DNA polymerase processing of DNA alkyl lesions. Nucleic Acids Res *25*, 1450-1457.

Eckert, K.A., Yan, G., and Hile, S.E. (2002). Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic human cells. Mol Carcinog *34*, 140-150.

Epstein, C.J., Martin, G.M., Schultz, A.L., and Motulsky, A.G. (1966). Werner's syndrome a review of its symptomatology, natural history, pathologic features, genetics and relationship to the natural aging process. Medicine *45*, 177-221.

Ermakova, O.V., Frappier, L., and Schildkraut, C.L. (1996). Role of the EBNA-1 protein in pausing of replication forks in the Epstein-Barr virus genome. J Biol Chem *271*, 33009-33017.

Fagagna, F.d.A.d., Reaper, P.M., Clay-Farrace, L., Fiegler, H., Carr, P., von Zglinicki, T., Saretzki, G., Carter, N.P., and Jackson, S.P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. Nature *426*, 194-198.

Fukuchi, K., Martin, G.M., and Monnat, R.J., Jr. (1989). Mutator phenotype of Werner syndrome is characterized by extensive deletions. Proc Natl Acad Sci U S A *86*, 5893-5897.

Gahn, T.A., and Schildkraut, C.L. (1989). The Epstein-Barr virus origin of plasmid replication, oriP, contains both the initiation and termination sites of DNA replication. Cell *58*, 527-535.

Gilson, E., and Geli, V. (2007). How telomeres are replicated. Nat Rev Mol Cell Biol *8*, 825-838.

Goddard, K.A., Yu, C.E., Oshima, J., Miki, T., Nakura, J., Piussan, C., Martin, G.M., Schellenberg, G.D., and Wijsman, E.M. (1996). Toward localization of the Werner syndrome gene by linkage disequilibrium and ancestral haplotyping: lessons learned from analysis of 35 chromosome 8p11.1-21.1 markers. American journal of human genetics *58*, 1286-1302.

Gottschling, D.E., and Zakian, V.A. (1986). Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA. Cell *47*, 195-205.

Gray MD, Shen JC, Kamath-Loeb AS, Blank A, Sopher BL, Martin GM, Oshima J, Loeb LA. The Werner syndrome protein is a DNA helicase. Nat Genet. 1997;17:100–103.

Gray, M.D., Wang, L., Youssoufian, H., Martin, G.M., and Oshima, J. (1998). Werner helicase is localized to transcriptionally active nucleoli of cycling cells. Exp Cell Res *242*, 487-494.

Greider, C.W., and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in tetrahymena extracts. Cell 43, 405-413.

Greider, C.W., and Blackburn, E.H. (1987). The telomere terminal transferase of tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell *51*, 887-898.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H., and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. Cell *97*, 503-514.

Hagelstrom, R.T., Blagoev, K.B., Niedernhofer, L.J., Goodwin, E.H., and Bailey, S.M. (2010). Hyper telomere recombination accelerates replicative senescence and may promote premature aging. Proc Natl Acad Sci U S A *107*, 15768-15773.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Harrigan, J.A., Wilson, D.M., 3rd, Prasad, R., Opresko, P.L., Beck, G., May, A., Wilson, S.H., and Bohr, V.A. (2006). The Werner syndrome protein operates in base excision repair and cooperates with DNA polymerase beta. Nucleic Acids Res *34*, 745-754.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Experimental Cell Research *25*, 585-621.

He, H., Multani, A.S., Cosme-Blanco, W., Tahara, H., Ma, J., Pathak, S., Deng, Y., and Chang, S. (2006). POT1b protects telomeres from end-to-end chromosomal fusions and aberrant homologous recombination. EMBO J *25*, 5180-5190.

Hemann, M.T., Rudolph, K.L., Strong, M.A., DePinho, R.A., Chin, L., and Greider, C.W. (2001a). Telomere Dysfunction Triggers Developmentally Regulated Germ Cell Apoptosis. Mol Biol Cell 12, 2023-2030.

Hemann, M.T., Strong, M.A., Hao, L.Y., and Greider, C.W. (2001b). The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. Cell *107*, 67-77.

Hendel, A., Ziv, O., Gueranger, Q., Geacintov, N., and Livneh, Z. (2008). Reduced efficiency and increased mutagenicity of translesion DNA synthesis across a TT cyclobutane pyrimidine dimer, but not a TT 6-4 photoproduct, in human cells lacking DNA polymerase eta. DNA Repair (Amst) *7*, 1636-1646.

Hile, S.E., and Eckert, K.A. (2004). Positive correlation between DNA polymerase alpha-primase pausing and mutagenesis within polypyrimidine/polypurine microsatellite sequences. J Mol Biol *335*, 745-759.

Hile, S.E., and Eckert, K.A. (2008). DNA polymerase kappa produces interrupted mutations and displays polar pausing within mononucleotide microsatellite sequences. Nucleic Acids Res *36*, 688-696.

Hile, S.E., Yan, G., and Eckert, K.A. (2000). Somatic Mutation Rates and Specificities at TC/AG and GT/CA Microsatellite Sequences in Nontumorigenic Human Lymphoblastoid Cells. Cancer Res *60*, 1698-1703.

Hockemeyer, D., Sfeir, A.J., Shay, J.W., Wright, W.E., and de Lange, T. (2005). POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. EMBO J *24*, 2667-2678.

Huang, S., Li, B., Gray, M.D., Oshima, J., Mian, I.S., and Campisi, J. (1998). The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease. Nature genetics *20*, 114-116.

Huang, S., Lee, L., Hanson, N.B., Lenaerts, C., Hoehn, H., Poot, M., Rubin, C.D., Chen, D.-F., Yang, C.-C., Juch, H., et al. (2006). The spectrum of WRN mutations in Werner syndrome patients. Human Mutation 27, 558-567.

Ivessa, A.S., and Zakian, V.A. (2002). To fire or not to fire: origin activation in Saccharomyces cerevisiae ribosomal DNA. Genes Dev *16*, 2459-2464.

Kamath-Loeb, A.S., Johansson, E., Burgers, P.M., and Loeb, L.A. (2000). Functional interaction between the Werner Syndrome protein and DNA polymerase delta. Proceedings of the National Academy of Sciences of the United States of America *97*, 4603-4608.

Kamath-Loeb, A.S., Loeb, L.A., Johansson, E., Burgers, P.M., and Fry, M. (2001). Interactions between the Werner syndrome helicase and DNA polymerase delta specifically facilitate copying of tetraplex and hairpin structures of the d(CGG)n trinucleotide repeat sequence. J Biol Chem *276*, 16439-16446.

Karlseder, J., Broccoli, D., Dai, Y., Hardy, S., and de Lange, T. (1999). p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. Science *283*, 1321-1325.

Karlseder, J., Smogorzewska, A., and de Lange, T. (2002). Senescence induced by altered telomere state, not telomere loss. Science *295*, 2446-2449.

Kim NW, Platyszek M, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW. (1994). Science, 266: 2011-2015

Kim, S.H., Kaminker, P., and Campisi, J. (1999). TIN2, a new regulator of telomere length in human cells. Nat Genet *23*, 405-412.

Kim, S., Parks, C.G., DeRoo, L.A., Chen, H., Taylor, J.A., Cawthon, R.M., and Sandler, D.P. (2009). Obesity and Weight Gain in Adulthood and Telomere Length. Cancer Epidemiology Biomarkers & Prevention 18, 816-820.

Kudlow, B.A., Kennedy, B.K., and Monnat, R.J., Jr. (2007). Werner and Hutchinson-Gilford progeria syndromes: mechanistic basis of human progeroid diseases. Nat Rev Mol Cell Biol *8*, 394-404.

Kusumoto, R., Dawut, L., Marchetti, C., Wan Lee, J., Vindigni, A., Ramsden, D., and Bohr, V.A. (2008). Werner protein cooperates with the XRCC4-DNA ligase IV complex in end-processing. Biochemistry *47*, 7548-7556.

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). Elevated telomere-telomere recombination in WRN-deficient, telomere dysfunctional cells promotes escape from senescence and engagement of the ALT pathway. Genes & Development *19*, 2560-2570.

Lee, J.Y., Yoon, J., Kihm, H.W., and Kim, D.S. (2008). Structural diversity and extreme stability of unimolecular Oxytricha nova telomeric G-quadruplex. Biochemistry *47*, 3389-3396.

Lei, M., Podell, E.R., and Cech, T.R. (2004). Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. Nat Struct Mol Biol *11*, 1223-1229.

Levy, M.Z., Allsopp, R.C., Futcher, A.B., Greider, C.W., and Harley, C.B. (1992). Telomere end-replication problem and cell aging. Journal of Molecular Biology *225*, 951-960.

Li, B., and Comai, L. (2001). Requirements for the Nucleolytic Processing of DNA Ends by the Werner Syndrome Protein-Ku70/80 Complex. Journal of Biological Chemistry *276*, 9896-9902.

Li, B., Oestreich, S., and de Lange, T. (2000). Identification of Human Rap1: Implications for Telomere Evolution. Cell *101*, 471-483.

Lillard-Wetherell, K., Machwe, A., Langland, G.T., Combs, K.A., Behbehani, G.K., Schonberg, S.A., German, J., Turchi, J.J., Orren, D.K., and Groden, J. (2004). Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2. Hum Mol Genet *13*, 1919-1932.

Lindner, S.E., and Sugden, B. (2007). The plasmid replicon of Epstein-Barr virus: mechanistic insights into efficient, licensed, extrachromosomal replication in human cells. Plasmid *58*, 1-12.

Lipps, H.J., and Rhodes, D. (2009). G-quadruplex structures: in vivo evidence and function. Trends in Cell Biology *19*, 414-422.

Liu, F.J., Barchowsky, A., and Opresko, P.L. (2009). The Werner syndrome protein functions in repair of Cr(VI)-induced replication-associated DNA damage. Toxicol Sci *110*, 307-318.

Liu, F.J., Barchowsky, A., and Opresko, P.L. (2010). The Werner syndrome protein suppresses telomeric instability caused by chromium (VI) induced DNA replication stress. PLoS ONE *5*, e11152.

Lombard, D.B., Beard, C., Johnson, B., Marciniak, R.A., Dausman, J., Bronson, R., Buhlmann, J.E., Lipman, R., Curry, R., Sharpe, A., *et al.* (2000). Mutations in the WRN Gene in Mice Accelerate Mortality in a p53-Null Background. Mol Cell Biol *20*, 3286-3291.

Luke, B., and Lingner, J. (2009). TERRA: telomeric repeat-containing RNA. EMBO J *28*, 2503-2510.

Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L., and Patel, D.J. (2006). Structure of the Human Telomere in K+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. Journal of the American Chemical Society *128*, 9963-9970.

Lyubchenko, Y.L. (2004). DNA structure and dynamics: an atomic force microscopy study. Cell Biochem Biophys *41*, 75-98.

Machwe, A., Xiao, L., Groden, J., Matson, S.W., and Orren, D.K. (2005). RecQ family members combine strand pairing and unwinding activities to catalyze strand exchange. The Journal of biological chemistry *280*, 23397-23407.

Maizels, N. (2006). Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat Struct Mol Biol *13*, 1055-1059.

Martinez, P., Thanasoula, M., Munoz, P., Liao, C., Tejera, A., McNees, C., Flores, J.M., Fernandez-Capetillo, O., Tarsounas, M., and Blasco, M.A. (2009). Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. Genes Dev *23*, 2060-2075.

McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in Zea Mays. Genetics *26*, 234-282.

McNees, C.J., Tejera, A.M., Martínez, P., Murga, M., Mulero, F., Fernandez-Capetillo, O., and Blasco, M.A. (2010). ATR suppresses telomere fragility and recombination but is dispensable for elongation of short telomeres by telomerase. The Journal of Cell Biology *188*, 639-652.

Mergny, J.L., Phan, A.T., and Lacroix, L. (1998). Following G-quartet formation by UV-spectroscopy. FEBS Lett *435*, 74-78.

Metcalfe, J.A., Parkhill, J., Campbell, L., Stacey, M., Biggs, P., Byrd, P.J., and Taylor, A.M. (1996). Accelerated telomere shortening in ataxia telangiectasia. Nat Genet *13*, 350-353.

Miller, K.M., Rog, O., and Cooper, J.P. (2006). Semi-conservative DNA replication through telomeres requires Taz1. Nature *440*, 824-828.

Mirkin, S.M. (2008). Discovery of alternative DNA structures: a heroic decade (1979-1989). Front Biosci *13*, 1064-1071.

Mohaghegh, P., and Hickson, I.D. (2002). Premature aging in RecQ helicase-deficient human syndromes. Int J Biochem Cell Biol *34*, 1496-1501.

Mohaghegh, P., Karow, J.K., Brosh Jr, R.M., Jr., Bohr, V.A., and Hickson, I.D. (2001). The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic Acids Res *29*, 2843-2849.

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). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A *85*, 6622-6626.

Muftuoglu, M., Oshima, J., von Kobbe, C., Cheng, W.-H., Leistritz, D., and Bohr, V. The clinical characteristics of Werner syndrome: molecular and biochemical diagnosis. Human Genetics.

Muller, H.J. (1938). The remaking of chromosomes. The Collecting Net *8*.

Murnane, J.P. (2010). Telomere Loss as a Mechanism for Chromosome Instability in Human Cancer. Cancer Research *70*, 4255-4259.

Neaves, K.J., Huppert, J.L., Henderson, R.M., and Edwardson, J.M. (2009). Direct visualization of G-quadruplexes in DNA using atomic force microscopy. Nucleic Acids Res *37*, 6269-6275.

O'Connor, M.S., Safari, A., Xin, H., Liu, D., and Songyang, Z. (2006). A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. Proceedings of the National Academy of Sciences *103*, 11874-11879.

Ogburn, C.E., Oshima, J., Poot, M., Chen, R., Hunt, K.E., Gollahon, K.A., Rabinovitch, P.S., and Martin, G.M. (1997). An apoptosis-inducing genotoxin differentiates heterozygotic carriers for Werner helicase mutations from wild-type and homozygous mutants. Hum Genet *101*, 121-125.

Opresko, P.L., Cheng, W.H., and Bohr, V.A. (2004). Junction of RecQ helicase biochemistry and human disease. J Biol Chem *279*, 18099-18102.

Opresko, P.L., Cheng, W.H., von Kobbe, C., Harrigan, J.A., and Bohr, V.A. (2003). Werner syndrome and the function of the Werner protein; what they can teach us about the molecular aging process. Carcinogenesis *24*, 791-802.

Opresko, P.L., Mason, P.A., Podell, E.R., Lei, M., Hickson, I.D., Cech, T.R., and Bohr, V.A. (2005). POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. J Biol Chem *280*, 32069-32080.

Opresko, P.L., Mason, P.A., Podell, E.R., Lei, M., Hickson, I.D., Cech, T.R., and Bohr, V.A. (2005). POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. The Journal of biological chemistry *280*, 32069-32080.

Opresko, P.L., Otterlei, M., Graakjaer, J., Bruheim, P., Dawut, L., Kolvraa, S., May, A., Seidman, M.M., and Bohr, V.A. (2004). The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. Mol Cell *14*, 763-774.

Opresko, P.L., Sowd, G., and Wang, H. (2009). The Werner Syndrome Helicase/Exonuclease Processes Mobile D-Loops through Branch Migration and Degradation. PLoS ONE *4*, e4825.

Opresko, P.L., von Kobbe, C., Laine, J.P., Harrigan, J., Hickson, I.D., and Bohr, V.A. (2002). Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. J Biol Chem *277*, 41110-41119.

Oshima, J., Huang, S., Pae, C., Campisi, J., and Schiestl, R.H. (2002). Lack of WRN results in extensive deletion at nonhomologous joining ends. Cancer Res *62*, 547-551.

Paeschke, K., Juranek, S., Simonsson, T., Hempel, A., Rhodes, D., and Lipps, H.J. (2008). Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates. Nat Struct Mol Biol *15*, 598-604.

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D., and Lipps, H.J. (2005). Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. Nat Struct Mol Biol *12*, 847-854.

Palm, W., and de Lange, T. (2008). How Shelterin Protects Mammalian Telomeres. Annual Review of Genetics *42*, 301-334.

Parris, C.N., and Seidman, M.M. (1992). A signature element distinguishes sibling and independent mutations in a shuttle vector plasmid. Gene *117*, 1-5.

Pennarun, G., Hoffschir, F., Revaud, D., Granotier, C., Gauthier, L.R., Mailliet, P., Biard, D.S., and Boussin, F.D. (2010). ATR contributes to telomere maintenance in human cells. Nucleic Acids Res *38*, 2955-2963.

Petersen, S., Saretzki, G., and von Zglinicki, T. (1998). Preferential accumulation of single-stranded regions in telomeres of human fibroblasts. Exp Cell Res *239*, 152-160.

Pichierri, P., Franchitto, A., Mosesso, P., and Palitti, F. (2001). Werner's syndrome protein is required for correct recovery after replication arrest and DNA damage induced in S-phase of cell cycle. Mol Biol Cell *12*, 2412-2421.

Pirzio, L.M., Pichierri, P., Bignami, M., and Franchitto, A. (2008). Werner syndrome helicase activity is essential in maintaining fragile site stability. J Cell Biol *180*, 305-314.

Platt, T.H., Tcherepanova, I.Y., and Schildkraut, C.L. (1993). Effect of number and position of EBNA-1 binding sites in Epstein-Barr virus oriP on the sites of initiation, barrier formation, and termination of replication. J Virol *67*, 1739-1745.

Poot, M., Gollahon, K.A., and Rabinovitch, P.S. (1999). Werner syndrome lymphoblastoid cells are sensitive to camptothecin-induced apoptosis in S-phase. Human Genetics *104*, 10-14.

PooT, M., Hoehn, H., Rünger, T.M., and Martin, G.M. (1992). Impaired S-phase transit of Werner syndrome cells expressed in lymphoblastoid cell lines. Experimental Cell Research *202*, 267-273.

Prescott, D.M. (1994). The DNA of ciliated protozoa. Microbiol Rev *58*, 233-267.

Puzianowska-Kuznicka, M., and Kuznicki, J. (2005). Genetic alterations in accelerated ageing syndromes: Do they play a role in natural ageing? The International Journal of Biochemistry & Cell Biology *37*, 947-960.

Ribeyre, C., Lopes, J., Boule, J.B., Piazza, A., Guedin, A., Zakian, V.A., Mergny, J.L., and Nicolas, A. (2009). The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. PLoS Genet *5*, e1000475.

Riethman, H. (2008). Human telomere structure and biology. Annu Rev Genomics Hum Genet *9*, 1-19.

Risitano, A., and Fox, K.R. (2003). Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexesâ€ Biochemistry *42*, 6507-6513.

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). Stabilization of quadruplex DNA perturbs telomere replication leading to the activation of an ATR-dependent ATM signaling pathway. Nucleic Acids Res *37*, 5353-5364.

Rocha, V., Devergie, G., SociÉ, G., Ribaud, P., EspÉrou, H., Parquet, N., and Gluckman, E. (1998). Unusual complications after bone marrow transplantation for dyskeratosis congenita. Br J Haematol *103*, 243-248.

Rossi, M.L., Ghosh, A.K., and Bohr, V.A. (2010). Roles of Werner syndrome protein in protection of genome integrity. DNA Repair (Amst) *9*, 331-344.

Roy, S., Tanious, F.A., Wilson, W.D., Ly, D.H., and Armitage, B.A. (2007). High-affinity homologous peptide nucleic acid probes for targeting a quadruplex-forming sequence from a MYC promoter element. Biochemistry *46*, 10433-10443.

Ruiz-Herrera, A., Nergadze, S.G., Santagostino, M., and Giulotto, E. (2008). Telomeric repeats far from the ends: mechanisms of origin and role in evolution. Cytogenet Genome Res *122*, 219-228.

Saharia, A., Guittat, L., Crocker, S., Lim, A., Steffen, M., Kulkarni, S., and Stewart, S.A. (2008). Flap endonuclease 1 contributes to telomere stability. Curr Biol *18*, 496-500.

Sakamoto, S., Nishikawa, K., Heo, S.J., Goto, M., Furuichi, Y., and Shimamoto, A. (2001). Werner helicase relocates into nuclear foci in response to DNA damaging agents and co-localizes with RPA and Rad51. Genes Cells *6*, 421-430.

Sallmyr, A., Tomkinson, A.E., and Rassool, F.V. (2008). Up-regulation of WRN and DNA ligase IIIalpha in chronic myeloid leukemia: consequences for the repair of DNA double-strand breaks. Blood *112*, 1413-1423.

Salvati, E., Leonetti, C., Rizzo, A., Scarsella, M., Mottolese, M., Galati, R., Sperduti, I., Stevens, M.F.G., D'Incalci, M., Blasco, M., et al. (2007). Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. The Journal of Clinical Investigation 117, 3236-3247.

Samani, N.J., and van der Harst, P. (2008). Biological ageing and cardiovascular disease. Heart 94, 537-539.

Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J., and Pluckthun, A. (2001). In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc Natl Acad Sci U S A *98*, 8572-8577.

Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J., and Aiyar, A. (2004). The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. J Virol *78*, 11487-11505.

Sfeir, A., Kosiyatrakul, S.T., Hockemeyer, D., MacRae, S.L., Karlseder, J., Schildkraut, C.L., and de Lange, T. (2009). Mammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication. Cell *138*, 90-103.

Sfeir, A.J., Chai, W., Shay, J.W., and Wright, W.E. (2005). Telomere-end processing the terminal nucleotides of human chromosomes. Mol Cell *18*, 131-138.

Shachar, S., Ziv, O., Avkin, S., Adar, S., Wittschieben, J., Reissner, T., Chaney, S., Friedberg, E.C., Wang, Z., Carell, T., *et al.* (2009). Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. EMBO J *28*, 383-393.

Shah, S.N., Opresko, P.L., Meng, X., Lee, M.Y., and Eckert, K.A. (2010). DNA structure and the Werner protein modulate human DNA polymerase delta-dependent replication dynamics within the common fragile site FRA16D. Nucleic Acids Res *38*, 1149-1162.

Shishkin, A.A., Voineagu, I., Matera, R., Cherng, N., Chernet, B.T., Krasilnikova, M.M., Narayanan, V., Lobachev, K.S., and Mirkin, S.M. (2009). Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell *35*, 82-92.

Shlyakhtenko, L.S., Potaman, V.N., Sinden, R.R., and Lyubchenko, Y.L. (1998). Structure and dynamics of supercoil-stabilized DNA cruciforms. J Mol Biol *280*, 61-72.

Sidorova, J.M., Li, N., Folch, A., and Monnat, R.J., Jr. (2008). The RecQ helicase WRN is required for normal replication fork progression after DNA damage or replication fork arrest. Cell Cycle *7*, 796-807.

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

Sowd, G., Wang, H., Pretto, D., Chazin, W.J., and Opresko, P.L. (2009). Replication protein A stimulates the Werner syndrome protein branch migration activity. J Biol Chem *284*, 34682-34691.

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

Sugden, B. (2002). In the beginning: a viral origin exploits the cell. Trends Biochem Sci *27*, 1-3.

Tauchi T, Shin-Ya K, Sashida G, Sumi M, Okabe S, et al. 2006.Telomerase inhibition with a novel G-quadruplex-interactive agent,telomestatin: in vitro and in vivo studies in acute leukemia. Oncogene 25:5719–5725

Tahara, H., Shin-Ya, K., Seimiya, H., Yamada, H., Tsuruo, T., and Ide, T. (2006). G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. Oncogene *25*, 1955-1966.

Takai, H., Smogorzewska, A., and de Lange, T. (2003). DNA damage foci at dysfunctional telomeres. Curr Biol *13*, 1549-1556.

Tang, J., Kan, Z.-y., Yao, Y., Wang, Q., Hao, Y.-h., and Tan, Z. (2008). G-quadruplex preferentially forms at the very 3' end of vertebrate telomeric DNA. Nucl Acids Res *36*, 1200-1208.

Tauchi T, Shin-Ya K, Sashida G, Sumi M, Okabe S, et al. 2006. Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. Oncogene 25:5719–5725.

Vallur, A.C., and Maizels, N. Distinct activities of exonuclease 1 and flap endonuclease 1 at telomeric g4 DNA. PLoS ONE *5*, e8908.

van Steensel, B., and de Lange, T. (1997). Control of telomere length by the human telomeric protein TRF1. Nature *385*, 740-743.

van Steensel, B., Smogorzewska, A., and de Lange, T. (1998). TRF2 Protects Human Telomeres from End-to-End Fusions. Cell *92*, 401-413.

Veldman, T., Etheridge, K.T., and Counter, C.M. (2004). Loss of hPot1 Function Leads to Telomere Instability and a cut-like Phenotype. Current Biology *14*, 2264-2270.

Verdun, R.E., and Karlseder, J. (2007). Replication and protection of telomeres. Nature *447*, 924-931.

Viguera, E., HernÃ¡ndez, P., Krimer, D.B., Boistov, A.S., Lurz, R., Alonso, J.C., and Schvartzman, J.B. (1996). The ColE1 Unidirectional Origin Acts as a Polar Replication Fork Pausing Site. Journal of Biological Chemistry *271*, 22414-22421.

Vorlickova, M., Chladkova, J., Kejnovska, I., Fialova, M., and Kypr, J. (2005). Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. Nucleic Acids Res *33*, 5851-5860.

Vulliamy, T.J., Knight, S.W., Mason, P.J., and Dokal, I. (2001). Very Short Telomeres in the Peripheral Blood of Patients with X-Linked and Autosomal Dyskeratosis Congenita. Blood Cells, Molecules, and Diseases *27*, 353-357.

Wang, C.-Y., and Sugden, B. (2008). Identifying a property of origins of DNA synthesis required to support plasmids stably in human cells. Proceedings of the National Academy of Sciences *105*, 9639-9644.

Wang, G., Christensen, L.A., and Vasquez, K.M. (2006b). Z-DNA-forming sequences generate large-scale deletions in mammalian cells. Proc Natl Acad Sci U S A *103*, 2677-2682.

Wang, G., and Vasquez, K.M. (2004). Naturally occurring H-DNA-forming sequences are mutagenic in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 13448-13453.

Wang, G., and Vasquez, K.M. (2006). Non-B DNA structure-induced genetic instability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *598*, 103-119.

Wang, H., Nora, G.J., Ghodke, H., and Opresko, P.L. (2011a). Single Molecule Studies of Physiologically Relevant Telomeric Tails Reveal POT1 Mechanism for Promoting G-quadruplex Unfolding. Journal of Biological Chemistry *286*, 7479-7489.

Wang, H., Nora, G.J., Ghodke, H., and Opresko, P.L. (2011b). Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. J Biol Chem *286*, 7479-7489.

Wang, J., Lindner, S.E., Leight, E.R., and Sugden, B. (2006c). Essential elements of a licensed, mammalian plasmid origin of DNA synthesis. Mol Cell Biol *26*, 1124-1134.

Watson, J.D. (1972). Origin of concatemeric T7 DNA. Nat New Biol *239*, 197-201.

Wentzensen, I.M., Mirabello, L., Pfeiffer, R.M., and Savage, S.A. (2011). The Association of Telomere Length and Cancer: A Meta-Analysis. Cancer Epidemiology Biomarkers & Prevention.

Wu, L., and Hickson, I.D. (2006). DNA Helicases Required for Homologous Recombination and Repair of Damaged Replication Forks. Annual Review of Genetics *40*, 279-306.

Wu, L., Multani, A.S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J.M., Bachilo, O., Pathak, S., Tahara, H., Bailey, S.M., *et al.* (2006). Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. Cell *126*, 49-62.

Wu, P., and de Lange, T. (2008). No Overt Nucleosome Eviction at Deprotected Telomeres. Mol Cell Biol *28*, 5724-5735.

Wu, X., and Maizels, N. (2001). Substrate-specific inhibition of RecQ helicase. Nucleic Acids Res *29*, 1765-1771.

Wu, Y., Shin-ya, K., and Brosh, R.M., Jr. (2008). FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA To Defend Genomic Stability. Mol Cell Biol *28*, 4116-4128.

Wyllie, F.S., Jones, C.J., Skinner, J.W., Haughton, M.F., Wallis, C., Wynford-Thomas, D., Faragher, R.G., and Kipling, D. (2000). Telomerase prevents the accelerated cell ageing of Werner syndrome fibroblasts. Nat Genet *24*, 16-17.

Xu, Y., Suzuki, Y., Ito, K., and Komiyama, M. (2010). Telomeric repeat-containing RNA structure in living cells. Proc Natl Acad Sci U S A *107*, 14579-14584.

Xu, Y., Ishizuka, T., Kurabayashi, K., and Komiyama, M. (2009). Consecutive Formation of G-Quadruplexes in Human Telomeric-Overhang DNA: A Protective Capping Structure for Telomere Ends. Angewandte Chemie International Edition 48, 7833-7836.

Yamaguchi, H., Calado, R.T., Ly, H., Kajigaya, S., Baerlocher, G.M., Chanock, S.J., Lansdorp, P.M., and Young, N.S. (2005). Mutations in TERT, the Gene for Telomerase Reverse Transcriptase, in Aplastic Anemia. New England Journal of Medicine *352*, 1413-1424.

Yang, Q., Xiang, J., Yang, S., Zhou, Q., Li, Q., Tang, Y., and Xu, G. (2009). Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: I. recognizing mixed G-quadruplex in human telomeres. Chem Commun (Camb), 1103-1105.

Yates, J.L., Warren, N., and Sugden, B. (1985). Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. Nature *313*, 812-815.

Yeager, T.R., Neumann, A.A., Englezou, A., Huschtscha, L.I., Noble, J.R., and Reddel, R.R. (1999). Telomerase-negative Immortalized Human Cells Contain a Novel Type of Promyelocytic Leukemia (PML) Body. Cancer Research 59, 4175-4179.

Yu, C.-E., Oshima, J., Fu, Y.-H., Wijsman, E.M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., *et al.* (1996a). Positional Cloning of the Werner's Syndrome Gene. Science *272*, 258-262.

Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DMA structures. Nature *350*, 718-720.

Zee, R.Y.L., Castonguay, A.J., Barton, N.S., Germer, S., and Martin, M. (2010). Mean leukocyte telomere length shortening and type 2 diabetes mellitus: a case-control study. Translational Research 155, 166-169

Zhang, M.L., Tong, X.J., Fu, X.H., Zhou, B.O., Wang, J., Liao, X.H., Li, Q.J., Shen, N., Ding, J., and Zhou, J.Q. Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. Nat Struct Mol Biol *17*, 202-209.

Zhang, P., Furukawa, K., Opresko, P.L., Xu, X., Bohr, V.A., and Mattson, M.P. (2006). TRF2 dysfunction elicits DNA damage responses associated with senescence in proliferating neural cells and differentiation of neurons. J Neurochem *97*, 567-581.

Zhao, J., Bacolla, A., Wang, G., and Vasquez, K.M. (2010). Non-B DNA structure-induced genetic instability and evolution. Cell Mol Life Sci *67*, 43-62.