INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9416002

Telomere protein and telomere protein homolog in *Euplotes crassus*

Wang, Wenlan, Ph.D.

The University of Nebraska - Lincoln, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

TELOMERE PROTEIN AND TELOMERE PROTEIN HOMOLOG IN *EUPLOTES CRASSUS*

by

Wenlan Wang

A DISSERTATION

Presented to the faculty of

The Graduate College in the University of Nebraska

In Partial Fulfillment of Requirement

For the Degree of Doctor of Philosophy

Major: Biochemistry

Under the Supervision of Professor Carolyn M. Price

Lincoln, Nebraska

December, 1993

DISSERTATION TITLE

Telomere Protein And Telomere Protein Homolog

In Euplotes Crassus

BY

Wenlan Wang

SUPERVISORY COMMITTEE:

	APPROVED	DATE
Signature	Carolyn Price	11/29/93
Typed Name	Carolyn M Price	
Signature	N. K. Gupta	11/29/93
Typed Name	N. K. Gupta	
Signature	Eric Davies	11/29/93
Typed Name	Eric Davies	
Signature	T. E. Elthon	11/29/93
Typed Name	T. E. Elthon	
Signature	W. H. Braunlin	11/29/93
Typed Name	W. H. Braunlin	
Signature		
Typed Name		

GRADUATE COLLEGE
UNIVERSITY OF NEBRASKA

TELOMERE PROTEIN AND TELOMERE PROTEIN HOMOLOG IN *EUPLOTES CRASSUS*

Wenlan Wang, Ph.D.

University of Nebraska, 1993

Advisor: Carolyn M. Price

Telomeres, the nucleoprotein complexes at the end of the chromosomes, consist of repeated DNA sequence and the associated proteins. While the telomeric DNA has been well characterized in many species, the studies of the telomere-associated proteins lag far behind.

Hypotrichous ciliates such as Euplotes and Oxytricha are particularly well suited for studying telomeres and telomere-binding proteins because they have enormous amounts of telomeres. Both Euplotes and Oxytricha telomere-binding proteins have been isolated and characterized. It appears that the two proteins comprise a very unusual category of DNA-binding proteins because they specifically bind to the 3' terminus of the telomeric DNA.

In order to learn more about the DNA-binding motif(s) of the Euplotes telomere-binding proteins, we isolated the gene encoding the Euplotes telomere-binding protein. We have shown that the gene encoding the Euplotes telomere-binding protein is highly homologous to the gene encoding the Oxytricha telomere-binding protein α subunit in

various regions. As these regions lie within the DNA-binding domain, they might make up the DNA-binding site(s).

While we were cloning the gene encoding the Euplotes telomere-binding protein, we also cloned a gene encoding a related protein - the telomere protein homolog. We have shown that the telomere protein homolog shares extensive amino acid sequence identity with the Euplotes and Oxytricha telomere-binding protein in their N-terminal DNA-binding domain. In the C-terminal region, however, the telomere protein homolog exhibits little sequence homology with Euplotes telomere-binding protein. To identify the native telomere protein homolog, we expressed the N-terminal and the C-terminal domains of the homolog protein and made antibodies against the expressed proteins. Using these antibodies we detected the homolog protein in the regions of the macronucleus where DNA replication is taking place. This subcellular location of the homolog protein indicates that the homolog protein might be involved in telomere replication.

My inspiration to continually pursue scientific answers comes from people suffering physically and emotionally from incurable cancers. I would like to encourage those people to have faith and know that I have committed my future research towards finding a cure for them.

To Mom Fengzhi, Dad Jiaquan, Qiuwei and Mary

ACKNOWLEDGEMENT

My experience at UNL was thoroughly enjoyable and will not be forgotten. The scientific knowledge and the English language I have learned will always be treasured. Many people have made this possible.

I am very indebted to my advisor, Dr. Carolyn M Price, for her guidance and support. Deep appreciation goes to my supervisory committee members: Dr. William H Braunlin, Dr. Naba K Gupta, Dr. Thomas E Elthon and Dr. Eric Davies for their time and help.

Thanks to Dr. Joris Vermeesch, a first born doctor from the lab, for many useful suggestions and enlightening discussions.

Rose Skopp for synthesizing oligonucleotides and helping with sequencing.

Marilynn Larson, Scott Perez and Devlina Lahiri for sharing bench space and their friendship. Also, the rest of the lab members; Jasmin Mayhew, Brent Premer, SuzAnne Emme, Nancy Rohe, Julie Bartos, Margret Scofield, Hee-jung Kim, Jongjin Lee, William Dewright, Angela Adams, Jacki Travers and Nathan Venkatesan. I will never forget the fun we had together.

Lixing Zhu and Lei Xiao, the two best friends back in China, for their encouragement and friendship.

Last but not the least, a special thanks goes to my husband, Qiuwei Xu, and my daughter, Mary Xu for their patience and encouragement.

TABLE OF CONTENTS

I. INTRODUCTION

Telomeres, the physical ends of eucaryotic chromosomes, are essential to cell growth and development (reviewed by Blackburn, 1991; Zakian, 1989; Biessmann and Mason, 1992). Since the pioneering work of Muller and McClintock (Muller, 1940; McClintock, 1941), it has been known that telomeres consist of tandemly repetitive DNA sequence and the associated protein(s). Telomeres exist to solve at least two problems: allowing complete replication of chromosomal ends and preventing chromosomes from end-to-end fusion. While the function of telomeres is well established, the detailed role of each telomere component still remains to be determined. Moreover, it is still poorly understood how the interaction of telomeres with each other and with other nuclear components determines their cellular functions.

A. FUNCTIONS OF TELOMERES

1. Maintenance of chromosome integrity

About fifty years ago, telomeres were implicated in the maintenance of chromosome integrity (Muller, 1940; McClintock, 1941). Now it is pretty clear that telomeres maintain chromosome integrity in at least two ways: preventing chromosomal end ligation and mediating complete replication of chromosomes.

Telomeres confer stability to chromosomes by preventing end-to-end fusion and exonucleolytic degradation. The importance of this property of telomeres has been well

documented as a result of studies of broken chromosomal termini. In early 1940, genetic and cytological evidence indicated that broken chromosome ends are very unstable and tend to undergo end-to-end fusions which lead to the formation of dicentric and ring chromosomes (Muller 1940; McClintock 1941). Studies of broken chromosomes from a variety of other species revealed that broken ends are not only subject to end-to-end fusion, but also easily degraded by nucleases. In addition, they tend to recombine with other portions of chromosomes (Aledo et al., 1989; Holzmann et al.; 1993; Schwartz et al.; 1989; Walker et al., 1991; Blackburn 1984). This abnormal behavior of broken chromosomes strongly suggested that the ends of intact chromosomes exist as specialized structures which serve to maintain chromosome stability.

Another important function of telomeres is to facilitate the complete replication of chromosomes (reviewed in Blackburn, 1991; Biessmann and Mason, 1992). DNA replication is achieved by DNA polymerases which require an RNA primer to initiate synthesis and can only synthesize DNA in the 5' to 3' direction (Watson 1972). These intrinsic properties of DNA polymerase pose a problem during replication of the 5' terminus of a linear DNA molecule. Subsequent removal of the RNA primer from the extreme end of the lagging strand will result in an unreplicated region at the 5' end of each daughter strand. In the absence of a mechanism to circumvent this problem, successive rounds of DNA replication would lead to progressive loss of the terminal sequences. Telomeres provide a mechanism to solve this replication problem.

2. Nuclear organization and gene expression

In addition to maintaining chromosome integrity, it now appears that telomeres are also involved in establishing and maintaining the three dimensional structure of the nucleus. This is most probably achieved through telomere-telomere association and/or telomere interactions with other nuclear components such as the nuclear matrix (reviewed by Gilson et al., 1993). For instance, during interphase in many types of cells, the chromosomes are arranged so that the centromere and telomeres are oriented towards the opposite poles of the nucleus (the Rabl orientation) (Gasser and Laemmli, 1987). Yet cells in meiotic prophase, often have the telomeres clustered together near the nuclear envelope, in a "bouquet" arrangement (Lima-de-Faria, 1983). Although the significance of these associations remains unknown, it has been proposed that interactions between telomeres or between telomeres and other nuclear components may affect gene expression or chromosome behavior (Zakian, 1989). In fact it has been clearly demonstrated that telomeres do affect the transcription of telomere proximal genes in Drosophila (Levis 1985), yeast (Gottschling et al, 1990) and Trypanosomes (reviewed by Pays and Steinert, 1988). In yeast and Drosophila, telomeres behave like other regions of heterochromatin and cause a decrease in the level of expression of genes which are positioned in their proximity (a phenomenon called position effect variegation). In contrast, transcription of genes encoding variant surface antigens in Trypanosoma becomes active only when they are placed near a telomere.

B. STRUCTURE OF TELOMERES

1. The DNA component

Telomeres are nucleoprotein complexes consisting of repeated DNA sequence and telomere-binding proteins. Both the telomeric DNA and telomere proteins are responsible for the ability of telomeres to maintain chromosome integrity.

Telomeric DNA is mostly composed of tandemly repeated sequences (reviewed in Zakian, 1989; Biessmann and Mason, 1992). Although the overall length of telomeric DNA varies widely (from 20 bp in Oxytricha to 150 kb in mouse), the unit length of each telomeric repeat is generally in the range of 6-10 bp. The sequence of the repeats is fairly conserved among species. In most organisms examined, the nucleotide composition of the two strands of telomeres is biased with one strand having a G-rich sequence which runs toward the 3' end of the chromosome. For instance, in Oxytricha and Euplotes the sequence of the telomeric DNA is (TTTTGGGG)n; Tetrahymena chromosomes terminate with the sequence (TTGGGG)n; plants have telomeric repeats of $(TTTAGGG)_n$; while in Trypanosomes and mammals telomeres consist of (TTAGGG)n repeats. In several well studied organisms the G-rich strand protrudes beyond the complementary C-rich strand to form a 12-16 base single-stranded 3' overhang (Klobutcher, et al., 1981; Henderson and Blackburn, 1989). This 3' overhang is believed to be a common feature of all telomeres (Henderson and Blackburn, 1989).

In vitro, synthetic oligodeoxynucleotides with the G-strand sequence can form G-quartet structures. The G-quartets are four-stranded structures that are formed by inter-

and intrastrand interactions involving Hoogsteen base pairing (Cech, 1988; Henderson et al., 1987; Williamson et al.; 1989). However, this G-quartet structure is not recognized by two known telomere-associated proteins, the Oxytricha telomerase and the Oxytricha telomere-binding protein (Raghuraman and Cech, 1990; Zahler et al., 1991). Thus, it is not known whether the G-quartet structure exists in vivo.

The telomeres of many higher eukaryotes have additional common elements proximal to the telomeric repeats. These elements comprise the subtelomeric region or the telomere associated sequences (reviewed in Zakian, 1989; Biessmann and Mason, 1992). The subtelomeric regions are usually quite long, although the absolute size varies from chromosome to chromosome. For example, in Plasmodium the subtelomeric region is ~100 kb long (Foote and Kemp, 1989) while in humans it is ~2600 kb long (Ellis and Goodfellow, 1989). Unlike the telomeric repeats, the sequences of the subtelomeric regions are not conserved. To date two major types of telomere associated sequences have been described in organisms such as Drosophila and humans: one of these consists of short, tandemly repeated sequences characteristic of satellite-DNA; the other consists of more complex, moderately repetitive sequences. It appears that the subtelomeric region is not essential to cells, although it may serve ancillary functions such as allowing frequent recombination or acting as a "buffer" zone between the chromosome end and the most distal gene (see review in Biessmann and Mason, 1992).

2. Telomere-binding proteins

The proteins preferentially bound to telomeres have been isolated from a number of organisms, such as Xenopus (Cardenas et al., 1993), Physarum (Coren et al., 1991), yeast (Shore and Nasmyth, 1987; Buchman et al., 1988; Longtine et al., 1989; Liu and Tye, 1991), Oxytricha (Price and Cech, 1987; 1989) and Euplotes (Price, 1990). So far all these telomere-binding proteins can be classified into two groups based on their binding patterns to the telomeres: those binding to internal stretches of telomeric DNA and those binding to the extreme ends of telomeric repeats (Price, 1992a).

A number of telomere proteins have been isolated that bind to internal telomeric DNA. Basically all these telomere proteins were identified by their in vitro binding activity to telomeric repeats. For example, a 50 kD protein, called TTAGGG repeat factor (TRF), present in nuclear extracts of human, mouse and monkey cells, shows specific binding activity to TTAGGG repeats in vitro (Zhong et al., 1992). Xenopus telomere end factor (XTEF) specifically recognizes the vertebrate telomeric repeat sequence TTAGGG when this sequence presented as a 3'-single stranded extension adjacent to double stranded DNA (Cardenas et al, 1993). PPT (Physarum polycephalum telomere-binding protein) is thought to coat the entire double-stranded (T_2AG_3) telomeric DNA (Coren et al., 1991), while yeast TBP α (telomere binding factor α) may bind at the junction between the $G_{1\text{-}3}T$ repeats and the subtelomeric X sequence (Liu and Tye, 1991). RAP1 is a multifunctional protein which recognizes a consensus sequence that occurs in upstream activator, silencer and telomeric DNA (Conrad et al., 1990; Lustig et al., 1990; Sussel and Shore, 1991). Abnormal expression of RAP1 can cause chromosome instability and deregulation of

telomere length (Conrad et al., 1990; Lustig et al., 1990). It appears that RAP1 performs its multiple roles as a result of interaction with different cellular proteins via its carboxyl terminus (Hardy et al., 1992; Kyrion et al., 1992; 1993).

Telomere proteins which bind to the extreme ends of chromosomes have been isolated from Oxytricha and Euplotes. The Oxytricha telomere protein is a 97 kD heterodimer consisting of 56 kD α subunit and 41 kD β subunit (Price and Cech, 1987; 1989). Studies of reconstituted Oxytricha telomeric complex revealed that the α subunit is the dominant DNA-binding moiety which determines the sequence specificity of the interaction. However, the β subunit is also required in order to achieve the full in vivo DNA-binding activity (Gray et al., 1991). The Euplotes telomere protein has been isolated as a 51 kD monomer (Price, 1990). This monomer resembles the Oxytricha telomere protein in that it binds telomeric DNA in a specific and salt-stable manner. The DNA-binding domain of Euplotes telomere protein has been mapped to an N-terminal 35 kD region of the protein (Price et al., 1992b). This region shares extensive amino acid sequence identity with the N-terminus of the a subunit of the Oxytricha protein (Wang et al., 1992).

C. TELOMERE LENGTH REGULATION

The length of telomeric DNA varies considerably between species. However, in a given organism the size of the telomeric region is fairly constant with the absolute length varying about a characteristic mean due to a dynamic equilibrium between

lengthening and shortening. Telomere length appears to be determined by physiological and genetic factors (Walmsley and Petes, 1985; Larson et al., 1987). For example, in continuous log-phase cultures of Trypanosomes, the telomeres grow at a rate of 6-10 bp per generation. Ultimately this growth results in a increase in both overall telomere length and length heterogeneity in the population (Bernards, 1983; Van der Ploeg, 1984). In vegetatively dividing Tetrahymena cells, the telomere length increases by 3-10 bp per generation during the first 300 log-phase divisions until a final maximum length is attained. This maximum length varies from strain to strain. Eventually subpopulations of Tetrahymena with shorter telomeres became dominant in the culture, this is probably because the cells with short telomeres have a slight growth advantage (Larson et al., 1987). In maize it appears that genetic factors play a dominant role in affecting average telomere length (Burr et al., 1992). It has been found that a remarkable telomeric length polymorphism exists between maize lines; within a given strain the number of telomeric repeats is stable, and no phenotypic polymorphisms have been observed.

Maintenance of telomere length appears to be important for proper cell growth and development. There is evidence suggesting that the length of the total telomeric repeats is a determinant of chromosome stability. For example, in yeast cells with the est1 mutation, it has been observed that a progressive decrease in telomere length is associated with cellular senescence. This senescence is apparent as a gradual reduction in cell viability and growth rate, and with a increase in cell death (Lundblad and Szostak, 1989). Senescent cultures of fibroblasts also display telomere shortening and often have a high

number of fused chromosomes (Harley et al., 1990; Benn, 1976; Sherwood et al., 1988). Moreover, it appears that the telomere shortening observed in some tumors may be related to the increased levels of telomere-telomere fusion (Holzmann et al., 1993; Smith and Yen, 1992; Adamson et al., 1992). Altered telomere length may also impair specific interactions between telomeric DNA and telomere proteins. In fact, increases in telomere length caused by overexpression of the telomere-binding protein RAP1 correlate with an increasing rate of chromosome loss and cell death (Conrad et al., 1990).

D. TELOMERE MAINTENANCE

Although the mechanisms of telomere length regulation are still not fully understood, it is apparent that telomere length is determined by a balance between activities that lengthen and shorten the telomeric DNA. The main shortening activity identified to date is incomplete DNA replication of the 5' end of the chromosome. The main lengthening activities are telomerase and recombination.

1. Incomplete replication

Strong evidence for the link between DNA replication and the loss of the most terminal 5' sequence has come from studies of Drosophila chromosomes that lack telomeres. In this type of mutant, it has been observed that chromosomes decrease in length at a rate of ~2 bp per generation (Biessmann et al., 1990). The same rate of reduction of ~ 2 bp per generation was also seen in the yeast tel1 mutant, a mutant thought to lack one of the replication elongation factors (Lustig and Petes, 1986). Further evidence

for the link between DNA replication and telomere length comes from studies of cdc17 mutant, there temperature sensitive yeast cells encode a deficient catalytic subunit of DNA polymerase I (Carson and Hartwell, 1985). In cdc17 cells, an increase in growth temperature from permissive to sublethal or the visa versa, resulted in an increase/decrease in telomere length over many generations. If a wild type CDC17 gene is back crossed into the mutant cells to rescue the defective gene, the long telomeres which are produced at elevated temperature in cdc17 mutants, will become shorter by 2 bp per generation. Given the biochemical properties of conventional DNA polymerases, removal of the last primer from the lagging strand will yield an eight-nucleotide gap in Drosophila (Kitani et al, 1984) and an eleven-nucleotide gap in S. cerevisiae (Singh and Dumas, 1984). If the telomere elongation mechanism is fully inhibited, the next two rounds of replication are expected to produce one strand of original length, two with an eight-nucleotide gap at the 5' terminus and one with a blunt-ended shortened by 8 bp in Drosophila and 11 bp in S. cerevisiae. Thus the average shortening of a linear molecule caused by removal of an 8-11 bp primer would be 2-2.7 bp per round of replication (Biessmann et al, 1990c). This calculated data agrees very well with the observed values from mutant studies (Biessmann and Mason, 1992).

2. Telomere elongation mechanisms

1). Telomerase

In order to avoid loss of genetic information, telomeres have to employ special mechanisms to counterbalance the continuous loss of DNA sequence that results from

incomplete replication. While many models have been proposed to explain the complete replication of telomeres, there is only experimental evidence for a few of proposed mechanisms. In one mechanism, an enzyme called telomere terminal transferase (or telomerase) elongates telomeres by adding species-specific telomeric repeats to the 3' end of the G-rich strand of the telomeric DNA (reviewed in Greider, 1990; Blackburn 1990a; 1990b; 1991). Telomerase is a ribonucleoprotein with an unusual reverse transcriptase activity (Greider and Blackburn, 1985; Greider and Blackburn, 1987). This enzyme contains its own internal RNA template to direct DNA synthesis (Greider and Blackburn, 1989; Shippen-Lentz and Blackburn, 1989). It appears that the telomerase is widespread among species as its activity has been detected in Tetrahymena thermophila, Euplotes crassus and Oxytricha nova as well as in human cell cultures (Greider and Blackburn, 1985; Shippen-Lentz and Blackburn, 1990; Zahler and Prescott, 1988; Morin, 1989; Counter et al., 1992). Although there is little similarity in the primary sequence of the telomerase RNA from different organisms, phylogenetic comparison of the telomerase RNAs from seven species of Tetrahymena revealed that the predicted secondary structure of the RNA is highly conserved. Based on this finding, it has been suggested that telomerase RNA may be the catalytically active component of telomerase activity (Romero and Blackburn, 1992). Compared with the RNA component of telomerase, very little information has been obtained about the telomerase protein. It appears that the protein moiety is very labile and hard to purify to homogeneity.

Although synthesis of telomeric repeats with the correct sequence is directed very

accurately by the intrinsic RNA template, the priming specificity of telomerase is relatively imprecise, as the enzyme will use one of a few related telomeric sequences as a primer (Harrington and Greider, 1991; Morin, 1991). This latter property of telomerase may be critical for "healing" of broken chromosomes by *de novo* telomere synthesis. During chromosome healing telomeric DNA is added directly to the non-telomeric sequences present at the end of the broken chromosome. For example, in humans chromosome healing via telomere addition has been clearly demonstrated in a patient with an a thalassemia mutation (Wilkie et al., 1990). In both ciliates and nematodes, *de novo* telomere addition following chromosome breakage and repair occur as a routine event during the life cycle of the organisms (Yu and Blackburn, 1991; Muller et al., 1991). During formation of the somatic nucleus, the large chromosomes from the germline nucleus are fragmented and much of the non-coding DNA is eliminated. The chromosome fragments that will be retained in the newly formed somatic nucleus are stabilized by the addition of telomeres to each end (Blackburn, 1986; Tobler et al., 1992).

2). Recombination

While telomerase is clearly responsible for the synthesis of telomeric DNA, at least in some organisms, recombination also seems to play an important role in telomere elongation (Zakian, 1989). It has been suggested that telomeres recombine very readily with each other because telomeric DNA is composed of repeated sequences. Recombination may occur between the telomeres on different chromosomes or between telomeres and regions of homologous sequences located elsewhere on the chromosome.