TELOMERES AND TELOMERE BINDING PROTEINS IN *ARABIDOPSIS THALIANA*

A Dissertation

by

YEVGENIY SHAKIROV

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2004

Major Subject: Biochemistry

TELOMERES AND TELOMERE BINDING PROTEINS IN

ARABIDOPSIS THALIANA

A Dissertation

by

YEVGENIY SHAKIROV

Submitted to Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Dorothy Shippen (Chair of Committee)	Thomas McKnight (Member)
Gary Kunkel (Member)	David Giedroc (Member)
Gregory Reinhart (Head of Department)	

May 2004

Major Subject: Biochemistry

ABSTRACT

Telomeres and Telomere Binding Proteins in *Arabidopsis thaliana*. (May 2004)

Yevgeniy Shakirov, B.S.; Ph.D., Kazan State University, Russia

Chair of Advisory Committee: Dr. Dorothy E. Shippen

Telomeres are important protein-DNA structures at the ends of linear eukaryotic chromosomes that are necessary to prevent chromosome fusions and exonuclease attack. We found that telomere tracts in *Arabidopsis* are fairly uniformly distributed throughout a size range of 2-9kb. Unexpectedly, telomeres in WS plants displayed a bimodal size distribution with some individuals exhibiting 4-8 kb telomeres and others 2-5 kb telomeres. We also examined the dynamics of telomere tracts on individual chromosome ends. Following the fate of telomeres in plants through successive generations, we found that the shortest telomeres were typically elongated in the subsequent generation, while the longest telomeres were usually shortened. Thus, telomere length homoeostasis is achieved through intermittent telomerase action on shorter telomeres to attain an optimal size.

Single-strand telomere binding proteins were also analyzed. Four major telomere binding protein complexes from cauliflower were identified and their DNA-binding properties characterized. The DNA-binding component of one of the complexes was purified and analyzed by mass-spectrometry. Peptide mass data was used to search for putative protein candidates from the *Arabidopsis thaliana* database.

Additionally, two *Arabidopsis* genes, AtPot1 and AtPot2, were identified and characterized. The genes encode two single-strand telomeric DNA binding proteins. AtPot1 and AtPot2 proteins can homo- and heterodimerize *in vitro*. Pot1 protein predominantly localizes to the nucleolus, whereas Pot2 is exclusively nuclear. Plants over-expressing full-length Pot1 and Pot2 proteins had no obvious phenotype, while over-expression of $P2^{DBD}$ and $P1^{\Delta DBD}$ caused moderate telomere shortening. Plants over-expressing $P2^{DBD}$ had severe morphological and reproductive defects, multiple chromosome abnormalities and aneuploidy. Over-expression of a chimeric protein $P2^{DBD}$-$P1^{\Delta DBD}$ led to rapid telomere shortening, confirming the involvement of *Arabidopsis* Pot proteins in telomere length maintenance. Intriguingly, telomerase in $P2^{DBD}$-$P1^{\Delta DBD}$-EYFP plants is inactivated, suggesting that Pot proteins are also involved in regulation of telomerase activity. The analysis of *Arabidopsis* telomeres and telomere binding proteins will provide additional information towards understanding the role of the telomeric nucleoprotein complex in eukaryotic chromosome biology.

DEDICATION

To my mother, for her enormous effort to raise me as a good person and for all of her suffering. To my wife, for her love, support, understanding and patience. To all of my relatives, for their help and belief in me. To all of my friends on both sides of the ocean. I would not be able to make it through without you. Thank you.

ACKNOWLEDGEMENTS

First of all, I am very thankful to Dr. Dorothy Shippen for providing me with the opportunity to conduct research in her laboratory and encouraging me to explore new areas of telomere biology. Her help and advice allowed me to fully develop as a young scientist. I am also thankful to members of my committee, Dr. David Giedroc, Dr. Gary Kunkel and Dr. Thomas McKnight for their suggestions, discussions and guidance. Dr. Thomas McKnight has been collaborating with our laboratory for a number of years and his scientific discussions and ideas have been influential in helping me develop my project. He also provided me with materials and instruments essential for my work. I learned a great deal from him about plants and plant biology.

I also want to acknowledge Dr. Peter Baumann for sharing with us some initial protein sequence information. Dr. Naohiro Kato (Dr. Eric Lam lab, Rutgers University) performed some of the GFP fluorescence microscopy and was also very helpful in developing our strategies.

I am also very thankful to former and current members of the Shippen laboratory: Dr. Karel Riha for a lot of technical advice and also for inspiring me to do many of the experiments presented here, Kalpana Kannan, Dr. Laurent Vespa, Rachel Idol, Nathan Osbun and especially Yulia Surovtseva for a lot of help with many experiments. I am also very grateful to Dr. Zemfira Karamysheva, Jennifer Bullock, Michelle Heacock, and J. Matthew Watson for their advice and feedback.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The physical ends of most eukaryotic chromosomes terminate in specialized protective structures known as telomeres. Telomeres are composed of repetitive DNA and associated proteins. Telomeres not only protect the terminal chromosomal DNA from degradation and end-to-end fusions, but also are involved in its replication. Shortening of the telomeric DNA tract below a minimal length (defined differently for different organisms) prevents continued cell proliferation. As a result, telomeres have been implicated in cellular aging and in oncogenesis (reviewed in Goytisolo and Blasco, 2002; Harley and Villeponteau, 1995).

Telomeric DNA consists of tandem repeats of short, guanine-rich sequence. For most plants, the repeat is TTTAGGG, and for humans is TTAGGG. The number of repeats varies among species. In some lower eukaryotes this number is precisely defined, whereas in most higher eukaryotes telomere length is maintained within a defined range, from several hundred nucleotides to many thousand nucleotides (reviewed in Chakhparonian and Wellinger, 2003). Although this species-specific number of telomere repeats is variable, a certain minimum number is necessary to constitute a functional telomere. In budding yeast, an average of 325 bp are necessary and sufficient

---

This dissertation follows the style and format of the Plant Cell.

to comprise a functional telomere (Shampay and Blackburn, 1988). In *Arabidopsis*, from which the first higher eukaryotic telomere was cloned (Richards and Ausubel, 1988), the telomere tracts are 2-7 kb in length (Richards et al., 1992; Riha et al., 2002) (Figure 1). In mammals, telomeres extend up to 150 kb (Kipling and Cooke, 1990). The common feature of all telomeres is the presence of a G-rich strand running 5'-3' towards the chromosomal end. This guanine-rich strand is longer than the complementary C-rich strand, forming a 3' single strand protrusion called the G-overhang.

G-overhangs have been detected in many organisms. In the ciliate *Euplotes crassus*, the overhang is only 14 nt long (Klobutcher et al., 1981), whereas in humans it is approximately 150 nt (Makarov et al., 1997). In plants, the overhang is present on at least half of the telomeric ends and is estimated to be 20-30 nt long (Riha et al., 2000). The G-overhang is an essential feature of a functional telomere, and its deficiency is associated with end-to-end chromosomal fusions (van Steensel et al., 1998). The overhang is also involved in the formation of t-loops, in which this single-stranded region of the telomere folds back and invades the duplex telomeric repeats, displacing more internally located G-rich repeats (Griffith et al., 1999) (Figure 2). T-loops are thought to provide extra protection to the telomere. Although this mechanism does not appear to be universal, t-loops were found in many evolutionary distant species, including plants and animals (Griffith et al., 1999; Murti and Prescott, 1999; Munoz-Jordan et al., 2001; Cesare et al., 2003).

Telomeres are dynamic structures that can shorten or lengthen. An equilibrium between telomere shortening and lengthening activities is thought to be established in each cell, resulting in the formation of species-specific length of the telomere tract

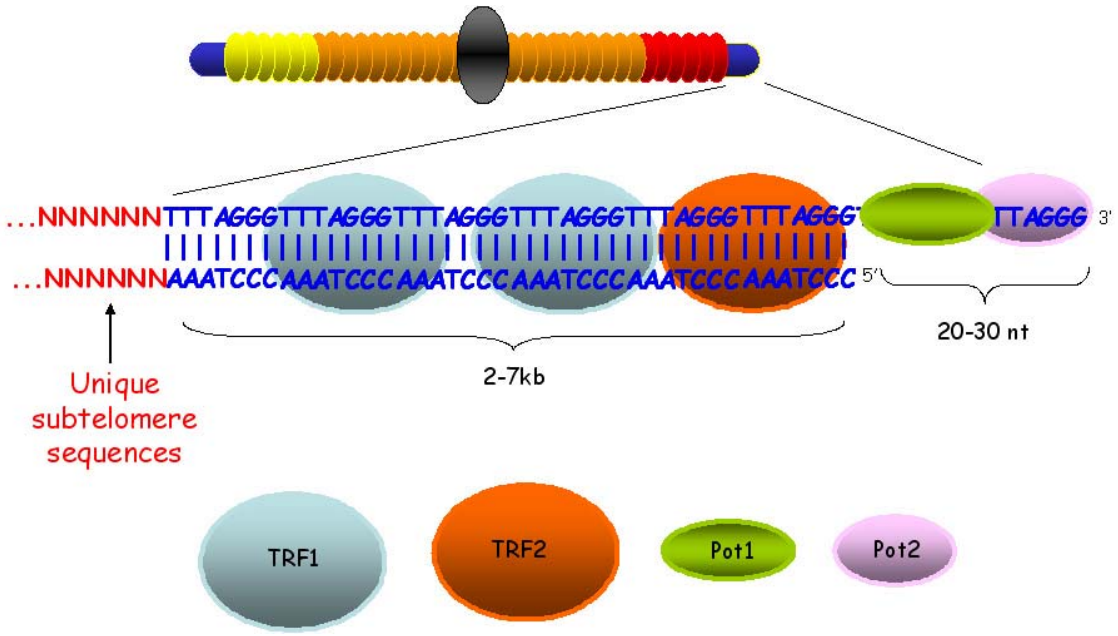

Figure 1. The structure of *Arabidopsis* telomeres.

Pot1, Pot2, TRF1 and TRF2 are putative telomere binding proteins.

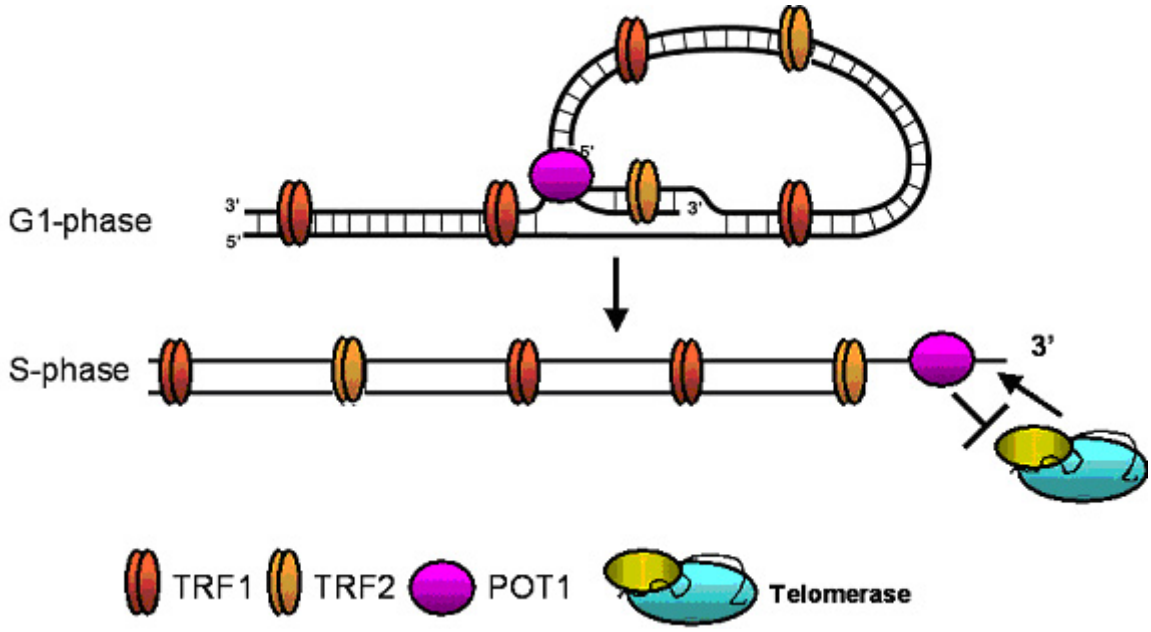

Figure 2. The t-loop model of telomere protection.

Telomeres are proposed to exist in the t-loop form through most of the cell cycle. In S-phase, the t-loop unfolds to provide access to conventional DNA replication machinery and telomerase. Only a small subset of the known telomere-associated proteins are shown here.

(reviewed in Greider, 1996). In wild-type cells, telomere shortening is thought to occur primarily as a result of incomplete replication of the DNA ends. The failure to completely replicate chromosomal ends originates in the semiconservative nature of DNA replication. When the last RNA primer from the extreme 5' terminus of the daughter strand is removed during the lagging strand synthesis, this single-stranded region can not be filled-in by the conventional DNA-replication machinery, leaving the newly synthesized telomeres a little shorter. This DNA amplification failure is called the "end-replication problem" (Olovnikov, 1973; Watson, 1972). Other, more active mechanisms of telomere shortening are also known. Examples include telomere rapid deletion (TRD) of exceptionally long telomeres (reviewed in Lustig, 2003) and telomere attrition caused by deficiency in key regulators of telomere length (Baumann and Cech, 2001). Recombination and active nuclease attacks are thought to be responsible for these rapid shortening events.

Telomere loss can be counteracted by the action of a specialized enzyme called telomerase (Greider and Blackburn, 1985). This enzyme contains two essential components: an RNA moiety and a reverse transcriptase (reviewed in Greider, 1996). The RNA subunit of telomerase contains a sequence complementary to the G-rich telomeric repeat. This sequence is used as a template for extending the G-overhang, which is performed by the reverse transcriptase subunit. The catalytic subunit of telomerase was designated TERT for telomerase reverse transcriptase. Although the two components described above are sufficient for telomerase activity *in vitro*, the holoenzyme is likely to contain additional components *in vivo*, that are involved in telomerase recruitment, activation or regulation of its activity *in vivo* (Gandhi and Collins, 1998; Greene and

Shippen, 1998; Lundblad and Szostak, 1989) (Figure 3). Telomerase-associated proteins have been cloned from several different species, including ciliates, yeast and mammals, but no conserved set of accessory proteins appears to be present in all organisms.

Genes encoding TERT subunits have been isolated from many organisms, including plants (Oguchi et al., 1999; Fitzgerald et al., 1999; Heller-Uszynska et al., 2002). Although all TERTs harbor several typical reverse transcriptase motifs and a telomerase-specific motif, they are not conserved outside of these regions. Similarly, RNA subunits from different organisms share almost no sequence similarity, but do form similar secondary structures (Lingner et al., 1994; Chen et al., 2000). Because of very little sequence conservation even between closely related organisms, genes encoding plant RNA subunits of telomerase have not yet been identified.

In many eukaryotes, telomerase expression and its activity are developmentally regulated. In the ciliate *Euplotes crassus*, new higher order telomerase complexes with distinct biochemical properties appear during development (Greene and Shippen, 1998), and both programmed ribosomal frame shifting and a switch between telomerase catalytic subunits appear to contribute to this change (Karamysheva et al., 2003). In humans, telomerase is only expressed in actively dividing tissues, such as germline and cancer cells (Wright et al., 1996), but not in most somatic cells. Similarly to that, telomerase is only detectable in reproductive or other actively dividing plant cells, such as flowers, root tips and callus (Fitzgerald et al., 1996; Heller et al., 1996; Riha et al., 1998).

Telomerase activity can be assayed *in vitro* using a single-stranded telomeric DNA primer, radio-labeled nucleotides and a cell-free extract (Greider and Blackburn,

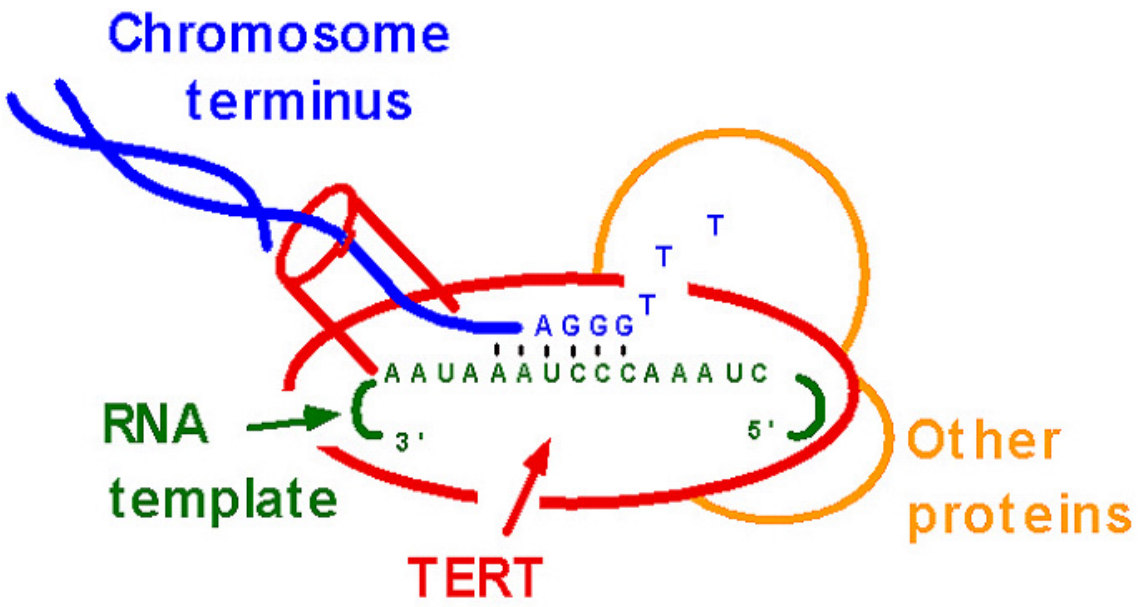

Figure 3. Telomerase is responsible for adding telomeric repeats.

1985). A more sensitive assay called the TRAP assay (telomere repeat amplification protocol) utilizes PCR for telomerase detection (Kim et al., 1994; Fitzgerald et al., 1996). In both the direct assay and the TRAP assay telomerase elongates the input primer, adding one nucleotide at a time to generate multiple tandem repeats of the telomeric sequence. When the products are resolved by PAGE, a ladder of bands is generated with the periodicity of the telomere repeat sequence. Although telomerase preferentially binds to telomeric sequence primers over non-telomeric sequences, many telomere-like sequences can be extended by the enzyme. Overall, telomerase has a general affinity for G-rich oligonucleotide sequences (Lee and Blackburn, 1993; Harrington and Greider, 1991).

Telomere length maintenance is crucial in the immortalization process. Since most somatic cells in higher eukaryotes are telomerase-negative, their telomeres shorten with each subsequent cell division. Because of this end-replication problem, normal human cells in culture divide only a finite number of times, after which they encounter a proliferative barrier (Hayflick, 1965). Over-expression of telomerase in such cells can overcome this barrier, lengthening the telomeres and extending the lifespan of the cells (Bodnar et al., 1998; Vaziri and Benchimol, 1998). Some cells do not respond to telomerase induction by lengthening their telomeres. Instead, telomeres continue to shorten and reach a new, shorter length which is maintained for some more cell divisions (Zhu et al., 1999; Yang et al., 1999).

The regulation of telomerase expression is complex and poorly understood. Transcription of the *hTERT* gene is likely to be the limiting factor regulating telomerase activity. The *hTERT* mRNA is not present in most somatic tissues and its expression is

elevated in most human cancer lines (Nakamura et al., 1997; Kilian et al., 1997; Meyerson et al., 1997). The *hTERT* promoter has been well-characterized and found to contain binding sites for many transcription factors. Transcription factor *c-Myc* can induce both *hTERT* expression and telomerase activity (Wang et al., 1998) by binding directly to the two E-box elements present in the *hTERT* promoter. Another transcription factor, *Sp1*, may also activate transcription of the *hTERT* gene. Mutations in *Sp1* binding sites abolish *hTERT* promoter activity (Kyo et al., 2000). Other regulators of *hTERT* expression may include *Mad1*, the E6 protein from human papilloma virus 16, *p53*, *WT-1* and many others.

Additional levels of telomerase regulation may include alternative splicing and other post-transcriptional modifications, such as phosphorylation. Various *hTERT* splice variants have been detected in different human tissues, suggesting a possible mechanism to produce alternative forms of *hTERT* molecules with somewhat different properties (Colgin et al., 2000; Yi et al., 2000). Protein phosphorylation is a common mechanism of regulating enzyme activity. In human cells, hTERT protein is regulated by phosphorylation. The tyrosin kinase c-ABL is able to phosphorylate hTERT both *in vitro* and *in vivo* (Kharbanda et al., 2000). Presumably, hTERT is phosphorylated in the cytoplasm in response to ionizing radiation. The phosphorylated hTERT protein then translocates to the nucleus, where it assembles into an active telomerase particle (Liu et al., 2001).

Telomerase-independent mechanisms for telomere maintenance have been previously reported (reviewed in Pardue and DeBaryshe, 1999). For example, telomeres in *Drosophila melanogaster* are composed of multiple copies of two retrotransposable

elements, HeT-A and TART. Transposition of HeT-A and TART yield arrays of repeats larger and more irregular than the repeats produced by telomerase (Pardue and DeBaryshe, 1999). In addition, certain telomerase-negative tumor cells and yeast use recombination-based mechanisms for chromosome end maintenance (Lundblad, 2002). Nonetheless, most eukaryotic cells depend on telomerase for chromosomal end maintenance. In the absence of telomerase, telomeres become progressively shortened triggering genome instability and cellular senescence (Shay and Wright, 2001; Forsyth et al., 2002; Riha et al., 2001; Artandi and DePinho, 2000; Riha et al., 2001).

Like most other eukaryotic organisms, *Arabidopsis* employs telomerase to maintain proper telomere length. The gene encoding the catalytic subunit of *Arabidopsis* telomerase (*AtTERT*) has been cloned (Fitzgerald et al., 1999; Oguchi et al., 1999) and shown to encode a typical reverse transcriptase with all the canonical telomerase protein motifs. The gene encoding the telomerase RNA subunit has not yet been cloned from any plant species. In *Arabidopsis*, as in humans, telomerase is only expressed in actively dividing tissues. As shown by the TRAP assay, telomerase activity is present in flowers, root tips and callus, but not vegetative tissues, such as leaves (Fitzgerald et al., 1996). The *Arabidopsis TERT* gene was the first gene shown to be involved in plant telomere biology. Disruption of this gene leads to progressive telomere shortening. For the first five or six generations, decreasing telomere length is the only apparent consequence of telomerase deficiency (Fitzgerald et al., 1999). However, from the sixth generation onward, major proliferation deficiency defects become apparent, including a decline in plant viability and growth (Riha et al., 2001). These defects result from progressively worsening genome instability as a consequence of chromosome fusions.

TELOMERE LENGTH REGULATION AND CHROMOSOME END PROTECTION IN CILIATES

Telomere length regulation is thought to be mediated by various proteins specifically associated with the telomeric DNA. This specialized group of proteins is called telomere end binding proteins (TEBP), and can be divided into two classes – those that bind along the length of the duplex region of the telomere and those that interact with the single-stranded G-overhang. Several examples of each class from evolutionary distant species have been well characterized both genetically and biochemically.

The first telomere binding protein was identified in the ciliate *Oxytricha nova*. This TEBP consists of two subunits, α and β, that form α–α homodimers or α–β heterodimers. The α–β heterodimer forms a very stable complex with the single-stranded G-overhang in 1:1:1 stoichiometry (Gray et al., 1991; Hicke et al., 1990). In the absence of β subunit, the α subunits form a dimer capable of less stable binding to the telomeric DNA (Fang et al., 1993). The α subunit of *O. nova* protein, as well as its homologue in *Euplotes crassus* (Price, 1990), binds the DNA via a specialized DNA-binding domain (N_TELO_alpha domain), which structurally folds into the oligosaccharide/oligonucleotide fold, known as OB fold (Murzin, 1993). The single-strand telomere binding proteins of ciliates are thought to stabilize telomeres by providing a cap at the extreme terminus.

SINGLE-STRAND TELOMERE BINDING PROTEIN IN BUDDING YEAST

Yeast telomeres have been shown to contain a 3' G-rich overhang of about 30 nucleotides in S-phase. A primary factor that binds to the G-overhang is the Cdc13p protein. Originally *CDC13* was identified as an essential gene involved in cell cycle control (Garvik et al., 1995). A temperature-sensitive allele of *CDC13*, *cdc13-1*, causes cell cycle arrest in the G2/M phase at the non-permissive temperature. In a later screen, *CDC13* was identified as one of four genes, mutations in which lead to the EST phenotype (ever shorter telomeres) (Nugent et al., 1996). A mutation in this gene, *cdc13-2 est*, results in phenotypes similar to those observed when the telomerase RNA subunit gene (TLC1) is deleted. In both cases, telomeres become progressively shorter eventually resulting in senescence. *CDC13* encodes a 924-amino acid protein of about 105 kDa (Garvik et al., 1995). Although analysis of its sequence does not identify any known DNA- or RNA-binding motifs, its DNA-binding domain can be localized to amino acids 451-693 (Wang et al., 2000). Gel mobility shift experiments showed that Cdc13p binds single-strand G-rich telomeric oligonucleotides with a binding affinity of 0.3 μM (Nugent et al., 1996). The protein is also capable of binding to human and ciliate telomeric sequences, though with much lower affinity. Interestingly, Cdc13p can also bind partially duplex substrates and does not require a free 3'-terminus for binding.

The *in vivo* and *in vitro* analysis of several Cdc13p mutations has provided insight into the protein's role at the telomeres. In *cdc13-1 ts* mutants, the C-strand of the telomere is rapidly lost at non-permissive temperature, resulting in a senescence phenotype. Interestingly, the *cdc13-2 est* strain exhibits none of the conditional lethality

or cell cycle arrest phenotypes of the *cdc13-1 ts* strain, and extensive single-strand telomeric DNA is not observed in this mutant. A *cdc13-2 est tlc1* double mutant exhibits the same degree of telomere shortening as *cdc13-2 est* or *tlc1* mutants alone, suggesting that both genes function in the same pathway. Since telomerase from *cdc13-2 est* is active *in vitro*, the *cdc13-2 est* mutation abolishes *CDC13* function that is required for regulation of telomerase *in vivo*, but not for enzyme activity *in vitro*. In contrast, the *cdc13-1 ts tlc1* double mutant had an exaggerated phenotype relative to the *tlc1* strain alone. Therefore, it was proposed that *cdc13-1 ts* alters a separate aspect of Cdc13p function. Based on these data Cdc13p was proposed to have two distinct roles at the telomere. First, Cdc13p physically protects the end of the chromosome, and second, it regulates telomerase access to the telomere (Nugent et al., 1996).

Much of our understanding of *CDC13* function in yeast cells comes from identification of Cdc13p-interacting proteins. One of these proteins, Stn1p, associates with Cdc13p by two-hybrid analysis (Grandin et al., 1997). An *STN1* deficiency causes accumulation of longer G-overhangs, and *stn1-13* cells exhibit abnormally long telomeres (Grandin et al., 1997). Another protein, Ten1p, was identified in a genetic screen as a suppressor of temperature-sensitive *stn1* mutations (Grandin et al., 2001). This protein physically associates with both Stn1p and Cdc13p, and defects in its binding to Stn1p are responsible for the long telomere phenotype of *stn1-13* mutants. In addition, several *ten1* mutations have been shown to cause telomerase-dependent telomere lengthening. Other temperature-sensitive mutants of *TEN1* accumulate single-stranded telomeric DNA (Grandin et al., 2001).

Based on the physical interaction between Cdc13p, Stn1p and Ten1p, these proteins are proposed to form a protective cap on the telomeres (reviewed in Evans and Lundblad, 2000). Therefore, one of the major functions of Cdc13p appears to be delivering this capping complex to chromosome ends. Each protein in this complex has a distinct role, with Cdc13p serving as the delivery vehicle and loading platform for associating factors, Stn1p being a major end capping protein and Ten1p assisting in this process. Mutations in one component of the complex can completely abolish activity of the complex as a whole.

CDC13 is also involved in positive regulation of telomerase access to the telomere. Positive regulation relies on the physical interaction between Cdc13p and Est1p, a non-catalytic component of telomerase (Qi and Zakian, 2000). Certain mutations in each protein that abolish their function also disturb the Cdc13p-Est1p interaction (Chandra et al., 2001). However, when the two mutations are simultaneously introduced into the same strain, telomere replication is completely restored. Analysis of the nature of these mutations showed that physical association of the two proteins depends on the presence of oppositely charged interacting residues (Pennock et al., 2001). The requirement for this type of protein-protein interaction can be bypassed if the 15 kDa DBD of Cdc13p protein is fused to Est1p in the absence of wild type Cdc13p protein. This experiment further demonstrated that Cdc13p is involved in telomerase recruitment *in vivo* (Pennock et al., 2001).

Interestingly, Ccd13p can also act as a negative regulator of telomere length. Certain mutations in *CDC13* lead to the telomere elongation phenotype. In these cells, the 3'-end overhangs on both strands are elongated as well. It is generally thought that in

addition to providing the capping function, Cdc13p-Stn1p-Ten1p complex coordinates leading and lagging strand DNA synthesis machineries, and mutations in any of these components disrupt this connection. While certain mutations in lagging strand replication machinery lead to similar telomere phenotypes (Adams Martin et al., 2000), there is evidence that Cdc13p and the catalytic subunit of polymerase α associate with each other *in vivo* (Qi and Zakian, 2000). It seems likely that the Cdc13p first recruits telomerase to the telomere and subsequently acts to limit telomerase action in response to the lagging strand DNA synthesis by the DNA polymerase α – primase complex.

RAP1P, A DOUBLE-STRAND TELOMERE BINDING PROTEIN IN BUDDING YEAST

In addition to single-strand terminus-specific binding proteins, telomeric DNA is bound by specific proteins along the length of the duplex telomeric region. In budding yeast, the major double-strand telomere DNA-binding protein is Rap1p (repressor-activator protein 1). In addition to binding telomeres, Rap1p is also involved in transcriptional control of many genes. Promoter regions of genes for many ribosomal proteins and glycolytic enzymes contain Rap1p binding sites, and *RAP1* function contributes to the transcriptional regulation of these genes (Hardy et al., 1992).

Most Rap1 protein molecules are concentrated at telomeres (Klein et al., 1992), where they bind DNA using two Myb-type helix-turn-helix (HTH) motifs (Konig et al., 1996) positioned in tandem in the middle of the polypeptide. Both HTH motifs bind independently to two repeated sequence elements 5'-GGGTGT-3' and 5'-GGTGT-3',

which are separated by 8 bp. This sequence arrangement corresponds to a simplified yeast telomeric repeat, TG_{1-3}.

RAP1 is an essential gene and encodes a negative regulator of telomere length. Every chromosome end in yeast contains about 18-20 high-affinity Rap1p binding sites. According to the protein-counting model of telomere length regulation, the number of telomere-bound Rap1p molecules determines whether a telomere will be elongated by telomerase (Marcand et al., 1997). This protein-counting feedback mechanism for telomere length regulation is independent of the number of telomeric repeats and relies solely on the number of telomere-bound Rap1 protein molecules. In wild-type yeast cells, the number of telomere-assembled Rap1p molecules is actively maintained at the same level. A telomere saturated with Rap1p molecules is prevented from elongation. When degradation or incomplete replication causes the loss of extra Rap1p-binding sites, the telomere is switched to a new state that favors its elongation. Once the number of the missing Rap1p-binding sites is restored, more Rap1 proteins can be bound and the telomere switches back to the telomerase-inaccessible state.

Once bound to the telomere, Rap1p recruits additional proteins to form a higher-order complex. Two sets of proteins, the Sir proteins (Moretti et al., 1994) and the Rif proteins (Wotton and Shore, 1997), associate with Rap1p on the telomeres via the Rap1 C-terminus. Sir3p and Sir4p form a complex with Sir2p to induce the transcriptional repression of subtelomeric genes (telomeric silencing) (Kyrion et al., 1992), while Rif1p and Rif2p are important for negative regulation of telomere length. When the RIF genes are deleted, telomeres lengthen in a telomerase-dependent manner (Wotton and Shore, 1997). Similarly, mutations in the conserved C-terminal domain of Rap1p lead to

telomerase-dependent telomere length increase (Kyrion et al., 1992). It is generally thought that Sir and Rif proteins bind to the same portion of Rap1p C-terminus and antagonize each other *in vivo* by competing for Rap1p binding. This competition is cell cycle regulated. It is known that occupancy of telomeres by the structural proteins Rap1p, Sir3p, Sir4p, Rif1p and Rif2p is lower in late G2/M than in G1 or S (Laroche et al., 2000). Also, telomerase acts in late S or G2/M but not in G1 (Marcand et al., 2000). This suggests that at least some components of the protein complexes at telomeres are dynamically assembled and disassembled, and Rap1p is a key regulatory factor of this machinery.

TELOMERE-ASSOCIATED FACTORS IN FISSION YEAST

Although the genome of *S. pombe* does contain an *S. cerevisiae* RAP1 sequence homologue, this protein does not directly bind telomeres in the fission yeast and instead associates with the telomeric chromatin via protein-protein interaction with Taz1p. Taz1p is the major double-strand telomere binding factor in the fission yeast *S. pombe*. The *taz1*$^{+}$ gene was identified in a one-hybrid screen for genes encoding double-strand telomere binding factors, and was shown to be a negative regulator of telomere length (Cooper et al., 1997). Taz1p binds to telomeres directly both in mitotic and meiotic cells and its function is required for a diverse range of telomere functions in fission yeast. Although *taz1*$^{+}$ function is not essential for vegetative growth under native conditions, its deficiency results in dramatically elongated telomeres, increased length of the G-overhang, and the loss of the telomere position effect (Cooper et al., 1997).

When arrested in G1 phase of the cell cycle, Taz1-deficient cells lose viability and accumulate telomere fusions through the Ku-dependent non-homologous end-joining (NHEJ) pathway (Ferreira and Cooper, 2001). In addition, the long telomeres in *taz1Δ* cells at cold temperatures become entangled with each other, leading to double-strand DNA breaks and chromosome missegregation (Miller and Cooper, 2003). These observations demonstrate that Taz1p normally contributes to the formation of the protective cap at the telomeres and restricts NHEJ from recognizing a chromosome terminus as an unrepaired double-strand break.

Taz1 function is also essential for progression into meiosis. In the absence of *taz1*$^{+}$, telomeres do not stably cluster at the spindle pole body, which leads to defective nuclear divisions (Cooper et al., 1997). Meiotic recombination and spore viability are greatly reduced in *taz1* mutants. Thus, Taz1 appears to play several important roles in the mitotic and meiotic yeast cells.

In terms of overall organization, Taz1p is similar to human TRF1 and TRF2 proteins (see below). It contains a single Myb-like domain at the C-terminus and a TRF homology domain in the central region. *In vitro,* Taz1p binds a fission yeast telomeric DNA fragment in a sequence specific manner (Spink et al., 2000). The binding requires Taz1p dimerization, which happens in the absence of DNA. The minimal DNA fragment bound by the full length Taz1p dimer contains two copies of the GGTTA sequence.

Taz1p regulates telomere length in fission yeast possibly by providing binding sites and recruiting other telomere length regulators, Rap1p and Rif1p, which may perform essential telomere functions (Kanoh and Ishikawa, 2001). *Schizosaccharomyces pombe* rap1^{+} and rif1^{+} were identified in a database search based on their similarity to

Saccharomyces cerevisiae RAP1 and *RIF1* (Kanoh and Ishikawa, 2001). In the mutant *rif1* strain, the telomeric DNA is 200 bp longer than in the wild type. Similarly, telomeres are significantly elongated in *rap1* mutants. Double mutants with *taz1* displayed telomere length similar to that of the single *taz1* mutant, suggesting that Rif1p and Rap1p function in a Taz1p-dependent manner. The *rap1 rif1* double mutant showed longer telomeres than either of the single *rap1* or *rif1* mutants, indicating that Rif1p and Rap1p function in separate telomere maintenance pathways. Consistent with that, both Rif1p and Rap1p can bind Taz1p in the absence of each other and do not interact in yeast two-hybrid screen. Both proteins require Taz1p for their telomere localization *in vivo*, but Taz1p can bind telomeres in the absence of Rif1p and Rap1p (Kanoh and Ishikawa, 2001). Therefore, it is thought that the primary role of Taz1p is to recruit Rif1p, Rap1p and, possibly, other telomere regulators.

TELOMERE BINDING PROTEINS IN HUMANS

Human telomeres are bound by two major double-strand telomere binding proteins, TRF1 and TRF2. Both proteins were identified on the basis of their ability to bind human telomeric oligonucleotides *in vitro* (Chong et al., 1995; Broccoli et al., 1997). As with most other double-strand telomere binding proteins, TRF1 and TRF2 contain a Myb-like helix-turn-helix DNA-binding domain at their C-terminus. TRF1 and TRF2 proteins share 56% identity in their Myb domain. The central dimerization domain is responsible for the formation of TRF1 and TRF2 homodimers (Bilaud et al., 1997). The most variable region of the proteins is the N-terminal domain, which is acidic in TRF1

and basic in TRF2. Both proteins specifically bind telomeric DNA *in vivo* (Broccoli et al., 1997; Bilaud et al., 1997). TRF1 dimer binds adjacent YTAGGGTTR sites on the DNA. The bound TRF1 dimers loop the sequence between the two half sites (Bianchi et al., 1997) and also have the ability to bring two telomeric tracts together (Griffith et al., 1998). Both TRF1 and TRF2 are extremely important for telomere length regulation (van Steensel and de Lange, 1997; Smogorzewska et al., 2000). Over-expression of TRF1 and TRF2 in human cell lines results in a gradual decline in telomere length, whereas over-expression of a dominant negative form of TRF1 leads to telomere elongation. These studies suggest that telomere-bound TRF1 is an inhibitor of telomere elongation.

One of the many important roles of TRF1 proteins at the telomere is the ability to interact with other telomere proteins, such as TIN2, Ku and tankyrase. The function of the conserved Ku heterodimer at the telomeres will be discussed below. Tankyrase is a poly (ADP-ribose) polymerase, containing 24 typical ankyrin repeats and an N-terminal acidic domain. The human tankyrase promotes telomere elongation by modifying TRF1 and inhibiting its binding to telomeric DNA (Smith et al., 1998). Over-expression of tankyrase induces ADP ribosylation of TRF1, releasing it from the telomere. Subsequently, the lack of a sufficient number of bound TRF1 molecules leads to gradual telomere elongation by telomerase.

Another telomeric protein, TIN2, interacts with the central domain of TRF1 and co-localizes with TRF1 at telomeres (Kim et al., 1999b). Over-expression of a dominant-negative TIN2 lacking a portion of its amino-terminal sequence induces telomere elongation only in telomerase-positive, but not telomerase-negative human cells, suggesting that TIN2 may also be a negative regulator of telomere elongation. It has been

proposed that TIN2 is an essential component of the TRF1 pathway and may act by promoting the formation of a large, multimeric complex at the telomeres (Kim et al., 1999b).

Recently, a novel TRF1-interacting protein called PinX1 was identified in a yeast two-hybrid screen (Zhou and Lu, 2001). Over-expression of PinX1 induces telomere shortening, leading to genome instability. PinX1 depletion results in elongation of telomeres. Remarkably, PinX1 was shown to inhibit telomerase activity both *in vitro* and *in vivo* by direct interaction with hTERT. Therefore, PinX1 is the first known telomerase inhibitor *in vivo*.

The second mammalian duplex telomere protein is TRF2. This protein binds along the length of the double-stranded telomere region with more than 100 protein molecules per chromosome end (Bilaud et al., 1997; Broccoli et al., 1997). Like TRF1 and TIN2, TRF2 is a negative regulator of telomere length (Smogorzewska et al., 2000). In recent years, TRF2 has also emerged as the major telomere end protection factor. Over-expression of a dominant-negative version of TRF2 results in the disappearance of endogenous TRF2 from telomeres, induction of p53-damage response pathway and chromosome end fusions (van Steensel et al., 1998; Karlseder et al., 1999). The end-to-end telomere fusions result from loss of the G-overhang. The G-overhang is an essential feature of functional telomeres, and its absence results in telomeres being recognized as double-strand DNA breaks. The cellular machinery responsible for repairing double-strand DNA breaks in higher eukaryotes is a multisubunit protein complex associated with the NHEJ pathway. This pathway is activated in cells lacking TRF2 and is thought to be responsible for fusing telomere ends in these mutants. In some cell types, TRF2

inhibition also induces apoptosis (Karlseder et al., 1999). Therefore, the loss of TRF2 from telomeres creates an altered chromosome end structure capable of activating the DNA damage response pathway and ultimately resulting in senescence or apoptosis.

TRF2 also protects critically shortened telomeres. In culture, primary human lung fibroblasts grow for 55 population doublings before reaching senescence with an average telomere length of 7 kb (Karlseder et al., 2002). Over-expression of wild-type TRF2 allows these cells to divide longer, with telomere length at senescence reaching only 4 kb (Karlseder et al., 2002). These cells also accumulate fewer chromosomal aberrations, suggesting that high levels of TRF2 protein provide better protection of chromosome ends. The extra stability provided by TRF2 over-expression could result from more efficient t-loop formation (Griffith et al., 1999). The current model suggests that in addition to TRF2's role in the inhibition of NHEJ processes at the natural chromosomal termini, this protein is also involved in chromosome end processing and overhang formation, which in turn are necessary to form stable t-loops (Karlseder, 2003; de Lange, 2002).

TRF2 also interacts with hRap1 and recruits it to telomeres. hRap1 protein was identified in a standard two-hybrid screen using TRF2 as bait (Li et al., 2000). Based on the sequence similarity, hRap1 was classified as a human ortholog of the *Saccharomyces cerevisiae* telomere protein Rap1. Human Rap1 is a 399 amino acid protein and contains significant sequence identity (24-25%) to three different domains in the yeast Rap1p. As in yeast, hRap1 harbors an N-terminal BRCT domain, but contains only one central DNA-binding Myb domain. Consistent with the data in yeast, human Rap1 is concentrated at telomeres throughout the cell cycle. Although hRap1 does not bind

telomeric DNA directly, it forms a complex with TRF2 on telomeric DNA. Over-expression of full-length hRap1 leads to a gradual telomere shortening, consistent with its role as a negative regulator of telomere length (Li et al., 2000).

DNA DAMAGE RESPONSE PROTEINS AS IMPORTANT COMPONENTS OF THE TELOMERIC COMPLEX

Double-strand breaks (DSB) in DNA are easily generated by reactive oxygen radicals, exposure to ionizing radiation and during V(D)J recombination in lymphocytes. In mammals, the DNA-PK complex, consisting of the Ku70/Ku80 heterodimer, DNA-PK_{cs}, DNA ligase IV and other factors, is responsible for repairing DNA double-strand breaks. The Ku heterodimer directly binds with high affinity to the broken DNA ends, and as such is a critical component of the NHEJ pathway (Ramsden and Gellert, 1998). In addition to its function in DNA repair, Ku has also been implicated in the maintenance of functional telomeres. In particular, yeast cells defective in either Ku subunit show a 60% loss of telomeric repeats, the loss of telomere clustering and telomeric silencing, and deregulation of the G-strand overhang (Boulton and Jackson, 1998; Gravel et al., 1998; Laroche et al., 1998). Fission yeast Ku mutants show elevated levels of subtelomeric DNA rearrangements (Baumann and Cech, 2000). Mammalian Ku can also bind to telomeric sequences (Bianchi and de Lange, 1999) and prevent end-to-end fusions (Bailey et al., 1999). In yeast, the Ku heterodimer is postulated to play a direct role in establishing a normal DNA end structure on chromosomes, conceivably by functioning as a terminus-binding factor. Furthermore, upon induction of chromosomal DNA damage,

Ku moves from the telomeres to DSB sites, suggesting a link between DNA repair and telomeres (Martin et al., 1999). Thus, it has been viewed as paradoxical that Ku should play important roles both in promoting NHEJ and in comprising functional telomeres, which strictly prohibit NHEJ. Nonetheless, the exact role of Ku at the telomeres appears to vary in different organisms.

In *Arabidopsis,* the Ku70/Ku80 heterodimer is also a major telomere length regulator. As discussed above, Ku plays a role in double-strand break DNA repair and in telomere protection and maintenance. Disruption of *Arabidopsis Ku70* gene leads to dramatic telomere lengthening, with telomeres of the second generation knock-out plants reaching up to 20 kb (Riha et al., 2002). This finding is in striking contrast to the situation in yeast, where telomeres shorten in the absence of Ku (Boulton and Jackson, 1998; Gravel et al., 1998; Polotnianka et al., 1998). Furthermore, in contrast to the situation in mammals (Espejel et al., 2002), Ku70-deficient *Arabidopsis* mutants do not show any chromosome fusions or rearrangements (Riha et al., 2002). Therefore, the function of the Ku heterodimer varies in different organisms.

Telomere elongation in Ku mutants depends on the presence of telomerase. In double *tert ku70 Arabidopsis* mutants, telomeres continue to shorten as in the single tert mutants, but the rate of shortening is dramatically accelerated (Riha and Shippen, 2003a). The same result was observed in double *tert ku80* mutants (Gallego et al., 2003), suggesting that in wild-type *Arabidopsis* Ku functions as a negative regulator of telomere extension by telomerase.

In addition to inhibiting telomerase, Ku has another function at telomeres. Both *ku70* and *tert ku70* mutant plants show increased length of the G-overhang, suggesting

that *Arabidopsis* Ku is also required for the maintenance of the telomeric C-strand or for coordination of leading and lagging strand DNA synthesis machineries at the telomeres (Riha and Shippen, 2003a). This finding is further supported by the accelerated loss of the telomere tract in the double *tert ku70* double mutants, consistent with a major problem with C-strand maintenance. In addition, unlike the situation in the single *ku70* mutants, *tert ku70* double mutants display abundant chromosome fusions, suggesting that *Arabidopsis* employs an alternative, Ku-independent mechanism to generate telomere end-to-end fusions (Riha and Shippen, 2003a).

Other proteins, shown to play a role at *Arabidopsis* telomeres are the components of the MRX complex (Rad50, Mre11, Xrs1), which is involved in the double-strand break DNA repair pathway (D'Amours and Jackson, 2002). Plants deficient in *RAD50* show meiotic and DNA repair defects and are sterile (Gallego et al., 2001). Although no telomere defects are observed in the mutant plants, dedifferentiated cells (callus) derived from *rad50* null mutants display progressive telomere shortening followed by premature crisis and cell death. Only a small fraction of the mutant cells survive, probably by activating a Rad50-independent mechanism of telomere extension. Such cells display much longer telomeres than wild-type (Gallego and White, 2001).

Like the situation in *rad50* mutant plants, an Mre11 deficiency also causes severe sensitivity to DNA damaging agents such as MMS and X-rays. In addition, Mre11-deficient plants showed increased telomere length, suggesting that the *Arabidopsis* MRX complex is involved in telomere length regulation mechanism (Bundock and Hooykaas, 2002).

SINGLE STRAND TELOMERE BINDING PROTEINS IN HIGHER EUKARYOTES AND FISSION YEAST

The G-overhang appears to be a universal feature of all eukaryotic telomeres, and G-overhang binding proteins are hypothesized to be present in all species across evolution. These proteins are thought to perform a variety of functions, including protection of telomere ends from degradation and double-strand break repair activities, as well as recruitment of telomerase and other factors needed for proper telomere length maintenance (reviewed in Wei and Price, 2003). Until recently, such proteins were only identified in lower eukaryotes. Although single-strand telomere binding activities were previously detected *in vitro* (Petracek et al., 1994; Kim et al., 1998; Lee et al., 2000), no genes encoding such factors were cloned from higher eukaryotes. In 2001, a database search revealed that the *S. pombe* genome contains an open reading frame with a significant degree of similarity to the α subunit of the ciliate telomere proteins (Baumann and Cech, 2001). This gene was designated *pot1*$^{+}$ for protection of telomeres. Detailed sequence analysis demonstrated that putative homologues of Pot1 proteins exist in other higher eukaryotes, including plants, chicken, yeast and fungi (Baumann and Cech, 2001; Baumann et al., 2002; Pitt et al., 2004; Wei and Price, 2004). Most of the sequence similarity is concentrated in the N-terminal TELO_N_alpha DNA-binding domain, which forms the known OB-fold structure characteristic of all single-strand telomere binding proteins (Murzin, 1993).

Much of our understanding of Pot1 function has come from the analysis of human and *S. pombe* Pot1 proteins. These proteins specifically bind to the G-rich telomeric

strand of the corresponding organism (Baumann and Cech, 2001) and not to the C-rich strand or to duplex telomeric repeats. Pot1 protein from *S. pombe* binds telomeric DNA in a highly cooperative manner with an apparent K_d of 83 nM. The minimal-length telomeric substrate for this protein is GGTTAC, which represents a simplified *S. pombe* telomeric repeat. The human Pot1 protein specifically binds to the telomeric nonamer TAGGGTTAG, with the optimal substrate having at least two full telomeric repeats. Unlike its ortholog in *S. pombe*, human Pot1 does not have a preference for the proximity of a 3' end, and can bind to internally positioned TTAGGG repeats (Loayza et al., 2004).

Immunolocalization studies in human cells indicate that Pot1 localizes to telomeres, and loss of the G-overhang causes a reduction in Pot1 binding (Baumann et al., 2002; Loayza and De Lange, 2003). Surprisingly, Pot1 was shown to associate not only with the G-overhang, but also with the duplex telomere region (Loayza and De Lange, 2003). The association with the double-stranded telomeric DNA is indirect and is mediated by protein-protein interactions with TRF1 (see below).

Pot1 protein in fission yeast is essential for proper telomere maintenance. Deletion of the *pot1*$^+$ gene causes immediate loss of telomeric and some subtelomeric DNA and cell death. The few surviving yeast cells manage to survive by circularizing all of their chromosomes (Baumann and Cech, 2001). The human Pot1 protein is also involved in telomere biology. Over-expression of either the full-length protein or a C-terminal fragment lacking the DNA-binding domain causes dramatic telomere lengthening (Colgin et al., 2003; Loayza and De Lange, 2003). In *pot1* mutant yeast cells, telomeric DNA is lost at a much faster rate than in a telomerase deficient strain, suggesting that Pot1p is more important for telomere maintenance in the short term than

telomerase. The human Pot1 homologue is expressed in all tissues, consistent with the Pot1's function as a housekeeping gene required to ensure the integrity of chromosome ends in all cells.

Although the exact mechanism is unknown, Pot1 seems to participate in the molecular communication process between telomerase and the negative TRF1-dependent telomere length regulation machinery (Loayza and de Lange, 2003). The current model predicts that the physical interaction between TRF1 and Pot1 affects the loading of Pot1 on the single-stranded telomeric DNA and transmits telomere length information to positively or negatively regulate telomerase at the telomere terminus.

The obvious telomere length defects observed in the Pot1 mutant human and yeast cells are not the only consequence of Pot1 deficiency. It has become increasingly clear that Pot1 is also involved in mitotic progression, probably due to the involvement of the telomeres in the establishment of the proper nuclear architecture. Deletion of the *S. pombe* $pot1^{+}$ gene results in a cell division defect leading to cell elongation and a high incidence of chromosome missegregation (Baumann and Cech, 2001). Similarly, a $pot1^{+}$ deficiency in *Aspergillus nidulans* leads to a variety of mitotic defects without cell cycle arrest, increased chromosomal instability and missegregation, and a loss of viability (Pitt et al., 2004). The combined data on Pot1 deficiencies in different organisms suggests that Pot1 provides a link between telomere structure and mitotic control.

ARABIDOPSIS AS A MODEL FOR TELOMERE BIOLOGY

Arabidopsis thaliana is a mustard weed plant with a very small genome (125 Mb), a short life cycle (only 6-8 weeks per generation) and well-established genetic tools. For years, *Arabidopsis* has been used as a model system to study plant development, floral morphogenesis and other aspects of plant biology. In 1988, the first higher eukaryotic telomere sequence TTTAGGG was cloned from *Arabidopsis* (Richards and Ausubel, 1988). Subsequently, maize, tomato and many other plants were shown to contain the same canonical telomere repeats (reviewed in Riha and Shippen, 2003b). Interestingly, members of the onion family (Alliaceae) contain human-like sequences TTAGGG (Fuchs, 1995), probably due to mutations in the template region of the telomerase RNA subunit. The length of the double-stranded telomere tract varies among different plant species. In *Arabidopsis*, telomeres are on average 2-7 kb, depending on ecotype. Although natural variation contributes to plant-to-plant difference in telomere length, the overall size of the telomere tract is very well controlled. There appears to be no difference in telomere length between cells in different *Arabidopsis* organs (Riha et al., 1998). The only exception is in callus, in which telomeres substantially elongate.

Arabidopsis telomeres also possess G-overhangs, which are on average 20-30 nucleotides long (Riha et al., 2000). The overhangs are predicted to exist on both chromosome ends, but can currently only be detected on 50% of all termini. Although no experiments have been done to address the existence of t-loops on *Arabidopsis* telomeres, the size of the telomere tract is likely to allow t-loop formation. Moreover, t-loops were shown to exist in another plant species, garden peas (Cesare et al., 2003).

Arabidopsis thaliana represents a good avenue for studying aspects of telomere biology (Riha and Shippen, 2003b). The genome sequence was completed in 2000 and it harbors many genes encoding proteins that are homologous to known telomere proteins from other model systems (*Arabidopsis* Genome Initiative, 2000). The availability of T-DNA insertion lines (http://signal.salk.edu/cgi-bin/tdnaexpress) allows for the identification of knock-outs of genes of interest. Also, unlike mammals, plants display an exceptional tolerance to genome instability and major stresses, such as DNA damage, chromosome fusions and telomere deregulation (Riha et al., 2001). Many of the genes involved in sensing and repair of these abnormalities are essential in other higher eukaryotes, rending it difficult or impossible to study mutations in these genes. In striking contrast, mutations in many *Arabidopsis* genes encoding functional homologues of these proteins are often not lethal and allow comprehensive analysis of their function. Double and triple mutants are also possible to generate in *Arabidopsis* by simple genetic crosses, permitting detailed analysis of multiple phenotypes and allowing one to determine their function within a complex pathway or network. Functional analysis of *Arabidopsis* genes may reveal new and essential information into the regulation of telomere homeostasis and other telomere-related processes. Given the strong correlation between many aspects of telomere biology in *Arabidopsis* and mammals, such information should be applicable to a wide range of organisms besides plants.

DISSERTATION OVERVIEW

Telomeres are essential components of eukaryotic chromosomes, providing protection against chromosome fusions and contributing to the overall maintenance and stability of the genome. Telomeric DNA packaged with specific telomere binding proteins forms a protective cap on the physical ends of eukaryotic chromosomes. The work presented in this dissertation describes telomere length regulation and the identification and analysis of single-strand telomere binding proteins in the model plant *Arabidopsis thaliana*.

Although *Arabidopsis* was the first higher eukaryotic organism from which a telomere was cloned, very little is known about the regulation of telomere length. In Chapter II, I show that wild-type *Arabidopsis* ecotypes display a two-fold variation in the length of their telomere tract. I demonstrate that plants from the WS ecotype exhibit a bimodal telomere size distribution. I also employed unique subtelomeric sequences to follow the fate of individual chromosome end for several successive plant generations. My findings indicate that telomere length homeostasis in *Arabidopsis* is achieved through the preferential action of telomerase on shorter telomeres in the population.

In Chapter III I describe the identification and analysis of two *Pot1*-like proteins in *Arabidopsis*, AtPot1 and AtPot2. I provide evidence that *Pot1* and *Pot2* genes are ubiquitously expressed and encode proteins capable of specific binding to single-stranded telomeric oligonucleotides *in vitro*. Co-immunoprecipitation and yeast two-hybrid experiments revealed that Pot1 and Pot2 can form homo- and hetero-oligomers. Additionally, using recombinant proteins we demonstrated that Pot1 and Pot2 proteins

show different subcellular localization, with Pot2 being a nuclear protein and Pot1 localizing predominantly to the nucleolus. Plants transformed with dominant-negative alleles of either *Pot1* or *Pot2* showed moderate telomere shortening, suggesting that both proteins are involved in telomere length control. Moreover, I demonstrated that *Arabidopsis* Pot2 is an important telomere capping protein, since over-expression of a chimeric protein consisting of a fusion of the DNA-binding domain of Pot2 and the remainder of Pot1 leads to dramatic telomere shortening and inhibition of telomerase. Thus, my data indicate that *Arabidopsis* Pot proteins perform a number of essential functions at telomeres.

Chapter IV describes my attempts to purify single-strand telomere binding proteins from *Brassica oleracia* (cauliflower), a close relative of *Arabidopsis thaliana*. I show that nuclear extracts from cauliflower contain proteins capable of specifically binding plant telomeric oligonucleotides. I provide evidence that these proteins have different molecular weights and DNA-binding characteristics. I also describe purification of one of these proteins and our attempts to obtain its peptide sequence. The data presented in this chapter demonstrate that single-strand telomere binding proteins are present in other plant species besides *Arabidopsis*. Finally, in the appendix I examine the substrate specificities of different plant telomerases and demonstrate three distinct modes of *de novo* telomere addition.

CHAPTER II

LENGTH REGULATION AND DYNAMICS OF INDIVIDUAL TELOMERE TRACTS IN WILD TYPE *ARABIDOPSIS*

INTRODUCTION

Telomeres are nucleoprotein complexes that distinguish the natural ends of chromosomes from damage-induced double-strand DNA breaks. Maintenance of the telomere cap is not only essential for genome stability, but is also required to promote the long-term proliferation capacity associated with immortalized and undifferentiated cell populations. Telomere architecture is well-conserved across evolution. In most plants, telomeric DNA consists of TTTAGGG repeats, with the extreme terminus of the G-rich strand forming a short 3' single-strand protrusion (Riha et al., 2000). Telomere length varies considerably among different eukaryotes, but in all cases telomeres are strictly maintained at a species-specific set point, and hence length homeostasis is achieved. Telomeres in the Columbia ecotype of *Arabidopsis* span 2-5 kb (Richards and Ausubel, 1988), while in tobacco telomeres are exceptionally long and reach 150 kb (Fajkus et al., 1995).

Although mechanisms governing telomere size are poorly understood, dynamic forces can both shorten and lengthen the repeat array (McEachern et al., 2000). Shortening occurs largely as a result of incomplete replication by conventional DNA replication machinery (Olovnikov, 1973; Watson, 1972). In the absence of a

compensating mechanism, a few nucleotides are lost from the 5' end of the daughter strand synthesized by the lagging strand machinery each time a cell divides. Telomeres can also be shortened through exonucleolytic degradation (Maringele and Lydall, 2002). If telomere loss proceeds unabated, the chromosome terminus will ultimately elicit a DNA damage checkpoint response leading to cell-cycle arrest. Cells that escape the arrest undergo illegitimate repair and face the consequences of genome instability. Fortunately, telomere shortening can be circumvented by the action of telomerase, a ribonucleoprotein reverse transcriptase that catalyzes the addition of G-rich telomere repeats onto the 3' overhang ends. Telomerase not only rebuilds shortened telomere tracts, but can also extend an existing array for a net increase in telomere length.

Telomere length differs not only between evolutionary distant species, but also within species of the same genera. For instance, telomeres in wild-derived mouse strains are similar in length to telomeres in humans (Hemann and Greider, 2000), while established inbred mouse strains have telomere lengths of approximately 40 kb (Zijlmans et al., 1997). How is telomere homeostasis achieved? Current models propose that telomere length is modulated by telomere-specific proteins that interact with either the 3' overhang or the duplex region of the telomere and regulate telomerase access to the terminus. Telomeres that reach an optimal length carry a full complement of telomere binding proteins and exist in a "closed" conformation largely inaccessible to telomerase (de Lange, 2002). In the absence of telomerase action, the tract gradually shortens over successive cell divisions, losing telomere protein binding sites, and as a consequence shifts to a more "open" conformation for telomerase.

While this model provides a useful framework for investigating aspects of telomere homeostasis, many questions remain unanswered. For example, how is the optimal telomere length established for different organisms? What is the fate of individual telomere tracts through successive generations? For instance, do individual telomeres tend to reside in a preset size territory or is the length of each tract dynamic? Finally, is the length of the telomere tract on homologous chromosome arms coordinately regulated? To begin to address some of these issues, we examined telomere length regulation in wild type *Arabidopsis*. *Arabidopsis* has emerged as a useful model for telomere biology (Riha and Shippen, 2003b). The first higher eukaryote whose telomeres were cloned and sequenced, *Arabidopsis* is unusual in that its subtelomeric DNA sequences are unique on eight of the ten chromosome arms. Thus, this genome is particularly well-suited for analysis of individual telomere tracts. In addition, previous studies indicate that telomere length varies among wild type *Arabidopsis* accessions (Richards et al., 1992; Riha et al., 2002), providing an opportunity to examine telomere length dynamics in different genetic settings.

Here we examine telomere length in 14 different *Arabidopsis* ecotypes. Not only do we detect significant size differences among these accessions, but we also uncover a striking bimodal size distribution of telomeres in individual WS plants. We also employ unique subtelomeric sequences as probes to follow the fate of individual chromosome ends in both Columbia and WS ecotypes through successive plant generations. Our data indicate that telomere length homoeostasis in *Arabidopsis* occurs through intermittent telomerase action on shorter telomeres to achieve an optimal, ecotype-specific size.

RESULTS

Ecotype-specific telomere lengths

To determine the extent of telomere length variation among wild type *Arabidopsis*, we performed terminal restriction fragment (TRF) analysis on 14 different *Arabidopsis* accessions. DNA samples were digested with *Tru1*1, which cleaves immediately adjacent to the telomeric DNA tract and then were hybridized with a $(TTTAGGG)_4$ probe. As previously noted, telomeres in Columbia and WS ecotypes were not the same size (Riha et al., 2002). Columbia telomeres ranged from 2-5kb in length and WS from 4-8kb (Figure 4A). Approximately half of the accessions we studied displayed the Columbia-type telomere length of 2-5kb, while the remainder were reminiscent of WS, harboring longer telomeres that extended up to 8kb (Figure 4B, Table 1). We conclude that telomeres in wild type *Arabidopsis* can span 2-8kb, with no obvious bias towards the short or long end of this range.

These initial experiments were conducted on pooled populations of *Arabidopsis* plants. To study telomere length regulation in more detail, we examined telomeres in individual plants. As expected, individual Columbia plants displayed a homogeneous profile of telomere length with the majority of plants bearing telomeres in the 2-4kb range (Figure 5A). Only occasionally were individuals with slightly longer or shorter telomere arrays observed (Figure 5A, compare lanes 1 and 7). A similar result was obtained for Landsberg erecta individuals (Richards et al., 1992). In this case, telomere tracts were even more regular and spanned 2-6kb (data not shown).

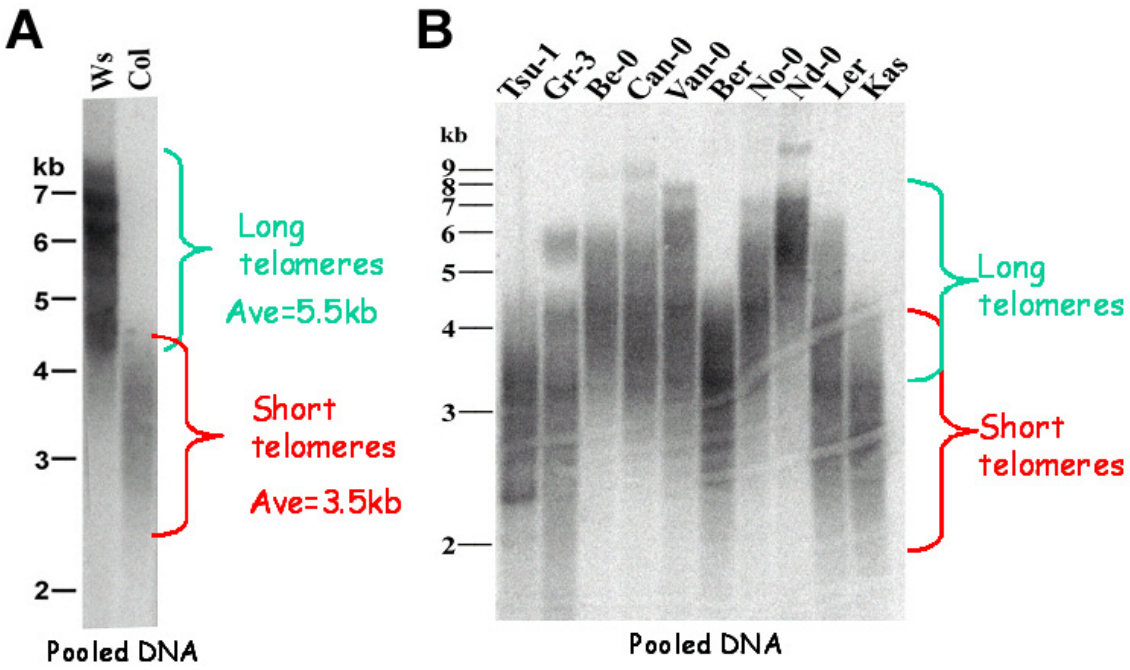

Figure 4. *Arabidopsis* ecotypes display bimodal telomere length distribution.

DNA from 10-12 individual plants from several ecotypes was harvested, pooled and subjected to TRF analysis. Ecotypes were divided into two groups, based on the length of their telomere tracts: short telomeres (2-5 kb) and long telomeres (3.5-8 kb). (A) TRF analysis of ecotypes often used to generate mutants by insertional mutagenesis, (B) TRF analysis for ten randomly selected ecotypes.

Table 1. Telomere length in wild-type *Arabidopsis* accessions

Ecotype	Telomere Length
Tsu-1	2-4.5
Gr-3	2-6
Be-0	3-6
Can-0	3-9
Van-0	3-8
Ber	2-4.5
No-0	3-6.5
Nd-0	3.5-9
Ler	2-6
Kas	2-4.5
La-0	3-7
Cvi-0	3.5-9
WS	2-8
Columbia	2-5

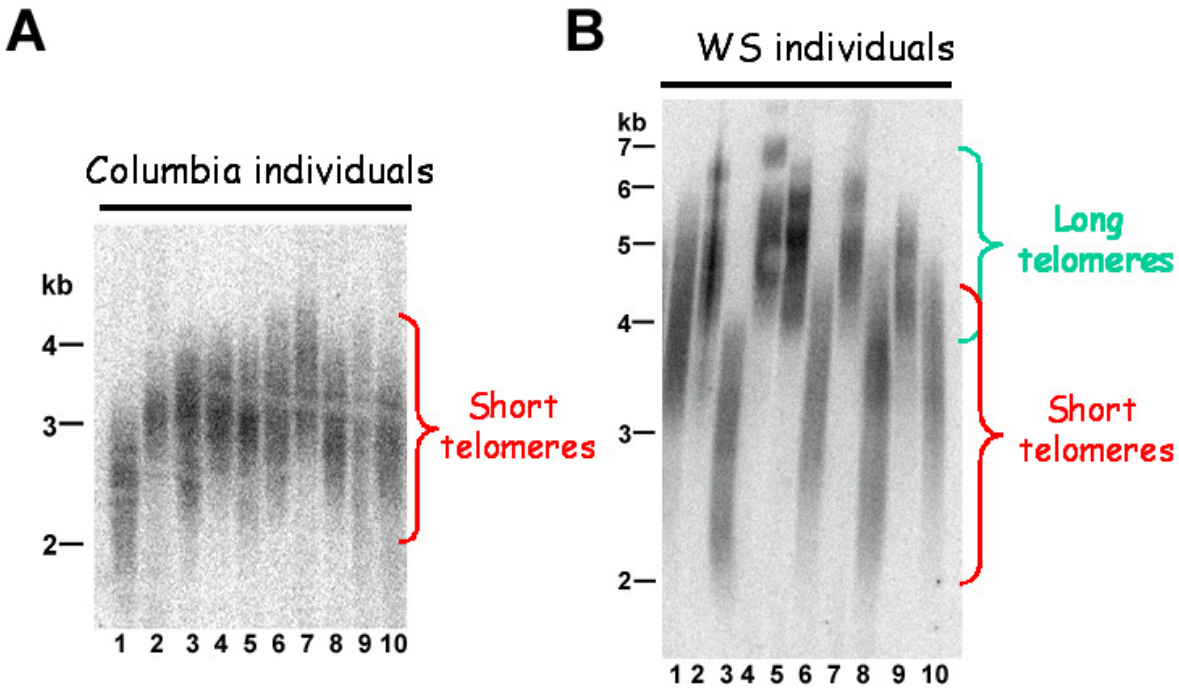

Figure 5. Telomere length distribution in individual Columbia and WS plants.

TRF analysis was performed on genomic DNA from ten individual plants of each ecotype. Plants of the WS ecotype display a bimodal distribution of telomere length, with short (2-5 kb) and long telomeres (4-7 kb). Columbia (A) or WS (B) individuals are shown.

A strikingly different result was observed with WS individuals. Instead of a single uniform profile, the telomeres of WS plants displayed a bimodal distribution with some individuals harboring Columbia-type telomeres of 2-5kb and others longer telomeres of 3.5-8kb (Figure 5B). All of the WS plants used for this analysis displayed the distinct WS morphology, ruling out cross-contamination with Columbia seeds. This bimodal length distribution was also observed in WS plants from Wisconsin's T-DNA insertion collection (data not shown). To determine whether other *Arabidopsis* accessions with long telomeres behaved like WS, TRF analysis was performed with individual plants from Nd-0 and Cvi-0 ecotypes. The telomere size distribution in Nd-0 and Cvi-0 plants was uniform among different individuals (data not shown), implying that WS is unusual among *Arabidopsis* accessions in harboring a bimodal distribution of telomere lengths.

Genetic analysis of telomere length variation in the WS ecotype

Given the striking variation in telomere length among WS individuals, we investigated whether telomere size was stably inherited in these lines. TRF analysis was performed on parents bearing either short (line 78-7) or long (line 71-13) telomeres and their progeny. For both lines, the telomere size in the parent was maintained for three successive generations (Figure 6A and data not shown). Similar results were obtained for Columbia plants; telomeres in the progeny differed from the parents by no more than 0.5 kb (data not shown). Thus, telomere size distribution in wild type WS and Columbia accessions is tightly regulated and heritable through successive generations.

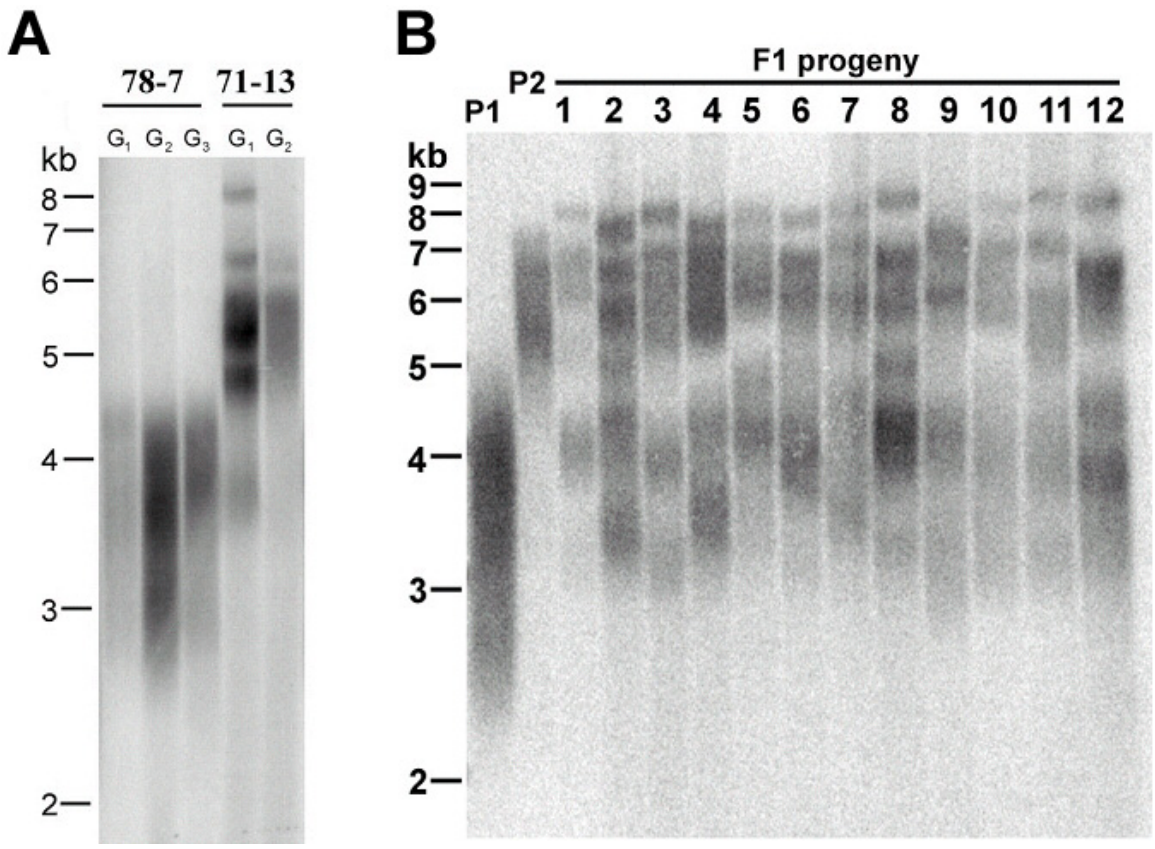

Figure 6. Telomere length is not established by a single dominant factor.

(A) TRF analysis of successive generations of WS individuals. Telomere length is stably maintained in both WS lines for at least 3 generations. 78-7, short telomere line; 71-13, long telomere line. (B) TRF analysis of F1 progeny from a cross between lines 78-7 (P1) and 71-13 (P2). F1 progeny display intermediate telomere length relative to their short and long telomere parents. (C) TRF analysis of F2 individuals. Telomeres do not segregate into long and short size distributions.

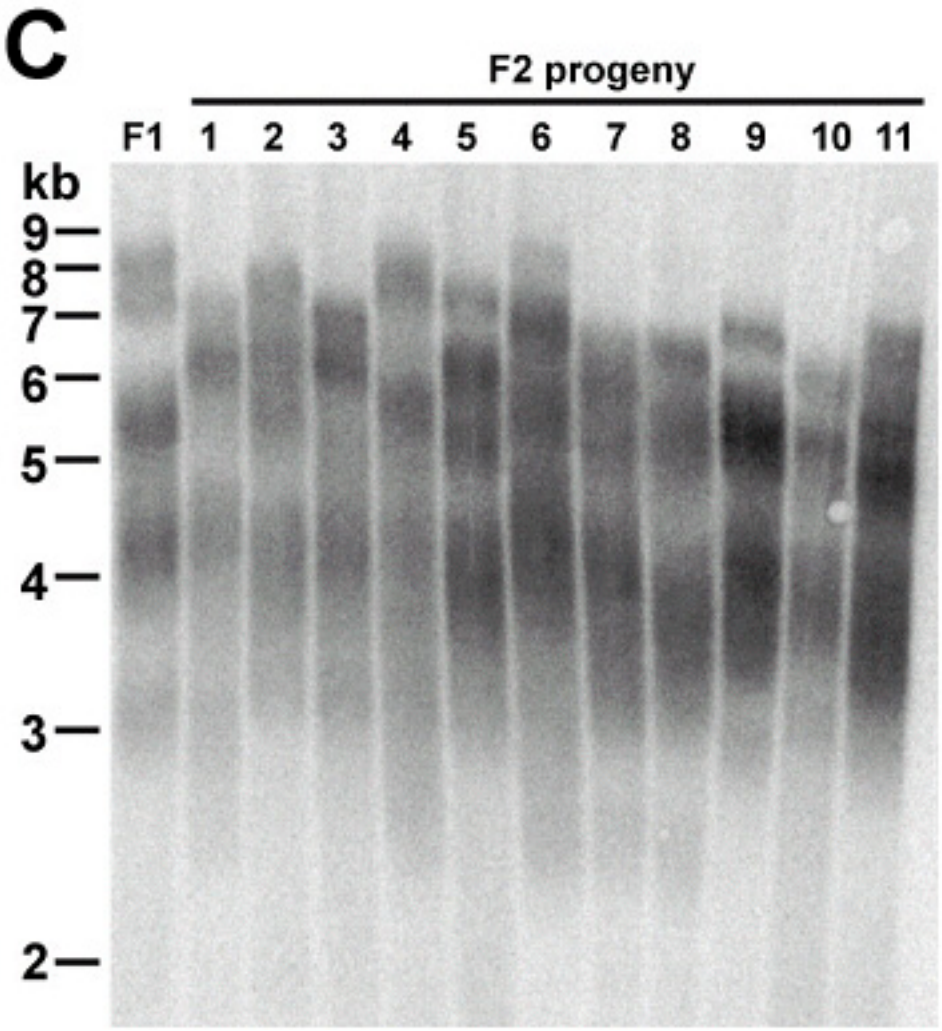

Fig.6. Continued.

To further explore the molecular basis for the bimodal size distribution among WS plants, we crossed individuals from lines 78-7 and 71-13 to obtain plants heterozygous with respect to parental telomere length. Telomeres in F1 progeny ranged from 2.7 to 8.5kb (Figure 6B). Instead of the more homogenous smear associated with parental DNA, telomeres in the progeny exhibited a more discrete banding pattern. We previously showed that individual telomere tracts in wild type Columbia plants span approximately 1kb (Riha et al., 2001 and see below). Hence, the banding profile observed in the F1 WS progeny likely reflects the increased size in overall distribution of telomeres in this genetic background.

If a single dominant factor were responsible for establishing telomere length in WS plants, we would expect segregation of telomere lengths in an F2 population back to the parental length. To test this prediction, F1 plants were self-pollinated to generate F2 progeny. Telomeres in the F2 plants displayed roughly the same size distribution as their F1 parent (Figure 6C). In 9/11 F2 individuals the longest telomeres were slightly shorter than in F1, and in approximately half of the F2 plants, the shortest telomeres were slightly shorter. Nevertheless, the overall range of telomere tracts was remarkably similar in F1 and F2 populations, arguing that like the situation in maize (Burr et al., 1992), telomere length in *Arabidopsis* is not controlled by a single genetic factor.

Telomere length dynamics at individual chromosome ends

While it is clear that the overall size of telomere tracts is maintained within pre-set boundaries, relatively little is known about how this size homeostasis is achieved for

individual telomere tracts. To investigate this question, we followed the fate of individual telomeres in parents and their progeny using chromosome end-specific probes. For analysis of individual chromosome ends, TRF analysis was performed on DNA isolated from whole plants that was digested with *Pvu*II and *Spe*I, which cleaves in the subtelomeric regions releasing the telomere tract plus unique sequences adjacent to it. In this study we used probes specific for the right arm of chromosome two (2R), and the right and left arms of chromosome 5 (5R and 5L, respectively).

As shown in Figure 7, a single discrete band was detected in most Columbia plants, indicating that the length of individual telomere tracts on homologous chromosomes is coordinately controlled throughout development. In a few plants, two (Figure 7A, lane 2) or even three bands (Figure 7A, lane 3) were observed. While two bands would be consistent with differentially sized telomeres on homologous chromosomes, the presence of three 5L bands in one plant, for example, implies the existence of different cell populations bearing 5L telomeres of different lengths. Although the size of individual telomere tracts is tightly regulated in a single plant, we noted dramatic variation in telomere length among unrelated individuals. For example, the 2R telomere in Columbia plant #3 was quite short (2.4 kb), but in plant #1 it was 1.3 kb longer (Figure 7B). Similarly, the 5L telomeres of plants #4 and 5 were much shorter than 5L in plant #1 (Figure 7A).

We also analyzed individual telomeres in WS plants, but in this case we compared telomere size in siblings derived from parents bearing either long or short telomeres. Overall, telomere lengths were more similar among siblings than unrelated individuals

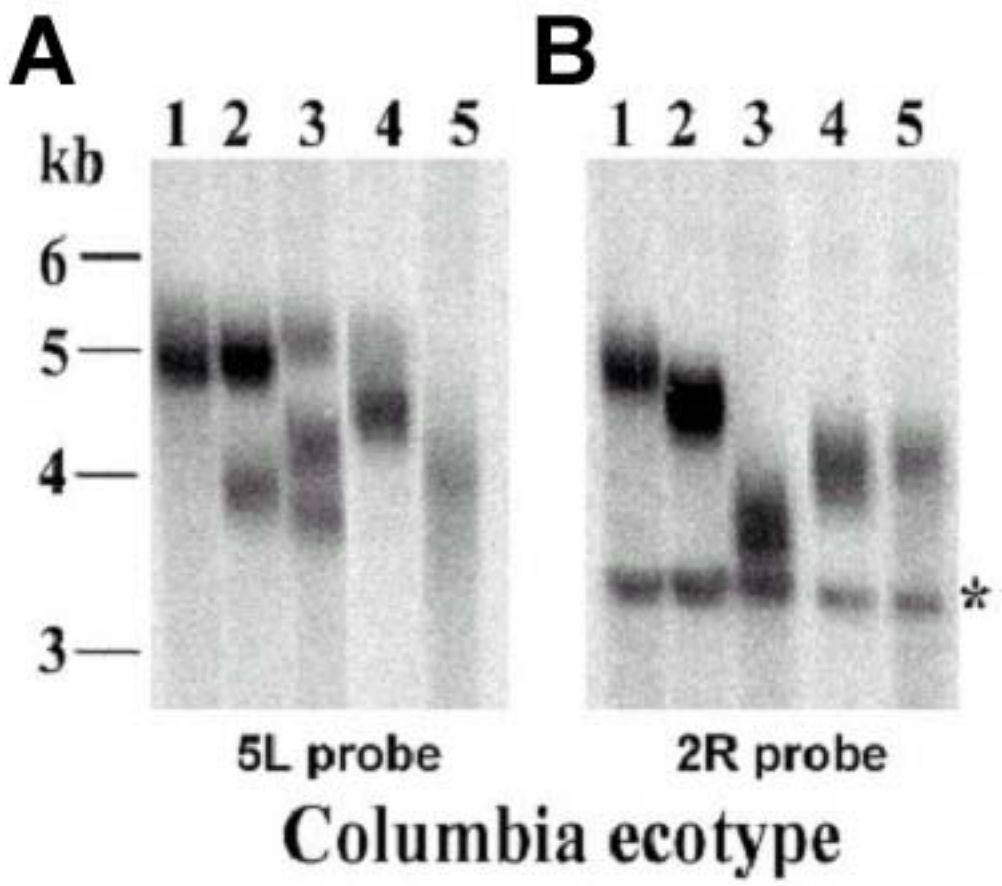

Figure 7. Individual telomeres in *Arabidopsis* occupy different size territories.

TRF analysis of individual chromosome arms from unrelated Columbia plants was conducted using probes specific for 5L (A) and 2R (B). Asterisk indicates interstitial band cross-hybridizing to 2R probe. TRF analysis of DNA from WS siblings from short and long telomere plants hybridized with a 5L (C) or 2R (D) probes are shown.

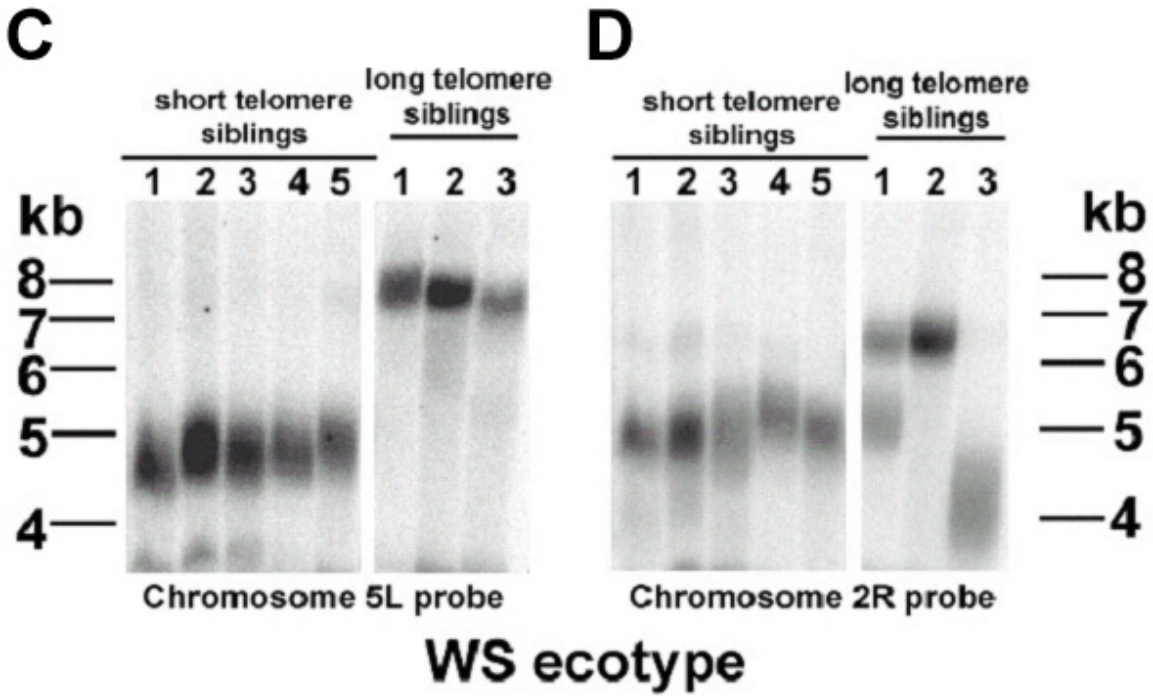

Figure 7. Continued.

(Compare Figure 7A and 7C). 2R and 5L telomeres in the siblings of the short telomere line (78-7) were remarkably uniform in size, with most tracts measuring between 3.3 and 3.7 kb. The same result was obtained for the 5L telomeres in siblings from long telomere line (71-13). In contrast, the 2R telomeres from this same line (71-13) varied dramatically in size among siblings. The 2R telomere in plant #3 was 2.5 kb shorter than in plant #2, while for plant #1 two discrete 2R telomere populations were observed, one corresponding to 2R in plant #2, and the other intermediate in size between 2R for plants #2 and #3 (Figure 7D). We conclude from these data that while global telomere length is strictly regulated within an ecotype-specific range, individual telomere tracts do not occupy a set size territory, and are subject to unique lengthening and shortening events.

Telomeres lengthen and shorten in progeny to achieve an optimal size

The data presented thus far demonstrate the dynamic nature of individual telomere tracts. However, the variations appeared to be stochastic and provided no insight into how the upper and lower size limits are established and maintained. To address this question, we performed a parent-progeny analysis to compare individual telomere tracts in Columbia and WS parents. Figure 8 shows the results for Columbia plants derived from different siliques of the same parent. The parental 5L telomere was 3.2 kb, but in essentially all of the progeny (from three different siliques), telomeres were lengthened for a net increase of 0.3 to 0.8kb (Figure 8A). In the majority of the progeny only a single prominent population of 5L telomeres was detected. However, in plants #2 and #5 from silique 3, two populations of 5L telomeres were observed, implying that the 5L telomere was

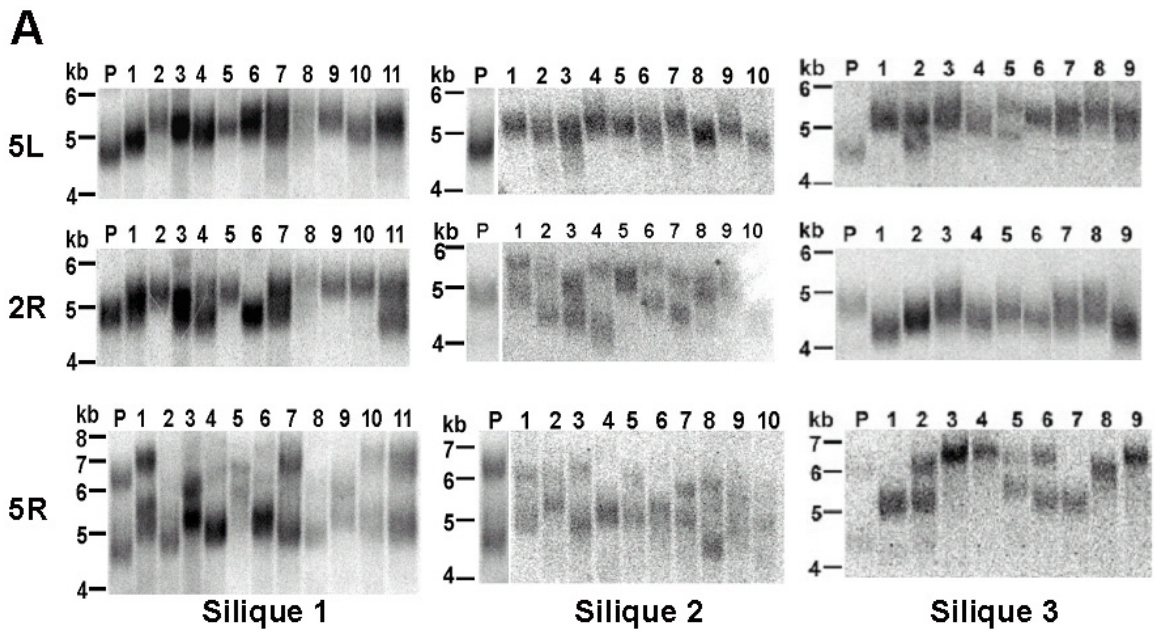

Figure 8. Parent-progeny analysis of Columbia plants.

(A) One Columbia plant (parent) was self-pollinated, and the progeny from three random siliques were chosen for the TRF using the probes indicated. Molecular size standards indicate the actual size of TRF fragments. To obtain the actual size of the telomere tract, the corresponding length of the subtelomeric sequence (1.3 kb for 2R, 1.5 kb for 5L, 2.6 kb for 5R) was subtracted from the sizes shown. (B) TRF analysis was performed on DNA from three successive generations of Columbia plants to measure 2R, 5L and 5R telomere lengths. The F1 lane shows results for DNA obtained from the progeny of the parent (P).

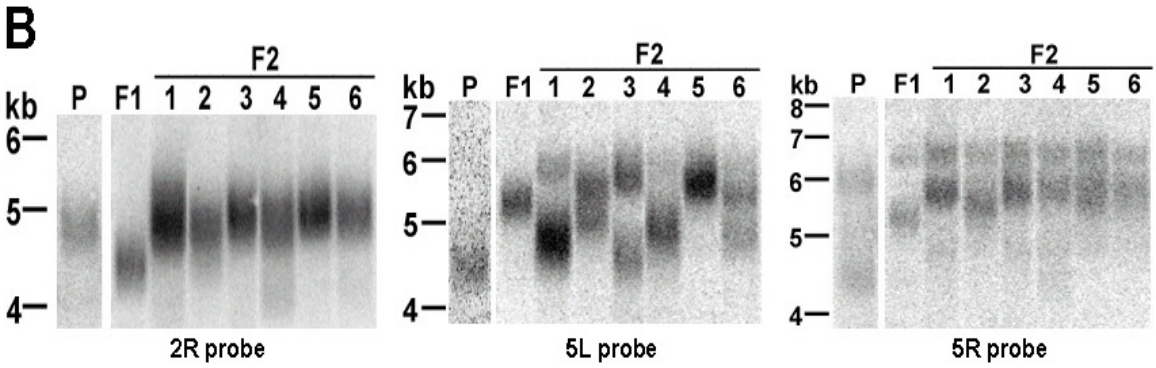

Figure 8. Continued.

subjected to differential processing in these plants. Nevertheless, the general trend was toward telomere lengthening to increase the size of the 5L telomere in the progeny relative to their parent.

Analysis of the 2R telomere yielded a different result. In this case, the parental telomere was 3.5 kb, ~300bp longer than the 5L telomere. Intriguingly, 2R telomeres in the progeny had strikingly different fates. For silique 1, 2R telomeres in a subset of the progeny were elongated (up to 0.6 kb), but this was not the case in all plants. In several individuals 2R telomeres were approximately the same size as in the parent, while the 2R telomeres in 6/10 progeny of silique 2 split into two populations with one telomere being shorter than the parent and one longer. Remarkably, for silique 3, all of the 2R telomeres were slightly shorter than the parent and ranged in size from 2.9 to 3.3 kb for a net decrease of 0.2-0.6 kb. This decrease corresponds to the amount of telomeric DNA lost per generation in a telomerase-deficient mutant (Riha et al., 2001), arguing that the 2R telomere was not an efficient substrate for telomerase action in this generation.

We were particularly interested in the fate of the 5R telomere. The parent displayed two distinct 5R telomere populations, one at 3.7 kb and second of only 1.9 kb, which represents the lower size limit of telomeres in all of the wild type accessions we examined. Strikingly, in all of the progeny, 5R telomeres were extended relative to this short parental telomere. Moreover, in silique 3 the 5R telomere converged in 6/9 individuals to comprise only a single size distribution that was up to 2 kb longer than the short parental telomere. These findings indicate that telomerase is acting preferentially on the shortest telomere in the population allowing it to enter a more favorable size range.

The dynamic nature of this telomere-measuring mechanism was even more evident when we followed the fate of telomeres through three consecutive generations (Figure 8B). The 3.5kb 2R telomere dropped in size in the second generation to 3kb but then in the third generation was restored to approximately the same size as in the first generation. In contrast, the 3.2 kb 5L telomere was extended to 3.7 kb in generation 2, but then in generation 3 telomeres in most of the progeny split into two size classes one longer and one shorter than their generation 2 parent. For 5R, two populations of telomeres were maintained throughout the two subsequent generations. However, the shorter telomere, which was only 1.9 kb in the first generation, was extended to 2.6 kb in the second generation and then to 3.1 kb in the third generation. Conversely, the longer 5R telomere was extended from 3.7 to 4kb in the second generation, but remained the same length in the third generation.

All together these data indicate that the length of telomere tracts in *Arabidopsis* are actively monitored and reset in each generation to maintain an optimal size. For the Columbia ecotype, this size appears to be 3-3.5kb. Telomeres shorter than this are likely to be acted on by telomerase in the next generation, while longer telomeres are less inclined to be telomerase substrates.

A new optimal telomere length in WS plants

Telomeres in WS accessions appeared to be subjected to the same type of measuring mechanism, but different length optima were observed depending on whether the plants were derived from a short (optima = 3.5kb) or long parent (optima = 6.0kb)

(data not shown). To explore how an ideal telomere length is established, we examined telomere dynamics in WS plants derived from the cross between short and long telomere parents (Figure 9).

As expected, the 2R telomere in each parent fell within predicted size range, 5.6kb for the long telomere parent and 4.0 kb for the short (Figure 9, lanes 1 and 2). In F1, two populations of 2R telomeres were detected, one population of the same size as the long telomere parent, and a second slightly longer than in the short telomere parent (Figure 9, lane 3). Thus, the broad telomere size distribution in F1 reflects the contribution of one short and one long telomere in the cross.

A strikingly different profile was observed in F2. In all but one of the plants, the two homologous 2R telomeres converged to become much shorter than the original long telomere parent with many approaching the size of the short telomere parent. In only one F2 plant (Figure 9, lane 11) was there evidence for significant telomere elongation. However, in this case, this telomere was 6.9 kb, still within the acceptable size range for the original long telomere parent. The second 2R band in this individual was slightly longer than the short telomere its F1 parent. For the most part, 2R telomeres in F2 ranged in size from 4.2 to 5.5kb. This observation suggests that a new broader set point had been established that is intermediate in size relative to the original long and short WS telomere parents. Therefore, we conclude that multiple genetic factors or epigenetic control of telomere length can contribute to the establishment of a new telomere length set point in these plants.

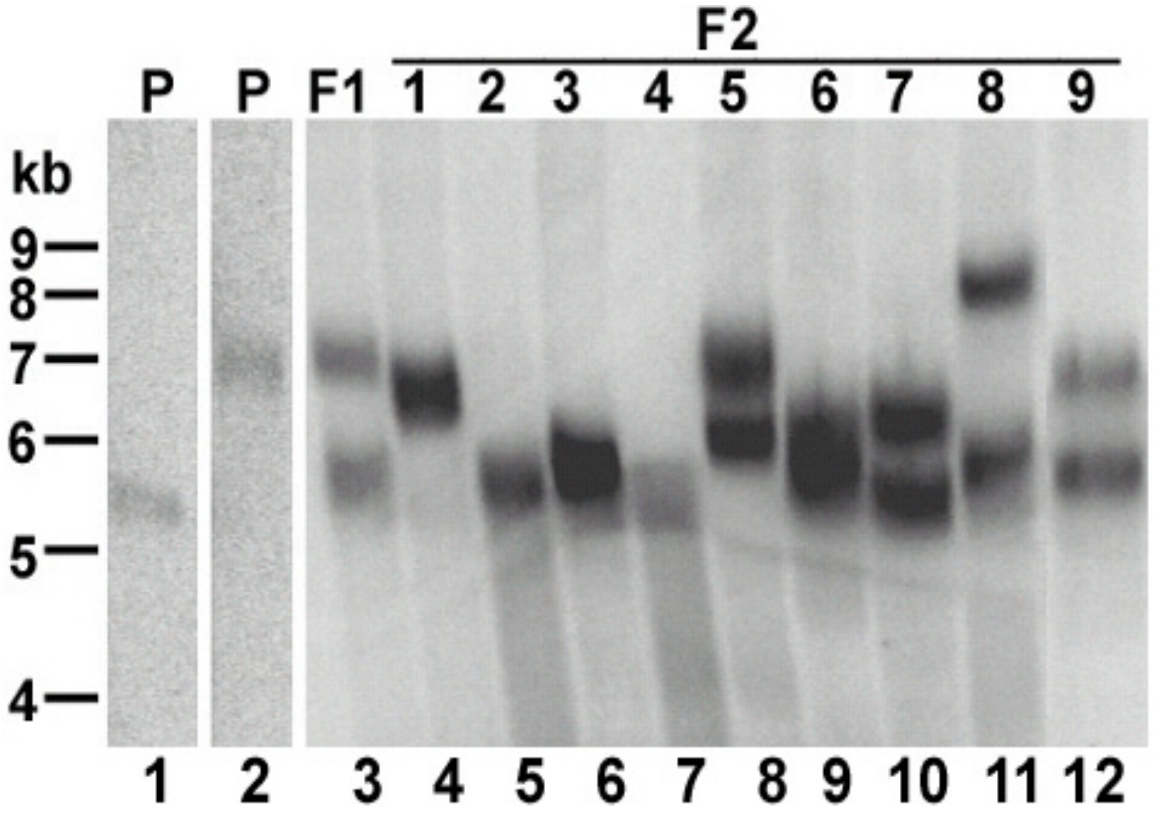

Figure 9. A new, wider acceptable telomere length range is established in the progeny of a cross between long and short telomere WS plants.

TRF analysis was used to measure 2R telomere lengths in three successive generations of WS plants. The F1 plant (lane 3) was derived from the cross between short (78-7, lane 1) and long (71-13, lane 2) telomere parents. F1 was self-pollinated to produce seeds for F2 plants (lanes 4-12).

DISCUSSION

Natural telomere length variation in *Arabidopsis* accessions

Eukaryotic organisms use telomeres as a general mechanism for chromosome end protection. Although the overall length of the telomeric tract varies from species to species, each organism maintains its telomeres within a defined, species-specific limit. The mechanisms dictating an acceptable telomere length for a given organism are unknown. It is also not currently clear how regulation of telomere length on a particular chromosome end contributes to the overall telomere maintenance within a cell. Since perturbations in the telomere length maintenance machinery can have profound influences on cell survival, it is important to understand how telomere length is established and maintained.

Arabidopsis thaliana appears to be an excellent model for investigating natural variation of telomere length. *Arabidopsis* accessions, collected over the years from various natural habitats, display a wide variety of evolutionary traits, and here we show that these variations include striking differences in telomere length set points (Richards et al., 1992; Riha et al., 2002; and this work). Since telomere length is stably inherited and maintained, *Arabidopsis* must employ an active mechanism to monitor telomere length and to ensure that telomeres are maintained within a defined ecotype-specific range.

Telomeres in *Arabidopsis* ecotypes fall into two size classes: short (defined as 2-5 kb Columbia-type telomeres) and long (3.5-8 kb as in Nd-0 and Cvi-0 group) telomere tract. Neither group appears to be prevalent, suggesting that natural variation of telomere lengths in *Arabidopsis* is widespread and may not be linked to other morphological traits

that distinguish ecotypes from each other. Telomeres that fall between 2-8 kb are acceptable to *Arabidopsis*.

Since most of the *Arabidopsis* insertional mutagenesis facilities employ WS and Columbia ecotypes, it was of interest from a practical standpoint to more thoroughly investigate telomere length regulation in these accessions. Telomeres in Columbia plants are homogeneous in their overall length, but this is not a situation for WS, where striking differences in telomere length can be observed among individual plants. Some WS plants display Columbia-type telomeres, ranging from 2 to 4 kb, but most harbor telomeres in the 3.5-8 kb size range, as previously reported in the literature. In our experience, short telomere WS plants are less common in the population than plants bearing longer telomeres. This may explain why in all of the published studies WS plants used to examine the role of telomere-related genes have had longer telomeres. Nonetheless, as more telomere-related genes are studied in the WS background, the distinct size classes of telomeres observed in wild-type plants must be considered. For example, mutant progeny derived from a heterozygous parent must be compared with their wild-type siblings and not with unrelated wild-type plants. Since natural plant-to-plant variation contributes to telomere length, such comparisons are necessary to accurately interpret the consequences of telomere-related mutations.

Telomere dynamics on individual chromosome ends

In contrast to the situation in yeast and mammals, popular models for telomere biology, chromosome arms in *Arabidopsis* harbor unique subtelomere sequences, which

can be used to assess behavior of individual telomeres throughout development of the entire plant or in a parent-progeny analysis. Two of the remaining three ends bear rDNA clusters. We exploited this feature of *Arabidopsis* chromosomes to examine the dynamics of individual ends. Our data demonstrate the dynamic nature of telomere maintenance and suggest that individual telomeres do not occupy a predetermined length territory. Instead, each telomere is free to move within the ecotype-specific size boundary. In addition, the length of telomere tracts on homologous chromosomes is coordinated, since in most cases we detected only one band corresponding to both homologues.

The regulated nature of telomere tracts in *Arabidopsis* is even more remarkable, considering that DNA for our analysis was isolated from entire plants. Since most cells are expected to have different proliferation histories, the length of an individual telomere is expected to vary significantly depending on how often telomerase engaged the telomere and how much telomeric DNA was added. This is the case for clonal population of human cells, which display a dramatic difference of up to 6 kb for a particular chromosome end (Baird et al., 2003). In contrast, we observed very limited length variation for a given telomere end and in most cases only one telomere-specific band was observed. The limited cell-to-cell variation of telomere lengths within the whole plant may reflect a low number of cell divisions required for generation of an *Arabidopsis* plant relative to cell divisions to maintain human cells in culture. In addition, the action of telomerase in all plant cells may be highly coordinated and predetermined by the size of the telomere tract on a chromosome following meiosis. This idea is further supported by the fact that individual chromosome arms in siblings tend to have similar telomere lengths compared to unrelated individuals. Although we occasionally observed

exceptions to this rule, most siblings display remarkably similar lengths of telomere tracts, suggesting that the fate of individual telomeres is largely predetermined by the size it inherits from the parent.

Our data suggest that telomerase can be quite processive *in vivo*, since extensions of up to 2 kb are possible within one generation. Whether this happens as a consequence of a single elongation event early in plant development or through multiple interactions throughout development is unknown. In *Arabidopsis*, telomerase appears to be preferentially recruited to the shortest telomeres. Our findings also indicate that longer telomeres shorten at a rate of approximately 200-500 bp per generation. This attrition occurs at the same rate as in a mutant lacking telomerase (Riha et al., 2001). Hence, it appears that some chromosome ends may not be substrates for telomerase throughout the entire lifespan of the plant. These data further imply that telomeres in wild-type *Arabidopsis*, in contrast to those in humans, are not subjects for nuclease attack.

Maintaining an optimal telomere length in *Arabidopsis*

Our parent-progeny analysis indicates that an optimal size of telomere tract is established and maintained in all *Arabidopsis* ecotypes. For Columbia, this optimal size is 3.0–3.5 kb. Telomeres are acted upon by two opposing forces driving them to this size. Telomeres shorter than 3.3 kb are prefentially extended by telomerase, while longer telomeres tend to be poor substrates for telomerase and passively drift down to the optimal length. We hypothesize that at the optimal size an equilibrium between telomere shortening and lengthening activities is reached, and the telomere has an equal chance of being elongated or shortened. When this optima is reached, telomere splitting can occur

to generate two subpopulations of shorter and longer sizes in the next generation. Telomere splitting would explain the presence of more than one terminus-specific band in some plants. Accordingly, the convergence of two chromosome end-specific bands into one would occur as a shorter and longer telomere are simultaneously brought closer in size to reach the optimal length. The concept of the optimal size range allows us to explain the otherwise seemingly stochastic nature of telomerase action.

The dynamic nature of telomere maintenance observed in wild-type *Arabidopsis* plants is remarkably similar to that described in the budding yeast *S. cerevisiae*. A protein-counting mechanism is proposed to govern telomere length maintenance in yeast (Marcand et al., 1997). In this model, Rap1p proteins bind the double strand region of the telomere tract and control telomerase accessibility. A full complement of Rap1p results in a more closed conformation of the chromosome terminus, which limits telomerase action on the chromosome terminus. Telomeres on this chromosome end progressively shorten due to the end replication problem resulting in the loss of telomeric nucleotides with each S-phase. As the telomere tract shortens, occupancy of Rap1 protein declines, favoring a more open conformation of the terminus and greater access by telomerase. In humans, TRF1 protein is proposed to mimick Rap1p activity in the regulation of telomere length. Although no Rap1p or TRF1 homologues have been shown to function at *Arabidopsis* telomeres, plants almost certainly harbor homologues for these proteins. Indeed, several proteins that fit the criteria for double-strand telomere binding proteins have been identified (Zentgraf, 1995; Zentgraf et al., 2000; Chen et al., 2001; Hwang et al., 2001).

Biphasic distribution of telomeres within the WS ecotype raises several interesting questions about the mechanisms that establish the species-specific telomere

size. Both populations of WS plants stably maintain their specific telomere lengths from generation to generation, arguing against unregulated telomere maintenance in WS ecotype. Since the majority of wild-type plants have telomeres in the longer range (4-8 kb), it seems likely that plants in the shorter telomere group have somehow reset their acceptable telomere tract range to the lower Columbia-type limit (2-5 kb).

Our inability to segregate short and long telomere lengths in the F2 progeny of a cross between short and long telomere WS parents implies that the establishment and maintenance of ecotype-specific telomere tract length is not regulated by a single genetic factor. Instead, a new intermediate set point of telomere length is established in F1 plants and subsequently maintained in F2 progeny. In these plants, telomeres coming from the parental lines are no longer restricted by the parental telomere set points of 2-4 or 4-8 kb. They are able to move freely within the newly established 2-8 kb range and occupy any position in this wider area. These results are different from those obtained for recombinant inbred maize lines, in which some of the progeny displayed telomeres shorter or longer than in their parents (Burr et al., 1992).

It is unlikely that several genetic factors have been simultaneously mutated in the short telomere plants to reset the telomere length control machinery. However, to test this possibility, analysis of further generations of the crossed plants will be necessary to allow segregation of such factors. Epigenetic control of the telomere length set point is also plausible. If telomere length is epigenetically established on each chromosome, the set point could persist throughout each plant's growth and development. This hypothesis predicts that any genetic factors such as putative homologous of mammalian TRF proteins would be involved in the maintenance of telomere length, but not in its

establishment. In light of the recent discovery that *S. pombe* has mechanistically distinguishable processes leading to the establishment and maintenance of the functional telomere complex (Sadaie et al., 2003; Garcia-Cao et al., 2004), it would be intriguing to test whether epigenetic control of telomere length occurs in higher eukaryotes.

MATERIALS AND METHODS

Plant material

Wild type *Arabidopsis thaliana* seeds (ecotype Wassilewskija) were purchased from Lehle seeds (Round Rock, USA), cat. No WT-08A-10. Parental WS lines 71-13 and 78-7 were randomly selected from this population based on the differences in the overall length of their telomeres. Wild type seeds for ecotypes Col-8, WS-4, Col-6, La-0, WS-2, LER, WS-1, Col-7 were obtained from ABRC (cat. # CS 60000, CS5390, CS8155, CS1299, CS 2360, CS 8581, CS 2223, CS 3731, respectively). *Arabidopsis* plants of ecotypes Cvi-0, Tsu-1, Nd-0, No-0, Can-0, Be-0, Ber, Van-0, Gr-3, Kas-1 were from a large greenhouse population maintained by Dr. Tom McKnight (Texas A&M University).

DNA isolation and TRF analysis

DNA from individual plants was extracted as described (Cocciolone and Cone, 1993). TRF analysis was performed with *Tru*1I (Fermentas) restriction enzyme and ^{32}P 5' end-labeled $(T_3AG_3)_4$ oligonucleotide as a probe (Fitzgerald et al., 1999). Single telomere analysis for the south or right arm of chromosome 2 (2R), north or left arm of chromosome 5 (5L) and south or right arm of chromosome 5 (5R) was performed as

follows: 1 μg of genomic DNA was digested with *Pvu*II and *Spe*I and DNA was separated by electrophoresis in a 0.8% agarose gel and blotted onto a nylon membrane. Telomere-adjacent DNA sequences were amplified with primers PAT51-5 (CAACATGGCCCATTTAAGATTGAACG) and PAT51-3 (CACATATATGTTTGTTGAGTGTCGC) for the 2R probe; TAS5R-F1 (TACGGTTTAGAGTTTAGGGT) and TAS5R-R1 (CGCTCTCATTGCGAGTGGTA) for the 5R probe; TAS5L-F2 (TGAGTTTGCATAAAGCGTCACG) and TAS5L-R2 (CGACAACGACGACGAATGACAC) for the 5L probe and were used for hybridization. Radioactive signals were scanned by a STORM PhosphorImager (Molecular Dynamics), and the data were analyzed by IMAGEQUANT software (Molecular Dynamics).

CHAPTER III

IDENTIFICATION AND ANALYSIS OF *ARABIDOPSIS* POT1 AND POT2 PROTEINS

INTRODUCTION

Telomeres are important protein-DNA structures at the ends of linear eukaryotic chromosomes that are necessary to prevent chromosome fusions and exonuclease attack. In most organisms, the DNA component of the telomeres consists of tandem repeats of short sequences, 5-9 nucleotides in length. The 5'-3' strand of the duplex telomere DNA is rich in guanines and terminates in a short single-strand extension called the G-overhang. The G-overhang has the ability to fold back and invade the duplex region of the telomere. This results in the formation of a displaced loop (D-loop) of single-stranded G-rich repeats, which are located more internally in the telomere than the overhang itself. This structure is called a t-loop. In the t-loop, the 3'- overhang is no longer single-stranded. Instead, it is predicted to base pair with the complementary C-strand, thus becoming structurally inaccessible to telomerase and other activities (Griffith et al., 1999). During the S phase of the cell cycle, the t-loop is thought to unfold, allowing the G-overhang to become available for binding by telomerase.

Telomeric DNA is packaged by specific proteins (McEachern et al., 2000) to form a unique nucleoprotein complex that acts as a protective cap over the end of the chromosome. This cap hides the DNA terminus from DNA repair machinery and

exonucleases (Blackburn, 2001; Chan and Blackburn, 2002; de Lange, 2002). At the same time, this cap also provides regulated access to the telomeres by telomerase, a specialized ribonucleoprotein enzyme, responsible for extending telomere length in most eukaryotes. In many organisms, regulated access to the telomere is achieved via conformational changes, which may constitute opening or closing of the t-loop.

Many proteins are known to bind to the double-stranded region of the telomeric DNA, providing stability, protection and regulating overall telomere length. The most well-studied double-strand telomere binding proteins are Rap1p from budding yeast (Giraldo and Rhodes, 1994), Taz1p from fission yeast (Cooper et al., 1997) and TRF1 and TRF2 from humans (Broccoli et al., 1997; van Steensel et al., 1998). All these proteins bind to telomeres directly through a MYB-type helix-loop helix DNA-binding domain (reviewed in Smogorzewska et al., 2000; de Lange, 2002).

Until recently, single-strand telomere binding proteins were known only in lower eukaryotes. In the hypotrichous ciliate *Oxytricha nova*, single-strand telomere end binding proteins (TEBP) consist of α and β subunits and can specifically bind to the 14 nt-long telomere overhang either as an α-α homodimer or an α-β heterodimer (Price and Cech, 1987). The α-β heterodimer utilizes four oligonucleotide/oligosaccharide binding folds (OB folds) for stable and specific DNA binding. Three of these OB folds are involved in DNA binding and one in protein-protein interactions (Horvath et al., 1998). The OB fold appears to be a structurally conserved motif used by most single-strand telomere binding proteins. These proteins are also found in other ciliates, such as *Stylonichia mytikis* (Fang and Cech, 1991) and *Euplotes crassus* (Wang et al., 1992).

Interestingly, the *E. crassus* genome encodes two proteins structurally related to the α subunit, but no β subunit.

In budding yeast, the G-overhang is bound by Cdc13p, an essential protein involved in a variety of important functions at the telomere. The main function of Cdc13p appears to be binding specifically to the overhang and serving as a loading platform for a number of other proteins, each of which performs a unique set of activities (Pennock et al., 2001). Through multiple protein-protein interactions, Cdc13p becomes a crucial factor for protecting the chromosome ends from degradation (Garvik et al., 1995; Pennock et al., 2001), and for recruiting and activating telomerase (Evans and Lundblad, 1999; Taggart et al., 2002). Although Cdc13p shares very little sequence similarity with TEBPs, its DNA-binding domain is also comprised of an OB fold, which is structurally similar to the OB folds of the TEBP α subunit (Horvath et al., 1998).

Recently, a distant relative of the ciliates TEBP, called Pot1 (protection of telomeres 1) was identified in *Schizosaccharomyces pombe* (Baumann and Cech, 2001). Pot1 protein shares weak sequence similarity with TEBPs through its 20 kDa N-terminal region, which also serves as the DNA-binding domain and exists as an OB fold (Lei et al., 2002; Lei et al., 2003). Deletion of the fission yeast *pot1*$^+$ gene leads to immediate loss of telomere and, to some extent, subtelomeric DNA and is immediately followed by cell death (Baumann and Cech, 2001). Only a few cells survive the loss of *pot1*$^+$, and these have circularized all their chromosomes. As a consequence of telomere loss, yeast cells lacking *pot1*$^+$ elongate, fail to divide and missegregate their chromosomes during the few initial rounds of cell division (Baumann and Cech, 2001), the latter feature a characteristic of end-to-end chromosome fusions.

Pot1 orthologs in other eukaryotes, such as primates, *Arabidopsis*, mouse (Baumann et al., 2002) and *A. nidulans* (Pitt et al., 2004) have also been identified. However, with the exception of the human Pot1, these proteins have not been extensively studied. In mammalian cells, hPot1 localizes to the telomeres, and its binding is reduced in cells that have lost the G-overhang (Baumann et al., 2002; Loayza and De Lange, 2003). The human Pot1 protein appears to have two important domains required for its function: an N-terminal DNA-binding domain similar to the OB folds of TEBP proteins from ciliates, and a protein interaction domain, located at the C-terminus. Over-expression of either the full length protein or the C-terminal fragment leads to extensive telomere lengthening, dependent on the presence of telomerase (Colgin et al., 2003; Loayza and De Lange, 2003). The C-terminal domain is involved in interactions with the major double-strand telomere binding protein, TRF1, and provides a direct link between TRF1 and telomerase (Loayza and De Lange, 2003). It has been proposed that human Pot1 inhibits telomerase in cis at chromosome termini and could play a role in the telomere length counting mechanism, performed by the TRF1 complex.

Arabidopsis thaliana is emerging as a powerful system for telomere biology (Riha and Shippen, 2003b). Unlike other genomes sequenced to date, the *Arabidopsis* genome contains two Pot1-like genes (Baumann et al., 2002). To learn more about how Pot1 functions in telomere length regulation and protection, we examined the role of these two proteins at the plant telomeres.

RESULTS

Identification of the *Arabidopsis* Pot1 and Pot2 genes

The genes encoding putative Pot1orthologs from *Arabidopsis thaliana* were identified using the *S. pombe* Pot1 protein (Baumann et al., 2002) as the query sequence in a BLAST search of the *Arabidopsis* database. *Arabidopsis* genes At2g05210, At5g06310 and At2g04395 showed strong similarity with the conserved ciliate telomere binding protein alpha subunit's N-terminal domain (Telo_bind_N, pfam02307), also present in *S. pombe* and human Pot1 proteins. The corresponding gene products were designated *Arabidopsis thaliana* Pot1, Pot2 and Pot3. The predicted *Pot3* gene is located 36 kb upstream of the *Pot1* gene on chromosome 2, suggesting a possible gene duplication event. Detailed sequence analysis of the *Pot3* gene indicated that the predicted gene product contains a truncated version of Pot1 protein, missing most of the coding sequence except for the N-terminal putative DNA-binding domain (see below). Multiple attempts to obtain cDNA for the *Pot3* gene failed (data not shown) suggesting that *Pot3* is a pseudogene.

Full-length cDNAs for *Pot1* and *Pot2* were obtained by RT-PCR using primers complementary to the predicted start and stop codons of both mRNAs. *Pot1* and *Pot2* genes harbor 9 and 10 exons, respectively. In contrast to the situation reported for the human Pot1 protein (Baumann et al., 2002), we found no evidence for alternative splicing of *Pot1* or *Pot2* mRNAs in *Arabidopsis* (data not shown).

The *Pot1* gene is predicted to encode a 421 amino acid polypeptide, which is 17% identical and 29% similar to human Pot1 protein, while the *Pot2* gene is predicted to

encode a 454 amino acid protein which is 13% identical and 25% similar to human Pot1 protein. Pot1 and Pot2 proteins are 26% identical and 41% similar to each other throughout the entire sequence (Figure 10A). The most conserved feature of both proteins is an N-terminally located Telo_bind_N DNA-binding domain (DBD), found in single-strand telomere binding proteins of unicellular ciliates and in Pot1 proteins of fission yeast and higher eukaryotes (Mitton-Fry et al., 2002). The least conserved portion of the proteins is their C-terminal extensions. The similarity between *Arabidopsis* proteins and all other known Pot1 homologues is even more apparent within this evolutionary conserved region (Figure 10B). With respect to the human Pot1 DBD, the DBD of *Arabidopsis* Pot1 is 22% identical and 38% similar, and the Pot2 DBD is 23% identical and 44% similar. The DBDs of *Arabidopsis* proteins are 29% identical and 49% similar to each other. Both *Arabidopsis* proteins lack the conserved T42 residue, but Pot1 harbors the conserved F65, which is important for DNA binding (Lei et al., 2003). Overall, both *Arabidopsis* proteins are phylogenetically closest to their mammalian orthologs and most distant to Pot1 proteins from yeast and fungi (Figure 10C).

Arabidopsis Pot1 **and** ***Pot2*** **genes are ubiquitously expressed at low levels**

To examine the expression profile, RT-PCR was performed on all *Arabidopsis* tissues. *Pot1* and *Pot2* mRNAs are present in all tissues analyzed (Figure 11), but are expressed at very low levels. Although we can not rule out intrinsic differences in the

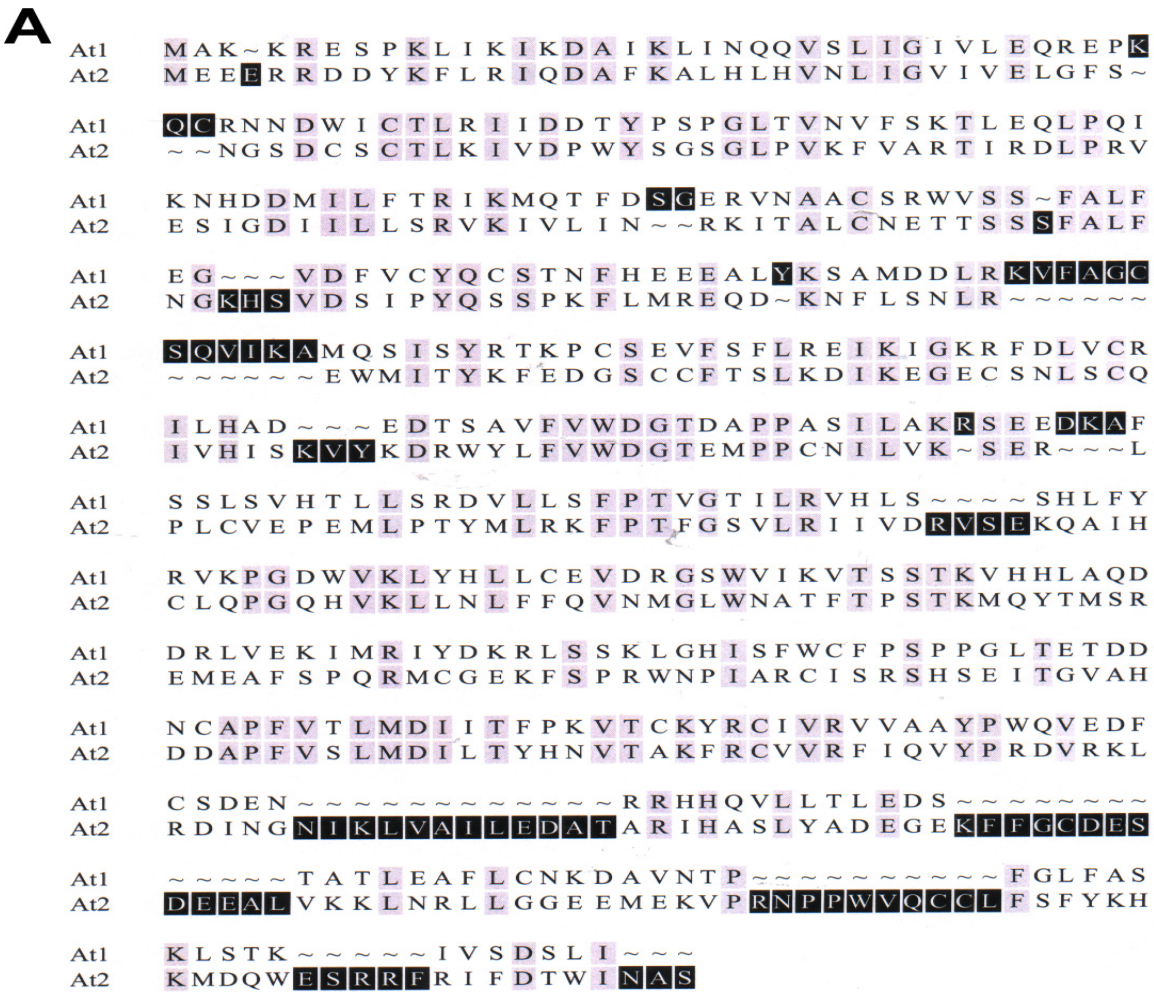

Figure 10. Alignment of Pot1 proteins.

(A) Alignment of full-length *Arabidopsis* Pot1 and Pot2 proteins. Identical residues are shown as black letters shaded in gray. Amino acid insertions present in only one protein are shown as white letters on black background. (B) Alignment of DNA-binding domains of Pot1 proteins from several species. Amino acids identical or similar in at least four sequences are shown as black letters shaded in gray, amino acids identical in at least four sequences are shown as white letters shaded black.The first and last amino acid positions are indicated for each sequence. Amino acids circled in green indicate *S. pombe* residues important for DNA binding (Lei et al., 2003), amino acids circled in red indicate corresponding residues in the consensus sequence. (C) Phylogenetic tree of Pot1 proteins. An – *Aspergillus nidulans*; Hs – *Homo sapiens*; Mm – *Mus musculus*; At1 – *Arabidopsis thaliana* Pot1, At2 - *Arabidopsis thaliana* Pot2; Ec - *Encephalitozoon cuniculi*; Nc – *Neurospora crassa*; Sp – *Schizosaccharamyces pombe.* The number at each node indicates the percentage of 1000 bootstrap replicates for statistical support.

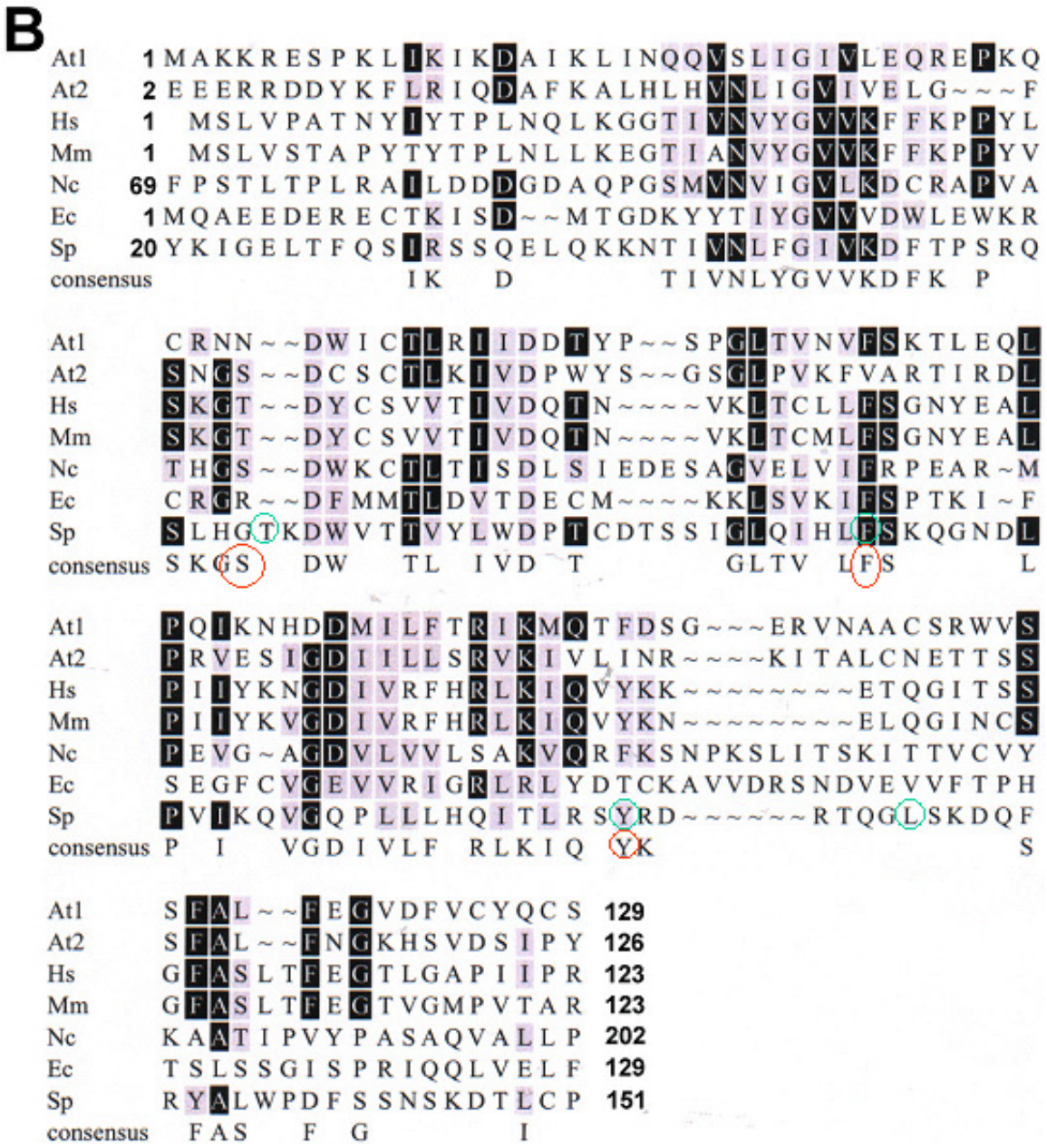

Figure 10. Continued.

C

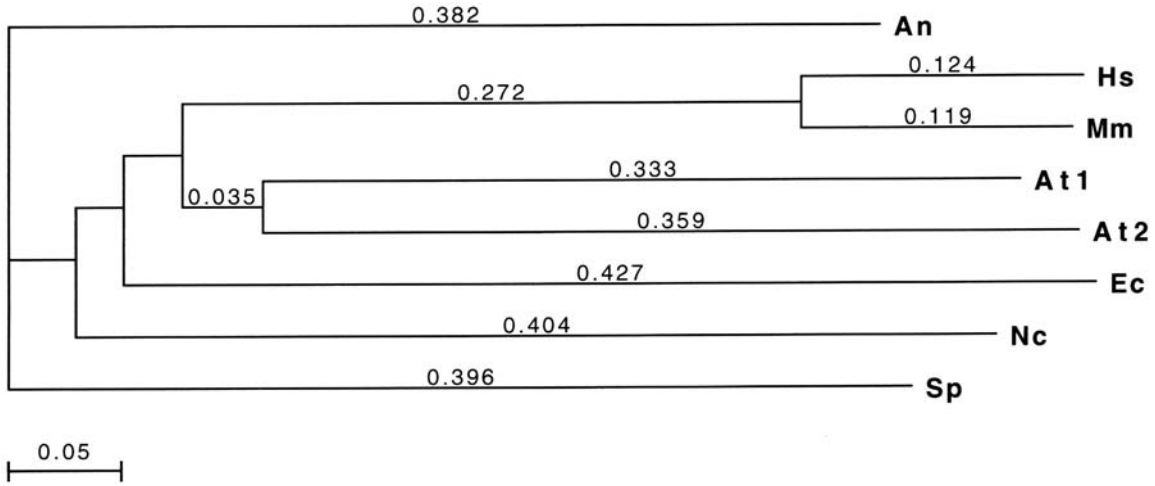

Figure 10. Continued.

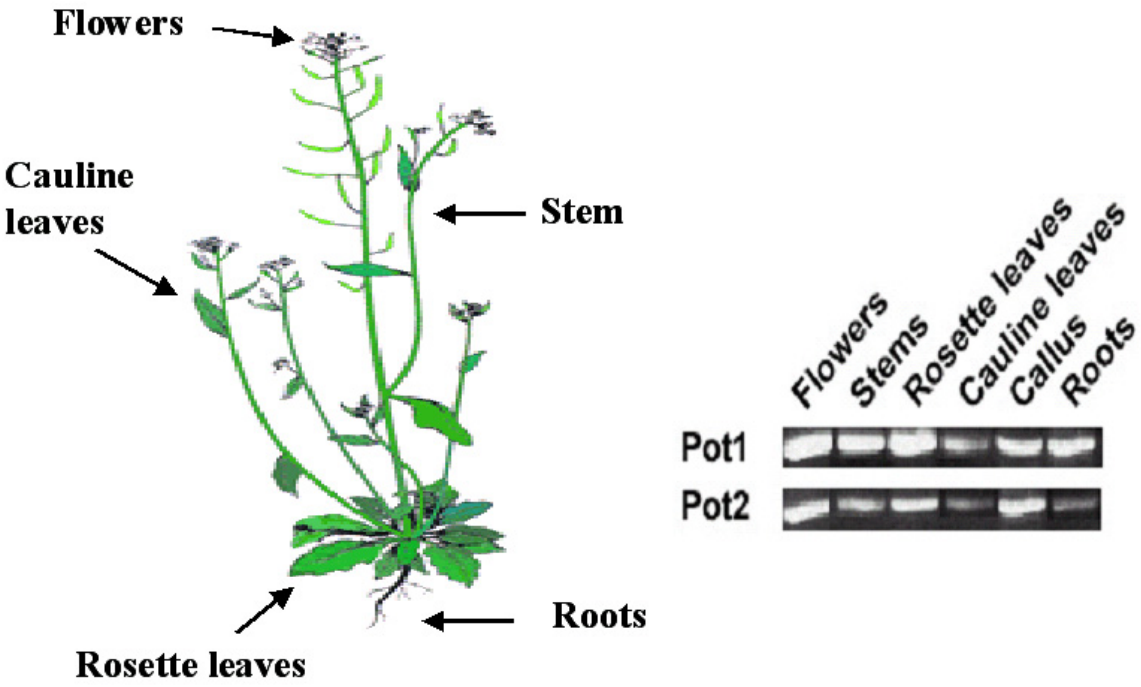

Figure 11. Expression of *AtPot1* and *AtPot2* mRNA in *Arabidopsis* tissues.

annealing efficiencies of primers used in our RT-PCR experiments, in general, *Pot2* mRNA is less abundant in all organs than *Pot1*, with especially low levels in vegetative tissues, such as rosette and cauline leaves, and stems. By contrast, actively dividing tissues, such as flowers, as well as dedifferentiated callus cells show increased levels of *Pot1* and *Pot2* mRNA expression. This ubiquitous pattern of expression is similar to that seen for human *Pot1* mRNA, which is present in all analyzed human tissues and cancer lines (Baumann et al., 2002). However, this is different from the pattern of *AtTERT* mRNA expression, which is only present in proliferating tissues (Fitzgerald et al., 1996). The ubiquitous presence of *Arabidopsis Pot1* and *Pot2* mRNAs suggests that these proteins could play a role at telomeres in all cells of the plant.

DNA binding properties of *Arabidopsis* Pot proteins

To examine DNA-binding specificity of Pot1, Pot2 and $P1^{DBD}$, recombinant proteins were over-expressed in *E. coli* and affinity purified by Ni-NTA affinity chromatography (Figure 12). The proteins were incubated with a variety of telomeric and non-telomeric substrate oligonucleotides, and binding was monitored by electrophoretic mobility shift assays. The sequence of all oligonucleotides used in this study is shown in Table 2. As shown in Figure 13, Pot1 bound telomeric oligonucleotides containing 6 repeats of *Arabidopsis* G-rich telomere sequence TTTAGGG and, but not an oligonucleotide corresponding to the C-rich telomeric strand or duplex telomeric DNA (Figure 13, lanes 2, 13, 14). The two observed protein-DNA complexes are specific, because binding was competed by an excess of cold telomeric oligonucleotide and not by a non-specific oligonucleotide of the same length (Figure 13, lanes 3-4 and 9-10). Pot1

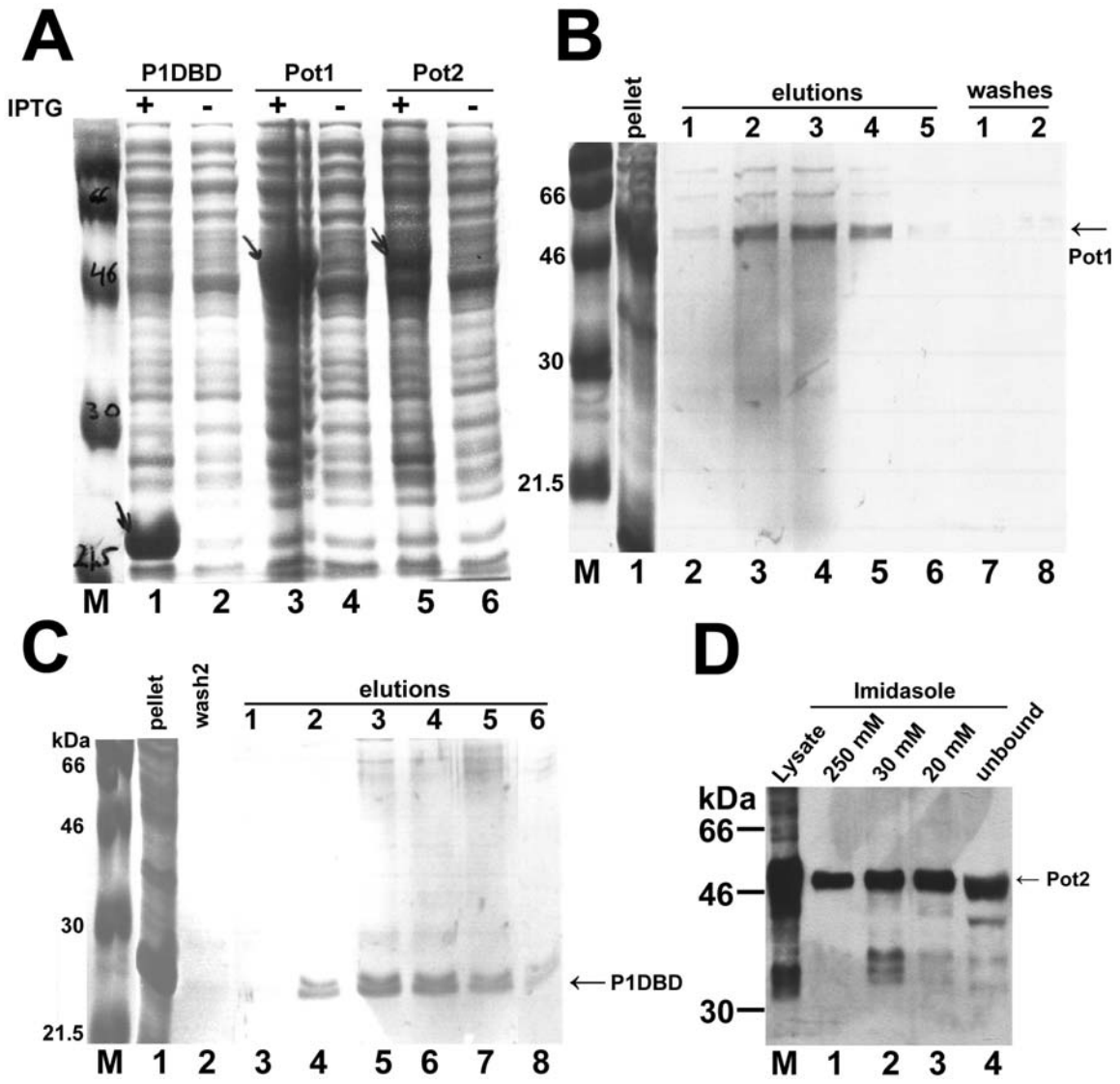

Figure 12. Expression and purification of Pot proteins.

(A) Over-production of Pot1, Pot2 and $P1^{DBD}$ proteins in *E.coli*. Lanes 2,4,6, proteins from *E.coli* cells grown in the absence of IPTG. The addition of IPTG (lanes 1,3,5) leads to the specific expression of the recombinant proteins. M- protein standards. All proteins were resolved by SDS-PAGE followed by staining with Coomassie. (B) Purification of Pot1 protein on a Ni-NTA column. The peak of Pot1 protein elutes off the column in fractions 2-4 (lanes 3-5). (C) Purification of $P1^{DBD}$ protein. The peak of the protein elutes in fractions 3-5. (D) Western blot of Ni-NTA-purified Pot2 protein. The protein leaks off the column and elutes in every fraction.

Table 2. Sequence of the oligonucleotides used in the gel-shift assays

Oligonucleotide name	Oligonucleotide sequence
TELO	AGCATGCAGC$(TTTAGGG)_6$
TELO-MID	AGCAT$(TTTAGGG)_6$GCAGC
TELO-UP	$(TTTAGGG)_6$AGCATGCAGC
NS	AGCATGCAGCTGGAGCTGTGATTCAGACGCTGCAC
5RNA	$(UUUAGGG)_5$
10NT-UP	AGCATGCAGC
C-rich	$(CCCTAAA)_5$
NON-TELO	$(CCCTAAA)_5$GCTGCATGCT
3TELO	$(TTTAGGG)_3$
4TELO	$(TTTAGGG)_4$
5TELO	$(TTTAGGG)_5$
6TELO	$(TTTAGGG)_6$
6HUMAN	$(TTAGGG)_6$
1MUT	$(TTTACGG)_5$

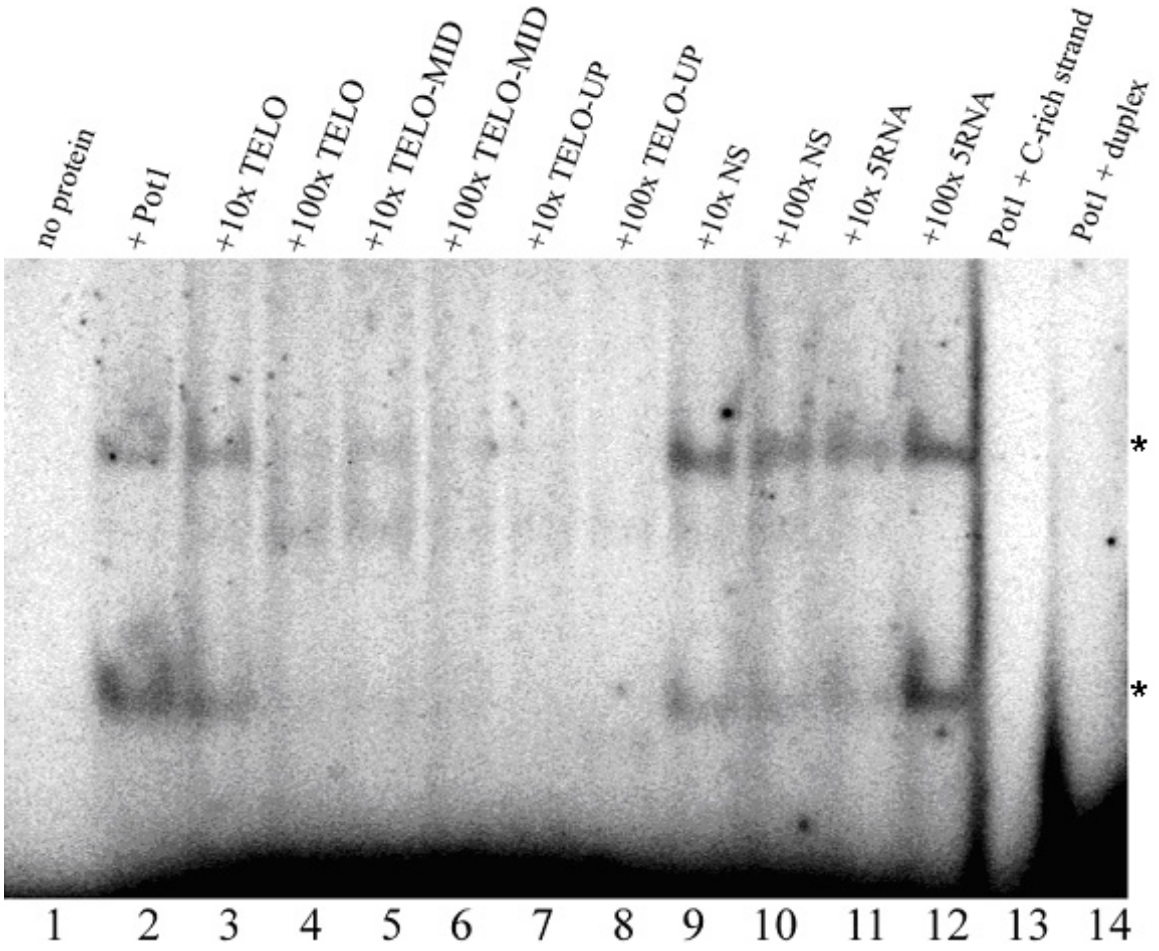

Figure 13. DNA-binding properties of recombinant Pot1 protein.

Electrophoretic mobility shift assays were performed. Lane 1, TELO oligonucleotide alone; lane 2, TELO plus Pot1 protein; lanes 3-12, addition of excess 10X (lanes 3,5,7,9,11) or 100X (lanes 4,6,8,10,12) cold competitor oligonucleotides; lane 13, labeled C-rich telomeric oligonucleotides plus Pot1; lane 14, labeled duplex telomeric substrates plus Pot1.

did not bind RNA; an RNA oligonucleotide of the telomeric sequence $(UUUAGGG)_5$ did not compete for binding (lanes 11,12). Since Pot1 proteins are proposed to bind both G-strand overhangs and the displaced G-strand DNA at t-loops (Baumann and Cech, 2001; de Lange, 2002), we tested whether *Arabidopsis* Pot1 requires telomeric DNA on the 3' end of the substrate (mimicking the G-overhang structure) or can bind any internal telomeric sequence. Unlike the *S. pombe* protein (Baumann and Cech, 2001), Pot1 did not show preference for free 3' end of the oligonucleotide. DNA substrate in which 10 non-telomeric nucleotides were placed either flanking or downstream of six telomeric repeats competed well for Pot1 binding (Figure 13, lanes 5-8), suggesting that *Arabidopsis* Pot1 may bind along the length of any long single-stranded substrate of telomeric sequence.

A truncated form of *Arabidopsis* Pot1, representing the 146 N-terminal amino acids ($P1^{DBD}$), was also able to specifically bind telomeric DNA. This DBD showed binding characteristics similar to the full-length Pot1 protein (Figure 14 and data not shown). The DNA-binding domain of *Schizosaccharomyces pombe* Pot1 also displays the same specificity for substrate binding as the full-length protein (Baumann and Cech, 2001). We next investigated minimum substrate length requirements for efficient DNA binding by $P1^{DBD}$ protein. Two retardation complexes were observed after incubation of $P1^{DBD}$ with telomeric oligonucleotide (Figure 14, lane 2). DNA binding was unaffected by the presence of a 100-fold excess of cold competitors harboring three-four repeats of telomeric sequence (lanes 3-4). On the contrary, oligonucleotides with five and six telomeric repeats competed well for binding, suggesting that $P1^{DBD}$ prefers to bind longer stretches of telomeric sequence (lanes 5-6). To further investigate the sequence

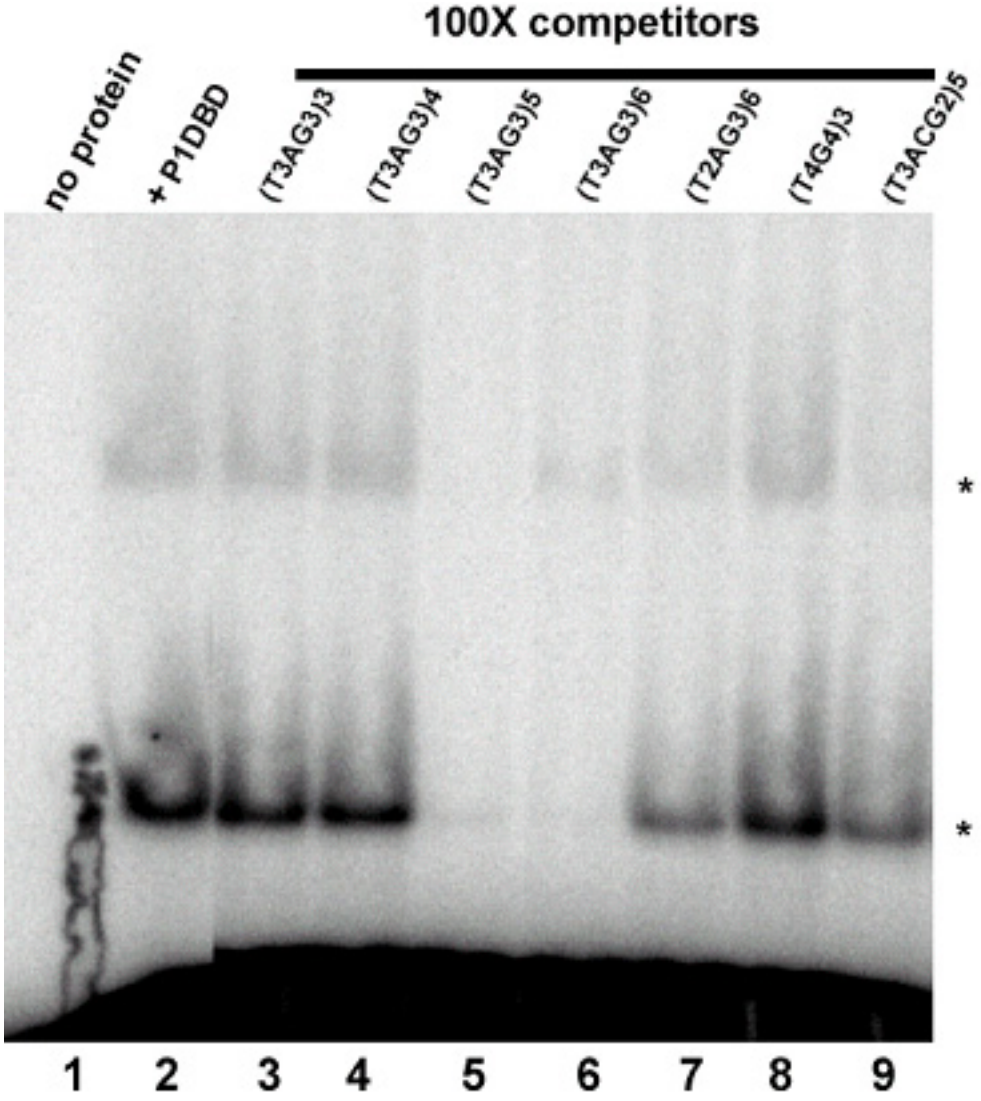

Figure 14. DNA-binding properties of recombinant $P1^{DBD}$ protein.

An electrophoretic mobility shift assay is shown. Lane 1, TELO oligonucleotide alone; lane 2, TELO plus $P1^{DBD}$ protein; lanes 3-9, addition of excess 100X cold competitor oligonucleotides.

specificity, we tested whether the G-rich strand of telomeric DNA from different species could serve as a substrate for binding. In our competition assays, 100-fold excess of the ciliate $(T_4G_4)_3$ sequence or the human sequence $(TTTAGGG)_6$ did not compete for $P1^{DBD}$ binding (Figure 14, lanes 7-8). In another test, adding a competitor with one G→C substitution within the *Arabidopsis* repeat also did not significantly reduce binding (Figure 14, lane 9). Thus, Pot1 exhibits a strong preference for plant telomeric DNA.

We next tested whether recombinant *Arabidopsis* Pot2 protein also binds telomeric substrates specifically. As shown in Figure 15, *Arabidopsis* Pot2 forms three separate complexes with telomeric substrate (Figure 15, lane 2). All three complexes were competed away by the addition of 150-fold excess of cold oligonucleotides in which the 6 telomeric repeats are placed on the 3' end, in the middle, and on the 5' end of the oligonucleotides (Figure 15, lanes 3-5, respectively). Overall, mobility shift experiments with Pot2 protein yielded results very similar to those seen for Pot1 protein.

Oligomerization of *Arabidopsis* Pot proteins

Telomere binding proteins engage in many protein-protein interactions, forming homo- and heterodimers and binding to other proteins in large multifunctional complexes (Peersen et al., 2002; Lustig, 2001). To determine whether *Arabidopsis* Pot1 and Pot2 form homo- and heterodimers, we performed a series of co-immunoprecipitation experiments with full-length Pot1 and Pot2 proteins expressed in rabbit reticulocyte lysate. Pot1 and Pot2 were expressed either as T7-tag fusions or ^{35S}Met labeled untagged proteins. Two differentially labeled proteins were combined in one reaction and then subjected to co-immunoprecipitation using T7 antibody agarose beads (Figure 16A). In

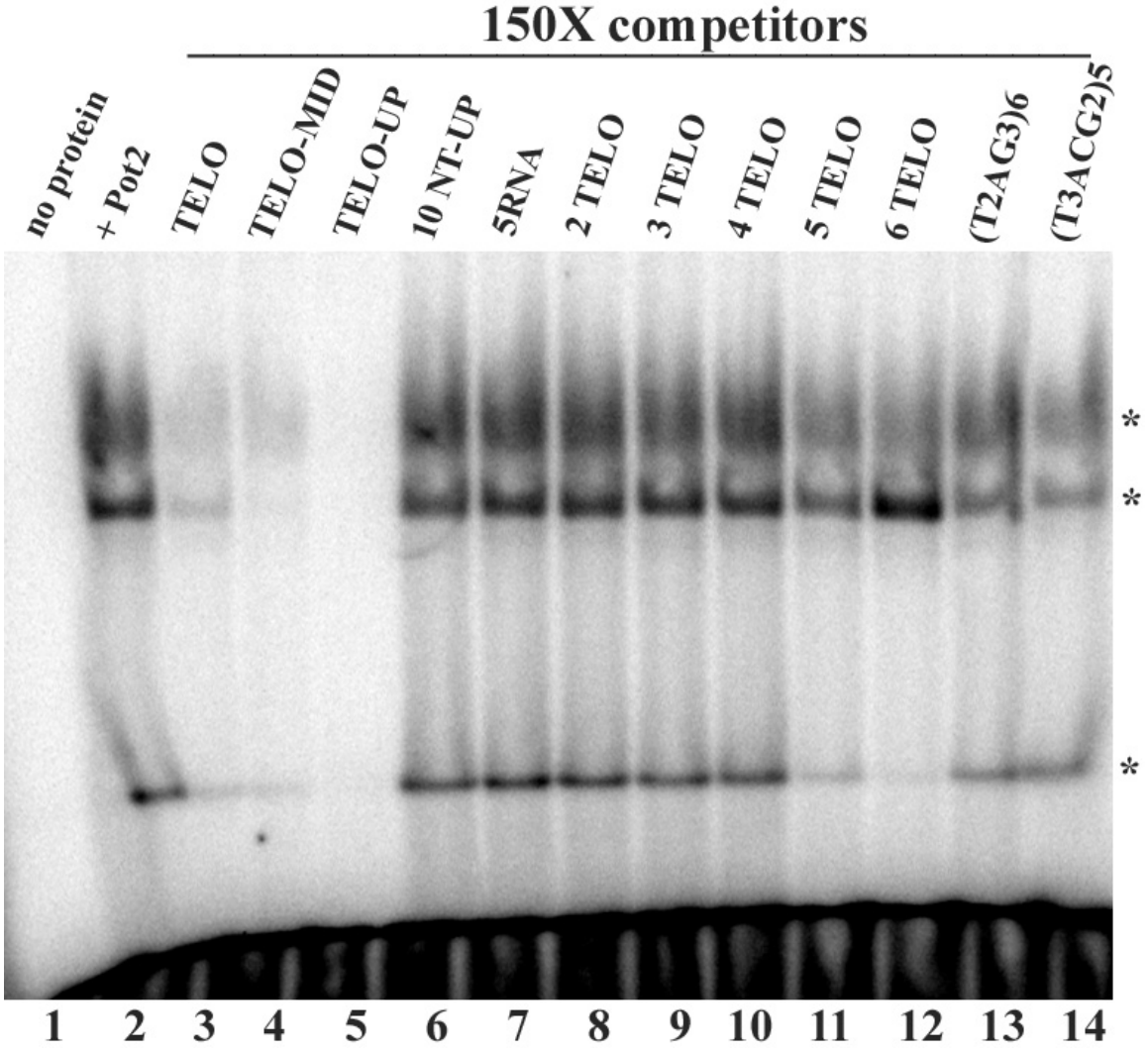

Figure 15. DNA-binding properties of recombinant Pot2 protein.

An electrophoretic mobility shift assay is shown. Lane 1, TELO oligonucleotide alone; lane 2, TELO plus Pot2 protein; lanes 3-14, addition of excess 150X cold competitor oligonucleotides.

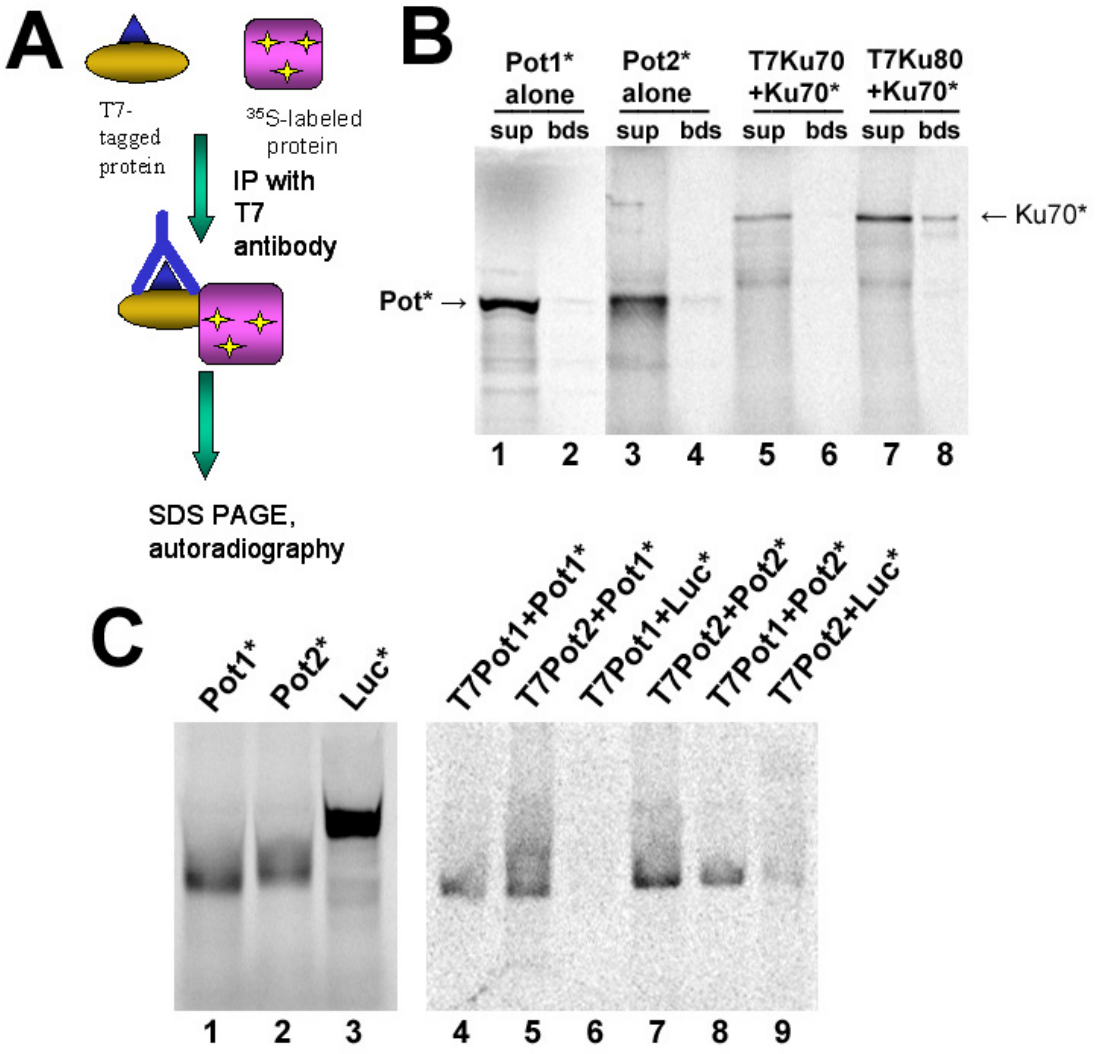

Figure 16. *In vitro* dimerization of Pot1 and Pot2 proteins.

(A) Schematic diagram of the pull-down assay. (B) Control pull-down assays. Radioactive Pot1 and Pot2 proteins expressed in rabbit reticulocyte lysate (lanes 1, 3) do not bind to beads in the absence of their T7-tagged interactors (lanes 2, 4). Radioactive *Arabidopsis* Ku70 protein forms a stable interaction with T7-tagged Ku80 protein (lane 7-8), but not with itself (lanes 5-6). (C) *Arabidopsis* Pot1 and Pot2 proteins homo- and hetero-dimerize *in vitro*. Radioactively labeled Pot1, Pot2 and luciferase proteins (lanes 1-3) were incubated with T7-tagged Pot1 and Pot2 proteins. Pot1 interacts with itself (lane 4) and Pot2 (lane 5), but not with luciferase (lane 6). Similarly, Pot2 interacts with itself (lane 7) and Pot1 (lane 8), but not with luciferase (lane 9).

general, if a T7-tagged protein interacts with its radioactively labeled partner, both proteins can be co-immunoprecipitated in the pull-down assay and then separated on an SDS-PAGE. The radioactively labeled interaction partner is visualized by autoradiography.

Arabidopsis Ku70 and Ku80 proteins were used as positive and negative controls. As expected, Ku70 associated with Ku80, but did not form homodimers (Figure 16B, lanes 5-8). As shown in Figure 16C, *Arabidopsis* Pot1 and Pot2 formed homo- and heterodimers. To address whether Pot1 and Pot2 proteins can interact with each other, we performed reciprocal pull-down assays using either Pot1 or Pot2 as a T7-tagged bait. Both ^{35S}Pot1 and ^{35S}Pot2 protein co-purified with their heterologous partners, indicating that Pot1 and Pot2 also have the ability to heterodimerize *in vitro* (Figure 16C, lanes 5 and 8). All Pot1 and Pot2 associations were specific as neither T7-tagged Pot1 nor Pot2 proteins immunoprecipitated radioactively labeled luciferase (Figure 16C, lanes 6 and 9). In addition, when radioactive Pot1 and Pot2 were incubated with T7 antibody agarose beads in the absence of their T7-tagged partner, only negligible background levels of binding were observed (Figure 16B, lanes 1-4).

The yeast two-hybrid assay provided further evidence for Pot1 and Pot2 homo- and heterodimerization. To confirm Pot1 and Pot2 interaction results obtained in our *in vitro* system, we wanted to use an alternative dimerization assay in a heterologous system. Both Pot1 and Pot2 proteins were expressed as Gal4 activation domain (GAD) and Gal4 DNA-binding domain (GDBD) fusions, and interactions were detected on selective plates. The yeast two-hybrid results were consistent with *in vitro* pull-down assays, and growth was observed in cells expressing GAD-Pot1 + GDBD-Pot1, GAD-

Pot2 + GDBD-Pot2, GAD-Pot1 + GDBD-Pot2 and GAD-Pot2 + GDBD-Pot1 fusions, but not in any negative control strains (data not shown). The complete list of all strains used in the study and the observed phenotypes is shown in Table 3. Pot1 and Pot2 interactions detected by yeast two-hybrid assay and by co-immunoprecipitation were reproducible, but relatively weak, which correlates well with the situation in chicken cells, where less than 5% of all chicken Pot1 protein molecules dimerize (C. Price, personal communication). Nonetheless, we conclude that *Arabidopsis* Pot1 and Pot2 proteins can form homo- and heterodimers *in vitro*. If oligomerization occurs *in vivo*, it may be a powerful mechanism to regulate protein functions.

Subcellular localization of *Arabidopsis* Pot proteins

Since Pot1 proteins bind the G-rich strand of telomeric DNA, they are expected to localize to the nucleus. Indeed, human Pot1 protein localizes to telomeres in interphase nuclei (Baumann et al., 2002). To determine the subcellular localization of *Arabidopsis* Pot1 and Pot2, we prepared EYFP-fusion Pot1 and Pot2 constructs and transiently expressed them from the 35S CaMV promoter in tobacco cells. As a negative control, we used EYFP alone, which gave a diffuse non-specific staining of the plasma membrane (Figure 17A). For positive controls we analyzed two previously described nuclear proteins, TGA5 (Xiang et al., 1997) and *Arabidopsis* Ku70 (Riha et al., 2002). Both proteins showed strong fluorescence signal exclusively in the nucleus (panels B and C).

As expected for a telomere protein, Pot2 localized to the nucleus (Figure 17, panel E). However, Pot1 showed a strikingly different localization. The majority of the Pot1 protein localized to the nucleolus, with the remainder diffusely staining the nucleus

Table 3. Yeast two-hybrid strains and the observed phenotypes of Pot1 and Pot2 interactions

Strain	Interaction	Selection	LacZ phenotype	Growth on –His medium
Pot1-pACT2 Pot2-pAS2-1	Heterodimerization	-Leu/-Trp	Blue	Yes
Pot2-pACT2 Pot1-pAS2-1	Heterodimerization	-Leu/-Trp	Blue	Yes
Pot1-pACT2 Pot1-pAS2-1	Homodimerization	-Leu/-Trp	Blue	Yes
Pot2-pACT2 Pot2-pAS2-1	Homodimerization	-Leu/-Trp	Blue	Yes
Pot1-pACT2	Negative control	-Leu	White	No
Pot2-pACT2	Negative control	-Leu	White	No
Pot1-pAS2-1	Negative control	-Trp	White	No
Pot2-pAS2-1	Negative control	-Trp	White	No
pACT2 pAS2-1	Negative control	-Leu/-Trp	White	No
Pot1-pACT2 pVA3-1	Negative control	-Leu/-Trp	White	No
Pot2-pACT2 pVA3-1	Negative control	-Leu/-Trp	White	No
pCL1	Positive control for β-Gal expression	-Leu	Blue	Yes
pVA3-1	Negative control	-Leu	White	No
pTD1-1	Negative control	-Trp	White	No
pVA3-1 pTD1-1	Positive control for interaction	-Leu/-Trp	Blue	Yes
pLAM5'-1	Negative control, lamin	-Trp	White	No
pAS2-1 Pot1-pACT2	Negative control	-Leu/-Trp	White	No
pAS2-1 Pot2-pACt2	Negative control	-Leu/-Trp	White	No

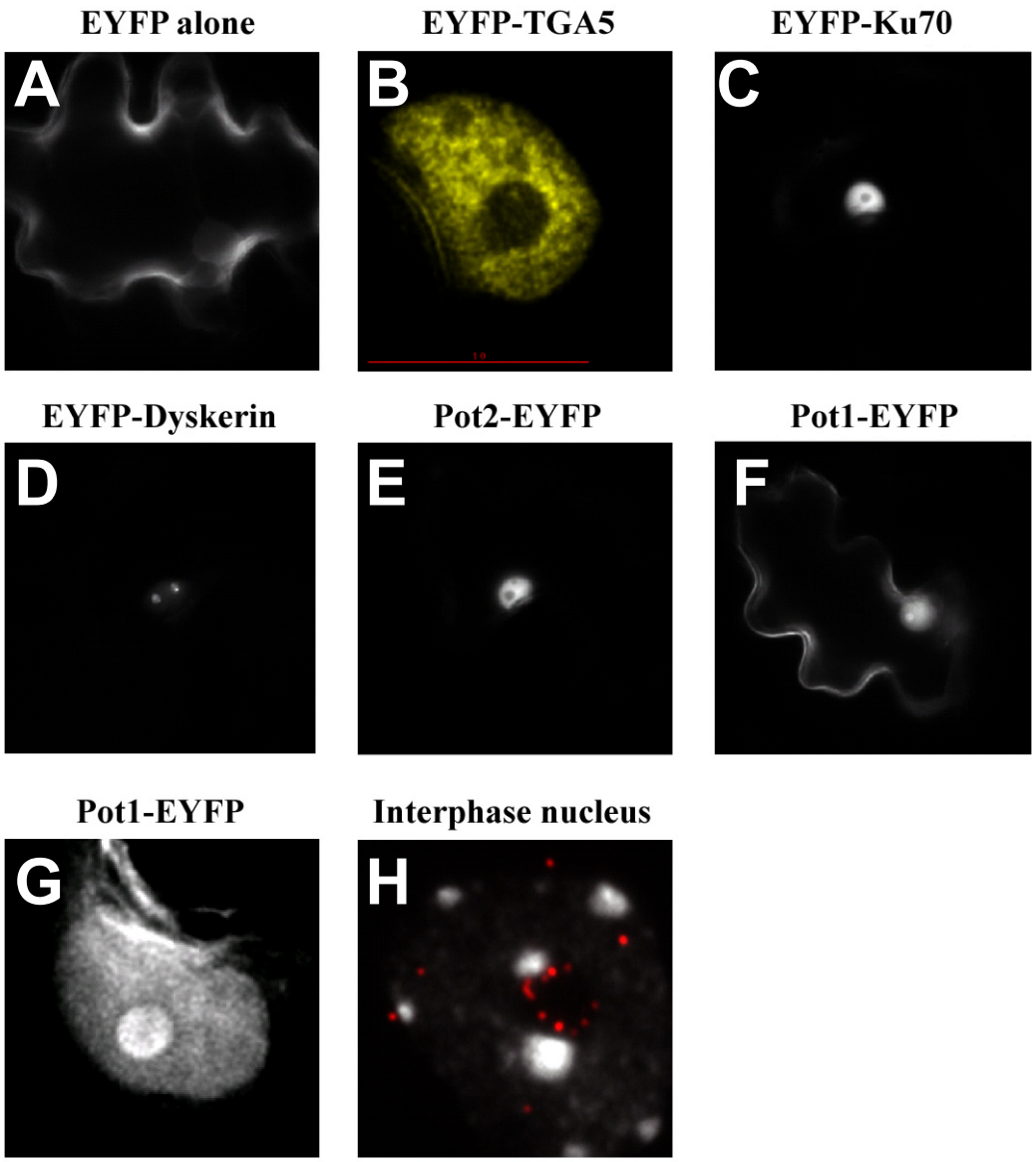

Figure 17. Cytological analysis of telomeres and telomere-associated proteins.

(A-G) Fluorescence microscopy of tobacco nuclei transiently transfected with EYFP (A), EYFP-TGA5 (B), EYFP-Ku70 (C), EYFP-Dyskerin (D), Pot2-EYFP (E), Pot1-EYFP (F, G). In all cases, the constructs were expressed from the 35S CaMV promoter. Panels B, G and H show a higher magnification of the nucleus. Pot1 protein localizes to the nucleus and is concentrated in the nucleolus. (H) FISH analysis of *Arabidopsis* telomeres using a biotin-UTP labeled TTTAGGG repeat probe detected by Avidin-Texas Red. The nucleus fills most of the frame with the telomeres clustering in the dark staining nucleolus. This work carried out in collaboration with Dr. Naohiro Kato and Dr. Eric Lam at Rutgers University.

(panels F, G). To verify that the nucleolar localization was specific, we examined previously characterized *Arabidopsis* protein, dyskerin (Maceluch et al., 2001). As expected for this protein, dyskerin localized exclusively to the nucleolus (panel D).

If *Arabidopsis* Pot proteins are indeed telomere binding factors *in vivo*, they are predicted to localize more specifically to plant telomeres. Experiments aimed at the direct co-localization of Pot proteins with telomeric DNA are impossible at the moment pending the development of antibodies specific for the endogenous *Arabidopsis* Pot1 and Pot2 proteins. However, the evident nucleolar localization of Pot1 is particularly interesting, as previous studies showed that *Arabidopsis* telomeres primarily localized to the nucleolus (Figure 17, panel J) in both meiotic and mitotic interphase (Armstrong et al., 2001). Localization of most telomeres and Pot1 protein, but not Pot2, to the same compartment of the cell is consistent with a role of Pot1 in telomere biology. Our subcellular localization experiments imply that Pot1 and Pot2 may have distinct functions in the cell.

Over-expression of AtPot1 and AtPot2 proteins in wild-type *Arabidopsis* plants

To examine the role of *Arabidopsis* Pot1 and Pot2 proteins *in vivo*, we attempted to obtain gene knock-outs. Unfortunately, appropriate T-DNA lines for these genes are currently unavailable. Therefore, we tested the effects of exogenous over-expression of Pot1, Pot2 or their truncation derivatives on telomere dynamics. For these studies, several transgene constructs were created for each gene: full-length, and an N-terminal and a C-terminal truncation. The C-terminal truncations contained the DNA-binding domains of the two proteins ($P1^{DBD}$ and $P2^{DBD}$), while the N-terminal truncations lacked the DBD

and corresponded to the remainder of the ORF ($P1^{\Delta DBD}$ and $P2^{\Delta DBD}$). The constructs were placed under the control of the CaMV 35S promoter to ensure robust expression (Figure 18). The presence of transgene mRNA was verified in all transformants by RT-PCR (Figure 18A,B and data not shown). Expression of full-length Pot1 and Pot2 proteins caused no defects in growth or development of the transgenic plants, and the length of their telomeres remained unchanged. However, initial experiments suggested that plants over-expressing portions of the Pot1 and Pot2 proteins, $P2^{DBD}$ and $P1^{\Delta DBD}$, displayed reproducible phenotypes, and these plants were further evaluated.

Over-expression of DNA-binding domain of Pot2 leads to severe morphological defects, moderate telomere shortening and telomere uncapping

Among many $P2^{DBD}$-over-expressing plants, several individuals showed distinct and unusual morphological changes at 3 weeks of age (Figure 19 A, B). Such plants had delayed growth and flowering time, and many lacked apical meristem dominance, producing multiple shoots. In many transformants, rosette leaves were small, wrinkled and curled-down (Figure 19C). Although numerous small siliques formed, they were essentially sterile, with each plant producing only few viable seeds. Several attempts to make reciprocal crosses of these mutants to wild-type plants failed, consistent with major defect in both male and female reproductive systems. Overall, 12.5% (6 out of 48) of the primary transformants with the $P2^{DBD}$ transgene displayed this phenotype.

Over-expression of $P2^{DBD}$ also resulted in an altered telomere phenotype. Although most $P2^{DBD}$-expressing plants had wild-type telomere length (Figure 20, lanes 3-4, 7-9, 11-12), those plants with severe morphological defects also showed telomere

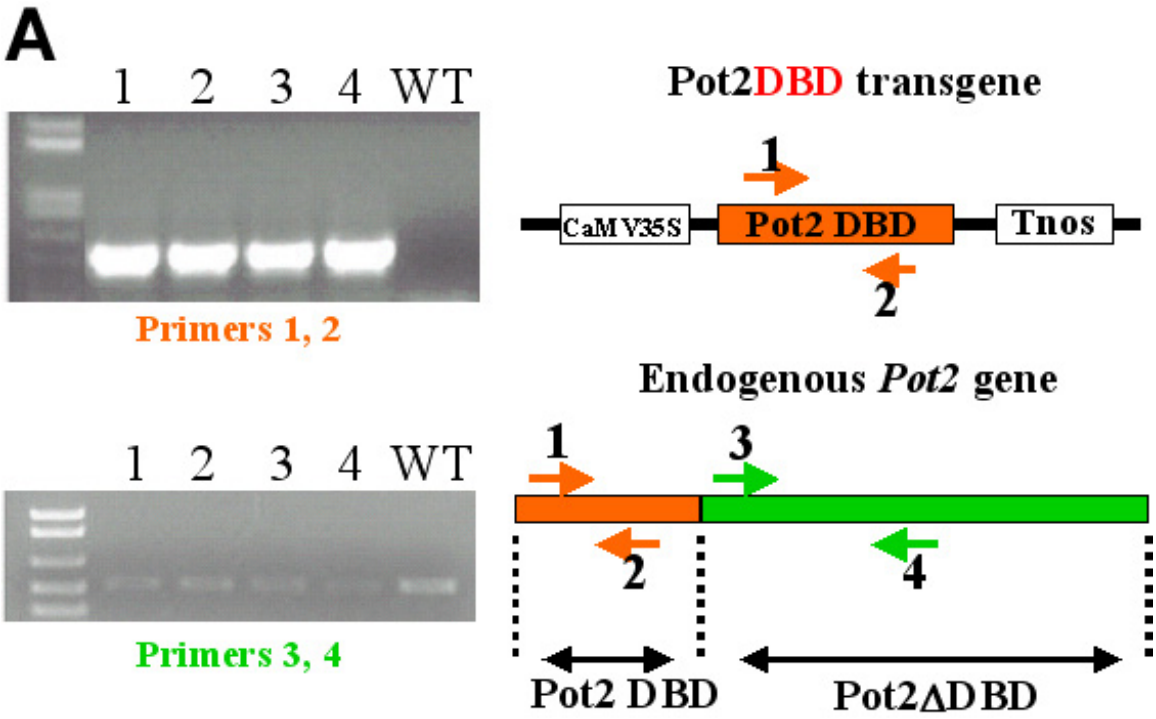

Figure 18. Analysis of transgene expression in transformed *Arabidopsis* plants.
RT-PCR results of $P2^{DBD}$ (A) or $P1^{\Delta DBD}$ (B) expression in transgenic plants (left) and schematic of the transgenes and their corresponding endogenous alleles (right). All transgenic and wild-type plants contain the endogenous *Pot1* and *Pot2* genes. (A) Primers 1 and 2 are specific for both $P2^{DBD}$ transgene and the endogenous *Pot2* allele. After 20 PCR cycles, these primers amplify $P2^{DBD}$ mRNA only in the transgenic plants (upper panel, lanes 1-4), but not in the wild-type plant, confirming that the $P2^{DBD}$ transgene is being over-expressed. The control primers 3 and 4 are only specific to the endogenous *Pot2* gene. After 40 cycles, these primers amplify the *Pot2* mRNA in all plants equally well (lower panel). (B) The control primers 1 and 2 are specific for the endogenous *Pot1* mRNA. After 40 cycles, these primers amplify *Pot1* mRNA in all plants equally well (upper panel). Primers 3 and 4 are specific for both the $P1^{\Delta DBD}$ transgene and the endogenous *Pot1* allele. After 20 PCR cycles, these primers amplify $P1^{\Delta DBD}$ mRNA only in the transgenic plants (lower panel, lanes 1-3), but not in the wild-type plant, confirming that the $P1^{\Delta DBD}$ transgene is being over-expressed.

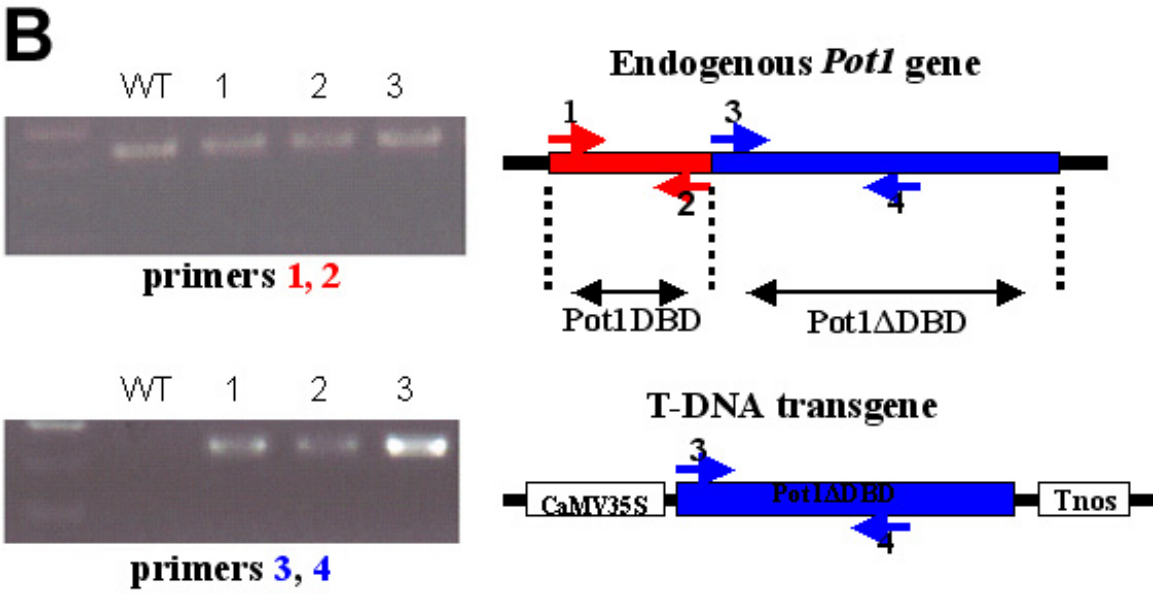

Figure 18. Continued.

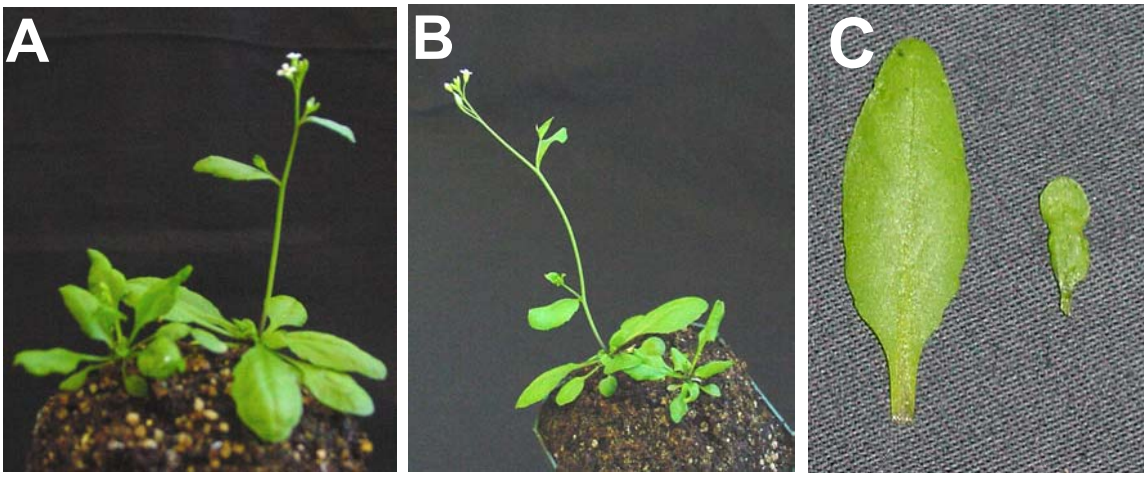

Figure 19. Morphological phenotypes of P2DBD mutant plants.

Growth delay of P2DBD-over-expressing mutants (A and B). Smaller plants (mutants) start flowering several days later, after the wild-type plants of the same age have already started setting the first seeds. (C) Leaf defects of P2DBD mutant plants (on the right) are shown. The wild-type rosette leaf is shown on the left.

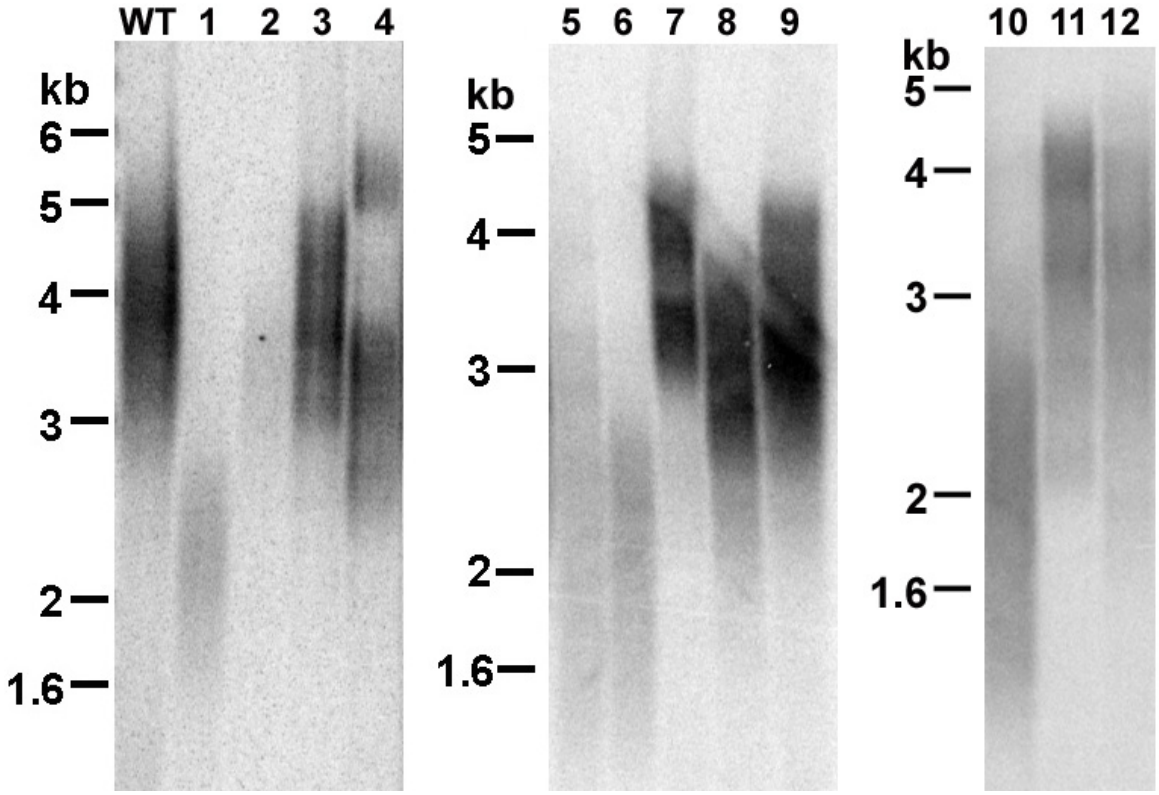

Figure 20. Over-expression of P2DBD leads to telomere shortening.

TRF analysis of the primary transformants. Some P2DBD transformants have wild-type telomere length (lanes 3-4, 7-9, 11-12), whereas other plants, which display morphological defects, also have shortened telomeres (lanes 1-2, 5-6, 10). WT- telomere length in wild-type Columbia plants.

shortening (Figure 20, lanes 1-2, 5-6, 10). Some of these plants lost up to 2kb of telomeric sequence in the first generation (compare WT lane with lanes 1 and 6), a rate of telomere shortening faster than in telomerase-deficient plants (Fitzgerald et al., 1999; Riha et al., 2001). TRAP assays revealed that telomerase remained active in $P2^{DBD}$ mutants, hence telomere shortening was not due to telomerase inactivation.

The severe morphological defects associated with $P2^{DBD}$ over-expression can not be explained by their relatively moderate telomere shortening. The extent of telomere shortening in these plants corresponds to G_2-G_3 TERT mutants, which are viable and show no morphological defects or genome instability (Fitzgerald et al., 1999). Instead, the phenotypes associated with $P2^{DBD}$ mutants correspond to the very late generation TERT mutants, which are sterile and exhibit multiple chromosome fusions. To further investigate the effects of $P2^{DBD}$ over-expression on cell proliferation capacity, we performed cytogenetical analysis of metaphase spreads of *Arabidopsis* chromosomes in these mutants. 8% of all metaphases contained anaphase bridges, consistent with the formation of the dicentric chromosomes as a result of telomere end-to-end fusions. Remarkably, 50% of all anaphases with fusions contained two and more bridges (Figure 21), indicative of massive genome instability.

The data presented above suggest that $P2^{DBD}$ protein may function as a dominant-negative version of the full-length Pot2 protein. We suspect that over-expression of the transgene results in the displacement of endogenous Pot2 protein from *Arabidopsis* telomeres. Without the C-terminal portion of the Pot2 protein, this truncated derivative is deficient in a telomere-related function. Since the length of the telomeres in $P2^{DBD}$ over-expressing mutants is not short enough to trigger uncapping, the high percentage of

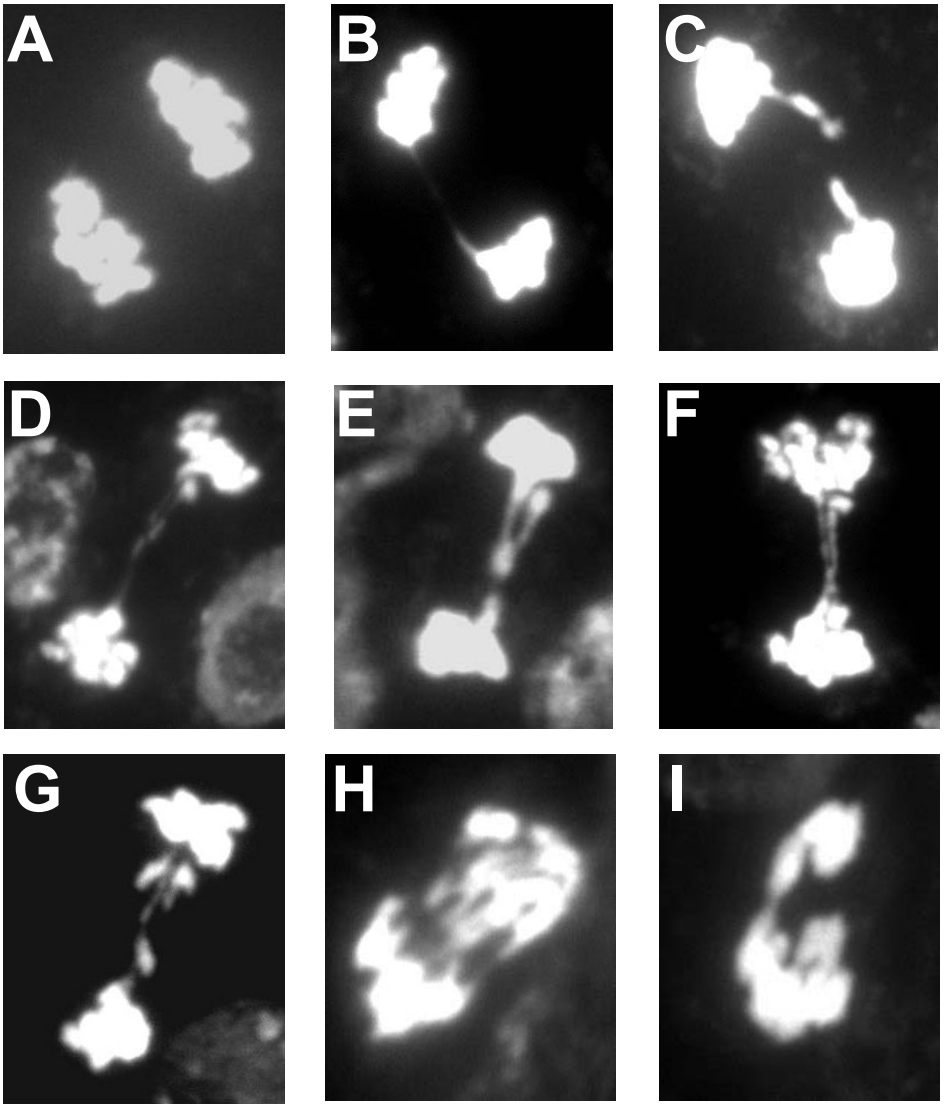

Figure 21. Cytogenetic defects in $P2^{DBD}$-over-expressing plants.

Anaphase bridges were observed in actively dividing mitotic tissues of pistils. Wild-type anaphase plane is shown in panel A. A single anaphase bridge (panel B), sometimes broken (panel C) is evident in some cells. Approximately 50% of all cells with fused chromosomes (panels D-F) harbor two or more anaphase bridges. In several cases, large pieces of chromosomes are trapped (panel G). Some cells contained an unusually high number of anaphase fusions, involving most chromosomes (panel H). In addition to multiple fusions, several cells displayed a clear chromosome mis-segregation phenotype, resulting in apparent aneuploidy (panel I).

anaphase bridges in these plants argues that Pot2 provides a protective capping function for the chromosome terminus. *Pot2* deficiency triggers telomere dysfunction and chromosome fusions, which ultimately leads to the proliferation defects and sterility associated with $P2^{DBD}$ mutants. Overall, the combined morphological and telomere phenotypes observed in $P2^{DBD}$ over-expression mutants suggest that *Arabidopsis* Pot2 is an important protein involved in telomere length regulation, telomere capping and general proliferation capacity of plant cells.

Over-expression of $P1^{\Delta DBD}$ results in shorter telomeres

Unlike plants over-expressing $P2^{DBD}$, $P1^{\Delta DBD}$ transformants were indistinguishable from wild-type plants and were fertile. Although most primary transformants had wild-type telomeres (Figure 22A, lanes 2, 4), a subset displayed reduced telomere length (Figure 22A, lanes 1, 3). No additional shortening was observed in the next generation of selfed mutants, with telomeres in some of the siblings returning back to the wild-type level. (Figure 22B, compare siblings 1-3 and 4-5; 7,9 and 6,8). These results were reproducible in the progeny of all analyzed primary transformants. This sibling-to-sibling variation in telomere length is not observed in wild-type plants (Figure 5A). Given the tight regulation of telomere length in *Arabidopsis*, telomere loss in this genetic background was not progressive, and later stabilized within the 1.5-5 kb range. We conclude from these data that over-expression of $P1^{\Delta DBD}$ leads to non-progressive telomere length shortening.

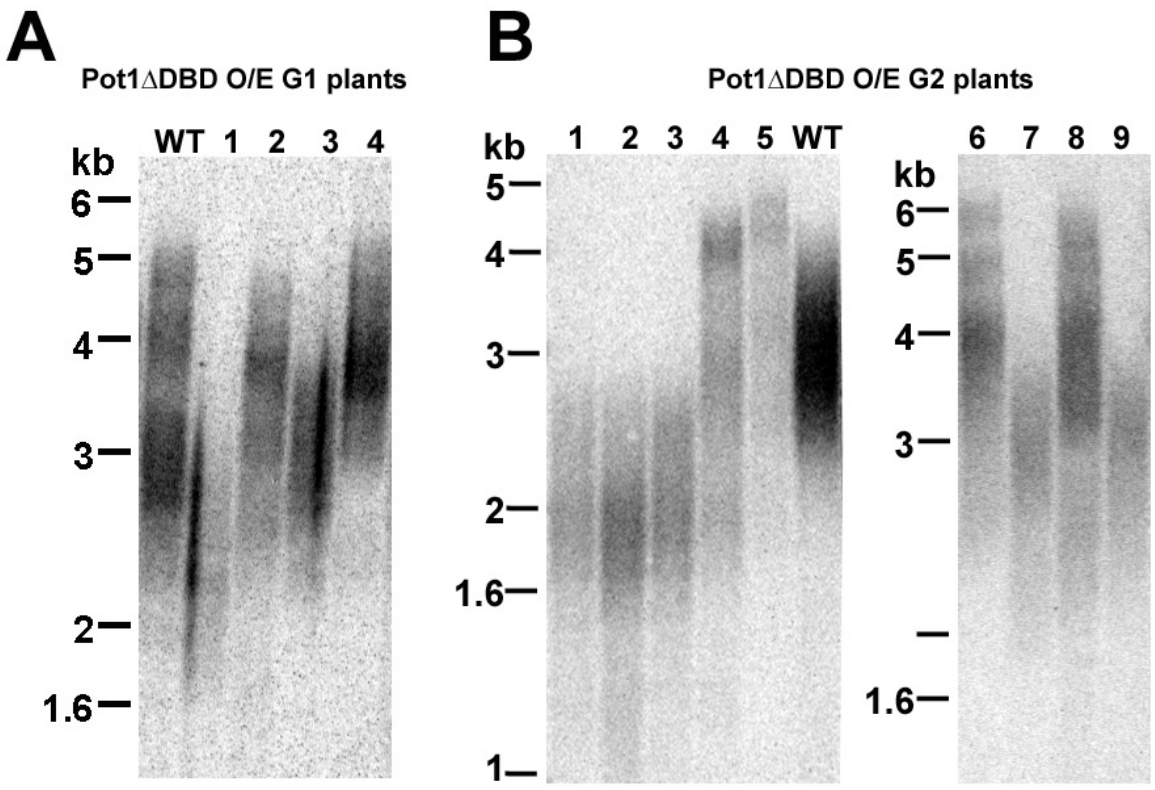

Figure 22. Over-expression of $P1^{\Delta DBD}$ leads to telomere shortening in the first generation of the mutants.

TRF analysis of $P1^{\Delta DBD}$ transformed plants. (A) Some $P1^{\Delta DBD}$ primary transformants have wild-type length telomeres (lanes 2, 4), whereas other plants have shortened telomeres (lanes 1, 3). (B) Analysis of the second generation (G2) of the mutant plants. DNA in lanes 1-5 comes from the progeny of the plant #1 from panel A. DNA in lanes 6-9 comes from the progeny of the plant #3 from panel A. No additional telomere shortening was observed in G2 plants, compared to their parents. WT- telomere length in wild-type Columbia plants.

Over-expression of $P2^{DBD}$-$P1^{\Delta DBD}$ chimeric protein leads to dramatically shortened telomeres and genome instability

Since Pot1 localizes to the same subnuclear compartment as telomeres, it seems likely that Pot1, and not Pot2, will be bound to telomeres throughout most of the cell cycle. If this was the case, chimeric fusions in which the DBD of one protein was swapped for the other would compromise function. To address this possibility, we created a chimeric construct, consisting of $P2^{DBD}$ and $P1^{\Delta DBD}$ cDNAs fused in frame to encode a single $P2^{DBD}$-$P1^{\Delta DBD}$ polypeptide, and placed it under control of the 35S promoter. As shown in Figure 23, over-expression of this chimeric protein led to a novel telomere phenotype. In four out of five primary transformants, telomeres decreased in size to below 2 kb, consistent with a major defect in telomere length maintenance. Telomeres in the first generation of these mutants were approximately the same size as telomeres in sixth or seventh generation TERT mutants (Riha et al., 2001). Thus, plants over-expressing $P2^{DBD}$-$P1^{\Delta DBD}$ lose telomeric DNA 6-7 times faster than telomerase-deficient plants.

As in TERT mutants, telomeres in $P2^{DBD}$-$P1^{\Delta DBD}$ over-expressing plants displayed a discrete banding profile, instead of the smear that is characteristic of telomeres in wild-type *Arabidopsis* . We previously showed in TERT mutants that these discrete bands represent individual telomere tracts. Hence, telomerase is primarily responsible for length heterogeneity at telomeres. Since telomeres in $P2^{DBD}$-$P1^{\Delta DBD}$ mutants shorten much faster than in TERT mutants, our data suggest that an active mechanism of telomere shortening, such as a nuclease attack, as opposed to the passive loss due to the end-replication problem, dramatically accelerates the loss of telomeric DNA in these mutants.

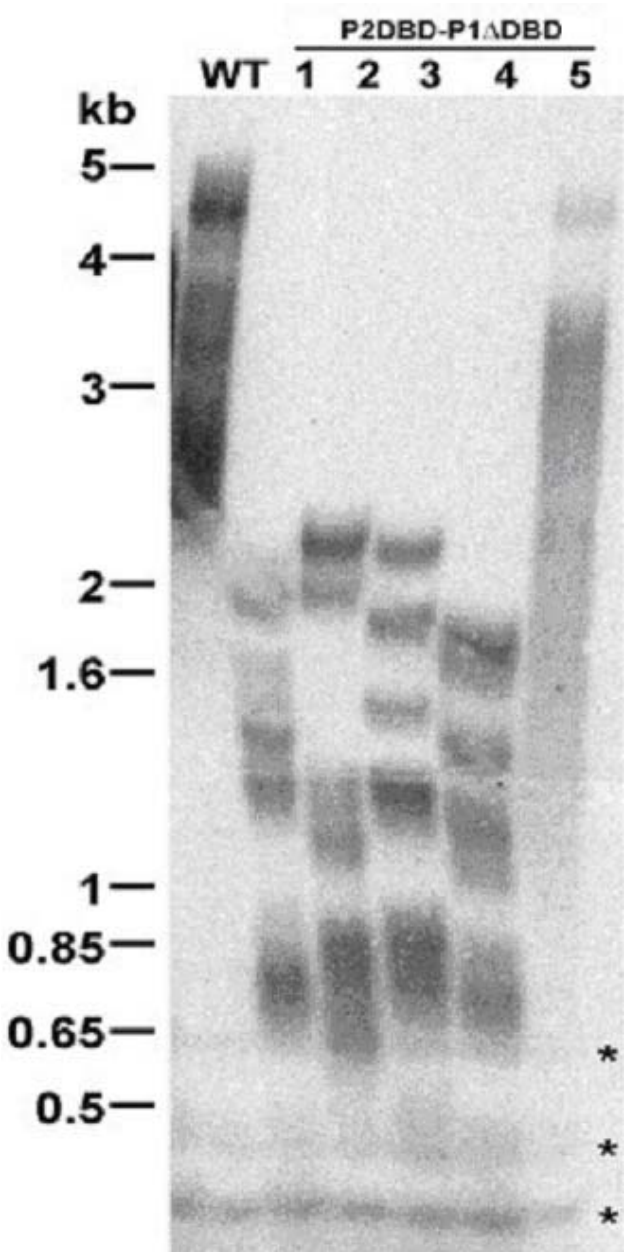

Figure 23. Telomere shortening in $P2^{DBD}$-$P1^{\Delta DBD}$ plants.

TRF analysis of the plants, transformed with $P2^{DBD}$-$P1^{\Delta DBD}$. Lanes 1-5, individual primary transformants. Asterisks indicate interstitial telomere signal. WT- telomere length in wild-type Columbia plants.

Cytogenetic analysis of $P2^{DBD}$-$P1^{\Delta DBD}$ mutants showed that ~5% of the metaphases contained anaphase bridges (Figure 24). Interestingly, most fusions had a different architecture than that observed in $P2^{DBD}$ over-expressing plants. Instead of long threads of pulled DNA in the middle of the anaphase plane, fusions in $P2^{DBD}$-$P1^{\Delta DBD}$ plants involved large, compact pieces of DNA, perhaps, even entire chromosomes trapped between two separating sets of DNA molecules (Figure 24, panels D-F). In addition to anaphase bridges, many metaphases showed other problems, such as unequal chromosome alignment and segregation, leading to aneuploidy in the daughter cells (Figure 24, panels H,I). These mitotic problems are similar to those observed in *A. nidulans pot1*$^{+}$-deficient cells, in which increased chromosome instability, segregation errors and a loss of viability are observed (Pitt et al., 2004). Analysis of the next generation of $P2^{DBD}$-$P1^{\Delta DBD}$ mutants will be crucial to understanding the consequences of such aberrations. The percentage of anaphase bridges observed in these plants is similar to that seen in G_6 TERT mutants. Together with the similarly comparable size of the telomeric DNA tract left in $P2^{DBD}$-$P1^{\Delta DBD}$ plants and G_6 TERT mutants, this data suggests that the observed anaphase bridges involve telomeric DNA and are triggered by critical shortening of the telomeres.

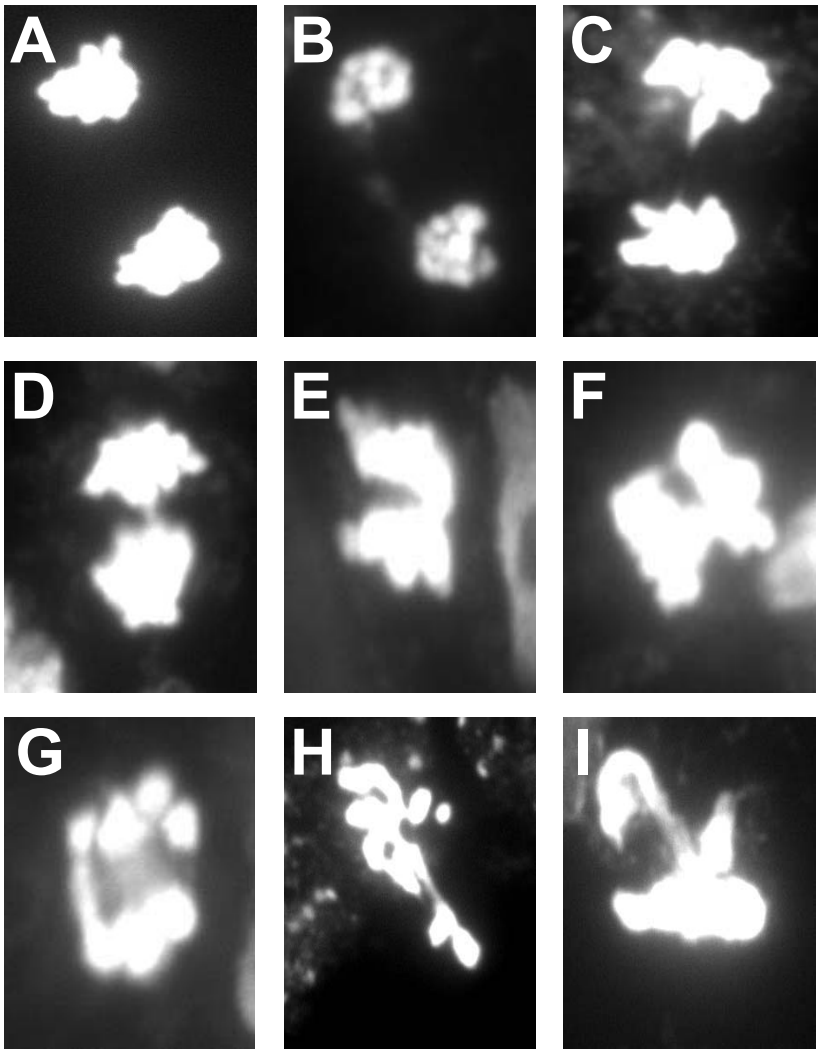

Figure 24. Cytogenetic defects in $P2^{DBD}$-$P1^{\Delta DBD}$ plants.

Anaphase bridges were observed in actively dividing mitotic tissues of pistils. Wild-type anaphase is shown in panel A. Some dividing nuclei contained a single anaphase bridge (panels B, C). Most chromosome fusions were evident in the beginning of the anaphase, before the chromosomes moved far away from each other (panels D-F). In many cases, anaphases contained an unusually high number of chromosome fusions, involving most chromosomes (panel G). In addition, many cells displayed a clear chromosome mis-segregation phenotype, resulting in obvious aneuploidy (panels H, I).

DISCUSSION

Arabidopsis harbors two *Pot1*-like genes

Telomeres are hypothesized to exist in three states: a "closed" conformation associated with the formation of a t-loop, that is inaccessible to telomerase, an "open" conformation, where the telomere is engaged with telomerase, and a Pot1-bound state, which may either be "open" or "closed" with respect to telomerase access. Although these conformations may exist at different stages of the cell cycle, they may not be mutually exclusive. The "closed" conformation is proposed to persist throughout most of the cell cycle, and is necessary to protect the telomeres from degradation and the action of DNA repair and recombination machinery (Griffith et al., 1999). This structure must be changed to an open conformation in the S-phase to allow for the replication of the DNA. At this point, proper telomere protection, as well as recruitment of telomerase, is especially important for the cell's ability to maintain the correct length of the telomere tract. Specific single-strand telomere DNA-binding proteins are thought to carry out many of these functions.

Until recently, single-strand telomere end-binding proteins were only known for ciliated protozoa and budding yeast. However, the availability of sequenced genomes has finally allowed identification of homologs in higher eukaryotes, which have been named Pot1. Pot1 genes were first identified in fission yeast and humans (Baumann and Cech, 2001) and later in many other diverse species, including macaque monkeys, mice, *Arabidopsis*, *N. crassa*, *E. cuniculi* and *A. nidulans* (Baumann et al., 2002; Pitt et al., 2004). Although the members of the Pot1 gene family are required for telomere length

regulation and protection (Baumann and Cech, 2001; Colgin et al., 2003; Loayza and De Lange, 2003), limited information is available on how these proteins work and whether they have additional functions.

To broaden our understanding of Pot1 function, we have identified and characterized Pot1 proteins in the small mustard weed *Arabidopsis thaliana*. In contrast to all other organisms studied (see below) *Arabidopsis* contains two *Pot1* genes, which we designated *Pot1* and *Pot2*. Both genes encode proteins with the conserved TELO_bind_N domain, previously shown to be involved in specific binding to single-stranded telomeric DNA. The overall amino acid sequence and domain structure of the Pot1 and Pot2 proteins is well conserved with their mammalian and fungal counterparts, indicating that at least some of the protein functions are likely to be similar.

Since both Pot1 and Pot2 proteins bind specifically to single-stranded G-rich telomeric DNA *in vitro*, strongly preferring plant telomeric sequences with very limited affinity for closely related sequences, we conclude that the DNA-binding properties of *Arabidopsis* Pot1 and Pot2 resemble those of the *S. pombe* Pot1 protein, and not the human Pot1 protein (Baumann and Cech, 2001). However, unlike the Pot1 protein from *S. pombe*, Arabidopsis Pot1 and Pot2 do not show a preference for a free 3'-end, suggesting that they have the potential to bind the G-strand overhang, as well as the G-rich telomere strand displaced by T-loop formation.

The purified DNA-binding domain of *S. pombe* Pot1 binds G-strand DNA in a highly cooperative manner so that only one shifted band is observed when the protein is incubated with an oligonucleotide that has multiple Pot1 binding sites (Lei et al., 2002). In contrast, binding of the full-length *Arabidopsis* Pot1 and Pot2 proteins to the G-strand

oligonucleotide with 6 *Arabidopsis* telomeric repeats gave rise to 2 and 3 specific complexes, respectively, suggesting that binding of plant proteins might be non-cooperative. In addition, although the *S. pombe* Pot1 DNA-binding domain binds tightly to a substrate of only 6 nucleotides (Lei et al., 2002), the binding sites of the full-length *Arabidopsis* Pot1 and Pot2 proteins appear to be longer, as the best binding substrates contain at least 5 telomeric repeats.

Pot1 and Pot2 can form homo- and heterodimers, raising the possibility of regulated interactions at telomeres. Both *Pot* genes are expressed in all analyzed plant tissues, consistent with a proposed role for them in telomere maintenance in all cells. Overall, the biochemical and genetic data presented in this chapter is consistent with the notion that Pot1-like proteins are key components of the telomere complex in *Arabidopsis*.

Arabidopsis as a unique model system for telomere biology

Arabidopsis is the only organism known to harbor two *Pot1*-like genes. This observation by itself is not completely surprising as approximately 35% of the *Arabidopsis* genome is thought to be duplicated (*Arabidopsis* Genome Initiative, 2000). In most cases, duplicated genes lie in well-characterized regions of chromosomes, known to have been subjected to major chromosome rearrangement events during evolution. This does not appear to be the case for *Pot1* and *Pot2* genes, which are located on regions of chromosomes that have not been duplicated in the past. Instead, the *Pot* genes appear to have evolved independently. Given the considerable differences among their

sequences, it is possible that they have distinct functions in telomere maintenance or regulation in *Arabidopsis*.

Arabidopsis Pot1 and Pot2 proteins localize to different subnuclear compartments, with Pot1 in the nucleolus and Pot2 excluded from the nucleolus, but still in the nuclear compartment. We suspect that the distinct localization of Pot proteins reflects different functions at telomeres. Unlike the situation in all other known organisms, including other plant species, *Arabidopsis* telomeres localize to the nucleolus throughout most of the cell cycle. This is an important difference, since telomere localization to the nuclear periphery in other eukaryotes is important for the establishment of proper nuclear architecture. Telomeres are also important for meiosis, where they cluster on the nuclear envelope (telomere bouquet) and move the diploid genetic material around so that homologous chromosomes can align and efficiently recombine. The fact that *Arabidopsis* telomeres localize to the nucleolus, and not to the nuclear periphery, may explain the apparent absence of the "telomere bouquet" in this organism.

We have not definitively showed that Pot proteins bind to chromosome ends *in vivo*, but based on their *in vitro* DNA binding preferences this seems likely. Therefore, we propose that Pot1 protein localizes to the telomeres in nucleolus and protects them from a variety of deleterious effects. Pot1 may be solely targeted to the nucleolus, or it may associate with telomeres during mitosis, when the nucleolus breaks down and then remains associated with telomeres when the nucleolus reforms. Pot2, on the other hand, is exclusively nuclear, and may associate with telomeres during M-phase. Alternatively, Pot2 may bind to the minor population of telomeres, which are not located in the nucleolus, throughout the cell cycle.

Possible functions of Pot proteins in *Arabidopsis*

Since no null mutants in *Pot1* or *Pot2* genes are currently available, we used an alternative approach to examine their function. We over-expressed full-length proteins; however, no phenotype was observed in such plants. We then turned to a dominant-negative approach, by over-expressing truncated forms of the Pot proteins. Since Pot2 does not localize to the nucleolus, we were surprised to see a number of major phenotypes in plants over-expressing the DNA-binding domain of Pot2. These phenotypes include severe sterility, morphological defects of leaves and moderate telomere shortening. These data argue that $P2^{DBD}$ does indeed function as a dominant-negative allele, and drives endogenous Pot2 molecules off the telomere by mass action, occupying their place at the terminus instead. Although the ability of $P2^{DBD}$ to bind telomeres has not yet been tested *in vitro*, the Pot1 DNA-binding domain is able to bind telomeres *in vitro* with even better efficiency than the full-length Pot1 protein, suggesting that this may be the case for $P2^{DBD}$ as well.

The $P2^{DBD}$ transgenic protein is missing approximately two thirds of the original sequence from its C-terminus, a portion predicted to be involved in protein-protein interactions (Colgin et al., 2003; Loayza and De Lange, 2003) and, therefore, is predicted to be incapable of interacting with its partners. The high incidence of anaphase bridges, as well as chromosome missegregation problems, appears to be a direct result of $P2^{DBD}$ over-expression, and may be responsible for the failure of the transgenic plants to maintain proper cell division, morphology and normal fertility. Interestingly, similar cytological phenotypes, such as chromosome missegregation and mitotic problems, are

also seen in the loss of function *pot1* mutants in *S. pombe* and *A. nidulans* cells (Baumann and Cech, 2001; Pitt et al., 2004). However, unlike the situation in an *S. pombe* deletion strain, over-expression of *Arabidopsis* $P2^{DBD}$ triggers only moderate telomere shortening, which can not explain the high incidence of end-to-end fusions in these cells. Our data suggest that that the endogenous Pot2 protein may be a part of the chromosome cap, an important structural component of the telomere complex whose main function is to prevent telomere-to-telomere fusions.

Unlike Pot2, *Arabidopsis* Pot1 localizes to the nucleolus. Interestingly, over expression of either full-length Pot1 protein or $P1^{DBD}$ has no effect on telomere length. On the other hand, over-expression of $P1^{\Delta DBD}$ causes moderate telomere shortening in the first generation of transgenic plants, which later stabilizes in the progeny. This truncated Pot1 polypeptide does not contain a DBD, therefore its effects on telomere length are not mediated by direct DNA binding. Presumably, $P1^{\Delta DBD}$ titrates away Pot1-interacting proteins from the telomeres, therefore disturbing the balance of telomere-lengthening and -shortening factors associated with the telomere itself. Alternatively, like the human $P1^{\Delta OB}$ (Loayza and De Lange, 2003), this truncated *Arabidopsis* polypeptide may still physically associate with telomeres via protein-protein interactions, influencing the telomere-counting mechanism proposed to regulate the length of the telomere tract. Future experiments will be necessary to determine whether all the Pot1 and Pot2 truncation derivatives localize to the same compartment as the endogenous proteins.

The most dramatic telomere shortening was seen in plants over-expressing chimeric $P2^{DBD}$-$P1^{\Delta DBD}$ constructs. This chimeric polypeptide corresponds to the Pot1 protein, harboring the DBD of Pot2. Remarkably, telomeres in the first generation of

these mutants were extremely short (less than 2 kb), approaching the size of telomeres seen in G_6-G_7 telomerase mutants. Consistent with the reduced telomere length in these mutants, about 5% of all anaphases contained bridges, the same frequency observed in G_6-G_7 TERT mutants (Riha et al., 2001). Thus, telomeres in chimeric mutants are becoming uncapped at a faster rate than in TERT mutants. Moreover, like many G_6-G_7 TERT plants, these mutants appear morphologically wild-type and are fertile.

One possibility is that the mutant chimeric proteins bind telomeres outside the nucleolus, thus delivering Pot1-associated functions to an inappropriate compartment. Another possibility is that without its own DBD, Pot1 can no longer properly regulate telomere length. Either way, our data demonstrate that *Arabidopsis* Pot1 and Pot2 proteins are involved in the regulation of telomere length, and Pot2 may also provide the telomere capping function. Future analysis of *Pot1* and *Pot2* knock-out lines can provide more useful insights into the function of these proteins *in vivo*.

Finally, it is curious that over-expression of the full-length human Pot1 or its C-terminal protein interaction domain causes significant telomere elongation (Colgin et al., 2003; Loayza and De Lange, 2003), while *Arabidopsis* mutants display the opposite phenotype, with telomeres in some cases losing over 50% of their overall wild-type tract in just one generation. This discrepancy may suggest that *Arabidopsis* and human Pot1 proteins may interact differently with telomere maintenance machinery. This difference in protein function is not unusual for factors involved in telomere biology. For example, Ku proteins are known to play different roles at yeast and plant telomeres. In yeast, Ku deficiency leads to telomere shortening, whereas in *Arabidopsis* such deficiency causes

significant telomere elongation (Boulton and Jackson, 1998; Gravel et al., 1998; Polotnianka et al., 1998; Riha et al., 2002).

The fact that *Arabidopsis* seems to "hide" telomeres in the nucleolus, where they can be exposed to a different environment than that on the nuclear periphery, would be consistent with different functions of Pot proteins. Alternatively, this apparent difference may simply be due to the fact that *Arabidopsis* possesses two Pot1-like proteins, which have separable and even perhaps opposite functions at telomeres. Either way, this finding once again confirms that different model organisms need to be studied in order to fully understand the role of telomeres and telomere binding proteins in a living cell.

The unexpected results of $P2^{DBD}$, $P1^{\Delta DBD}$ and $P2^{DBD}$-$P1^{\Delta DBD}$ over-production clearly demonstrate that *Arabidopsis* Pot proteins are involved in telomere biology *in vivo*, most likely performing several different functions at plant telomeres. Still, the mechanisms behind all of the observed phenotypes, and especially those leading to such a great telomere loss in $P2^{DBD}$-$P1^{\Delta DBD}$ plants are unknown. Future work will shed light on how over-production of this chimeric protein compromises telomere length regulation in *Arabidopsis*.

MATERIALS AND METHODS

Plant materials, transformation and growth conditions

Wild-type *Arabidopsis* seeds were purchased from the *Arabidopsis* Biological Resource Center (Ohio State University, Columbus), cold-treated overnight at 4°C, then placed in the environmental growth chamber and grown under a 16/8-hr light/dark

photoperiod at 23°C. To obtain Pot1 and Pot2 over-expressing mutants, *Pot1* and *Pot2* cDNAs were amplified by PCR with primers complementary to the start and stop codons of each gene, and then inserted into a binary vector pCBK05 (Riha et al., 2002) to allow expression from a 35S CaMV promoter. The constructs were then introduced into *Agrobacterium tumefaciens* strain GV3101. Transformation of the wild-type plants was performed by the *in planta* method (Riha et al., 2002). T1 primary transformants were selected on 0.5 BM medium supplemented with 20 mg/l of phosphinothricine (Crescent Chemical), genotyped and analyzed by RT-PCR for transgene expression.

cDNA synthesis

Total mRNA was extracted from 0.1–0.5 g of plant tissue using Tri Reagent solution (Sigma). *Pot1* and *Pot2* cDNAs were synthesized from total leaf RNA using Superscript II reverse transcriptase (Gibco). Primers complementary to the stop codons of each cDNA were incubated with 2 μg of total RNA in the supplied buffer at 65°C for 5 min. Reverse transcription was carried with 100U of Superscript II at 42°C for 50 min. RNA was degraded with RNase H (USB). The coding regions of *Pot1* and *Pot2* were then amplified with Ex-Taq polymerase (Takara). PCR products were cloned into a pCR2.1-TOPO vector (Invitrogen) and sequenced.

Expression and purification of recombinant proteins in *E.coli*

Pot1, *Pot2* and $P1^{DBD}$ cDNAs were cloned in-frame in the pET28a vector (Novagen) using *BamH*I and *Xho*I sites, and were then introduced into the ER2566 *E.coli* strain (New England Biolabs). Cells were grown in 0.5 L culture flasks until they reached

a density of 0.4 OD units. Protein expression was induced for 8-16 hr with 0.1 mM IPTG at 20°C. Cell pellets were collected following centrifugation and resuspended in lysis buffer (50mM NaH_2PO_4, adjusted to pH 8.0 with NaOH, 500 mM NaCl, 10 mM imidazole, 1% NP-40). Cells were then lysed by sonication. For protein purification, Ni-NTA resin (Novagen) was used, following manufacturer's protocol with minor adjustments. Briefly, 15 ml cell lysate was incubated with 2 ml of resin for 1-1.5 h at 4°C on a rotating wheel, and the sample was poured into a disposable 15-ml column (PIERCE). Unbound sample was collected, and the resin washed 4 times with 4 ml wash buffer (same as lysis buffer plus 40 mM imidazole and no NP-40). The bound protein was eluted 4-8 times with 0.5 ml elution buffer (same as lysis buffer plus 250 mM imidazole and no NP-40) each. The eluted protein was then analyzed by SDS-PAGE. Elution was also monitored by Western blotting with T7-tag antibodies, following manufacturer's recommendations. After purification, the proteins were dialyzed against 50 mM TrisHCl pH 8.0, 0.5 mM EDTA, 0.5 mM DTT, 0.2 M KCl, 10% glycerol.

Electrophoretic mobility shift assays

All radioactive oligonucleotides used in the gel-shift assays were labeled with T4 polynucleotide kinase (New England Biolabs) and γ-^{32}P-ATP (3000 Ci/mmol; Amersham Pharmacia), and purified through nucleotide removal disposable columns (Qiagen). The binding reactions were performed in 50 μl in the following buffer: 25 mM Hepes pH 7.5, 1 mM EDTA, 100 mM NaCl, 5% glycerol, 0.5 μg *Hae*III-cut *E.coli* DNA and 0.25 nM probe. 5μg of the protein was added last, and the binding reaction was incubated for 30

min at room temperature. For competition experiments, protein was added to the reactions containing up to 150-fold molar excess of unlabeled competitor oligonucleotide. Electrophoresis was performed in 5% polyacrylamide gels run in 1X TBE for 4 hours at 150V at RT. The gels were dried on Whatman DE81 paper at 80°C, exposed to phosphorimager screens and analyzed using ImageQuant software.

***In vitro* dimerization assays**

Full-length *Pot1* and *Pot2* cDNA were inserted into the *BamH*I–*Xho*I restriction sites in the expression vector pET28a (Novagen) to produce fusion proteins with an N-terminal His-T7 tag. The cDNAs were also cloned into pCITE4a vector (Novagen) using the same restriction sites to produce untagged proteins. All obtained constructs and the control vector encoding untagged luciferase were transcribed and translated separately in a TnT-coupled rabbit reticulocyte lysate (Promega) with or without [^{35}S]L-methionine (Amersham). Translation was stopped with cyclohexamide (4 ng/µl) before ^{35}S-labeled proteins were mixed with T7-tagged unlabeled proteins in a ratio of 3:1 and incubated at 30°C for 15 min. Proteins were immunoprecipitated with T7 antibodies as described (Bryan et al., 2000). Precipitate and supernatant fractions were analyzed by SDS–PAGE and autoradiography.

Yeast two-hybrid experiments

Pot1 and *Pot2* cDNA were cloned into pAS2-1 and pACT2 vectors (CLONTECH) to obtain fusion constructs for both the GAL4 DNA-binding domain (GDBD) and the GAL4 Activation domain (GAD), respectively. GDBD-Pot1 and

GDBD-POT2 were then transformed into yeast strain Y190 (CLONTECH) using the lithium acetate technique, and plated on –Trp medium to recover transformants. The activation domain constructs were similarly introduced into these strains and plated onto –Leu/-Trp medium to recover double transformants that contained both the GDBD and the GAD constructs. These double transformants were then replica-plated onto medium lacking Leucine, Tryptophan and Histidine to test for interaction between the GDBD- and GAD-fusion proteins, as the growth on –His plates is indicative of such interaction. The strains were also checked for the presence of another reporter activity, β-galactosidase expression, which was used to confirm the interactions.

Telomere analysis

DNA from individual plants was extracted as described (Cocciolone and Cone, 1993). TRF analysis was performed with *Tru*1I (Fermentas) restriction enzyme and ^{32}P 5' end-labeled $(T_3AG_3)_4$ oligonucleotide as a probe (Fitzgerald et al., 1999). Anaphase spreads were prepared from pistils as described previously (Riha et al., 2001).

CHAPTER IV

IDENTIFICATION OF SINGLE-STRAND TELOMERE BINDING PROTEINS IN CAULIFLOWER

INTRODUCTION

Telomeres are specialized nucleoprotein structures at the ends of linear eukaryotic chromosomes that provide capping functions and allow for the complete replication of chromosome ends. Most telomeres consist of simple G-rich, 6-8 nucleotide tandem repeats, ranging from several tens of bases to several tens of thousands nucleotides in size. For example, in most genetic backgrounds, *Arabidopsis* telomeres are composed of 2-5 kb tandem arrays of TTTAGGG repeats. The G-rich strand of the telomere is synthesized by telomerase, which utilizes its own intrinsic RNA template for telomeric DNA synthesis. Telomeric DNA is associated with specific telomere binding factors, which play critical roles in telomere maintenance. Telomere binding proteins are also involved in regulation of telomerase activity.

Structurally, telomeres have two distinct regions – double-stranded telomeric DNA accounting for most of the telomeric tract, and a short G-rich single-strand protrusion on the telomere terminus called the G-overhang. Two different classes of telomere binding proteins are able to interact with these regions. Both classes of proteins exhibit strong sequence specificity, tightly binding to telomeric DNA sequence of that organism, but not to others. Proteins that bind to double-stranded regions of telomeres have been found in yeast (Rap1p, Taz1p), mammals (TRF1, TRF2) and other eukaryotes

(reviewed in Kanoh and Ishikawa, 2003). Some of these proteins provide a protective capping function for the telomeres (van Steensel and de Lange, 1997; van Steensel et al., 1998). Other telomere binding proteins negatively regulate telomerase by a protein-counting mechanism in which the number of the bound protein molecules is precisely measured, permitting or restricting telomerase access to the telomeres (Marcand et al., 1997). Another role for telomere proteins is exemplified by the mammalian TRF2 protein which directly participates in telomere looping, a process in which the distal telomeric repeats are folded back onto the rest of the telomere tract to form the t-loop (Griffith et al., 1999).

Thus far, no plant proteins have been shown to bind telomeric DNA *in vivo*. However, several groups have reported identification of such proteins that bind telomeric DNA *in vitro*. Most of the genes encoding putative double-strand telomere binding proteins from plants were isolated based on their sequence homology to known human and yeast telomere proteins. Such proteins display a single Myb-like DNA binding motif at their C-terminus, similar to that present in human TRF1 and TRF2 proteins (Yang et al., 2003; Yu et al., 2000, Hwang et al., 2001; Chen et al., 2001). Additionally, a novel plant-specific SMH1 protein from maize was also isolated based on its ability to bind double-stranded telomeric DNA *in vitro* (Marian et al., 2003). In *Arabidopsis*, TRP1 and TBP1 proteins specifically bind double-stranded telomeric DNA *in vitro* (Chen et al., 2001; Hwang et al., 2001). In addition, the nuclear protein ATBP1, capable of binding both double- and single-stranded telomeric DNA (Zentgraf, 1995), was also identified. This 67 kDa ATBP1 protein identified by Zentgraf was later found to associate via protein-protein interactions with another, smaller protein ATBP2 during the onset of

senescence. This higher-order protein complex was only able to bind double-stranded telomeric DNA (Zentgraf et al., 2000). Taken together, these data suggest that the *Arabidopsis* genome may encode several telomere binding proteins.

The second group of telomere binding proteins consists of factors that bind single-stranded telomeric G-overhangs. Some of these proteins, particularly those from ciliated protozoa, have been extensively studied biochemically. In *Oxytricha nova*, a heterodimeric protein containing α- and β- subunits tightly binds to the overhang (Gottschling and Zakian, 1986; Price and Cech, 1987), forming a ternary complex with a 1:1:1 stoichiometry (Fang and Cech, 1993). Both subunits are required for maximal binding, but the α-subunit can also bind the overhang alone (Gray et al., 1991). In *Euplotes*, a telomere binding protein with homology to the α-subunit appears to have a similar role at the telomeres (Price et al., 1992). In budding yeast, the major single-strand telomere binding protein is Cdc13p (Nugent et al., 1996). Cdc13p is a multifunctional protein: it protects telomeres from degradation, positively and negatively regulates telomerase recruitment to the telomeres, and provides a link between leading and lagging strand replication machineries (reviewed in Evans and Lundblad, 2000). In fission yeast and mammals, the functional homologue of Cdc13p is the Pot1 protein, which provides telomere protection against nucleases and serves as a direct link between telomere-length counting machinery and telomerase (Baumann and Cech, 2001; Loayza and De Lange, 2003).

Several published reports indicate that single-strand telomere binding proteins exist in plants. Nuclear extracts from rice and mung bean, and from *Chlamydomonas reinhardtii*, contain activities capable of specifically interacting with single-stranded

telomeric oligonucleotides in gel-shift assays (Petracek et al., 1994; Kim et al., 1998; Lee et al., 2000). Recently, an RNA-binding protein STEP1 was purified from *Arabidopsis* nuclear extracts and shown to bind to telomeric DNA *in vitro* (Kwon and Chung, 2003).

Arabidopsis thaliana has proven to be an excellent model organism to study telomere biology in higher eukaryotes (Riha and Shippen, 2003b). It seems clear that plants share most components of telomere maintenance machinery with mammals and other eukaryotes. Furthermore, *Arabidopsis* has a completely sequenced genome and well-established genetic tools, providing an excellent opportunity to study the role of telomere proteins *in vivo*. However, the limitation of using *Arabidopsis* is difficulty in getting large quantities of biochemical material. In this work, we decided to take advantage of the close evolutionary relationship between *Arabidopsis thaliana* and cauliflower (*Brassica oleracia*), which are members of the same plant family. Unlike *Arabidopsis*, cauliflower provides an abundant source of plant tissue, from which large, preparative quantities of nuclear extract can easily be made. The genomes of both plants are thought to have substantial similarity and identity on both the nucleotide and protein levels. Therefore, our rationale was to purify telomere binding activities from cauliflower, isolate and sequence the proteins by mass-spectrometry, and use the obtained peptide data to search the *Arabidopsis* database for proteins with the same peptide composition as their putative cauliflower homologues.

We identified four major telomere binding protein complexes from cauliflower and characterized their DNA-binding properties. Here, we show that these complexes have distinct molecular weights and binding affinities towards various telomeric substrates. The DNA-binding component of one of the complexes was purified and

analyzed by mass-spectrometry. Peptide mass data was used to search for putative protein candidates from the *Arabidopsis thaliana* database.

RESULTS

Identification of single-strand telomere binding activities in cauliflower nuclear extracts

To learn more about the components of the telomere complex in plants, we looked for proteins in cauliflower that can bind oligonucleotides corresponding to the G-rich telomere overhang. To determine whether cauliflower nuclear extract contains factors capable of specific binding to single-stranded telomere substrates, gel retardation assays were performed with a ^{32}P-labeled telomeric $(T_3AG_3)_5$ oligonucleotide (see Table 2 for the complete list of all oligonucleotides used in this study and their sequences). As shown in Figure 25, the cauliflower nuclear extract contained single-strand binding activities, which were able to shift the telomeric DNA probe. Four major complexes were detected (Figure 25, lane 2), designated A, B, C and D from top to bottom. A less abundant and less stable complex, designated by the asterisk, was also detected in some extract preparations. We verified that a protein was responsible for this DNA binding activity by adding protease K to the binding reaction and showing that complex formation was completely abolished. In contrast, addition of RNase A had no effect on binding (data not shown). No complex formation was observed with a C-rich telomeric oligonucleotide $(C_3TA_3)_5$ or duplex telomere repeats (data not shown).

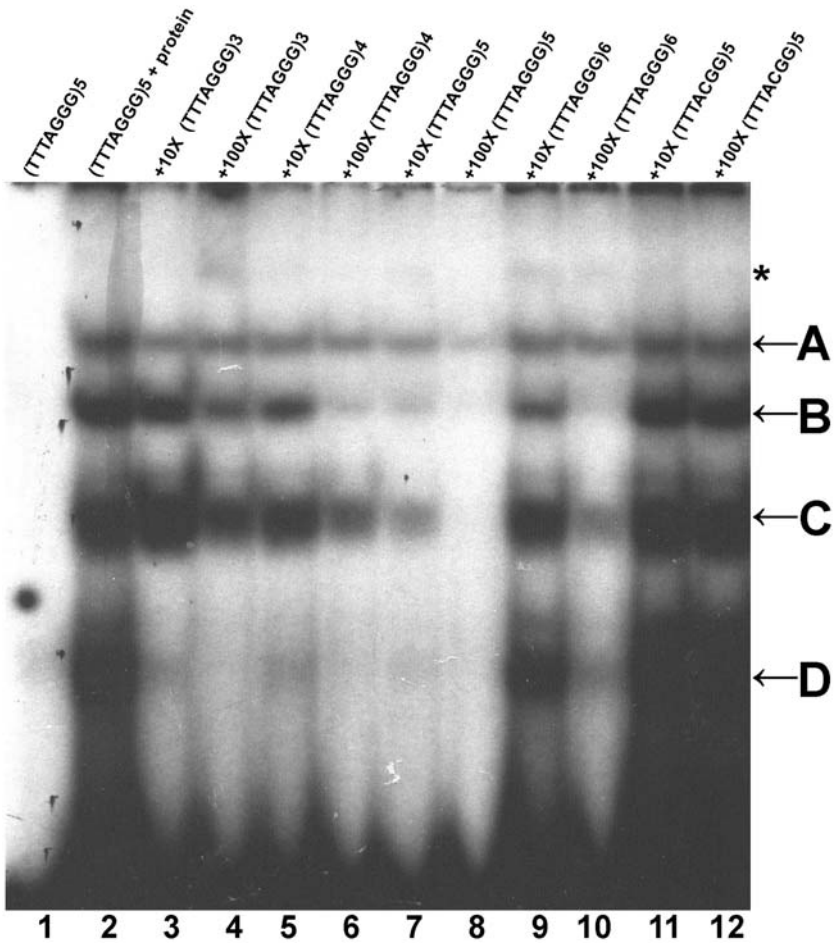

Figure 25. Identification of four specific telomeric DNA-binding complexes in cauliflower nuclear extracts.

Electrophoretic mobility shift assay. Lane 1, $(TTTTAGGG)_5$ oligonucleotide alone, lane 2, $(TTTTAGGG)_5$ plus nuclear extract. lanes 3-12, addition of 10X (lanes 3,5,7,9,11) or 100X (lanes 4,6,8,10,12) excess cold competitor oligonucleotides with various number of telomeric repeats (lanes 3-10) or with a point mutation in the repeat sequence (lanes 11-12). The four major complexes are designated A to D. The asterisk designates a minor band of lower mobility.

The binding of the cauliflower proteins to telomeric DNA was highly specific. Addition of up to 100-fold molar excess of cold $(T_3ACG_2)_5$ oligonucleotide did not abolish binding (Figure 25, lanes 11-12). Since the nucleotide sequence of this competitor is only one nucleotide different from the *Arabidopsis* telomeric sequence T_3AG_3, the data argue that the binding activity we have identified is specific for telomeric DNA and would not likely bind the degenerate telomere repeats found in the subtelomeric and intergenic regions of plant chromosomes.

To further examine the specificity of DNA binding, we performed competition experiments with oligonucleotides corresponding to human $(TTAGGG)_5$ and *Oxytricha nova* $(T_4G_4)_3$ sequences. Complexes C and D were abolished with the human telomere repeat competitor, while complexes A and B showed only minor reduction in binding (Figure 26, lanes 5-6). No competition was observed with the ciliate telomeric DNA (Figure 26, lanes 7-8). This result demonstrates that the four observed complexes have different DNA-binding properties, with the two upper bands being most specific for plant DNA. The results presented in Figure 26 also indicate that all the cauliflower proteins involved in telomeric DNA binding prefer closely related telomeric substrates and do not bind sequences from evolutionary distant species.

To determine the minimal number of telomeric repeats necessary for complex formation, we performed competition assays with oligonucleotides containing three, four, five and six telomeric repeats. Cauliflower extract was incubated with radiolabeled $(T_3AG_3)_5$ oligonucleotide and up to 100-fold molar excess of cold competitor. Oligonucleotides with more telomeric repeats showed progressively better competition. Three telomeric repeats showed minimal competition for the A, B and C complexes

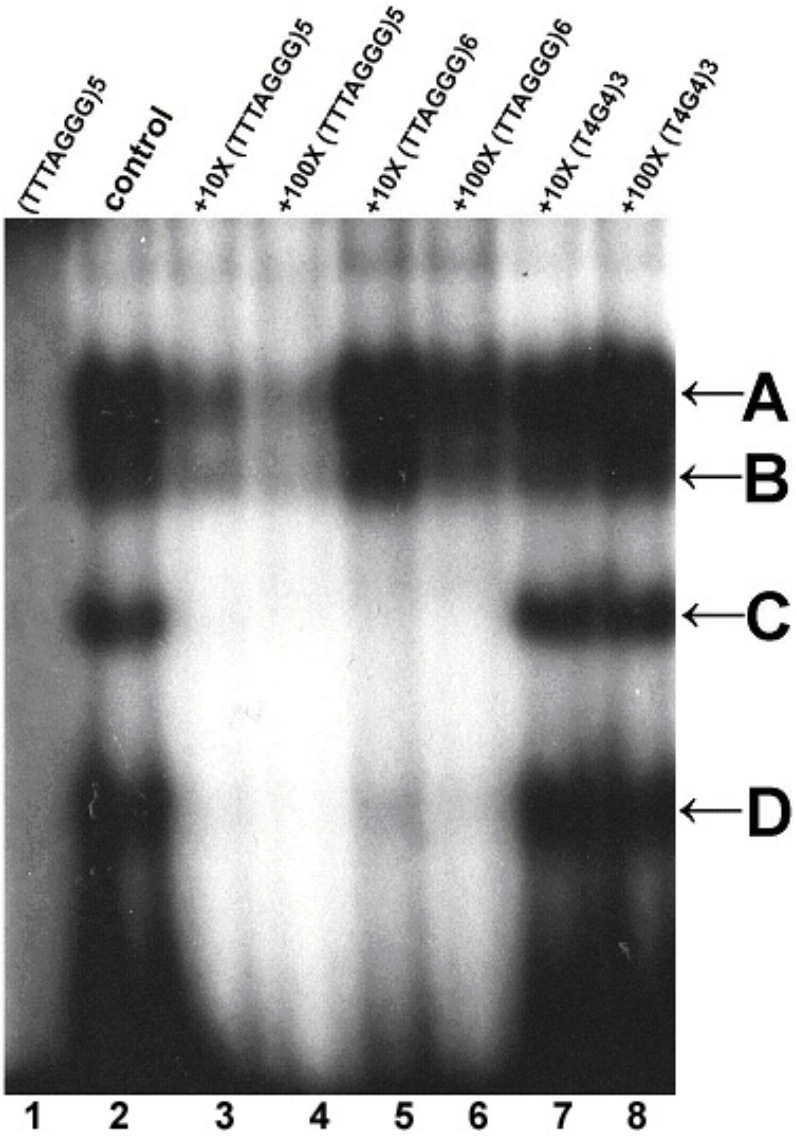

Figure 26. Binding preference of cauliflower proteins to plant telomeric sequences.

Lane 1, $(TTTTAGGG)_5$ oligonucleotide alone, lane 2, $(TTTTAGGG)_5$ plus nuclear extract, lanes 3-8, addition of 10X (lanes 3, 5, 7) or 100X (lanes 4, 6, 8) excess cold competitor DNA oligonucleotides with plant (lanes 3-4), human (lanes 5-6) and ciliate (lanes 7-8) telomere repeats.

(Figure 25, lanes 3-4), while the five telomeric repeats showed the best competition (Figure 25, lanes 5-8). Complex D was competed away with all of the cold telomeric oligonucleotides, suggesting that it requires the least number of telomeric repeats for binding. Notably, an oligonucleotide with six telomeric repeats was not as efficient a competitor as was an oligonucleotide with five telomeric repeats (Figure 25, lanes 9-10). One possible explanation is that the $(T_3AG_3)_6$ oligonucleotide may undergo conformational changes in solution to form a secondary structure that prevents efficient protein binding. Since telomeric oligonucleotides are known to form "G-quartet" structures *in vitro* (Sundquist and Klug, 1989), it is conceivable that $(T_3AG_3)_6$ oligonucleotide assumes a similar conformation. Overall, the results presented in Figures 25 and 26 indicate that cauliflower nuclear extracts contain telomere binding protein activities that can bind single-stranded G-rich substrates.

Several telomere binding proteins prefer to bind to a 3' terminus (Sheng et al., 1995; Baumann and Cech, 2001). To test if this is also the case for the cauliflower proteins, we performed competition experiments with oligonucleotides containing six telomeric repeats located in the middle or at either the 5' or 3' end of the DNA. Both TELO-UP (5' position) and TELO-MID competed much better for binding than the TELO (3' position) (Figure 27, lanes 3-8), suggesting that telomere binding proteins in cauliflower do not have preference for the free 3'-overhang. This result is similar to that observed for rice single-strand telomere binding proteins (Kim et al., 1998).

In *Arabidopsis*, RNA-binding proteins are known to interact with plant single-stranded telomeric DNA (Kwon and Chung, 2003). In other species, proteins with RNA recognition motifs display both telomeric DNA and RNA binding activities (Lin and

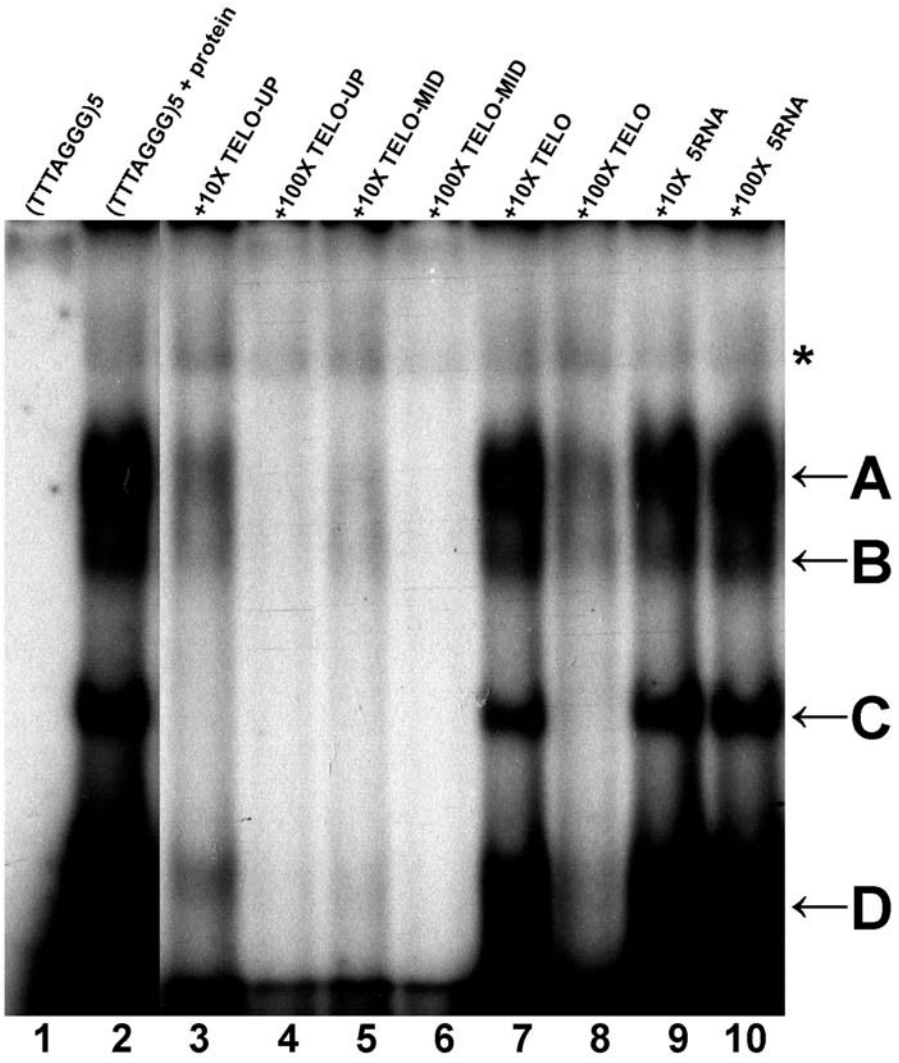

Figure 27. Effect of telomeric DNA position on oligonucleotide binding.

Lane 1, $(TTTTAGGG)_5$ oligonucleotide alone, lane 2, $(TTTTAGGG)_5$ plus nuclear extract. Lanes 3-10, addition of 10X (lanes 3, 5, 7, 9) or 100X (lanes 4, 6, 8, 10) excess cold competitor DNA oligonucleotides with differentially positioned telomeric repeats (lanes 3-8) or RNA oligonucleotide of telomeric sequence (lanes 9-10).

Zakian, 1994). To test whether the telomeric binding activities from cauliflower have the ability to bind the plant telomere cognate RNA $(U_3AG_3)_5$ sequence, competiton experiments were performed. No competition was observed when up to 100-fold cold $(U_3AG_3)_5$ oligonucleotide was added to the reaction (Figure 27, lanes 9-10). Thus, the cauliflower proteins bind specifically to DNA, but not RNA substrates.

Preliminary titration experiments demonstrated that intensities of each shifted band increased as protein concentration increased, suggesting that each complex is composed of unique and stable protein components (data not shown). The four major protein-DNA complexes observed in our experiments may represent several different telomere binding proteins, each of which formed a separate shifted band on the gel. Alternatively, one single DNA-binding protein may be responsible for shifting all bands in the assay. In this scenario, higher order complexes may form with multimers of a single protein by incorporation of additional factors. To address this possibility, we subjected cauliflower nuclear extracts to size fractionation on a Superose 12 column. Each fraction was analyzed for the telomere binding activity. As shown in Figure 28, the peak of fractions containing A and D complexes eluted in fractions 32 and 33 (50-80 kDa). Fraction 30 (150-160 kDa) contained proteins necessary for complex C formation. Interestingly, complex B had two peaks – in fractions 30 and 33 (150-160 kDa and 50-70 kDa, respectively). An additional high molecular weight complex, indicated by the asterisk, was formed in fractions 28-29 (170-200 kDa). These results indicate that the protein complexes involved in telomeric DNA binding have distinct molecular weight, and may have different protein compositions. Alternatively, the DNA-binding component

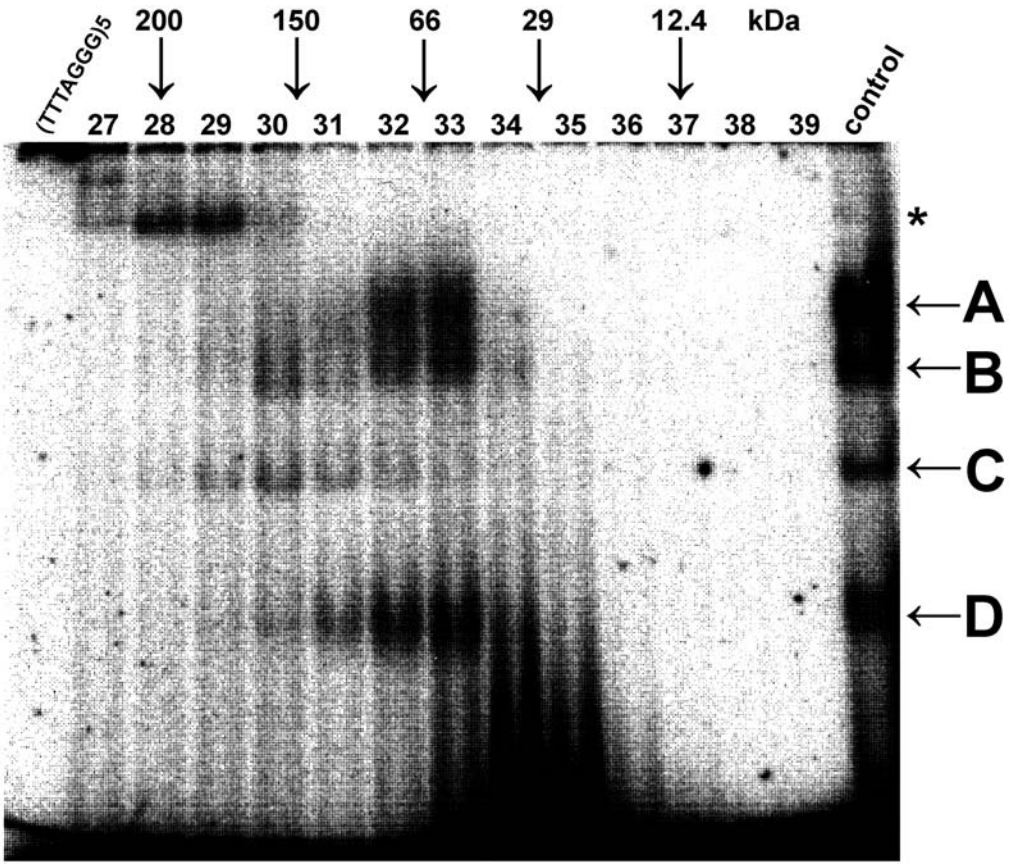

Figure 28. Superose 12 fractionation of cauliflower nuclear proteins.

Gel-shift assays were performed with fractions 27-39. The position of molecular weight protein markers are shown on the top of the gel. Unfractionated nuclear extract was used as a control in the reaction.

of the complexes might miltimerize either before or after association with the DNA, resulting in increased molecular weight of the shifted complexes.

Purification of cauliflower telomere binding proteins

To purify telomere binding proteins from cauliflower we designed a purification scheme consisting of four major steps: ammonium sulfate precipitation, preparative isoelectric focusing, size exclusion chromatography and DNA-affinity purification. As shown in Figure 29A, most of the telomere binding activity from cauliflower nuclear extract precipitated at 60% ammonium sulfate saturation. The protein sample from this fraction was dialyzed and fractionated by preparative isoelectric focising in the Rotofor system (Bio-Rad). As shown in Figure 29B, most DNA-binding complexes appeared in fractions 14 and 15 (pH 4.4 - 4.5), and the second peak of complex B also appeared in fractions 17 and 18 (pH 4.7 – 4.8). The intensities of complexes C and D were significantly reduced, perhaps due to dissociation of the protein components of these complexes.

To further purify complexes A and B, samples from fractions 14 and 15 were combined and subjected to size-exclusion chromatography on a Superose 12 column. The peak of complex A eluted in fractions 25-26 (40-70 kDa) and the peak of complex B eluted in fractions 22-23 (120-170 kDa, Figure 30A). To estimate the apparent molecular weight of the DNA-binding component of complex A, fractions 20-28 from Superose 12 column were used in a cross-linking assay. Proteins were allowed to interact with ^{32}P-labeled $(T_3AG_3)_5$, and then were irradiated with UV light to form covalent protein-DNA

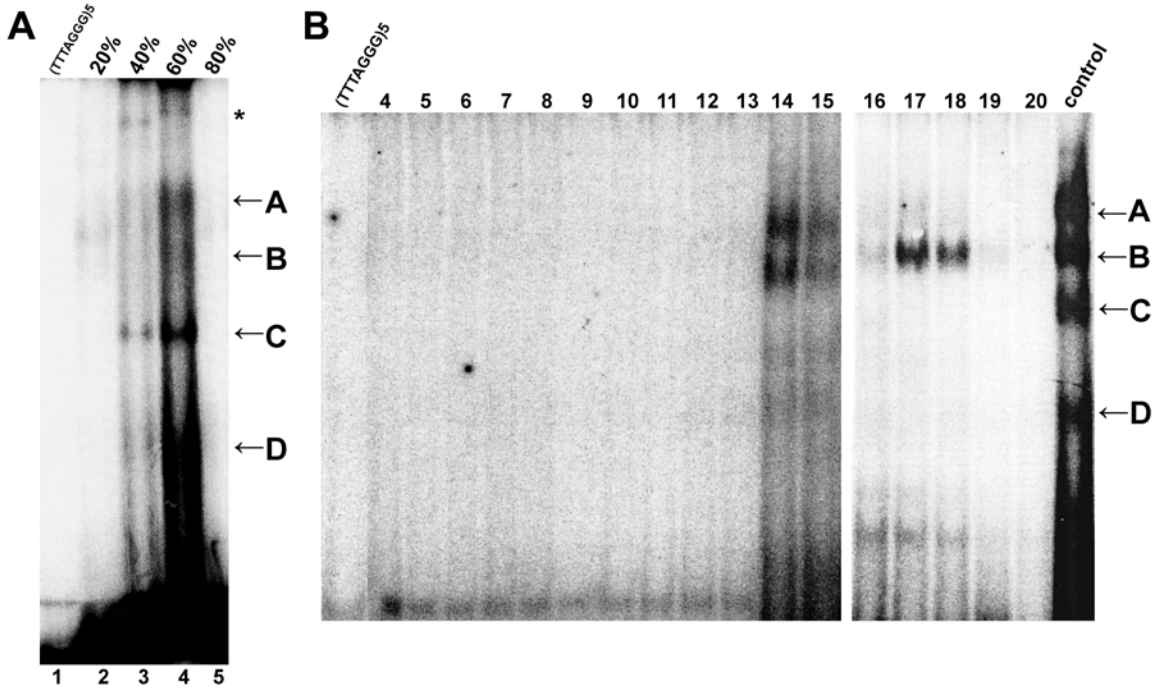

Figure 29. Purification of telomere binding proteins from cauliflower nuclear extract.

(A) Gel-shift assay following ammonium sulfate precipitation of cauliflower proteins. Most telomere binding activity precipitates at 60% saturation. (B) Fractionation in a Rotofor column. Proteins in fractions 4-20 separated by their isoelectric point on a Rotofor column were subjected to a gel-shift assay. Complexes A and B eluted in fractions 14 and 15 (pI 4.4 and 4.5). The second peak of complex B was observed in fractions 17 and 18 (pI 4.7-4.8). Unfractionated nuclear extract was used as control.

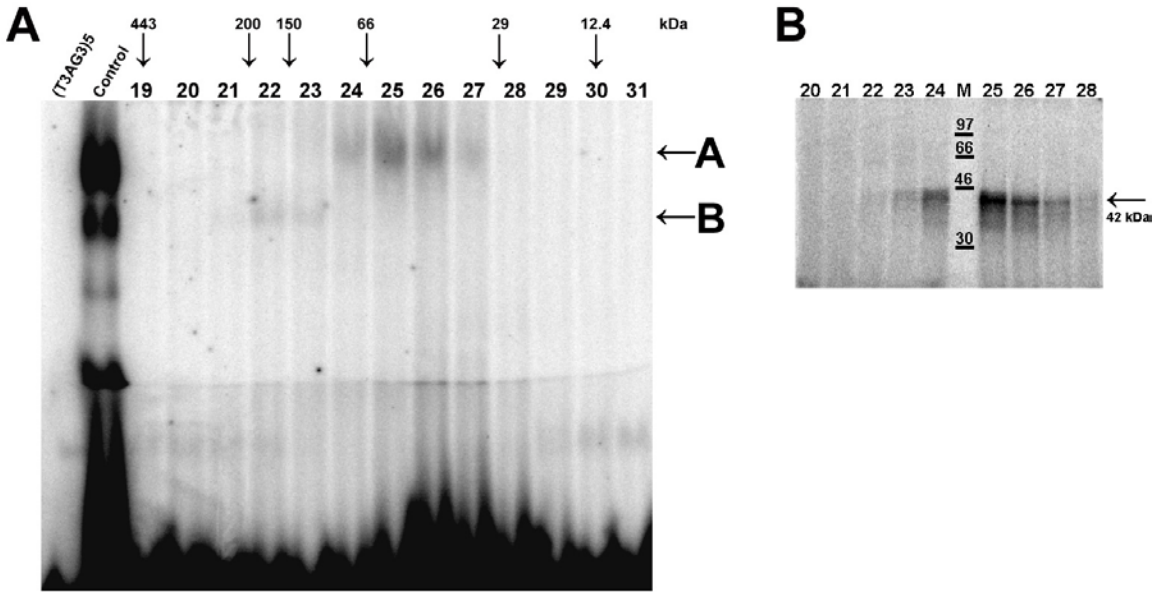

Figure 30. The cauliflower telomere binding protein in complex A has a molecular weight of less than 40 kDa.

(A) Gel-shift assay following Superose 12 fractionation of partially purified proteins in complexes A and B. Molecular weight protein markers are shown on the top of the gel. Unfractionated nuclear extract was used as a control in the reaction. Complex A elutes in fractions 25 and 26 (40-70 kDa). (B) SDS-PAGE of proteins in Superose 12 fractions 20-28 crosslinked to the radioactively labeled telomeric DNA. Only one band of 42 kDa is seen across the fractions.

bonds. The cross-linked fractions were subjected to SDS-PAGE. Only one major cross-linked band was visible on the gel across the fractions, with most of the DNA-binding activity was present in fraction 25 (Figure 30B). This result is consistent with the results for size-exclusion chromatography. The estimated size of the band is around 42 kDa, which includes both the protein and the DNA oligonucleotide cross-linked to it. Based on these results, we conclude that the size of the putative telomere binding protein is less than 30 kDa.

To further purify telomere binding proteins, fractions 25 and 26 were pooled and subjected to DNA-affinity purification. Figure 31A shows that complex A was successfully purified and eluted off the column in the 0.5 M NaCl fraction. Complex B was apparently lost during the purification. To visualize proteins eluted with 0.5 M NaCl, we performed SDS-PAGE and stained the fractions with silver. Although the sample applied to the affinity column contained multiple polypeptides (Figure 31B, lanes 2 and 4), only a few of them were bound to the DNA. Moreover, the 0.5 M NaCl elution sample contained at least three easily detectable bands in the size range below 30 kDa, which are not present in the 2M NaCl elution sample (Figure 31B, compare lanes 1 and 3). Since the 2M NaCl elution sample contains no detectable telomere binding activity (Figure 31A), these three bands in the 0.5 M NaCl fraction represent good candidates for mass-spectrometry analysis. Therefore, all three bands (25, 18 and 16 kDa, designated ES-1, ES-2 and ES-3, respectively) were excised from the gel, digested with trypsin and analyzed by MALDI-TOF. For a negative control, a portion of the gel without any protein bands was also excised and used as a blank.

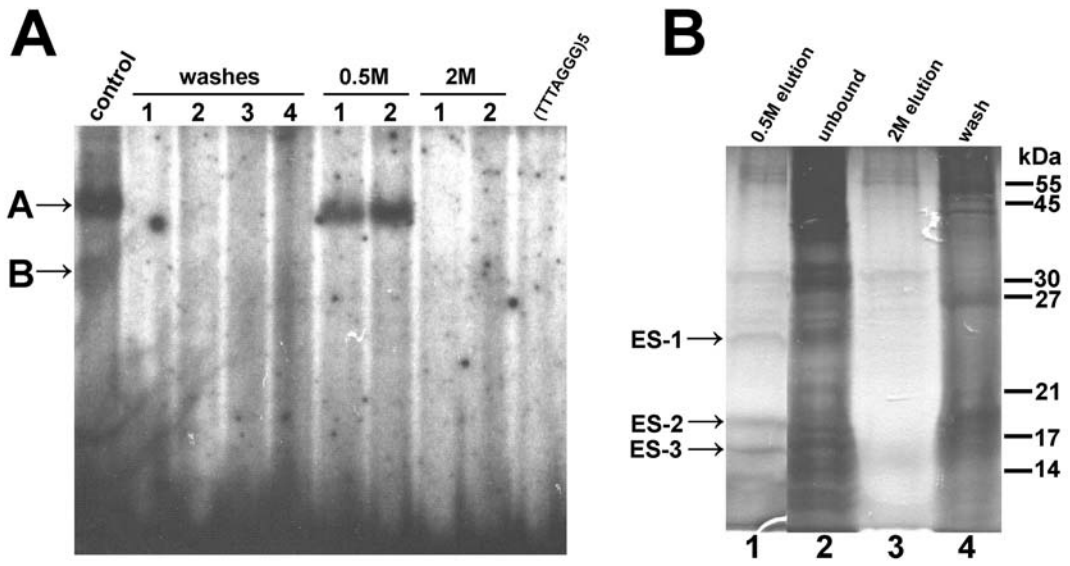

Figure 31. DNA-affinity purification and silver-staining of cauliflower telomere binding proteins.

(A) Gel-shift assay after the final purification step of cauliflower telomere binding proteins. The affinity-purified telomere binding activity elutes in 0.5 M NaCl elution fractions, but not in 2 M NaCl elution or wash fractions. Proteins from the second purification step (Rotofor column) were used as a control in the reaction. (B) SDS-PAGE and silver-staining of affinity-purified proteins. Three protein bands below 30 kDa are seen in the 0.5 M NaCl elution fraction (lane 1), but not in the 2M elution fraction (lane 3).

The initial mass-spec analysis identified a number of peptide masses for ES-1 and ES-3 samples. Sample ES-2 was similar in peptide composition to the blank (data not shown) and was not used further in the study. We used the MS-Fit program from the ProteinProspector package (http://prospector.ucsf.edu) to search the *Arabidopsis* database for proteins whose hypothetical trypsin digestion could produce such a range of peptide masses. For the ES-3 protein sample, the best putative candidate was the At2g14470 gene product, which matched 5 out of 18 peptide masses. The At2g14470 gene encodes a 144 kDa protein with putative DNA helicase domain and very low similarity to the DNA repair and recombination protein Pif1 from *S. cerevisiae*. Other candidates included heat shock transcription factor HSF4 (At4g36990) and ABC transporter protein (A t2g13610). None of these proteins are likely to be telomere binding proteins *in vivo*.

For ES-1, two putative proteins (GenBank accessions # 18395275 and 11346161) matched 3 out of 13 available peptide masses. These proteins are encoded by genes At1g23710 and At5g16810, respectively. The At1g23710 gene product is an expressed protein with predicted molecular weight of 33.6 kDa and isoelectric point 4.9 with unknown function and no obvious protein motifs. The At5g16810 gene product is a hypothetical protein of 38.6 kDa and pI 8.3, also with no obvious protein domain or known function. If shown to bind telomeric DNA *in vivo* or *in vitro*, these proteins will represent a novel class of telomere binding proteins.

DISCUSSION

The telomere is a conserved structure with multiple roles in cell biology. In most eukaryotes, the telomeric sequence is quite similar and usually consists of short repeats with 2-4 adjacent guanines on the DNA strand that extends in the 5'-3' direction to form the G-overhang. Telomere function is achieved through multiple interactions between the DNA component of the telomere and various proteins that either directly bind to it or associate with the telomere via protein-protein interactions. The functions of telomeres are conserved throughout evolution, with the most important being chromosome protection against degradation and end-to-end fusions. Given the fact that telomeres provide an excellent solution to chromosome maintenance problems, it is not surprising that most organisms also have evolved similar proteins for their telomere maintenance machinery. Both double- and single-strand telomere binding proteins have been conserved throughout evolution. Double-strand telomere binding proteins, identified in budding and fission yeast and also in humans, are negative regulators of telomerase, and function by counting the number of existing telomere repeats on each chromosome end. Single-strand telomere binding proteins, on the other hand, seem to have multiple functions at the telomeres, both negatively and positively influencing telomere extension by telomerase, as well as providing a direct link between leading and lagging strand DNA replication machineries. Although sequence conservation between single-strand telomere binding proteins is less profound, all of these proteins appear to utilize a specific secondary structure, the oligosaccharide/oligonucleotide fold (OB fold), for efficient binding to telomeric DNA.

Given the overall evolutionary conservation of telomere proteins, it is likely that terminus-specific factors will be conserved in closely related species of plants. Although *Arabidopsis* is the plant model of choice for genetics, its close relative cauliflower is a plentiful biochemical source of telomerase and telomere binding proteins. In this study, we used a gel mobility shift assay to detect proteins in cauliflower nuclear extracts that bind G-rich single-strand oligonucleotide of telomeric sequence $(T_3AG_3)_5$. We detected four major protein complexes capable of specific binding to telomeric DNA. Competition experiments using different telomeric and non-telomeric oligonucleotides revealed that the cauliflower telomere binding proteins do not have a preference for a free 3' overhang and prefer to bind longer single-stranded regions of telomeric sequence. This finding suggests that these proteins bind along the length of the single-stranded overhang or to the displaced G-rich strand in the t-loop structure. We have also shown that telomeric DNA from evolutionary distant organisms such as ciliates can not be bound by cauliflower proteins. At the same time, the observed complexes have different affinities towards a more closely related telomere sequence from humans, with two out of four complexes being efficiently competed away by this DNA. This result implies that the cauliflower telomere binding proteins may recognize degenerate telomeric sequences dispersed throughout the plant genome.

Although the exact composition and identity of the telomere binding complexes observed in our gel-shift assays has not been determined, their strict substrate specificity suggests that plants possess specific single-strand telomere binding proteins capable of protecting chromosomal ends *in vivo*. The ability of cauliflower nuclear extracts to produce 4 biochemically distinct protein-DNA complexes could represent different steps

in the assembly of a protective structural cap on the chromosome terminus. Alternatively, these complexes may contain different proteins that play unique roles at the telomeres during various phases of the cell cycle or plant development. With this respect, telomere binding proteins specific to senescing *Arabidopsis* leaves have already been previously described (Zentgraf et al., 2000).

It is tempting to speculate that some of the telomere binding proteins observed in our assays are cauliflower orthologs of *Arabidopsis* Pot1 and Pot2 proteins. Since these two plant species are closely related, cauliflower is likely to possess *Pot1*-like genes. With the future development of antibodies specific to *Arabidopsis* Pot1 and Pot2 proteins, it will be possible to determine whether or not such proteins exist in cauliflower and whether the activities we have purified actually correspond to Pot1 and Pot2.

MATERIALS AND METHODS

Cauliflower nuclear extract preparation

To obtain large quantities of nuclear extract, up to 12 cauliflower heads were ground in a Waring blendor with 2 ml of grinding buffer (25 mM MES, pH 6.0, 10 mM $MgCl_2$, 5 mM $NaCl_2$, 0.5 M sucrose, 40% glycerol, 14 mM β-mercaptoethanol) per gram of tissue. The resulting mass was homogenized with a Polytron, and the extract filtered through cheesecloth and then through Miracloth (Calbiochem). The suspension was spun for 10 min at 1 krpm at 4^{o}C and then for 30 min at 3 krpm in a JA-14 rotor. To break chloroplasts, the pellet was resuspended several times in 10 ml wash buffer (25 mM MES, pH 6.0, 10 mM $MgCl_2$, 5 mM $NaCl_2$, 0.5 M sucrose, 25% glycerol, 0.1% Triton X-

100, 1.4 mM β-mercaptoethanol) and spun for 30 min at 4°C in a JA-20 rotor at 5 krpm. The final pellet was resuspended in extraction buffer (50 mM Tris HCl pH 7.8, 10 mM $MgCl_2$, 2M NaCl, 0.6 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, 1 mM DTT) and left overnight at 4°C for protein extraction. Nuclear extract was then dialyzed against dialysis buffer (50 mM Tris HCl pH 7.8, 5 mM $MgCl_2$, 40 mM KCl, 0.2 mM EDTA, 0.01 mM PMSF, 1 mM DTT, 20% glycerol) and cleared from DNA by DNase I digestion. Aliquots were quick-frozen in liquid nitrogen and stored at -80 °C until needed.

Electrophoretic mobility shift assay

Different concentrations of cauliflower nuclear extract were mixed with 0.5 ng of ^{32}P-labeled $(TTTAGGG)_5$ oligonucleotide in DNA-binding buffer (20 mM TrisHCl pH 8.0, 50 mM NaCl, 10 mM MgCl, 1mM EDTA, 1 mM DTT, 5% glycerol, 15 μg *Hae*III-digested *E .coli* DNA), and incubated at RT for 15 min. The complexes were then separated on 5% polyacrylamide gel (acrylamide: bisacrylamide 29:1) for 4 h at 150 volts in 1X TBE at RT, dried and exposed to film or PhosphorImager screens.

Protein purification and cross-linking assays

Ammonium sulfate was added to the cauliflower nuclear extract in 20% increments to 60% saturation at 4°C with stirring. The precipitate was collected by centrifugation, resuspended in TMG-50 buffer (50 mM Tris-HCl, pH 8.0, 5 mM $MgCl_2$, 10% glycerol) and dialyzed against dialysis buffer. The sample was then mixed with ampholytes (3-5 pH range) and subjected to isoelectric focusing on the Rotofor system (Bio-Rad) according to the manufacturer's instructions. Fractions with telomere binding

activity were pooled together and size-fractionated on a Superose 12 column (Pharmacia) in 20 mM Tris-HCl pH 8.0, 10% glycerol, 150 mM NaCl, 1 mM $MgCl_2$. The eluted sample was affinity-purified as described below. Activity was monitored at all stages by gel-shift assays. For DNA cross-linking assays, partially purified protein samples from Superose 12 column chromatography were incubated with radioactively labeled $(TTTAGGG)_5$ as described above for 15 min, and then cross-linked in a UV Stratalinker 1800 (Stratagene) for 15 min, followed by 10% SDS-PAGE. The gels were dried and exposed to PhosphorImager screens.

DNA-affinity purification and mass-spec analysis of cauliflower proteins

Partially purified telomere binding proteins from cauliflower (500 µg total protein) were mixed with 0.5 nmole of biotinylated $(TTTAGGG)_5$ oligonucleotide in 1 ml of WB buffer (20 mM Tris-acetate pH 7.5, 10 mM MgCl2, 1 mM EDTA, 50 mM NaCl, 5% glycerol, 1 mM DTT, 0.05% NP-40), incubated on ice for 30 min and added to 150 µl of UltraLink Immobilized NeutrAvidin Plus beads (Pierce), previously blocked for 15 min with WB buffer containing 0.01% NP-40, 1 mg/ml leupeptin, 0.05 mg/ml glycogen. The binding reaction was allowed to proceed for 1 h at 4^oC. The beads were then washed 3 times with 1 ml of cold WB buffer and eluted twice by 200 µl of elution buffer (20 mM Tris-HCl pH 8.0, 2M NaCl, 1 mM EDTA, 5% glycerol, 0.05% NP-40) at room temperature. The eluted samples were dialysed against 200 ml of dialysis buffer at 4^oC overnight. If higher concentrations of the proteins were needed, samples were dialyzed against solid polyvinylpyrrolidone (MW 40,000) at 4^oC for 2h.

For mass-spectrometry analysis, affinity-purified protein samples were run on SDS-PAGE and stained with a non-fixing silver stain (PIERCE). The candidate protein bands were cut out of the gel, digested with trypsin and subjected to mass-spectrometry analysis at the Texas A&M University mass-spectrometry laboratory. ProteinProspector package was used to identify *Arabidopsis* protein candidates.

CHAPTER V

SUMMARY AND DISCUSSION

Telomeres are essential features of most eukaryotic chromosomes, and analysis of their structure offers many clues towards understanding aspects of cellular senescence, aging and cancer. The past twenty years have provided an enormous amount of data on how telomeres perform their function. Although most of the initial work was performed on lower eukaryotes, it is becoming increasingly clear that many aspects of telomere biology are shared by organisms across evolution. In recent years, mammals have become a model of choice for studying telomeres in the context of complex genomes. These studies have significantly advanced our knowledge of telomere function, regulation and architecture. The identification of the mammalian telomere binding proteins (TRF1, TRF2, hPot1) has allowed scientists to exploit new avenues for understanding the relationship between telomere end protection, cell cycle, DNA damage responses, senescence and apoptosis. The discovery of t-loops in human cells demonstrated that complex organisms employ unusual and elegant solutions to chromosome end protection.

Although human and mouse represent organisms of choice for studies aimed at more medical-oriented investigations, such as analysis of cancer cells and mechanisms leading to tumor formation, these organisms also harbor various intrinsic drawbacks. For example, generation of gene knock-outs necessary to elucidate gene function is almost impossible in human cells. Moreover, while construction of gene knock-outs is possible

for mice, this process is very laborious and time-consuming. In addition, disruption of many telomere-related genes in the mouse system is lethal, making it impossible to study their functions.

The model plant organism *Arabidopsis thaliana* offers an alternative to more traditional higher eukaryotic models (Riha and Shippen, 2003b). The genome of this small mustard weed has been completely sequenced and annotated. The plant itself is very easy to grow, has a short generation time and is easily manipulated on the genetic level. T-DNA insertions containing various constructs can be easily introduced into the *Arabidopsis* genome to generate gene knock-out or activation lines, suitable for genetic and biochemical analysis. Most importantly, *Arabidopsis* shares much similarity with mammals in terms of general telomere organization, telomerase expression and responses to genome instability. In addition, *Arabidopsis* is unique among other eukaryotes in its exceptional tolerance to genome instability. Because of its great plasticity in development and genome organization, *Arabidopsis* can survive in the absence of telomerase and a number of DNA-damage response proteins. Similar deficiencies are not tolerated by mammals, perhaps because p53 activation by DNA damage prevents further cell division. This major difference is very important from the genetic point of view and makes *Arabidopsis* an ideal system to investigate the function of many essential genes. The knowledge obtained from elucidation of gene function in *Arabidopsis* should be applicable to a wide variety of other systems.

In addition to many similarities in telomere-associated activities between plants and other higher eukaryotes, *Arabidopsis* is also emerging as a powerful model to study certain aspects of telomere biology specific to plants. For example, it is becoming

increasingly clear that regulation of telomerase activity in plants is connected to auxin metabolism. In addition, factors regulating telomerase expression in *Arabidopsis* may also contribute to plant drought tolerance (S.Ren and T.McKnight, personal communication). Overall, analysis of *Arabidopsis* telomeres and their function can significantly contribute to our understanding of many aspects of telomere biology across evolution.

In this study, we analyzed several different aspects of telomere biology in *Arabidopsis.* First, telomere length was measured in different wild-type *Arabidopsis* accessions. We found that *Arabidopsis* ecotypes harbor different lengths of telomere tracts. One group of accessions displays Columbia-style telomeres in the range of 2-5 kb. The second group has much longer telomeres, almost twice the size of Columbia telomeres. We found that WS ecotype is unique among other accessions in that it displays a bimodal telomere size distribution. Approximately 15-20% of all plants randomly selected from this ecotype displayed short telomeres, while most individuals had longer telomeres. This finding raises important questions about the mechanism regulating acceptable telomere tract size in plants. It also demonstrates that natural variation can be a significant factor influencing telomere length in a given plant, and needs to be taken into consideration when analyzing telomere length mutants. This finding is especially important since many mutations in genes involved in telomere biology may only cause minor telomere length fluctuations. Very careful analysis of these mutants will be necessary to correctly establish a link between these genes and telomere length homeostasis.

Currently, *Arabidopsis* is the only model system in which the dynamic nature of telomere homeostasis can be analyzed in great details. Unlike other organisms studied to date, seven out of ten chromosome ends in *Arabidopsis* harbor unique subtelomeric sequences immediately adjacent to the telomere tracts for which specific DNA probes can be amplified. These probes allow analysis of telomere length on individual chromosome arms. Therefore, the fate of an individual telomere can be followed throughout plant development or in parent-progeny analysis. Using subtelomeric probes, we discovered that the length of each telomere is not fixed, but fluctuates within an ecotype-specific range, which we call the set point. Our data indicate that *Arabidopsis* does not have a predetermined shortest telomere end. Instead, the dynamic nature of telomere lengthening and shortening gives all telomere ends an equal opportunity to be the shortest or the longest in the population. Our data also demonstrated for the first time that telomeres on homologous chromosomes are coordinately regulated throughout plant development. In addition, we found that telomere lengths in siblings are more similar than in unrelated plants, suggesting that parental telomere length predetermines individual telomere fate in progeny.

Telomere length homeostasis is regulated through the action of telomerase. This enzyme is responsible for adding telomere repeats to chromosome ends to circumvent telomere shortening caused by the end-replication problem. We demonstrated that *Arabidopsis* telomerase preferentially extends the shortest telomere in the population. We also showed that *Arabidopsis* accessions have a preferred size range for the telomeres, which we call the optimal size range. For example, in the Columbia ecotype, while the overall telomere tract range is 2-5 kb, the optimal size for any given telomere appears to

be 3-3.5 kb. Telomeres shorter than 3 kb serve as a good substrate for telomerase, which tends to extend this end to the optimal size. Telomeres longer than 3.5 kb are usually not acted on by telomerase and slowly drift down in size. The concept of the optimal size range allows us to explain the otherwise seemingly stochastic nature of telomerase action. If a telomere is at the optimal size point, the shortening and lengthening forces have the same strength, and the telomere has an equal chance of being shortened or extended. We hypothesize that at this point the telomere may be divided into two populations, one acted on by telomerase and one not. Thus, telomere splitting reflects extensions and shortening events that either happened to the homologous chromosomes or to both chromosomes in different parts of the plant body. Future work should focus on the differences in telomere length between various plant organs. In some plants, we observed dramatic telomere extensions for a particularly short parental chromosome arm up to 2 kb in the progeny. We would like to know when telomerase action happens. Does telomerase act in meiosis or throughout plant development? Is this telomere lengthening a result of a single extension event or does telomerase act on the same chromosome arm in each cell to coordinately extend the telomere to the optimal size range? These questions can be addressed by following the fate of the same chromosome arm through plant development. DNA samples should be taken from leaves, stems, flowers and other plant tissues and analyzed for the changes in the telomere length. These experiments can also address whether recombination plays a role in the telomere maintenance or if telomerase and the end-replication problem are the major players affecting telomere length in *Arabidopsis*.

Another important question that still remains to be answered concerns the mechanisms responsible for the establishment of species-specific telomere length limit. It

is obvious from our studies that closely related ecotypes of the same species may have up to a two-fold difference in the acceptable telomere length, but the genetic determinants for this variation are not known. A single genetic factor is not responsible for all the telomere length differences observed in plants. Instead, it is likely that many separate genetic factors contribute to the heterogeneity of acceptable telomere length. If this is the case, mapping these genes may be possible to do by crossing individuals from different ecotypes and analyzing the progeny of the cross. Although the influence of certain genes on telomere homeostasis is well documented in yeast, for example, epigenetic factors may also contribute to the complexity of the telomere length control. It is known that epigenetics can play a major role at and near the telomeres. For example, telomere silencing, caused by rearrangement in the chromatin composition of the telomeres can lead to genetically heritable silencing of genes positioned near the telomeres. Therefore, a careful study of telomeric chromatin and factors known to modify it may shed light on the mechanisms of telomere length control and regulation in *Arabidopsis*.

The identification of Pot1, the first single-strand telomere binding protein in higher eukaryotes has had a tremendous influence on telomere research in all species. Pot1, as well as its sequence and functional homolog in budding yeast and ciliates, provides a general solution to the telomere end protection problem. In all species studied to date, telomeres are protected by a Pot1-like capping protein, which binds to the G-overhang. This protein provides physical protection for telomere ends from degradation and double strand break repair activities, and facilitates protein-protein interactions with other members of the telomere complex. In *Arabidopsis*, two *Pot1*-like genes were identified by a homology database search. This finding was unexpected, since all other

model organisms have only a single *Pot1* gene. Although over 30% of the *Arabidopsis* genes are thought to be duplicated (*Arabidopsis* Genome Initiative, 2000), this does not appear to be the case for *Pot1* and *Pot2* genes, which are located on regions of chromosomes that have not been duplicated in the past. Instead, it seems likely that these genes evolved to have different functions.

Both *Pot1* and *Pot2* genes are ubiquitously expressed. Although both proteins show similar DNA-binding characteristics *in vitro*, they localize to different subnuclear compartments, with Pot2 found exclusively in the nucleus, and Pot1 predominantly in the nucleolus. This finding supports the idea that *Arabidopsis* Pot1 and Pot2 have distinct functions in the cell. Telomeres in *Arabidopsis*, unlike those of other known species, localize to the nucleolus. Since both Pot1 and telomeres localize to the same compartment, we suspect that Pot1 binds telomeres throughout the cell cycle. During the M-phase, when the nucleolus disappears, telomeres are expected to be released into the nucleus where they become free to initiate homologous chromosome pairing. We hypothesize that during this phase of the cell cycle telomeres are bound by the nucleus-localized Pot2.

Other questions that remain to be answered include determining whether *Arabidopsis* Pot proteins actually bind telomeres *in vivo.* We would also like to know whether there is a dynamic switch between Pot1- and Pot2-bound telomere states. Is it cell cycle controlled? The development of Pot1 and Pot2 specific antibodies will allow us to perform immunodetection of Pot proteins *in vivo*, and determine whether they colocalize with telomeres.

Our current data indicate that *Pot2* mRNA is less abundant in all tissues than *Pot1* mRNA. One intriguing possibility is that Pot2 protein is only present in M-phase. To test this, variations in Pot1 and Pot2 protein levels during the cell cycle will need to be analyzed. Fluorescence resonance energy transfer (FRET) is a powerful technique for studying molecular interactions inside living cells (Sekar and Periasamy, 2003). This technique is commonly used in the laboratory of our collaborators, Dr. Naohiro Kato and Dr. Eric Lam (Rutgers University). FRET analysis of Pot1 and Pot2 interactions could be performed by Dr. Kato to establish whether these proteins can homo- and heterodimerize *in vivo*. Our yeast two-hybrid and *in vitro* co-immunoprecipitation experiments suggest that this may be the case, but *in vivo* analysis will be most informative in understanding biological significance of these interactions. Oligomerization of Pot proteins could serve as a switch between Pot1- and Pot2-bound states, leading to important changes in the structure of the telomere cap. Such studies could also help us understand why *Arabidopsis* is the only known organism whose telomeres reside in the nucleolus.

From studies done in humans and fission yeast it is clear that Pot1 proteins perform two important functions: protection of telomere tracts from degradation and regulation of telomerase accessibility to chromosome ends. In fission yeast, a $pot1^{+}$ deficiency leads to immediate telomere loss and cell death. Although knock-outs of the human Pot1 gene are currently unavailable, over-expression of a full-length or dominant-negative version of Pot1 protein leads to significant telomere elongation without any noticeable deleterious effects on cell viability. Knock-outs of the *Arabidopsis Pot1* and *Pot2* genes are also currently unavailable and identification of such mutants is a priority

for future research. Given that more T-DNA insertions in the *Arabidopsis* genome are being generated, such knock-outs should be available in the near future.

Since Pot1 is an essential gene in other organisms, Pot1 and Pot2 deficient *Arabidopsis* mutants may not be viable. If this is the case, point mutation or deletion mutants should be generated. Currently, we are in the process of analyzing four Pot2 and one Pot1 TILLING mutants (http://tilling.fhcrc.org:9366), which carry point mutations in conserved amino acids. In addition to its value as a powerful technique to generate loss-of-function alleles, the TILLING technology also has the potential to uncover separation of function mutants, which will be instrumental in analyzing different aspects of Pot1 and Pot2 function at the telomeres.

Over-expression is another powerful technique commonly used to uncover the role of telomere binding proteins (Smogorzewska et al., 2000; Colgin et al., 2003). Over-expression of full-length Pot1 and Pot2 proteins in *Arabidopsis* caused no apparent detrimental phenotype, suggesting that the levels of these proteins may not be a rate-limiting step in the telomere protection mechanism. Previous studies of human Pot1 protein demonstrated that a dominant-negative approach may be a valuable tool in understanding Pot1 function. Therefore, we constructed N- and C-terminal truncations of Pot1 and Pot2, and expressed them under the control of the strong viral 35S promotor. Although some of the transgenes ($P1^{DBD}$ and $P2^{\Delta DBD}$) showed no phenotype, others displayed clear telomere deregulation problems. For example, plants over-expressing $P1^{\Delta DBD}$ showed moderate telomere shortening, indicating that *Arabidopsis* Pot1 is involved in telomere length control. Some of the plants over-expressing $P2^{DBD}$ also showed telomere shortening. Remarkably, these plants were sterile and displayed a high

incidence of chromosome fusions, detected cytogenetically as anaphase bridges. This discovery clearly demonstrated that *Arabidopsis* Pot2 is a part of the capping complex involved in telomere protection against double strand DNA break repair activities, since over-expression of $P2^{DBD}$ causes a telomere cap deficiency, manifested by the chromosome fusion formation.

The $P2^{DBD}$ construct appears to be a dominant-negative version of the full-length Pot2. What is the mechanism leading to telomere cap deficiency in $P2^{DBD}$ mutants? Several models can be proposed to explain these phenotypes. First, $P2^{DBD}$ molecules may occupy all or most Pot2 binding sites on the telomeres, driving the endogenous Pot2 protein off the telomere by mass action. If this is the case, *in vivo* characterization of $P2^{DBD}$ subcellular localization and its colocalization with telomeres will need to be performed to address whether $P2^{DBD}$ binds telomeres. Second, if $P2^{DBD}$ itself does not bind telomeres, it may dimerize with the endogenous Pot2 protein and sequester it from the telomeres. In this case, telomeres will not be bound by Pot2 at all, leaving them exposed to the DNA repair machinery. To address this possibility, *in vitro* dimerization assays and FRET analysis of Pot2-$P2^{DBD}$ interactions will need to be performed. In addition, disruption of Pot2-telomere colocalization will also need to be addressed by immunostaining and FISH analysis.

One of the most intriguing phenotypes of $P2^{DBD}$ over-expression is the severe sterility of these mutants. In contrast to hundreds and even thousands of seeds usually collected from the wild-type plants, only 20 to 30 viable seeds are produced by some of the $P2^{DBD}$ over-expressing mutants. Since even late generations of telomerase mutants with remarkably short telomeres are still fertile, this striking phenotype of $P2^{DBD}$ plants

suggests that in addition to a telomere capping deficiency, this mutation also leads to problems in meiosis. Pot1 in fission yeast and fungi is known to be involved in mitotic control (Baumann and Cech, 2001; Pitt et al., 2004). Therefore, it is tempting to propose that *Arabidopsis* Pot2 protein is also involved in telomere-related meiotic control, perhaps through properly organizing the telomere for meiotic pairing. Analysis of meiotic events in $P2^{DBD}$ over-expressing plants will need to be performed in order to elucidate the possible function of Pot2 in meiosis.

Taken together, the distinct phenotypes provide strong evidence that Pot1 and Pot2 perform different functions in the cell (Figure 32). In our search for more severe dominant-negative alleles of Pot1 and Pot2, we created chimeric fusions of Pot1 and Pot2 domains hoping to simultaneously compromise the functions of both proteins. To our satisfaction, transgenic plants over-expressing $P2^{DBD}$-$P1^{\Delta DBD}$ displayed a variety of severe phenotypes leading to telomere length deregulation and inactivation of telomerase *in vivo*. The first generation of $P2^{DBD}$-$P1^{\Delta DBD}$ plants showed dramatic telomere shortening, comparable to the G6-G7 telomerase mutants. Therefore, the rate of telomere shortening in such plants was 6-7 times greater than in the mutants lacking telomerase. Thus, Pot proteins in the short term are much more critical for telomere stability than telomerase.

What is the mechanism of rapid telomere shortening in the $P2^{DBD}$-$P1^{\Delta DBD}$ mutants? The chimeric $P2^{DBD}$-$P1^{\Delta DBD}$ polypeptide corresponds to Pot1 protein whose DBD was replaced by Pot2 DBD. Our preliminary data indicate that this chimeric protein localizes to the nucleolus (data not shown), just like the full-length Pot1 itself. It is

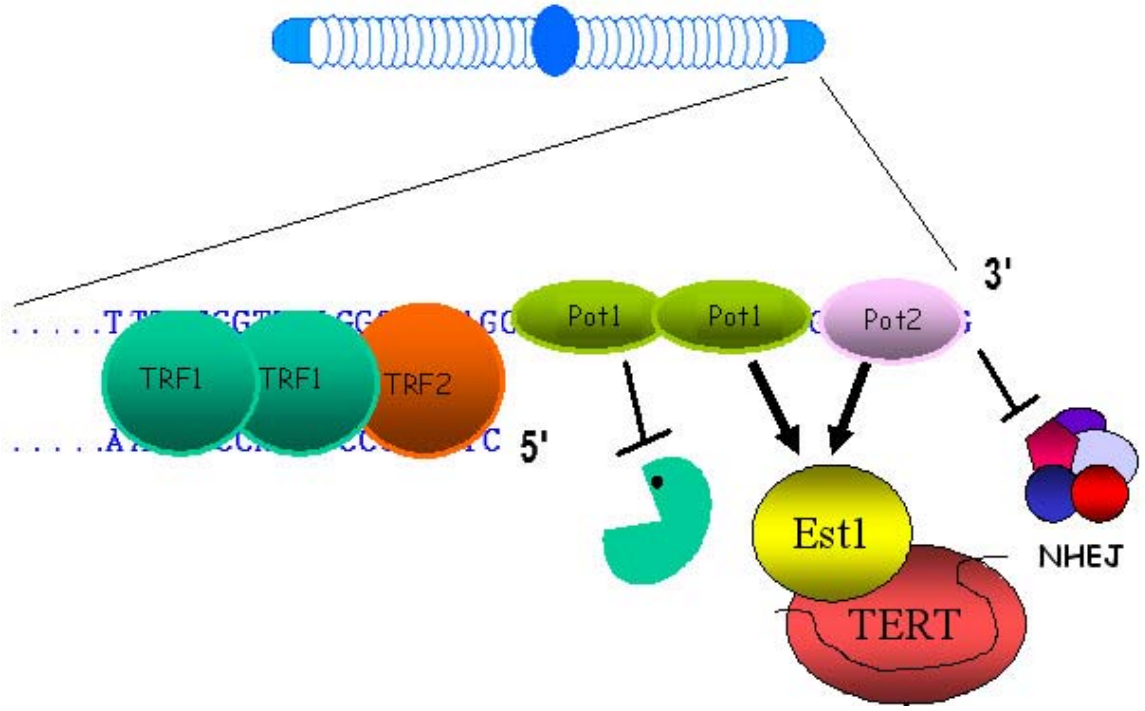

Figure 32. Proposed functions of single strand telomere binding proteins.

Both Pot1 and Pot2 proteins are involved in telomere length control, possibly by regulating telomerase access to telomeres. Est1 and TERT are components of telomerase. Pot1 protein is likely to protect telomeres from nuclease attack. Pot2 protein is also a component of the protective cap, restricting access of the non-homologous end joining (NHEJ) machinery to the chromosome ends. Also shown are the double-strand telomere binding proteins TRF1 and TRF2, which remain to be identified.

possible that in spite of the high degree of sequence similarity between the two DBDs, they have different DNA-binding properties *in vivo*. If the chimeric protein improperly binds to the telomeres, it may not be able to provide sufficient protection against nuclease attack. Alternatively, it may titrate the endogenous Pot1 protein from the telomeres, essentially exposing chromosome ends to unregulated telomere shortening activities. More careful analysis of $P2^{DBD}$-$P1^{\Delta DBD}$ subcellular localization and its interaction with the endogenous proteins will be necessary to fully understand these phenotypes.

Interestingly, separate over-expression of both $P2^{DBD}$ and $P1^{\Delta DBD}$ leads to only moderate telomere length deregulation, while over-expression of $P2^{DBD}$-$P1^{\Delta DBD}$ causes an exaggerated synergistic effect, more dramatic than in either single mutant. Since the chimeric protein contains $P2^{DBD}$ and $P1^{\Delta DBD}$ together as two parts of the same polypeptide, it is necessary to establish whether the observed phenotypes can be recapitulated by co-expression of $P2^{DBD}$ and $P1^{\Delta DBD}$ independently in the same cell, or if these two truncated proteins must be physically linked in one polypeptide to cause the observed telomere shortening. To test this, mutants over-expressing $P2^{DBD}$ and $P1^{\Delta DBD}$ can be crossed and their progeny analyzed.

One more major phenotype of $P2^{DBD}$-$P1^{\Delta DBD}$ mutants requires more detailed analysis in the future. When analyzing telomere length in these mutants, we noticed that their telomeres appeared not as a smear characteristic of wild-type plants, but as distinct bands reminiscent of telomeres in telomerase-deficient plants. The unique banding profile of *Arabidopsis* telomeres is a hallmark of telomerase deficiency (Riha et al., 2001), indicating that telomeres in $P2^{DBD}$-$P1^{\Delta DBD}$ mutants are not substrates for telomerase. This observation prompted us to test whether telomerase is active in these plants. TRAP assay

performed with proteins extracted from $P2^{DBD}$-$P1^{\Delta DBD}$ plants revealed that telomerase in these mutants was unable to elongate the DNA primer substrate (Y. Surovtseva and D. Shippen, unpublished). Moreover, when protein extract from the wild-type plants was mixed with extract from $P2^{DBD}$-$P1^{\Delta DBD}$ mutants, no telomerase activity could be detected (data not shown). We conclude that extracts from $P2^{DBD}$-$P1^{\Delta DBD}$ mutants contain a soluble and efficient telomerase inhibitor, capable of preventing telomere elongation by telomerase both *in vivo* and *in vitro*. These results are especially exciting since no natural factor capable of inactivating telomerase *in vivo* has been previously identified in any model system, including humans.

If the $P2^{DBD}$-$P1^{\Delta DBD}$ protein is indeed inhibiting telomerase *in vivo*, several mechanisms can be proposed to explain this (Figure 33). First, the chimeric protein may inhibit telomerase indirectly by tightly binding to telomeric DNA and preventing its binding by telomerase. Alternatively, $P2^{DBD}$-$P1^{\Delta DBD}$ may inhibit telomerase directly by physically interacting with the reverse transcriptase subunit or another component of telomerase holoenzyme. Both models can be tested in the future. The ability of $P2^{DBD}$-$P1^{\Delta DBD}$ to bind telomeres will be analyzed *in vitro* by gel-shift assays and *in vivo* by immunolocalization coupled with telomere FISH. Physical interaction of $P2^{DBD}$-$P1^{\Delta DBD}$ with TERT can be tested by coimmunoprecipitation, FRET analysis and yeast two-hybrid experiments.

Although the implications of discovering an *in vivo* telomerase inhibitor are significant, many more experiments will need to be performed to fully elucidate the functions of Pot1 and Pot2 in *Arabidopsis* and the mechanism of telomerase inhibition by $P2^{DBD}$-$P1^{\Delta DBD}$. Since the G-overhang is an essential component of a functional

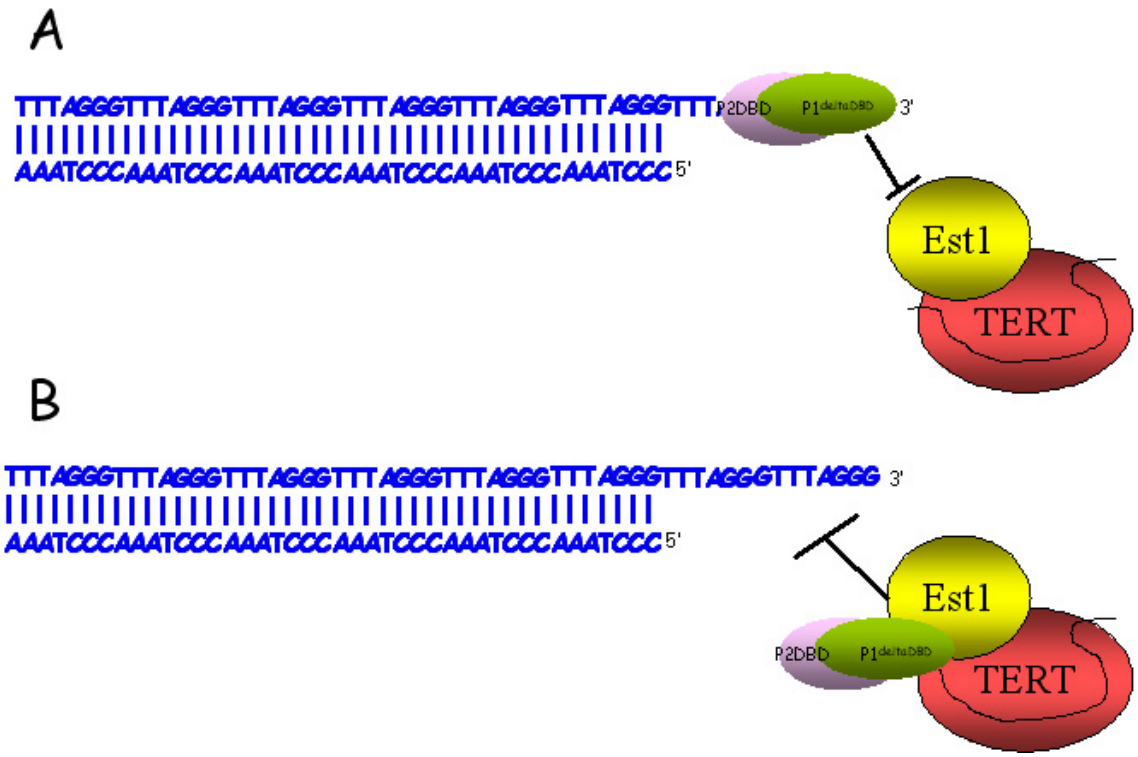

Figure 33. Proposed models of telomerase inhibition by $P2^{DBD}$-$P1^{\Delta DBD}$.

The chimeric protein may bind to the G-overhang and prevent telomerase from binding to it (A) or it can physically inhibit telomerase in the absence of telomeric DNA (B).

telomere, its length will need to be measured in all of the mutants mentioned above. The current literature indicates that Pot1-like proteins are unlikely to function on their own, but rather are likely to exist in a large multi-subunit complex, that performs multiple functions at the telomeres. Other components of the Pot1 and Pot2 complexes in *Arabidopsis* will need to be discovered through a number of approaches including yeast two-hybrid and reverse genetics. Most importantly, since human Pot1 protein was shown to act downstream of the major double-strand telomere binding protein TRF1, its homologues in *Arabidopsis* will need to be identified. Finally, the genetic interactions of Pot1 and Pot2 with other telomere-related factors, such as TERT, Ku and Mre11, will also need to be established.

In conclusion, the research presented here significantly adds to the wealth of knowledge gathered previously on *Arabidopsis* telomere biology. The results obtained here together with future analysis of *Arabidopsis* telomeres and telomere-associated factors will undoubtedly advance our understanding of telomere biology in higher eukaryotes overall, uncovering new mechanisms that regulate eukaryotic telomeres. Further characterization of *Arabidopsis* Pot1 and Pot2 proteins may also help us to understand the intriguing relationship between telomeres, aging and cancer development.

REFERENCES

Adams Martin, A., Dionne, I., Wellinger, R.J., and Holm, C. (2000). The function of DNA polymerase alpha at telomeric G tails is important for telomere homeostasis. Mol Cell Biol **20,** 786-796.

***Arabidopsis* Genome Initiative.** (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408,** 796-815.

Armstrong, S.J., Franklin, F.C., and Jones, G.H. (2001). Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. J Cell Sci **114,** 4207-4217.

Artandi, S.E., and DePinho, R.A. (2000). A critical role for telomeres in suppressing and facilitating carcinogenesis. Curr Opin Genet Dev **10,** 39-46.

Bailey, S.M., Meyne, J., Chen, D.J., Kurimasa, A., Li, G.C., Lehnert, B.E., and Goodwin, E.H. (1999). DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. Proc Natl Acad Sci U S A **96,** 14899-14904.

Baird, D.M., Rowson, J., Wynford-Thomas, D., and Kipling, D. (2003). Extensive allelic variation and ultrashort telomeres in senescent human cells. Nature Genetics **33,** 203-207.

Baumann, P., and Cech, T.R. (2000). Protection of telomeres by the Ku protein in fission yeast. Mol Biol Cell **11,** 3265-3275.

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., Podell, E., and Cech, T.R. (2002). Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing. Mol Cell Biol **22,** 8079-8087.

Bednenko, J., Melek, M., and Shippen, D.E. (1998). Reiterative dG addition by *Euplotes crassus* telomerase during extension of non-telomeric DNA. Nucleic Acids Res **26,** 3998-4004.

Bednenko, J., Melek, M., Greene, E.C., and Shippen, D.E. (1997). Developmentally regulated initiation of DNA synthesis by telomerase: evidence for factor-assisted *de novo* telomere formation. EMBO J **16,** 2507-2518.

Bianchi, A., and de Lange, T. (1999). Ku binds telomeric DNA *in vitro*. J Biol Chem **274,** 21223-21227.

Bianchi, A., Smith, S., Chong, L., Elias, P., and de Lange, T. (1997). TRF1 is a dimer and bends telomeric DNA. EMBO J **16,** 1785-1794.

Bilaud, T., Brun, C., Ancelin, K., Koering, C.E., Laroche, T., and Gilson, E. (1997). Telomeric localization of TRF2, a novel human telobox protein. Nature Genetics **17,** 236-239.

Blackburn, E.H. (2001). Switching and signaling at the telomere. Cell **106,** 661-673.

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.

Bottius, E., Bakhsis, N., and Scherf, A. (1998). *Plasmodium falciparum* telomerase: *de novo* telomere addition to telomeric and nontelomeric sequences and role in chromosome healing. Mol Cell Biol **18,** 919-925.

Boulton, S.J., and Jackson, S.P. (1998). Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. EMBO J **17,** 1819-1828.

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

Bryan, T.M., Goodrich, K.J., and Cech, T.R. (2000). A mutant of *Tetrahymena* telomerase reverse transcriptase with increased processivity. J Biol Chem **275,** 24199-24207.

Bundock, P., and Hooykaas, P. (2002). Severe developmental defects, hypersensitivity to DNA-damaging agents, and lengthened telomeres in *Arabidopsis* MRE11 mutants. Plant Cell **14,** 2451-2462.

Burr, B., Burr, F.A., Matz, E.C., and Romero-Severson, J. (1992). Pinning down loose ends: mapping telomeres and factors affecting their length. Plant Cell **4,** 953-960.

Cesare, A.J., Quinney, N., Willcox, S., Subramanian, D., and Griffith, J.D. (2003). Telomere looping in *P. sativum* (common garden pea). Plant J **36,** 271-279.

Chakhparonian, M., and Wellinger, R.J. (2003). Telomere maintenance and DNA replication: how closely are these two connected? Trends in Genetics **19,** 439-446.

Chan, S.W., and Blackburn, E.H. (2002). New ways not to make ends meet: telomerase, DNA damage proteins and heterochromatin. Oncogene **21,** 553-563.

Chandra, A., Hughes, T.R., Nugent, C.I., and Lundblad, V. (2001). Cdc13 both positively and negatively regulates telomere replication. Genes Dev **15,** 404-414.

Chen, C.M., Wang, C.T., and Ho, C.H. (2001). A plant gene encoding a Myb-like protein that binds telomeric GGTTAG repeats *in vitro*. J Biol Chem **276,** 16511-16519.

Chen, J.L., Blasco, M.A., and Greider, C.W. (2000). Secondary structure of vertebrate telomerase RNA. Cell **100,** 503-514.

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

Cline, J., Braman, J.C., and Hogrefe, H.H. (1996). PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res **24,** 3546-3551.

Cocciolone, S.M., and Cone, K.C. (1993). Pl-Bh, an anthocyanin regulatory gene of maize that leads to variegated pigmentation. Genetics **135,** 575-588.

Cohn, M., and Blackburn, E.H. (1995). Telomerase in yeast. Science **269,** 396-400.

Colgin, L.M., Baran, K., Baumann, P., Cech, T.R., and Reddel, R.R. (2003). Human POT1 facilitates telomere elongation by telomerase. Curr Biol **13,** 942-946.

Colgin, L.M., Wilkinson, C., Englezou, A., Kilian, A., Robinson, M.O., and Reddel, R.R. (2000). The hTERTalpha splice variant is a dominant negative inhibitor of telomerase activity. Neoplasia **2,** 426-432.

Collins, K. (1999). Ciliate telomerase biochemistry. Annu Rev Biochem **68,** 187-218.

Collins, K., and Gandhi, L. (1998). The reverse transcriptase component of the *Tetrahymena* telomerase ribonucleoprotein complex. Proc Natl Acad Sci U S A **95,** 8485-8490.

Cooper, J.P., Nimmo, E.R., Allshire, R.C., and Cech, T.R. (1997). Regulation of telomere length and function by a Myb-domain protein in fission yeast. Nature **385,** 744-747.

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.

de Lange, T. (2002). Protection of mammalian telomeres. Oncogene **21,** 532-540.

Eckert, K.A., and Kunkel, T.A. (1990). High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res **18,** 3739-3744.

Espejel, S., Franco, S., Rodriguez-Perales, S., Bouffler, S.D., Cigudosa, J.C., and Blasco, M.A. (2002). Mammalian Ku86 mediates chromosomal fusions and apoptosis caused by critically short telomeres. EMBO J **21,** 2207-2219.

Evans, S.K., and Lundblad, V. (1999). Est1 and Cdc13 as comediators of telomerase access. Science **286,** 117-120.

Evans, S.K., and Lundblad, V. (2000). Positive and negative regulation of telomerase access to the telomere. J Cell Sci **113 Pt 19,** 3357-3364.

Fajkus, J., Kovarik, A., Kralovics, R., and Bezdek, M. (1995). Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*. Mol Gen Genet **247,** 633-638.

Fang, G., and Cech, T.R. (1993). *Oxytricha* telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits. Proc Natl Acad Sci U S A **90,** 6056-6060.

Fang, G., Gray, J.T., and Cech, T.R. (1993). *Oxytricha* telomere-binding protein: separable DNA-binding and dimerization domains of the alpha-subunit. Genes Dev **7,** 870-882.

Fang, G.W., and Cech, T.R. (1991). Molecular cloning of telomere-binding protein genes from *Stylonychia mytilis*. Nucleic Acids Res **19,** 5515-5518.

Ferreira, M.G., and Cooper, J.P. (2001). The fission yeast Taz1 protein protects chromosomes from Ku-dependent end-to-end fusions. Molecular Cell **7,** 55-63.

Fitzgerald, M.S., McKnight, T.D., and Shippen, D.E. (1996). Characterization and developmental patterns of telomerase expression in plants. Proc Natl Acad Sci U S A **93,** 14422-14427.

Fitzgerald, M.S., Riha, K., Gao, F., Ren, S., McKnight, T.D., and Shippen, D.E. (1999). Disruption of the telomerase catalytic subunit gene from *Arabidopsis* inactivates telomerase and leads to a slow loss of telomeric DNA. Proc Natl Acad Sci U S A **96,** 14813-14818.

Forsyth, N.R., Wright, W.E., and Shay, J.W. (2002). Telomerase and differentiation in multicellular organisms: turn it off, turn it on, and turn it off again. Differentiation **69,** 188-197.

Fuchs, J.A.B., Schubert I. (1995). Telomere sequence localization and karyotype evolution in higher plants. Plant Syst. Evol. **196,** 227-241.

Gallego, M.E., and White, C.I. (2001). RAD50 function is essential for telomere maintenance in *Arabidopsis*. Proc Natl Acad Sci U S A **98,** 1711-1716.

Gallego, M.E., Jalut, N., and White, C.I. (2003). Telomerase dependence of telomere lengthening in Ku80 mutant *Arabidopsis*. Plant Cell **15,** 782-789.

Gallego, M.E., Jeanneau, M., Granier, F., Bouchez, D., Bechtold, N., and White, C.I. (2001). Disruption of the *Arabidopsis* RAD50 gene leads to plant sterility and MMS sensitivity. Plant J **25,** 31-41.

Gandhi, L., and Collins, K. (1998). Interaction of recombinant *Tetrahymena* telomerase proteins p80 and p95 with telomerase RNA and telomeric DNA substrates. Genes Dev **12,** 721-733.

Garcia-Cao, M., O'Sullivan, R., Peters, A.H., Jenuwein, T., and Blasco, M.A. (2004). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nature Genetics **36,** 94-99.

Garvik, B., Carson, M., and Hartwell, L. (1995). Single-stranded DNA arising at telomeres in cdc13 mutants may constitute a specific signal for the RAD9 checkpoint. Mol Cell Biol **15,** 6128-6138.

Gilley, D., Lee, M.S., and Blackburn, E.H. (1995). Altering specific telomerase RNA template residues affects active site function. Genes Dev **9,** 2214-2226.

Giraldo, R., and Rhodes, D. (1994). The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J **13,** 2411-2420.

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.

Goytisolo, F.A., and Blasco, M.A. (2002). Many ways to telomere dysfunction: *in vivo* studies using mouse models. Oncogene **21,** 584-591.

Grandin, N., Reed, S.I., and Charbonneau, M. (1997). Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. Genes Dev **11,** 512-527.

Grandin, N., Damon, C., and Charbonneau, M. (2001). Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. EMBO J **20,** 1173-1183.

Gravel, S., Larrivee, M., Labrecque, P., and Wellinger, R.J. (1998). Yeast Ku as a regulator of chromosomal DNA end structure. Science **280,** 741-744.

Gray, J.T., Celander, D.W., Price, C.M., and Cech, T.R. (1991). Cloning and expression of genes for the *Oxytricha* telomere-binding protein: specific subunit interactions in the telomeric complex. Cell **67,** 807-814.

Greene, E.C., and Shippen, D.E. (1998). Developmentally programmed assembly of higher order telomerase complexes with distinct biochemical and structural properties. Genes Dev **12,** 2921-2931.

Greider, C.W. (1996). Telomere length regulation. Annual Review of Biochemistry **65,** 337-365.

Greider, C.W. (1998). Telomerase activity, cell proliferation, and cancer. Proc Natl Acad Sci U S A **95,** 90-92.

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

Griffith, J., Bianchi, A., and de Lange, T. (1998). TRF1 promotes parallel pairing of telomeric tracts *in vitro*. Journal of Molecular Biology **278,** 79-88.

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.

Hardy, C.F., Balderes, D., and Shore, D. (1992). Dissection of a carboxy-terminal region of the yeast regulatory protein RAP1 with effects on both transcriptional activation and silencing. Mol Cell Biol **12,** 1209-1217.

Harley, C.B., and Villeponteau, B. (1995). Telomeres and telomerase in aging and cancer. Curr Opin Genet Dev **5,** 249-255.

Harrington, L.A., and Greider, C.W. (1991). Telomerase primer specificity and chromosome healing. Nature **353,** 451-454.

Hayflick, L. (1965). The limited *in vitro* lifetime of human diploid cell strains. Experimental Cell Research **37,** 614-636.

Heller, K., Kilian, A., Piatyszek, M.A., and Kleinhofs, A. (1996). Telomerase activity in plant extracts. Mol Gen Genet **252,** 342-345.

Heller-Uszynska, K., Schnippenkoetter, W., and Kilian, A. (2002). Cloning and characterization of rice (*Oryza sativa* L) telomerase reverse transcriptase, which reveals complex splicing patterns. Plant J **31,** 75-86.

Hemann, M.T., and Greider, C.W. (2000). Wild-derived inbred mouse strains have short telomeres. Nucleic Acids Res **28,** 4474-4478.

Hicke, B.J., Celander, D.W., MacDonald, G.H., Price, C.M., and Cech, T.R. (1990). Two versions of the gene encoding the 41-kilodalton subunit of the telomere binding protein of *Oxytricha nova*. Proc Natl Acad Sci U S A **87,** 1481-1485.

Horvath, M.P., Schweiker, V.L., Bevilacqua, J.M., Ruggles, J.A., and Schultz, S.C. (1998). Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. Cell **95,** 963-974.

Hwang, M.G., Chung, I.K., Kang, B.G., and Cho, M.H. (2001). Sequence-specific binding property of *Arabidopsis thaliana* telomeric DNA binding protein 1 (AtTBP1). FEBS Lett **503,** 35-40.

Kanoh, J., and Ishikawa, F. (2001). spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. Curr Biol **11,** 1624-1630.

Kanoh, J., and Ishikawa, F. (2003). Composition and conservation of the telomeric complex. Cell Mol Life Sci **60,** 2295-2302.

Karamysheva, Z., Wang, L., Shrode, T., Bednenko, J., Hurley, L.A., and Shippen, D.E. (2003). Developmentally programmed gene elimination in *Euplotes crassus* facilitates a switch in the telomerase catalytic subunit. Cell **113,** 565-576.

Karlseder, J. (2003). Telomere repeat binding factors: keeping the ends in check. Cancer Letters **194,** 189-197.

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

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.

Keohavong, P., and Thilly, W.G. (1989). Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci U S A **86,** 9253-9257.

Kharbanda, S., Kumar, V., Dhar, S., Pandey, P., Chen, C., Majumder, P., Yuan, Z.M., Whang, Y., Strauss, W., Pandita, T.K., Weaver, D., and Kufe, D. (2000). Regulation of the hTERT telomerase catalytic subunit by the c-Abl tyrosine kinase. Curr Biol **10,** 568-575.

Kilian, A., Bowtell, D.D., Abud, H.E., Hime, G.R., Venter, D.J., Keese, P.K., Duncan, E.L., Reddel, R.R., and Jefferson, R.A. (1997). Isolation of a candidate human telomerase catalytic subunit gene, which reveals complex splicing patterns in different cell types. Hum Mol Genet **6,** 2011-2019.

Killan, A., Heller, K., and Kleinhofs, A. (1998). Development patterns of telomerase activity in barley and maize. Plant Mol Biol **37,** 621-628.

Kim, B., Ayran, J.C., Sagar, S.G., Adman, E.T., Fuller, S.M., Tran, N.H., and Horrigan, J. (1999a). New human immunodeficiency virus, type 1 reverse transcriptase (HIV-1 RT) mutants with increased fidelity of DNA synthesis. Accuracy, template binding, and processivity. J Biol Chem **274,** 27666-27673.

Kim, J.H., Kim, W.T., and Chung, I.K. (1998). Rice proteins that bind single-stranded G-rich telomere DNA. Plant Mol Biol **36,** 661-672.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L., and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science **266,** 2011-2015.

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

Kipling, D., and Cooke, H.J. (1990). Hypervariable ultra-long telomeres in mice. Nature **347,** 400-402.

Klein, F., Laroche, T., Cardenas, M.E., Hofmann, J.F., Schweizer, D., and Gasser, S.M. (1992). Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. Journal of Cell Biology **117,** 935-948.

Klobutcher, L.A., Swanton, M.T., Donini, P., and Prescott, D.M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc Natl Acad Sci U S A **78,** 3015-3019.

Konig, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996). The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. Cell **85,** 125-136.

Kreiter, M., Irion, V., Ward, J., and Morin, G. (1995). The fidelity of human telomerase. Nucleic Acids Symp Ser**,** 137-139.

Kunkel, T.A., and Soni, A. (1988). Mutagenesis by transient misalignment. J Biol Chem **263,** 14784-14789.

Kurenova, E.V., and Mason, J.M. (1997). Telomere functions. A review. Biochemistry (Mosc) **62,** 1242-1253.

Kwon, C., and Chung, I.K. (2003). Interaction of an *Arabidopsis* RNA-binding protein with plant single-stranded telomeric DNA modulates telomerase activity. J Biol Chem **279,** 12812-12818.

Kyo, S., Takakura, M., Taira, T., Kanaya, T., Itoh, H., Yutsudo, M., Ariga, H., and Inoue, M. (2000). Sp1 cooperates with c-Myc to activate transcription of the human telomerase reverse transcriptase gene (hTERT). Nucleic Acids Res **28,** 669-677.

Kyrion, G., Boakye, K.A., and Lustig, A.J. (1992). C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*. Mol Cell Biol **12,** 5159-5173.

Laroche, T., Martin, S.G., Tsai-Pflugfelder, M., and Gasser, S.M. (2000). The dynamics of yeast telomeres and silencing proteins through the cell cycle. J Struct Biol **129,** 159-174.

Laroche, T., Martin, S.G., Gotta, M., Gorham, H.C., Pryde, F.E., Louis, E.J., and Gasser, S.M. (1998). Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. Curr Biol **8,** 653-656.

Lee, J.H., Kim, J.H., Kim, W.T., Kang, B.G., and Chung, I.K. (2000). Characterization and developmental expression of single-stranded telomeric DNA-binding proteins from mung bean (*Vigna radiata*). Plant Mol Biol **42,** 547-557.

Lee, M.S., and Blackburn, E.H. (1993). Sequence-specific DNA primer effects on telomerase polymerization activity. Mol Cell Biol **13,** 6586-6599.

Lei, M., Baumann, P., and Cech, T.R. (2002). Cooperative binding of single-stranded telomeric DNA by the Pot1 protein of *Schizosaccharomyces pombe*. Biochemistry **41,** 14560-14568.

Lei, M., Podell, E.R., Baumann, P., and Cech, T.R. (2003). DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. Nature **426,** 198-203.

Li, B., Oestreich, S., and de Lange, T. (2000). Identification of human Rap1: implications for telomere evolution. Cell **101,** 471-483.

Lin, J.J., and Zakian, V.A. (1994). Isolation and characterization of two *Saccharomyces cerevisiae* genes that encode proteins that bind to (TG1-3)n single strand telomeric DNA *in vitro*. Nucleic Acids Res **22,** 4906-4913.

Lingner, J., Hendrick, L.L., and Cech, T.R. (1994). Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. Genes Dev **8,** 1984-1998.

Liu, K., Hodes, R.J., and Weng, N. (2001). Cutting edge: telomerase activation in human T lymphocytes does not require increase in telomerase reverse transcriptase (hTERT) protein but is associated with hTERT phosphorylation and nuclear translocation. J Immunol **166,** 4826-4830.

Loayza, D., and De Lange, T. (2003). POT1 as a terminal transducer of TRF1 telomere length control. Nature **424,** 1013-1018.

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

Lundblad, V. (2002). Telomere maintenance without telomerase. Oncogene **21,** 522-531.

Lundblad, V., and Szostak, J.W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. Cell **57,** 633-643.

Lustig, A.J. (2001). Cdc13 subcomplexes regulate multiple telomere functions. Nat Struct Biol **8,** 297-299.

Lustig, A.J. (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. Nature Reviews Genetics **4,** 916-923.

Maceluch, J., Kmieciak, M., Szweykowska-Kulinska, Z., and Jarmolowski, A. (2001). Cloning and characterization of *Arabidopsis thaliana* AtNAP57--a homologue of yeast pseudouridine synthase Cbf5p. Acta Biochim Pol **48,** 699-709.

Magnenat, L., Tobler, H., and Muller, F. (1999). Developmentally regulated telomerase activity is correlated with chromosomal healing during chromatin diminution in *Ascaris suum*. Mol Cell Biol **19,** 3457-3465.

Makarov, V.L., Hirose, Y., and Langmore, J.P. (1997). Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. Cell **88,** 657-666.

Marcand, S., Gilson, E., and Shore, D. (1997). A protein-counting mechanism for telomere length regulation in yeast. Science **275,** 986-990.

Marcand, S., Brevet, V., Mann, C., and Gilson, E. (2000). Cell cycle restriction of telomere elongation. Curr Biol **10,** 487-490.

Marian, C.O., Bordoli, S.J., Goltz, M., Santarella, R.A., Jackson, L.P., Danilevskaya, O., Beckstette, M., Meeley, R., and Bass, H.W. (2003). The maize Single myb histone 1 gene, Smh1, belongs to a novel gene family and encodes a protein that binds telomere DNA repeats *in vitro*. Plant Physiol **133,** 1336-1350.

Maringele, L., and Lydall, D. (2002). EXO1-dependent single-stranded DNA at telomeres activates subsets of DNA damage and spindle checkpoint pathways in budding yeast yku70Delta mutants. Genes Dev **16,** 1919-1933.

Martin, S.G., Laroche, T., Suka, N., Grunstein, M., and Gasser, S.M. (1999). Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. Cell **97,** 621-633.

McClintock, B. (1941). The stability of broken ends of chromosomes of *Zea mays*. Genetics **26**, 234-282.

McEachern, M.J., Krauskopf, A., and Blackburn, E.H. (2000). Telomeres and their control. Annu Rev Genet **34,** 331-358.

McKnight, T.D., Fitzgerald, M.S., and Shippen, D.E. (1997). Plant telomeres and telomerases. A review. Biochemistry (Mosc) **62,** 1224-1231.

Melek, M., and Shippen, D.E. (1996). Chromosome healing: spontaneous and programmed *de novo* telomere formation by telomerase. Bioessays **18,** 301-308.

Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddle, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q., Bacchetti, S., Haber, D.A., and Weinberg, R.A. (1997). hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. Cell **90,** 785-795.

Miller, K.M., and Cooper, J.P. (2003). The telomere protein Taz1 is required to prevent and repair genomic DNA breaks. Molecular Cell **11,** 303-313.

Mitton-Fry, R.M., Anderson, E.M., Hughes, T.R., Lundblad, V., and Wuttke, D.S. (2002). Conserved structure for single-stranded telomeric DNA recognition. Science **296,** 145-147

Moretti, P., Freeman, K., Coodly, L., and Shore, D. (1994). Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. Genes Dev **8,** 2257-2269.

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

Murti, K.G., and Prescott, D.M. (1999). Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. Proc Natl Acad Sci U S A **96,** 14436-14439.

Murzin, A.G. (1993). OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. EMBO J **12,** 861-867.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). Telomerase catalytic subunit homologs from fission yeast and human. Science **277,** 955-959.

Nugent, C.I., Hughes, T.R., Lue, N.F., and Lundblad, V. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. Science **274,** 249-252.

Oguchi, K., Liu, H., Tamura, K., and Takahashi, H. (1999). Molecular cloning and characterization of *AtTERT*, a telomerase reverse transcriptase homolog in *Arabidopsis thaliana*. FEBS Lett **457,** 465-469.

Olovnikov, A.M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. Journal of Theoretical Biology **41,** 181-190.

Pardue, M.L., and DeBaryshe, P.G. (1999). Telomeres and telomerase: more than the end of the line. Chromosoma **108,** 73-82.

Peersen, O.B., Ruggles, J.A., and Schultz, S.C. (2002). Dimeric structure of the *Oxytricha nova* telomere end-binding protein alpha-subunit bound to ssDNA. Nat Struct Biol **9,** 182-187.

Pennock, E., Buckley, K., and Lundblad, V. (2001). Cdc13 delivers separate complexes to the telomere for end protection and replication. Cell **104,** 387-396.

Petracek, M.E., Konkel, L.M., Kable, M.L., and Berman, J. (1994). A *Chlamydomonas* protein that binds single-stranded G-strand telomere DNA. EMBO J **13,** 3648-3658.

Pitt, C.W., Moreau, E., Lunness, P.A., and Doonan, J.H. (2004). The *pot1+* homologue in *Aspergillus nidulans* is required for ordering mitotic events. J Cell Sci **117,** 199-209.

Polotnianka, R.M., Li, J., and Lustig, A.J. (1998). The yeast Ku heterodimer is essential for protection of the telomere against nucleolytic and recombinational activities. Curr Biol **8,** 831-834.

Price, C.M. (1990). Telomere structure in *Euplotes crassus*: characterization of DNA-protein interactions and isolation of a telomere-binding protein. Mol Cell Biol **10,** 3421-3431.

Price, C.M. (1999). Telomeres and telomerase: broad effects on cell growth. Curr Opin Genet Dev **9,** 218-224.

Price, C.M., and Cech, T.R. (1987). Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. Genes Dev **1,** 783-793.

Price, C.M., Skopp, R., Krueger, J., and Williams, D. (1992). DNA recognition and binding by the *Euplotes* telomere protein. Biochemistry **31,** 10835-10843.

Qi, H., and Zakian, V.A. (2000). The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. Genes Dev **14,** 1777-1788.

Ramsden, D.A., and Gellert, M. (1998). Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. EMBO J **17,** 609-614.

Richards, E.J., and Ausubel, F.M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell **53,** 127-136.

Richards, E.J., Chao, S., Vongs, A., and Yang, J. (1992). Characterization of *Arabidopsis thaliana* telomeres isolated in yeast. Nucleic Acids Res **20,** 4039-4046.

Riha, K., and Shippen, D.E. (2003a). Ku is required for telomeric C-rich strand maintenance but not for end-to-end chromosome fusions in *Arabidopsis*. Proc Natl Acad Sci U S A **100,** 611-615.

Riha, K., and Shippen, D.E. (2003b). Telomere structure, function and maintenance in *Arabidopsis*. Chromosome Res **11,** 263-275.

Riha, K., Fajkus, J., Siroky, J., and Vyskot, B. (1998). Developmental control of telomere lengths and telomerase activity in plants. Plant Cell **10,** 1691-1698.

Riha, K., McKnight, T.D., Griffing, L.R., and Shippen, D.E. (2001). Living with genome instability: plant responses to telomere dysfunction. Science **291,** 1797-1800.

Riha, K., Watson, J.M., Parkey, J., and Shippen, D.E. (2002). Telomere length deregulation and enhanced sensitivity to genotoxic stress in *Arabidopsis* mutants deficient in Ku70. EMBO J **21,** 2819-2826.

Riha, K., McKnight, T.D., Fajkus, J., Vyskot, B., and Shippen, D.E. (2000). Analysis of the G-overhang structures on plant telomeres: evidence for two distinct telomere architectures. Plant J **23,** 633-641.

Sadaie, M., Naito, T., and Ishikawa, F. (2003). Stable inheritance of telomere chromatin structure and function in the absence of telomeric repeats. Genes Dev **17,** 2271-2282.

Sekar, R.B., and Periasamy, A. (2003). Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. J Cell Biol **160,** 629-633.

Shampay, J., and Blackburn, E.H. (1988). Generation of telomere-length heterogeneity in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A **85,** 534-538.

Shay, J.W., and Wright, W.E. (2001). Aging. When do telomeres matter? Science **291,** 839-840.

Sheng, H., Hou, Z., Schierer, T., Dobbs, D.L., and Henderson, E. (1995). Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. Mol Cell Biol **15,** 1144-1153.

Smith, S., Giriat, I., Schmitt, A., and de Lange, T. (1998). Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. Science **282,** 1484-1487.

Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M.R., Schnapp, G., and de Lange, T. (2000). Control of human telomere length by TRF1 and TRF2. Mol Cell Biol **20,** 1659-1668.

Spink, K.G., Evans, R.J., and Chambers, A. (2000). Sequence-specific binding of Taz1p dimers to fission yeast telomeric DNA. Nucleic Acids Res **28,** 527-533.

Sprung, C.N., Reynolds, G.E., Jasin, M., and Murnane, J.P. (1999). Chromosome healing in mouse embryonic stem cells. Proc Natl Acad Sci U S A **96,** 6781-6786.

Sundquist, W.I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature **342,** 825-829.

Taggart, A.K., Teng, S.C., and Zakian, V.A. (2002). Est1p as a cell cycle-regulated activator of telomere-bound telomerase. Science **297,** 1023-1026.

Tsujimoto, H., Usami, N., Hasegawa, K., Yamada, T., Nagaki, K., and Sasakuma, T. (1999). *De novo* synthesis of telomere sequences at the healed breakpoints of wheat deletion chromosomes. Mol Gen Genet **262,** 851-856.

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.

Vaziri, H., and Benchimol, S. (1998). Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. Curr Biol **8,** 279-282.

Wang, J., Xie, L.Y., Allan, S., Beach, D., and Hannon, G.J. (1998). Myc activates telomerase. Genes Dev **12,** 1769-1774.

Wang, M.J., Lin, Y.C., Pang, T.L., Lee, J.M., Chou, C.C., and Lin, J.J. (2000). Telomere-binding and Stn1p-interacting activities are required for the essential function of *Saccharomyces cerevisiae* Cdc13p. Nucleic Acids Res **28,** 4733-4741.

Wang, W., Skopp, R., Scofield, M., and Price, C. (1992). *Euplotes crassus* has genes encoding telomere-binding proteins and telomere-binding protein homologs. Nucleic Acids Res **20,** 6621-6629.

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

Wei, C., and Price, M. (2003). Protecting the terminus: t-loops and telomere end-binding proteins. Cell Mol Life Sci **60,** 2283-2294.

Wei, C., and Price, C.M. (2004). Cell cycle localization, dimerization, and binding domain architecture of the telomere protein cPot1. Mol Cell Biol **24,** 2091-2102.

Weinrich, S.L., Pruzan, R., Ma, L., Ouellette, M., Tesmer, V.M., Holt, S.E., Bodnar, A.G., Lichtsteiner, S., Kim, N.W., Trager, J.B., Taylor, R.D., Carlos, R., Andrews, W.H., Wright, W.E., Shay, J.W., Harley, C.B., and Morin, G.B. (1997). Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. Nature Genetics **17,** 498-502.

Wotton, D., and Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. Genes Dev **11,** 748-760.

Wright, W.E., Piatyszek, M.A., Rainey, W.E., Byrd, W., and Shay, J.W. (1996). Telomerase activity in human germline and embryonic tissues and cells. Developmental Genetics **18,** 173-179.

Xiang, C., Miao, Z., and Lam, E. (1997). DNA-binding properties, genomic organization and expression pattern of TGA6, a new member of the TGA family of bZIP transcription factors in *Arabidopsis thaliana*. Plant Mol Biol **34,** 403-415.

Yang, J., Chang, E., Cherry, A.M., Bangs, C.D., Oei, Y., Bodnar, A., Bronstein, A., Chiu, C.P., and Herron, G.S. (1999). Human endothelial cell life extension by telomerase expression. J Biol Chem **274,** 26141-26148.

Yang, S.W., Kim, D.H., Lee, J.J., Chun, Y.J., Lee, J.H., Kim, Y.J., Chung, I.K., and Kim, W.T. (2003). Expression of the telomeric repeat binding factor gene NgTRF1 is closely coordinated with the cell division program in tobacco BY-2 suspension culture cells. J Biol Chem **278,** 21395-21407.

Yi, X., White, D.M., Aisner, D.L., Baur, J.A., Wright, W.E., and Shay, J.W. (2000). An alternate splicing variant of the human telomerase catalytic subunit inhibits telomerase activity. Neoplasia **2,** 433-440.

Yu, E.Y., Kim, S.E., Kim, J.H., Ko, J.H., Cho, M.H., and Chung, I.K. (2000). Sequence-specific DNA recognition by the Myb-like domain of plant telomeric protein RTBP1. J Biol Chem **275,** 24208-24214.

Yu, G.L., Bradley, J.D., Attardi, L.D., and Blackburn, E.H. (1990). *In vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. Nature **344,** 126-132.

Zentgraf, U. (1995). Telomere binding proteins of *Arabidopsis thaliana*. Plant Mol Biol **27,** 467-475.

Zentgraf, U., Hinderhofer, K., and Kolb, D. (2000). Specific association of a small protein with the telomeric DNA-protein complex during the onset of leaf senescence in *Arabidopsis thaliana*. Plant Mol Biol **42,** 429-438.

Zhou, X.Z., and Lu, K.P. (2001). The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor. Cell **107,** 347-359.

Zhu, J., Wang, H., Bishop, J.M., and Blackburn, E.H. (1999). Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. Proc Natl Acad Sci U S A **96,** 3723-3728.

Zijlmans, J.M., Martens, U.M., Poon, S.S., Raap, A.K., Tanke, H.J., Ward, R.K., and Lansdorp, P.M. (1997). Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats. Proc Natl Acad Sci U S A **94,** 7423-7428.

APPENDIX

DIFFERENT MODES OF *DE NOVO* TELOMERE FORMATION BY PLANT TELOMERASES*

INTRODUCTION

Eukaryotes ensure the integrity of their linear chromosomes by capping the ends with telomeres. The telomeric nucleoprotein complex protects chromosomes from end-to-end fusion, recombination and exonucleolytic degradation (Kurenova and Mason, 1997) and allows continued cell proliferation in a dedifferentiated or undifferentiated state (Price, 1999). Telomeric DNA in plants consists of tandem arrays of TTTAGGG repeats (Richards and Ausubel, 1988; Riha and Shippen, 2003b). The predominant mechanism for generating and sustaining telomeric DNA is through the action of telomerase, a specialized ribonucleoprotein reverse transcriptase that adds telomeric repeats onto chromosome ends to replenish terminal DNA sequences not effectively duplicated by conventional DNA replication machinery (Greider, 1996). Telomerase expression is tightly regulated in higher eukaryotes (reviewed in Greider, 1998; McKnight et al., 1997).

Both the telomerase RNA moiety and the telomerase reverse transcriptase (TERT) are necessary for enzymatic activity (Collins and Gandhi, 1998; Weinrich et al., 1997). While TERT has been characterized in a variety of different organisms, including *Arabidopsis* (Fitzgerald et al., 1999), the telomerase RNA subunit has not yet been identified from any plant species. Nevertheless, studies in other organisms demonstrate that this molecule provides a templating sequence complementary to the G-rich strand of the telomere (Greider, 1996). Telomeric DNA synthesis is initiated when the G-rich single-stranded 3' terminus of a natural chromosome end (or a corresponding DNA oligonucleotide for the *in vitro* reaction) hybridizes to the telomerase RNA template and is elongated by the addition of telomeric repeats. Once the end of the templating domain is copied, the 3' terminus of the DNA is repositioned back at the beginning of the template for another round of repeat synthesis.

In addition to maintaining pre-existing tracts of telomeric DNA, telomerase can form telomeres *de novo* on non-telomeric DNA. This property is distinct from telomere maintenance, as the DNA substrate cannot form an extended hybrid with the telomerase RNA template. Barbara McClintock (McClintock, 1941) coined the term “chromosome healing” to describe the acquisition of telomeric function at non-telomeric ends resulting from chromosome breakage. We now know that the healing corresponds to *de novo* telomere formation.

We are examining the mechanism of chromosome healing in plants using an *in vitro* telomerase assay. Our previous studies showed that cauliflower telomerase can initiate telomere repeat synthesis on DNA oligonucleotide primers lacking any complementarity to the predicted telomerase RNA template. In contrast, carrot

telomerase requires primers carrying at least two nucleotides of telomeric sequence on the 3' terminus (Fitzgerald et al., 1996). Here we expand this survey to include telomerases from five additional monocots and dicots. Our data uncover three distinct classes of telomerase enzymes that vary not only in their recognition and utilization of non-telomeric DNA sequences, but also in the fidelity with which they form new tracts of telomeric DNA.

RESULTS

An *in vitro* chromosome healing assay for plant telomerases

The chromosome healing function of telomerases from *Arabidopsis thaliana*, soybean, *Silene latifolia*, sorghum, and maize was examined using a modified version of TRAP (telomere repeat amplification protocol) (Kim et al., 1994; Fitzgerald et al., 1996). TRAP works in two steps: telomerase adds telomeric repeats onto a forward primer, then the products are amplified by PCR using a reverse primer complementary to the telomere repeat. The forward primer in our standard TRAP assays is a 21-mer oligonucleotide, AG_3T_3-G_3, containing the telomeric sequence AGGGTTT at its 5' terminus, 11 non-telomeric nucleotides in the middle and three guanine residues at the 3' terminus. The three 3' terminal dG residues can form Watson-Crick base pairs with the rC_{6-8} residues in the predicted telomerase RNA template for plants (Figure 34A). The reverse transcriptase subunit then catalyzes the addition of TTTAGGG repeats by copying the RNA templating domain. Elongation products are amplified by PCR using the reverse primer $(C_3TA_3)_3$.

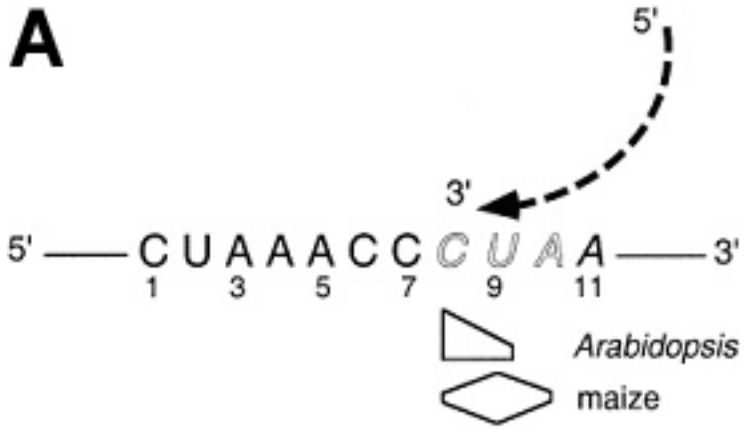

Figure 34. *De novo* telomere synthesis activities of *Arabidopsis*, soybean and *Silene* telomerases.

(A) Diagram of the predicted RNA template for plant telomerases. The sequence permutation is based on experimentally defined templating domains for ciliate and mammalian telomerase (Chen et al., 2000; Collins, 1999). The alignment region of the template (open italics) and the preferred positioning sites for *Arabidopsis* and maize telomerases are indicated. The dashed line represents annealing of a DNA primer to the template. The shapes below the RNA template represent 3' terminus alignment preferences. (B-D) TRAP assay results. Assays were performed with the primers indicated using extracts from soybean (B), *Arabidopsis* (C) and *Silene* (D). $(C_3TA_3)_3$ does not serve as a substrate for *de novo* telomere formation (Fitzgerald et al., 1996).

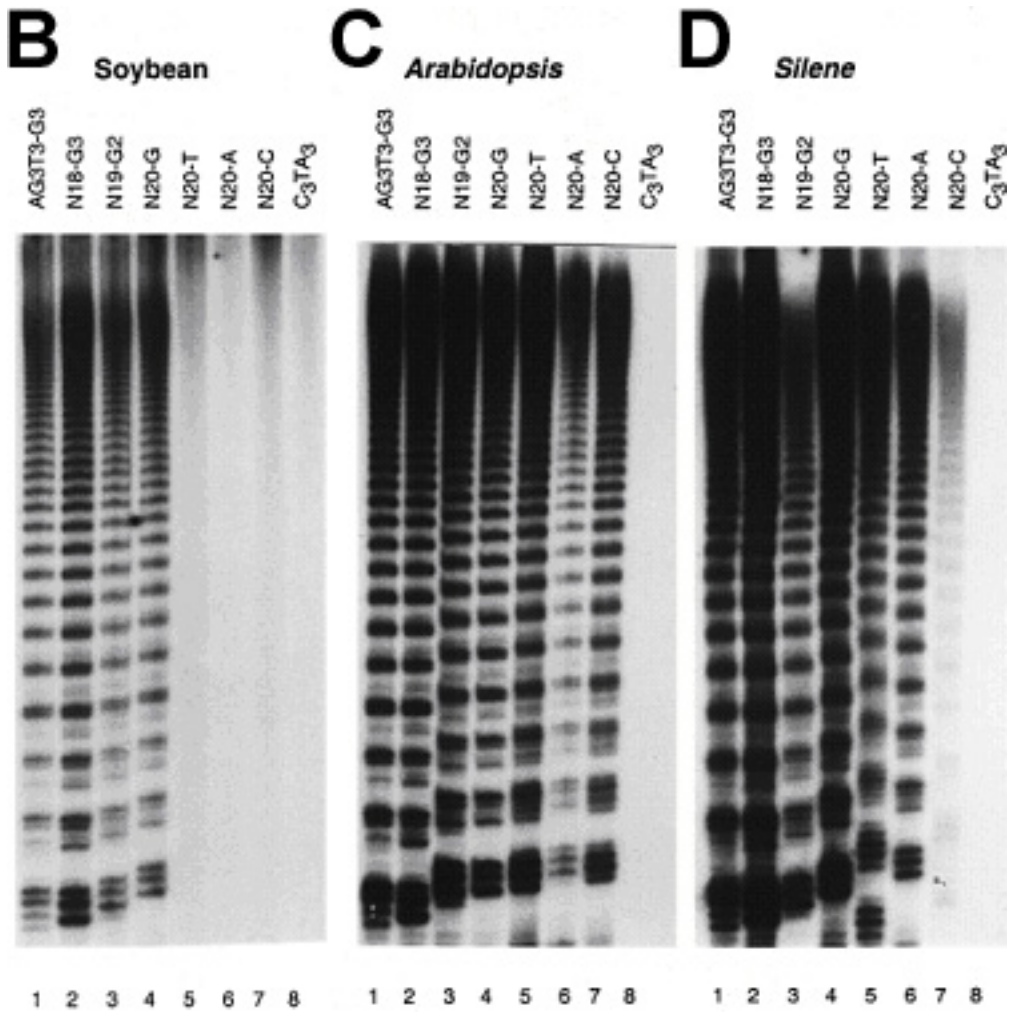

Figure 34. Continued.

TRAP assays were performed with a series of primers that lacked the 5' telomeric cassette and terminated in different 3' nucleotides. For example, the primer N19-G2 is composed of 19 nucleotides of non-telomeric sequence and two dG residues at its 3' terminus. Similarly, the primer N20-C contains 20 nucleotides of non-telomeric sequence and terminates in a dC residue. In contrast to other primers in this study, the 3' terminus of the N_{20}-C primer cannot form any base pairs with the predicted template sequence, and hence elongation of this primer 3' terminus must occur in the absence of hybridization to the RNA (Figure 34A). We have previously shown that the telomeric C-strand sequence $(C_3TA_3)_3$ does not serve as a substrate for *de novo* telomere formation (Fitzgerald et al., 1996).

Soybean telomerase exhibits strict substrate specificity

Soybean telomerase efficiently extended primers AG_3T_3-G_3, N_{18}-G_3, N_{19}-G_2, and N_{20}-G (Figure 34B, lanes 1-4). As the number of guanine residues at the primer 3' end was reduced from three, to two, to one, the reaction products became offset from each other by one nucleotide at a time. The strongest bands in the reaction product profile reflect a "pause", or product release, as the end of the RNA template is copied into DNA (Greider, 1996). Therefore, an offset in banding profiles indicates that the register of telomeric repeats synthesized is shifted, a result expected if primer 3' ends are positioned at different sites on the RNA template during the first round of elongation. Thus, the data imply that primers terminating in dG residues align appropriately with $rC_{6\text{-}8}$ in the RNA template (Figure 34A).

In contrast to other telomerases in our study, the soybean telomerase did not extend primers terminating in dT, dA or dC (Figure 34B, lanes 5-7). The N_{20}-G primer 3' terminus, in contrast to N_{20}-T and N_{20}-A, can form up to two Watson-Crick base pairs with the RNA template (AG/UC) (see Materials and Methods). Therefore, the soybean telomerase may require primers carrying a minimum of two nucleotides of 3' terminal complementarity to the RNA template. Alternatively, a 3'-terminal dG residue could be necessary for extension by the soybean telomerase. This possibility seems unlikely, however, as mammalian and protozoan telomerases efficiently elongate primers terminating in dG or T residues (Bottius et al., 1998; Greider, 1996).

***Arabidopsis* and maize telomerases use preferred sites on the RNA template for extension of non-telomeric DNA.**

As with the soybean telomerase, the *Arabidopsis* and maize enzymes efficiently elongated AG_3T_3-G_3, N_{18}-G_3, N_{19}-G_2, and N_{20}-G primers (Figure 34C, lanes 1-4; data not shown), apparently aligning the 3' terminal nucleotides with rC_{6-8} in the RNA template (Figure 34A). However, in contrast to soybean, these enzymes also elongated N_{20}-T, N_{20}-A and N_{20}-C (Figure 34C, lanes 5-7; data not shown), indicating that *Arabidopsis* and maize telomerases have a much more relaxed specificity for DNA recognition. The banding profiles obtained in these reactions were either the same as for N_{20}-G (Figure 34C, lanes 5 and 7) or offset by one nucleotide (Figure 34C, compare lanes 5-7 with lane 4; data not shown), suggesting that the primers aligned preferentially at rC_8 or rU_9 in the RNA.

The *Silene* and sorghum telomerases do not use preferred sites on the RNA template to initiate *de novo* telomere formation

The *Silene* and sorghum telomerases displayed properties that set them apart from the other enzymes in our study. Both enzymes extended primers terminating in dG, dA, dT or dC residues (Figure 34D and data not shown), although the *Silene* telomerase reproducibly elongated the N_{20}-C primer with much lower efficiency (Figure 34D, lane 7). For both enzymes, the banding products generated with the N_{20}-A primer were offset by one nucleotide relative to the N_{20}-G products (Figure 34D, lanes 4 and 6**;** data not shown), suggesting that the 3' terminus of N_{20}-A aligned with the rU_9 residue in the RNA template. However, in striking contrast to reactions with the *Arabidopsis* and maize telomerases, the N_{20}-T and N_{20}-C products generated by *Silene* and sorghum were grossly offset from one another and from the N_{20}-G primer (Figure 34D, lanes 4-7; data not shown), indicating that these primers were not positioned together at a preferred site on the RNA template.

Sequence analysis of TRAP products

To learn more about the products of the *in vitro* chromosome healing assays, we cloned and sequenced several TRAP products from each reaction. Each clone was sequenced at least twice to verify that errors had not been introduced during sequencing reactions. Since soybean telomerase was incapable of extending the N_{20}-T, N_{20}-A and N_{20}-C primers, the data set for this enzyme is substantially smaller.

Primers terminating in a dG residue

For all the telomerases, the majority of reactions carried out with AG_3T_3-G_3 initiated with the addition of TTTAGGG, an outcome predicted by Watson-Crick base-pairing of the primer 3' terminus with rC_{6-8} on the RNA template (Table 4; Figure 34A). Unexpectedly, all five of the maize and two of the sorghum AG_3T_3-G_3 reactions incorporated one or two extra dG residues before dT's were added, an error termed G-slippage.

The N_{20}-G primer was also extended in a similar manner by all telomerases (Table 5). Theoretically, the 3'-terminal nucleotide of N_{20}-G can pair with any residue in the rC triad (Figure 34A, positions 6 to 8). However, rC_8 should be preferred because the N_{20}-G primer terminates in AG allowing base pairs to form with rU_9 and rC_8. Sequence data support this prediction: in all cases the N_{20}-G primer was extended initially by GGTTTAG (Table 5).

The primer terminating in dA

As described above, the N_{20}-A primer was not extended by soybean telomerase. However, telomerases from all other species, except *Arabidopsis*, extended this primer in a similar manner (Table 6). In all maize, *Silene* and sorghum products, the first nucleotides added to the primer were GGG, suggesting that the 3'-terminal dA was positioned across from the rU_9 residue in the RNA template (Figure 34A). Again, this observation supports our interpretation of the TRAP banding profile for this primer. In reactions with the *Silene* telomerase, the N_{20}-A banding profile was offset by one

Table 4. Sequence of TRAP products from the $AG_3T_3N_{11}G_3$ primer

Plant	Clone	Sequence[1]
Arabidopsis	#4	GGGtttaggg(tttaggg)$_6$
	#5	GGGtttaggg(tttaggg)$_{13}$
	#7	GGGtttaggg(tttaggg)$_{21}$
	#9	GGGtttaggg(tttaggg)$_{10}$
	#12	GGGtttaggg(tttaggg)$_{11}$
Maize	#6	GGG**G**tttaggg(tttaggg)$_7$
	#8	GGG**G**tttaggg(tttaggg)$_7$
	#50	GGG**G**tttaggg(tttaggg)$_4$
	#51	GGG**GG**tttaggg(tttaggg)$_6$
	#53	GGG**G**tttaggg**TT**aggg(tttaggg)$_{11}$
Silene	#60	GGGtttaggg(tttaggg)$_6$
	#62	GGGtttaggg (tttaggg)$_2$
	#65	GGGtttaggg(tttaggg)$_3$
Sorghum	#87	GGG**G**tttaggg(tttaggg)$_{14}$
	#88	GGG**G**tttaggg(tttaggg)$_9$ttt**T**ggg (tttaggg)$_4$
	#89	GGG(tttaggg)$_4$
	#91	GGGtttaggg(tttaggg)$_5$
	#92	GGGtttaggg(tttaggg)$_5$
Soybean	#6	GGGtttaggg(tttaggg)$_{34}$
	#50	GGGtttaggg(tttaggg)$_4$
	#71	GGGtttaggg(tttaggg)$_2$
	#73	GGGtttaggg(tttaggg)$_5$
	#74	GGGtttaggg(tttaggg)$_2$

[1] Primer sequence is shown in upper case and errors are indicated by bold, underlined upper case

Table 5. Sequence of TRAP products from the N_{20}-G primer

Plant	Clone	Sequence[1]
Arabidopsis	#9	CAGgg(tttaggg)$_{17}$
	#10	CAGgg(tttaggg)$_{18}$
	#20	CAGgg(tttaggg)$_{4}$
	#23	CAGgg(tttaggg)$_{20}$
	#26	CAGgg(tttaggg)$_{8}$
Maize	#1	CAGggtttaggg(tttaggg)$_{6}$
	#4	CAGggtttaggg(tttaggg)$_{26}$
	#50	CAGggtttaggg(tttaggg)$_{3}$
	#57	CAGggtttaggg(tttaggg)$_{14}$
Silene	#20	CAGggttt<u>T</u>aggg(tttaggg)$_{7}$
	#22	CAGggtttagggttt<u>T</u>aggg(tttaggg)$_{6}$
	#66	CAGggtttaggg(tttaggg)$_{16}$
	#67	CAGggtttaggg(tttaggg)$_{9}$
Sorghum	#4	CAGggtttagggttt<u>T</u>agggtttaggg ttt<u>T</u>agggttt<u>T</u>aggg(tttaggg)$_{6}$
	#6	CAGggtt<u>A</u>aggg(tttaggg)$_{9}$
	#50	CAGggtt<u>A</u>aggg(tttaggg)$_{10}$
	#93	CAGggtttaggg(tttaggg)$_{2}$
	#94	CAGggtttaggg(tttaggg)$_{11}$
Soybean	#1	CAGgg (tttaggg)$_{13}$
	#15	CAGgg (tttaggg)$_{3}$
	#42	CAGgg**<u>G</u>**tttaggg(tttaggg)$_{6}$
	#75	CAGggtttaggg(tttaggg)$_{8}$
	#76	CAGggtttaggg(tttaggg)$_{5}$

[1] Primer sequence is shown in upper case and errors are indicated by bold, underlined upper case

Table 6. Sequence of TRAP products from the N_{20}-A primer

Plant	Clone	Sequence[1]
Arabidopsis	#3	CAAgggtttagggttt**T**aggg(tttaggg)$_3$**TT**agggtttaggg
	#7	CAAgg(tttaggg)$_8$
	#30	CAAgg(tttaggg)$_{35}$
	#36	CAAgg(tttaggg)$_8$
	#44	CAAggg(tttaggg)$_{38}$
Maize	#10	CAAggg(tttaggg)$_3$tt**A**aggg
	#11	CAAgggtttaggg(tttaggg)$_{35}$
	#13	CAAgggtttaggg(tttaggg)$_{23}$
	#14	CAAgggtttaggg(tttaggg)$_{26}$
	#50	CAAggg(tttaggg)$_{17}$
Silene	#30	CAAgggtttagggttt**T**aggg(tttaggg)$_2$
	#31	CAAggg(tttaggg)$_4$ttt**T**aggg(tttaggg)$_{15}$
	#32	CAAgggtttagggttt**T**agggtttaggg)$_3$**TT** agggtttagggttt**T**aggg
	#33	CAAgggtttaggg(tttaggg)$_8$
Sorghum	#1	CAAggg (tttaggg)$_{14}$
	#13	CAAgggtttaggg(tttaggg)$_3$
	#14	CAAgggttt**T**aggg(tttaggg)$_{11}$
	#15	CAAgggtttaggg(tttaggg)$_2$
	#16	CAAggg (tttaggg)$_4$

[1] Primer sequence is shown in upper case and errors are indicated by bold, underlined upper case

nucleotide relative to N_{20}-G (Figure 34D, lane 4), a primer whose 3' terminus is predicted to align with rC_8 in the template (Figure 34A).

In two of the *Arabidopsis* clones (#30 and #36) the first nucleotides added were GG rather than GGG (Table 6), which implies that the primer was initially positioned across from rC_8. This finding was unexpected since such positioning overrides normal base pairing between the primer 3'-terminal dA and rU_9 in the template. However, in conjunction with sequence data from N_{20}-C clones (see below), this observation supports a model in which the *Arabidopsis* rC_8 acts as a default or preferred site for primer alignment, bypassing normal Watson-Crick base pairing. In the N_{20}-A case, additional primer-template stability can also be derived from pairing of the rU_9 position to the dA residue adjacent to the primer 3' terminus.

The primer terminating in dT

Sequence data from the N_{20}-T TRAP reactions yielded several unexpected results (Table 7). The 3' terminus of N_{20}-T is expected to anneal to either of the rA_{10} or rA_{11} residues in the RNA template (Figure 34A). However, this primer annealed to a variety of positions on the RNA, with maize, *Silene* and sorghum telomerases employing at least two different annealing sites for primer binding. Sorghum telomerase was the most variable: the N_{20}-T primer was extended by a different telomere repeat permutation in four of the six clones sequenced. The variability in N_{20}-T positioning was also evident in the TRAP banding profiles from *Silene* telomerase. While N_{20}-A generated a product profile with three strong bands per repeat, the N_{20}-T products contained four to six strong

Table 7. Sequence of TRAP products from the N_{20}-T primer

Plant	Clone	Sequence[1]
Arabidopsis	#1	CATggttt**T**aggg(tttaggg)$_4$
	#5	CATgg(tttaggg)$_{31}$
	#30	CATggttt**T**aggg(tttaggg)$_3$
	#34	CATgg(tttaggg)
	#35	CATgg(tttaggg)$_{25}$
Maize	#5	CATgggtttaggg(tttaggg)$_{28}$
	#6	CATtagggtttaggg(tttaggg)$_{11}$
	#7	CATagggtttaggg(tttaggg)$_5$
	#8	CATagggtttaggg(tttaggg)$_9$**TC**tttaggg(tttaggg)$_{13}$
	#23	CATagggtttaggg(tttaggg)$_{12}$
Silene	#25	CATtagggtttagggttt**T**aggg(tttaggg)$_{22}$
	#26	CATagggtttaggg(tttaggg)$_3$
	#27	CATagggtttaggg**TT**aggg
	#28	CATagggtttaggg(tttaggg)$_3$
Sorghum	#8	CATgggttt**T**aggg(tttaggg)$_8$
	#9	CATggttt**T**ggg**G**(tttaggg)$_2$
	#12	CATttagggtttagggtttagggtt**A**aggg**TT**aggg
	#30	CATggtt**A**agggttt**T**aggg(tttaggg)$_2$
	#31	CATttagggtttagggttt**T**aggg(tttaggg)$_3$
	#35	CATagggttt**T**agggtttagggtttag**A**g(tttaggg)$_3$

[1] Primer sequence is shown in upper case and errors are indicated by bold, underlined upper case

bands per repeat (Figure 34D, lanes 5 and 6). This increased number of "pause" sites in each repeat is consistent with alignment of N_{20}-T at multiple sites on the RNA template.

A completely different situation is seen for *Arabidopsis* clones. All five clones sequenced initiate with two dG's, implying that in each case the primer was positioned across from rC_8. As for N_{20}-A, this position was able to override normal base pairing of N_{20}-T primer to rA_{10} and rA_{11}. We can not exclude the possibility that by sequencing random clones we missed some alternative primer positioning in the case of the N_{20}-T primer for *Arabidopsis* (as suggested by Figure 34C, lane 5). However, that all five clones sequenced show the same distinctive properties is remarkable.

The primer terminating in dC

Since the 3' terminus of the N_{20}-C primer cannot base pair with any residues in the telomerase RNA template, sequence analysis of these TRAP products should reveal whether a true preferred alignment position is used in the initial round of primer elongation. For *Arabidopsis*, all of N_{20}-C, N_{20}-T and a subset of the N_{20}-A reactions resulted in incorporation of GGTTTAG (Table 8). These data imply that *Arabidopsis* telomerase preferentially aligns non-telomeric DNA on the rC_8 position in the RNA template (Figure 34A) to override canonical Watson-Crick base pairing interactions at other sites in the RNA template.

The maize telomerase also utilized a preferred alignment position for the N_{20}-C primer. For all five clones, three dG residues were initially added (Table 8), implying that N_{20}-C invariably anealed on the rU_9 position (Fig. 34A). Intriguingly, this interaction was very unstable, since 4/5 clones displayed G-slippage where a fourth dG

Table 8. Sequence of TRAP products from the N_{20}-C primer

Plant	Clone	Sequence[1]
Arabidopsis	#10	CACggtttaggg(tttaggg)$_{23}$
	#11	CACggtttaggg(tttaggg)$_{5}$
	#12	CACggttt**T**aggg(tttaggg)$_{7}$
	#19	CACggtttaggg(tttaggg)$_{21}$
	#20	CACggtttaggg(tttaggg)$_{15}$
Maize	#15	CACggg**G**tttaggg(tttaggg)$_{4}$
	#16	CACggg**G**(tttaggg)$_{4}$
	#17	CACggg**G**tttaggg(tttaggg)$_{24}$
	#18	CACggg**G**tttaggg(tttaggg)$_{42}$
	#19	CACgggtttaggg(tttaggg)$_{17}$
Silene	#35	CACtttaggg(tttaggg)$_{10}$ttt**T**aggg
	#38	CACttagggttt**T**aggg(tttaggg)$_{5}$
Sorghum	#19	CACtttaggg(tttaggg)$_{3}$
	#20	CACggtttaggg(tttaggg)$_{15}$
	#22	CACtttaggg(tttaggg)$_{16}$
	#24	CACtttaggg(tttaggg)$_{2}$
	#51	CACttaggg(tttaggg)$_{16}$

[1] Primer sequence is shown in upper case and errors are indicated by bold, underlined upper case

was added to the primer prior to TTTA (see Table 8). Nevertheless, the data suggest that Maize and *Arabidopsis* use a similar mechanism for positioning non-telomeric DNA in the telomerase active site.

By contrast, N_{20}-C elongation reactions catalyzed by *Silene* and sorghum telomerases were much more variable. Though the data set for *Silene* was limited, the enzyme added either TTAGGG or TTTAGGG. Of the five N_{20}-C sorghum reactions sequenced, three clones initiated with TTTAGGG, one clone with GGTTTAG, and one clone with TTAGGG (Table 8). Thus, the sequence data in combination with the gel based TRAP assay results strongly suggest that the *Silene* and sorghum telomerases do not use a preferred site for anealing non-telomeric DNA on the RNA template.

The fidelity of plant telomerases

Our large data set of TRAP sequences allowed us to gauge the fidelity of different plant telomerases and provided further support for distinct classes of telomerase enzymes. As shown in Tables 4-8, many of the TRAP products contained imperfect telomere repeats, even reactions primed by AG_3T_3-G_3, the most telomere-like sequence.

Errors in TRAP products from all telomerases were most prevalent in the first two repeats added (Tables 4-8). Incorporation of one or two extra dG residues (G-slippage) was a common error (Table 9). G-slippage was confined to reactions with AG_3T_3-G_3, N_{20}-G and N_{20}-C (Tables 4, 5 and 8). Thirteen of the fourteen G-slippage events occurred with monocot telomerases, and ten of these were with the maize enzyme. Technically, addition of one or two dG residues to the AG_3T_3-G_3 primer (Table 4), whose 3' terminus could form three G-C base pairs with the RNA template, could reflect

Table 9. Error distribution in the sequenced clones

Plant species	Nucleotides sequenced	Error rate	Clones with errors/total clones	Errors in first 2 repeats/total errors	G-slippage	T-slippage	Misincorporation
Arabidopsis	2762	$1.8X10^{-3}$	4/25 (16%)	4/5	0	5	0
Maize	2736	$5.1X10^{-3}$	11/24 (45.8%)	11/14	10	1	3
Silene	1102	$1.0X10^{-2}$	9/17 (52.9%)	6/11	0	11	0
Sorghum	1586	$1.2X10^{-2}$	12/26 (46.2%)	12/19	3	9	7
Soybean*	641	$1.6X10^{-3}$	1/10 (10%)	1/1	1	0	0

*Soybean telomerase extended only a subset of primers.

positioning at an alternative site in the template. However, since the N_{20}-G primer, which can form a maximum of two Watson-Crick base pairs (AG) with the template, was properly positioned in 22/23 TRAP clones we sequenced (Table 5), the data strongly argue that addition of an extra dG residue to the AG_3T_3-G_3 primer represents a G-slippage event.

T-slippage (addition or omission of a dT residue) was also observed with all the telomerase reactions except those for soybean. The most common error overall, T-slippage accounted for 52% of the mistakes (Table 9). Although T-slippage events occurred throughout the newly synthesized sequence, this error was strongly biased towards the first two repeats added (Tables 4-8). We counted the rare addition or omission of an A residue in sorghum clones #9 (Table 7) and #88 (Table 4) as misincorporation events rather than slippage. Other examples of misincorporation of dA (Table 5, sorghum clones #50 and #6; Table 6, maize clone #10; Table 7, sorghum clones #30, 35 and 12) or dC (Table 7, maize clone 8) were observed. In maize clone #8 (Table 7) the extra dT and dC were counted as individual mistakes.

Approximately 10% of the soybean TRAP clones and 16% of the *Arabidopsis* clones harbored at least one imperfect telomeric repeat (Table 9, also see Tables 4-8 for sequences of each clone). In contrast, from 45% to 52% of the clones from *Silene,* maize and sorghum reactions contained a deviant telomeric repeat. For maize, the majority (10/14) of such clones harbored G-slippage errors in the first two repeats, while the errors generated by the *Silene* and sorghum telomerases were primarily T-slippage events later in the elongation reaction. Remarkably, the overall error rate for *Silene* and sorghum telomerases 1.0×10^{-2} and 1.2×10^{-2}, respectively) was over six times greater than that of

Arabidopsis and soybean (1.8×10^{-3} and 1.6×10^{-3}, respectively), and at least twice that for maize telomerase (5.1×10^{-3}). Pairwise chi-square calculations on the error rates of elongation (ignoring errors in the first two error-prone repeats) demonstrated that *Silene* and sorghum telomerases both had significantly higher error rates than either *Arabidopsis* or maize telomerases ($P<0.05$). In contrast, the differences in error rates between telomerases in the same class (maize vs. *Arabidopsis* and *Silene* vs. sorghum) were not significant, providing further support for distinct classes of plant telomerases.

DISCUSSION

Although the primary function of telomerase is to maintain pre-existing tracts of telomeric DNA, the enzyme can efficiently add telomeric repeats onto non-telomeric DNA *in vitro* or onto broken chromosome ends *in vivo* (Melek and Shippen, 1996). *In situ* hybridization studies demonstrated that new telomeres are added onto broken chromosome ends in plants, although for at least wheat and maize, breakage-fusion-bridge cycles occur prior to telomere addition (McClintock, 1941; Tsujimoto et al., 1999). Intriguingly, the newly synthesized wheat telomeres contain an increased number of aberrant repeats (Tsujimoto et al., 1999), suggesting that telomerase fidelity is compromised during *de novo* telomere formation.

In this study we investigated the mechanism of new telomere formation in five different monocot and dicot species, focusing our efforts on understanding how non-telomeric DNA interacts with the telomerase active site. Three different mechanisms can be envisioned for positioning non-telomeric DNA 3' ends on the telomerase RNA

template: limited Watson-Crick base pairing; “default” or preferred positioning mediated by specific protein-RNA interactions in the telomerase active site; and random positioning. Here we show that all three mechanisms are utilized by plant telomerases.

Class I telomerases

The soybean and carrot enzymes typify Class I telomerases that exhibit a stringent requirement for base pairing (this study; (Fitzgerald et al., 1996). These enzymes position a primer 3' terminus on the RNA template strictly by Watson-Crick base pairing interactions; primers that lack sufficient 3’ terminal telomeric nucleotides are not extended. Substrates bearing a few residues of complementarity to the template may also be required for the wheat telomerase as telomeres are invariably formed on broken chromosome ends containing two to six nucleotides of telomeric sequence (Tsujimoto et al., 1999). Likewise, analysis of chromosome healing events in mammals, *Ascaris lumbricoides* and *Plasmodium falciparum* suggest that short stretches of telomeric DNA are necessary for efficient new telomere formation (Melek and Shippen, 1996; Bottius et al., 1998; Magnenat et al., 1999; Sprung et al., 1999).

Class II telomerases

We defined the *Arabidopsis*, maize and cauliflower telomerases as Class II enzymes. These telomerases efficiently elongate non-telomeric DNA 3’ ends by positioning the DNA at a preferred site in the RNA template (this study; (Fitzgerald et al., 1996). For *Arabidopsis*, this site is rC_8, although rU_9 also can be utilized at a lower efficiency (Figure 34A). The preferred alignment position for maize telomerase is offset

by one nucleotide from *Arabidopsis*, where rU_9 is greatly preferred, although rA_{10} and rC_8 can be used (Figure 34A). While maize telomerase showed a clear preference for this region of the RNA template, non-telomeric DNA primer positioning on maize template is considerably less stable than in *Arabidopsis*. Nine of the ten maize TRAP clones obtained with the N_{20}-C and AGT-GGG primers initiated in G-slippage events compared to none of the *Arabidopsis* clones (Tables 7 and 4).

Preferred positioning sites for non-telomeric DNA have been defined for ciliate and yeast telomerases (Kramer and Haber, 1993; Melek and Shippen, 1996; Wang and Blackburn, 1997; Wang et al., 1998). *Plasmodium* telomerase also exhibits default positioning of non-telomeric 3' ends, when those ends are adjacent to a stretch of telomeric sequence (Bottius et al., 1998). The advantage of such a precise mechanism for active site placement in telomerases from most lower eukaryotes and plants is unclear. In the ciliate *Euplotes crassus*, default positioning of non-telomeric DNA is confined to cells undergoing development; cells in the vegetative stage of the life-cycle require primers containing terminal telomeric sequence (Bednenko et al., 1997). We detected no difference in primer elongation by maize embryo and endosperm extracts (Grace, Fitzgerald and Shippen, unpublished data), indicating that the higher level of telomerase activity associated with maize embryos (Killan et al., 1998); Grace, Fitzgerald and Shippen, unpublished data) accounts for the increased chromosome healing capacity of this tissue (McClintock, 1941).

Class III telomerases

Enzymes from *Silene* and sorghum constituted a third class of enzymes. These telomerases annealed primers to a variety of non-adjacent sites on the RNA template, incorporating a medley of different sequence permutations (Tables 7 and 8). To the best of our knowledge, the *Silene* and sorghum enzymes are the first examples of telomerases that randomly position non-telomeric DNA primers in the active site. We postulate that this property contributes to the higher overall error rate for these enzymes (Table 9).

It is intriguing that both monocot and dicot species are represented in each telomerase category. The remarkable degree of diversity not only suggests that the active site of telomerase continued to evolve after monocots and dicots diverged, but also that the architecture of *Silene* and sorghum enzyme active sites may be distinct from other known telomerases.

Telomerase fidelity during de novo telomere formation

Our analysis of TRAP product clones revealed that a variety of aberrant telomere repeats were incorporated during elongation of non-telomeric primers. Unfortunately, the paucity of sequence information for plant telomeres precludes an accurate estimate of telomerase fidelity *in vivo* during telomere maintenance, much less chromosome healing. Nevertheless, three lines of evidence argue that the errors we observe reflect the action of distinct plant telomerases. First, the overall rate for sorghum telomerase, the most error-prone telomerase in our study, is 1.2×10^{-2}, over two to four orders of magnitude higher than the error-rate reported for Taq polymerase (Cline et al., 1996; Keohavong and Thilly, 1989; Eckert and Kunkel, 1990). In addition, we measured the error rate of Taq

polymerase in synthesizing telomeric repeats by performing a PCR amplification reaction with *Arabidopsis* clone #5 (see Table 7). Seven random clones containing 34 telomere repeats each (1666 total nucleotides) were sequenced, and none of them contained an error (data not shown). Since the number of nucleotides sequenced in these control experiments exceeded the entire data sets for sorghum and *Silene* telomerases, we conclude that mistakes introduced by Taq polymerase during the PCR amplification step of TRAP have no significant impact on telomerase error rates in our study.

Second, the majority of errors were clustered in the first two repeats added to the primer. Mistakes introduced by Taq polymerase are expected to occur equally through the amplified sequence. It should also be noted that our TRAP reaction conditions favor products with perfect telomere repeat arrays; mismatches between the newly synthesized telomeric repeats and the reverse primer would not be amplified (Fitzgerald et al., 1996; Fitzgerald and Shippen, unpublished results). Thus, our data set may under-represent the number of errors introduced.

Third, statistically significantly more errant repeats were detected in clones from sorghum and *Silene* than in *Arabidopsis* and maize (Table 9). Although each extract was prepared in the same manner, we cannot rule out the possibility that contaminants in our crude extracts influence enzyme fidelity. Nevertheless, we believe it is noteworthy that the range of errors we detected for plant telomerases correspond to other reverse transcriptases. For example, the error rate of *Arabidopsis*, soybean and maize telomerases is comparable to that reported for human telomerase ($2x10^{-3}$) (Kreiter et al., 1995) and AMV reverse transcriptase ($2.8x10^{-3}$) (Kim et al., 1999a) whereas the higher error rate for *Silene* and sorghum telomerases resembles HIV-1 reverse transcriptase

($2.5x10^{-2}$) (Kim et al., 1999a). Fourth, the error rates we observed are consistent with the telomerase categories we designated. The overall error rate for class III (*Silene* and sorghum) telomerases was approximately 5 to 7 times higher than that for class II (*Arabidopsis* and maize*)* and class I (soybean) enzymes. Together, these results argue that plant telomerases have inherently different levels of enzyme fidelity.

The majority of mistakes we detected (34/50) were clustered in the first two repeats added to the primer, implying that enzyme interactions with non-telomeric DNA are inherently less stable than interactions with tracts of telomeric DNA. By far, the most prevalent errors corresponded to the addition of extra dG or dT residues, reflecting a phenomenon called template slippage (Kunkel and Soni, 1988). Also known as pseudo-templated polymerization, template slippage occurs on sequences characterized by short (1-5 nucleotide) repeats. Template slippage is well-documented for ciliate telomerases (Gilley et al., 1995; Yu et al., 1990) and is postulated to be the primary mechanism for generating irregular telomere repeats in *Saccharomyces* (Cohn and Blackburn, 1995). In *Euplotes*, reiterative dG addition is confined to elongation of non-telomeric 3' ends (Bednenko et al., 1998). Lateral instability of the primer-template interaction is thought to cause unpairing and reassociation of the DNA and RNA strands, in some cases leading to strand mispairing and re-exposure of the rC residues for reiterative copying (Bednenko et al., 1998). Since none of the primers in our study (including the AG_3T_3-G_3 primer) perfectly mimic telomeric DNA tracts *in vivo*, it is not surprising that a significant number of the TRAP products (28%) initiated by reiterative dG synthesis. Maize telomerase provided the most striking example of dG slippage with more than 38% of the clones harboring this error.

The most common error overall was T-slippage, resulting in the formation of TTAGGG and TTTTAGGG repeats. Interestingly, for most telomerases, we observed dramatically increased flexiblity in positioning N_{20}-T relative to N_{20}-A, with the terminal T in N_{20}-T participating in numerous non-Watson-Crick interactions (Tables 6 and 7). This observation indicates that primer T-template A base pairing is inherently less stable than primer A-template U base pairing. Recent studies in wheat suggest that T-slippage occurs at a high frequency *in vivo* at chromosome break sites with 21.3% of the telomeric repeats added comprised of the sequence TTAGGG (Tsujimoto et al., 1999). Thus, the high occurrence of T-slippage we detected *in vitro* may reflect a biologically relevant phenomenon.

In all three classes of enzymes, there was a correlation between the stringency of the initial interaction with the primer and the error rate for extension past the first two repeats. Class I telomerase from soybean had the most exact requirement for primer-template interaction. Although this stringency reduced the sample size for soybean telomerase, we found no errors for this enzyme past the first repeat. Class II telomerases from *Arabidopsis* and maize annealed primers to default sites and had modest elongation error rates. Class III telomerases from *Silene* and sorghum bound primers at many positions along the template and also had the highest elongation error rates. This correlation implies that there is a trade-off between the ability to add telomeric DNA to random ends resulting from chromosome breakage and the fidelity of this added sequence. The three classes of telomerase do not correlate with phylogenetic relationships, suggesting that the active site of telomerase is able to evolve rapidly. The rapid evolution of telomerase is evident in analysis of the telomerase RNA subunit.

Telomerase RNA sequences are exceptionally divergent and only recently has a model for vertebrate telomerase RNA secondary structure been proposed (Chen et al., 2000). Increased chromosome breakage would favor plants with the ability to heal many of these breaks, even at the expense of increased errors, whereas relatively stable genomes may favor plants with more faithful telomerases. Studies of a greater variety of both plant and animal telomerases are needed to fully define the mechanisms of this unusual reverse transcriptase.

MATERIALS AND METHODS

Preparation of plant extracts

Arabidopsis thaliana (Columbia) seedlings were grown for seven days in MS medium prior to preparation of extracts. *Silene latifolia* seedlings were grown in water for three days. Soybean (*Glycine max*), sorghum (*Sorghum bicolor*-cultivar Btx623) seeds were sterilized in 20% bleach for 10 min, rinsed, and imbibed in water for 1 hr. Soybean seeds were sandwiched between paper towels wetted in 100 μM $CaCl_2$ and grown in the dark for 3 days at 29°C. The terminal 15 mm of hypocotyl root tips were cut off and used for extract preparation. Sorghum seeds were sandwiched between wetted paper towels and grown in the dark at room temperature for three days. The epicotyls were then excised and extracted. Extracts from maize (*Zea mays L.*) were prepared from the developing kernals of "Hi II" variety. Kernels were the F2 cross from F1 plants of inbred parents A and B. Developing ears were harvested 20 days after self-pollination. Crude extracts containing telomerase activity were prepared from the

excised plant tissues as previously described (Fitzgerald et al., 1996). Protein concentrations in extracts ranged from 0.25 to 2.0 mg/ml.

TRAP assays

Telomerase was detected by a modified version of the telomere repeat amplification protocol (TRAP) (Fitzgerald et al., 1996). Reactions with *Arabidopsis* were performed with 1.0 μg of protein from seven day seedlings, soybean reactions with 1 μg of protein from six-day hypocotyl, *Silene* reactions with 0.5 μg of protein three-day-old seedlings and maize reactions with 25 ng of protein from 20 day embryo extracts. Primers were obtained from Gibco-BRL and gel purified before use. Following TRAP, samples were phenol extracted, ethanol precipitated, and resolved on 12% sequencing gels which were dried and subjected to autoradiography. Exposures were typically for overnight.

In each set of TRAP assays, reactions were performed with one of the following forward primers: AG_3T_3-G_3 (AGGGTTTAACTACGCGATGGG), N_{20}-G, N_{20}-A, N_{20}-T or N_{20}-C. N represents the sequence CACTATCGACTACGCGATCA and the 3' terminal nucleotide is either G, T, A or C. For the primers N_{19}-G_2 and N_{18}-G_3, the 3' terminal residues of N_{20} were replaced by two or three G residues, respectively. In every assay, the reverse primer had the sequence $(C_3TA_3)_3$.

Sequencing TRAP products

TRAP assays were performed as described (Fitzgerald et al., 1996) except that the final 72°C extension time was lengthened to 15 min. Following the PCR step, 2 µl from each reaction were used in ligation reactions for insertion into the TOPO TA pCR2.1 cloning vector (Invitrogen). The plasmids were transformed into TOP10F' cells and blue-white screening was used to identify transformants. Those plasmids with 40-300 bp insertions were transformed into *E. coli* DH5□ cells and then isolated with Qiaprep Spin Miniprep Kits (Qiagen). The insertions were sequenced from the M13 forward and reverse primers using ABI Big-Dye Terminators in the presence of 14% DMSO. Five random TRAP clones were sequenced from each extension reaction except in the case of the maize reaction with N_{20}-G (four clones) and the sorghum reaction with N_{20}-T (six clones). Only two to four positive clones per primer were recovered from reactions with *Silene* telomerase. Clones having less than six repeats were not included in the study, since the final three repeats correspond to the reverse primer. Accordingly, the last three repeats were removed from all sequences presented here. The Chi-Square test was applied in pairwise comparisons to determine if any significant difference was found in error rates among the various telomerases. A p-value ≤ 0.05 was considered significant.

VITA

YEVGENIY SHAKIROV

Permanent address:
67 K.Marx Str, Apt.36, Zelenodolsk, Tatarstan 422520, Russia

Education

Ph.D., Biochemistry, 1997-2004, Texas A&M University – College Station, Department of Biochemistry and Biophysics

Ph.D., Microbiology, 1997-2000, Kazan State University, Department of Microbiology, Kazan, Russia

B.S., Microbiology, 1992-1997, Kazan State University, Department of Microbiology, Kazan, Russia

Honors and awards:

Russian Presidential Scholarship Award to study abroad, 1996

George Soros Foundation Student Award, 1995; 1996

Publications:

Sedivy, J.M., Shippen, D.E., Shakirov, E.V. (2003). Surprise ending. Nature Genetics **33**, 114-6.

Fitzgerald, M.S., Shakirov, E.V., Hood, E.E., McKnight, T.D., Shippen, D.E. (2001). Different modes of *de novo* telomere formation by plant telomerases. Plant J **26**, 77-87.