Identification and characterization of G4-DNA binding proteins from Tetrahymena thermophila

Ted Patrick Schierer
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Molecular Biology Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/10979

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
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.
Identification and characterization of G4-DNA binding proteins from

*Tetrahymena thermophila*

by

Ted Patrick Schierer

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department: Zoology and Genetics
Interdepartmental Major: Molecular, Cellular, and Developmental Biology

Approved:

In Charge of Major Work

For the Interdepartmental Major

For the Major Department

For the Graduate College

Iowa State University
Ames, Iowa

1995
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 1: A PROTEIN FROM TETRAHYMENA THERMOPHILA THAT SPECIFICALLY BINDS PARALLEL-STRANDED G4-DNA</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER 2: PURIFICATION AND CHARACTERIZATION OF TGP, A PROTEIN FROM TETRAHYMENA THAT BINDS TO PARALLEL-STRANDED G4-DNA</td>
<td>36</td>
</tr>
<tr>
<td>GENERAL CONCLUSIONS</td>
<td>56</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>58</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>62</td>
</tr>
<tr>
<td>APPENDIX A: IDENTIFICATION AND CHARACTERIZATION OF A PUTATIVE TELOMERE END-BINDING PROTEIN FROM TETRAHYMENA THERMOPHILA</td>
<td>63</td>
</tr>
</tbody>
</table>
ABSTRACT

Alternative DNA conformations have been known to exist since before the elucidation of the double helix. The study of alternative DNA conformations formed by telomeric sequences has led to detailed studies of biologically relevant DNA capable of forming unusual structures. In order to understand the biological role of alternative DNA conformations formed by telomeric DNA, proteins which bind these structures have been identified and characterized.

The initial portion of my research involved the identification of a protein from Tetrahymena thermophila which binds preferentially to parallel-stranded G-quartet structures. Competition and binding assays demonstrated that the protein, TGP (Tetrahymena G4 binding protein), binds to parallel G-DNA structures but not to antiparallel G-DNA structures or to non-G-DNA duplexes and single stranded oligonucleotides.

This initial identification and characterization was followed up by attempts to purify TGP. Cation exchange and Y4-affinity chromatography partially purified two proteins (83 and 50 kDa) that coelute with peak TGP activity. UV cross-linking analysis confirmed the involvement of the 83 kDa protein as a main component of the TGP/G4-DNA complex.
INTRODUCTION

Dissertation Organization

The following introduction will provide an overview of unusual DNA structures and the proteins which bind these structures. The overview addresses the overall goal of determining the biological function of unusual DNA structures and G-DNA binding proteins. The main body of the dissertation following the introduction is composed of two papers. The first of these two papers has been published in *Biochemistry* and the second is a manuscript in preparation and discusses my efforts to purify and characterize the protein identified in the first paper. In addition, a paper published in *Molecular and Cellular Biology* in which I conducted initial studies which were followed by Hong Sheng is included in Appendix A. In particular, I designed the DNA probe (one that mimics the natural end of the telomere) that was successful in identifying TEP whereas two previous attempts by others were unsuccessful. The conclusion found after the second paper summarizes the results of my research and relates the results to the overall goal of determining the biological relevance of my research. Research topics worthy of further investigation are also addressed in the conclusion. References for the introduction and conclusion are cited after the conclusion.

I am the first author on both papers included in the main body of the dissertation and I completed all (first paper) or most of the work described. In paper 2, I used a gel filtration and an S-Sepharose column poured by Zhen Hou. In addition, Luming Niu assisted me in the calibration of the gel filtration column and Biorex 70 chromatography.

Unusual DNA structures and G-DNA nomenclature

It has been known for many years that DNA is capable of adopting more than one conformation. Fiber diffraction studies showed that hydrated DNA adopted a different conformation (B-DNA) than dehydrated DNA (A-DNA) even before the double helix model of DNA was understood (Rich, 1993). In subsequent years, it was shown that DNA can adopt a triple helix (Felsenfeld, 1957; Hoogsteen, 1963), a quadruple helix (Gellert et al., 1962; Kang et al., 1992) and a left-handed helix (Wang, 1979) and parallel-stranded duplexes (Rippe et al., 1992). The main structural feature of quadruple helices formed by guanine-rich
sequences, the G-quartet, was first proposed by Gellert et al. (1962). The G-quartet is a cyclic, hydrogen-bonded array (Figure 1) in which each guanine is hydrogen bonded on both the Hoogsteen face and the Watson-Crick face. It was the only model sufficient to explain the remarkable stability of the gels formed by guanylic acid (GMP) (Gellert et al., 1962). It was also predicted that the G-quartets would stack upon one another in DNA structures since their large planar surfaces would result in strong van der Waals attractions.

![Figure 1. a G-quartet](image)

DNA structures containing G-quartets are generally referred to as G-DNA. If the structure contains four independent parallel strands, it is referred to as G4-DNA (Sen & Gilbert, 1990). When referring to the stoichiometry of a G-DNA complex, the terms monomer, dimer and tetramer are used to describe structures mediated by one, two and four molecules respectively. However, tetraplex and quadruplex are terms used to describe the number of bases involved in mediating formation of the structure. Thus, all G-quartet structures can be referred to as
tetraplexes or quadruplexes. For example, the intramolecular structure formed by d(T2G4)4 (Figure 3) is referred to as a monomeric quadruplex.

Telomeric G-DNA structures

Interest in the study of G-quartet structures has increased dramatically in recent years since it was discovered that biologically relevant DNA sequences such as telomeres (Henderson et al., 1987) and immunoglobulin switch regions (Sen & Gilbert, 1988) form unusual DNA structures. Since that time, several different G-rich DNA or RNA sequences having potential biological roles have been tested for their ability to form G-DNA or G-RNA structures. HIV genomic RNA and the fragile X d(CGG)n repeats are two recent examples which will be discussed below.

Structural studies involving telomeric sequences far outnumber studies performed on sequences from non-telomeric locations. Telomeric sequences have important features which are conserved across a wide range of species from humans to single-celled eukaryotes (Figure 2, reviewed by Blackburn & Szostak, 1984). It was later shown that these conserved features, namely repetitive blocks of contiguous guanines separated by blocks of sequence usually rich in thymine, are important for structure formation (Figure 2). Over 30 different telomeric oligonucleotides have now been examined for their structural characteristics. The first study to find that telomeric oligonucleotides form unusual DNA structures was in 1987 (Henderson et al., 1987). Non-denaturing gels were used to show that telomeric oligonucleotides form compact structures which migrate anomalously compared to unstructured oligonucleotides. In addition, NMR evidence suggested that guanine-guanine base-pairs mediated formation of these compact structures.

![Repetitive, G-blocked sequences](AACCCCTGGGGTTGGGGTTGGGG)

Figure 2. Conserved features of telomeric sequences that influence structure
The initial study by Henderson et al. (1987) was followed up by dozens of subsequent structural studies of different telomeric oligonucleotides including two X-ray crystal structures (Kang et al., 1992; Laughlan et al., 1994) and several NMR studies (Smith & Feigon, 1992; Smith & Feigon, 1993; Aboul-ela et al., 1992; Wang & Patel, 1992; Wang & Patel, 1993; Gupta et al., 1993. The first follow-up studies used DMS (dimethylsulfate) protection assays to show that the N-7 of each guanine was protected from methylation indicating their involvement in Hoogsteen base-pairs (Williamson et al., 1989; Sundquist & Klug, 1989; Sen & Gilbert, 1990). The finding of Hoogsteen base-pairs between G's led to the proposal of the G-quartet model for telomeric DNA structures in which G-quartets are stacked upon each other and stabilized by monovalent cations coordinated between the G-quartet planes (Figure 3; Williamson et al., 1989). This model was later confirmed by X-ray crystallography (Kang et al., 1992; Laughlan et al., 1994).

![G-tetraplex structural diversity](image)

Subsequent studies showed that important features of telomeric sequences which influence the structure formed include the number of contiguous guanines and the number of thymines. It was determined that G-quartets cannot exist in isolation from other quartets and therefore at least two contiguous guanines are necessary for the formation of G-quartet structures in sequences containing at least two blocks of guanines (Jin et al., 1990). Sequences having only one block of guanines need at least three contiguous G's to form intermolecular G-quartet
structures (Sen & Gilbert, 1992). In addition, the number of T's between the blocks of guanines determines whether the structure will be an antiparallel hairpin dimer or a parallel-stranded tetramer when sequences contain at least 2 blocks of guanines (Balagurumoorthy et al., 1992; Guo et al., 1993). At least two T's are necessary to allow formation of a hairpin foldback dimer in sequences with two telomeric repeats. If only one T is present between the G-blocks, the structure formed is a parallel-stranded tetramer (Balagurumoorthy et al., 1992; Guo et al., 1993).

**Parallel vs Antiparallel Telomeric G-DNA**

All parallel-stranded G-quartet structures studied to date are right-handed helices with entirely anti glycosidic conformations. The NMR solution structure of five different single G-block molecules has been determined. These molecules include dT4G4 (Gupta et al., 1993; *Oxytricha* telomeric repeat), dTG4T (Aboul-ela et al., 1992), dT2G4T (Wang & Patel, 1993), T2G4 (Wang & Patel, 1992; *Tetrahymena* telomeric repeat), and T2AG3 (Wang & Patel, 1992; human telomeric repeat). They differ only in the number of T's flanking the G-block except for the human telomeric repeat which contains an adenine. In each structure, each of the guanines is involved in G-quartets which are stacked upon one another. The thymine residues do not form quartets and only the T layer closest to the G-quartet region is influenced by stacking interactions (Gupta et al., 1993; Wang & Patel, 1993). However, the number of flanking T's does influence the stability of single G-block structures. Increasing the number of T's adjacent to the G-block results in a corresponding decrease in Tm (Guo et al., 1993).

Antiparallel G-DNA structures are characterized by alternating syn and anti glycosidic conformations along each strand in the quadruplex (Kang et al., 1992; Smith and Feigon, 1992). Geometric restraints require two adjacent strands in a quadruplex to have opposite glycosidic conformations if the strands are antiparallel as is the case in a hairpin dimer structure (Williamson, 1993). T's form the loop portion of hairpin dimers and intramolecular quadruplexes (Figure 3). If the number of T's between the G-blocks is less than two, a loop cannot be formed and the molecule will form a parallel-stranded structure (Balagurumoorthy et al., 1992; (Guo et al., 1993).
Direct comparison of the stability of parallel versus antiparallel structures was not possible without taking the different subunit stoichiometries of the complexes into consideration (antiparallel hairpin dimer vs parallel tetramer). To address this problem, Lu et al. (1993) connected two dT2G4 molecules using a 5'-p-5' UrJiage to create a parallel-stranded hairpin dimer to compare directly to the antiparallel hairpin dimer formed by G4T4G4. This study revealed that the parallel-stranded structure is thermodynamically more stable having a Tm almost 20°C higher than the antiparallel structure (Lu et al., 1993). Thus, the parallel-stranded form is the most stable structure for telomeric DNA.

Telomeric C-strand structures

In addition to the G-rich strand of the telomere, cytosine-rich telomeric sequences have recently been shown to adopt quadruplex structures via self recognition (Ahmed & Henderson, 1992; Gehring et al., 1993). Both the NMR solution structure of TC5 (Gehring et al., 1993) and the X-ray crystal structure of C4 (Chen et al., 1994) agree in their major features. Cytosine quadruplexes formed by molecules having a single block of C's are mediated by C·C+ base-pairs instead of base-quartets. Low pH is necessary to stabilize C-tetraplexes because one of the C's involved in C·C+ base-pair is protonated (Figure 4a, Ahmed & Henderson, 1992). The quadruplex is formed by the intercalation of 2 parallel-stranded duplexes. Although the strands of a duplex involved in a base-pairing arrangement are parallel, the two intercalated duplexes run antiparallel to each other (Figure 4b; Gehring et al., 1993; Chen et al., 1994).

![Figure 4. a. C·C+ base pair b. C-quadruplex formed by dTC5.](image-url)
The ability of both telomeric strands to form structures mediated by self-recognition suggests that both strands may participate in cellular processes in which alternative DNA structures have been implicated such as meiotic chromosome pairing (Sen & Gilbert, 1988), control of gene expression (Smith et al., 1989), and recombination (White et al., 1993). The biological relevance of alternative DNA structures will be discussed in more detail below.

**Role of cations in G-quartet structures**

Several studies have demonstrated the importance of monovalent cations for formation of G-quartet structures. All telomeric DNA structures formed in the presence of K⁺ or Na⁺ have been found to be mediated by G-quartets. Only in the presence of Li⁺ is a simple hairpin foldback formed lacking G-quartets (Choi & Choi, 1994). K⁺ and Na⁺ are thought to stabilize G-quartet structures by coordinating the 8 carbonyl oxygens in the quartets above and below the cavity where the cation rests (Sundquist and Klug, 1989). Although both K⁺ and Na⁺ mediate formation of G-DNA structures, there are significant differences in the structural characteristics and stability of G-DNA formed in K⁺ verses those formed in Na⁺. K⁺ provides almost 20 °C greater thermal stability compared to Na⁺ (Jin et al., 1992; Hardin et al., 1991; Xu et al., 1993). The enhanced thermal stability provided by K⁺ correlates with the preferential binding of K⁺ at specific sites on G-DNA quadruplexes (Xu et al., 1993). In addition, different structures are formed in K⁺ and Na⁺ by the same molecule (Guo et al., 1993). In Na⁺, two-repeat telomeric oligonucleotides adopted CD spectra characteristic of antiparallel quadruplexes while those formed in K⁺ adopted the spectra of a parallel-stranded quadruplex (Guo et al., 1993).

Divalent cations such as Sr²⁺ and Ba²⁺ have also been shown to facilitate formation of G-DNA structures but at concentrations 100 fold less than their monovalent counterparts (10 mM vs 1 M; Venczel & Sen, 1993). Furthermore, concentrations of Mg²⁺ in the physiological range (5-10 mM) facilitate formation of these structures (Zahler et al., 1991; Schierer & Henderson, 1994). In support of the important role of Mg²⁺, it was clearly shown to affect the assembly of a novel G-DNA structure termed G-wires. G-wires grow up to 5x longer in Na⁺/Mg²⁺ than in Na⁺ or K⁺ alone (Marsh et al., 1995). Concentrations of K⁺ and Na⁺ needed for formation of G-DNA structures (1 mM for K⁺; 100 mM for
Na⁺) are also within the physiological range. The ability of G-DNA structures to form in physiological salt concentrations supports the possibility of their existence in vivo.

**G-DNA binding proteins and possible biological roles**

Two approaches have been attempted to address the biological significance of G-DNA structures. One approach involves making antibodies against G-quartet structures and using these antibodies to detect G-quartet structures in vivo as has been done with Z-DNA (Wittig et al., 1989; Rahmouni & Wells, 1989). So far, this approach has been unsuccessful (Charles Hardin, Department of Biochemistry, North Carolina State University, Raleigh, North Carolina 27695). The second approach involves the identification and characterization of proteins which bind to G-DNA structures. This is the approach taken in my research and by several other labs. At least 8 different proteins have been identified in a variety of organisms which bind to a variety of G-DNA structures. Among these, the strongest functional evidence comes from a yeast nuclease which binds to a parallel-stranded G-DNA structure (Liu et al., 1993). A homozygous deletion of the gene for this protein (KEM 1) blocks meiosis at the 4N stage. FACS (Fluorescence Activated Cell Sorting) analysis revealed that the mutants underwent premeiotic DNA synthesis but arrested before spore wall formation (Tishkoff et al., 1991). The KEM 1 protein also has strand exchange activity (Kolodner et al., 1987). Taken together, these results suggest a role for KEM 1 in meiotic recombination.

On the basis of the phenotypes described above, Liu and Gilbert (Liu & Gilbert, 1994) proposed a model for the involvement of the KEM 1 protein in meiotic recombination. They propose that a parallel-stranded G4-DNA structure is formed by the G-rich strands of the bivalent and bound by the KEM 1 protein. The nuclease activity of KEM 1 then cleaves strands to be recombined and the strand exchange activity follows to transfer chromosomal segments between homologues. The proposed model may be the mechanism responsible for recombination of G-rich sequences such as telomeres and immunoglobulin switch regions. In support of the proposal that G4-DNA structures have a role in meiotic recombination, telomeric sequences have been found to strongly stimulate meiotic recombination (White et al., 1993). Furthermore, RAP 1, which has been shown
to bind telomeres in vivo (Conrad et al., 1990) and G-DNA structures (Giraldo & Rhodes, 1994), stimulates meiotic recombination (White et al., 1991).

Once antibodies to the KEM 1 protein are made, cytological studies can be performed to test whether this protein is associated with the synaptonemal complex during meiosis. Furthermore, antibodies to this protein may be useful in disrupting meiotic recombination in wild type cells. Evidence such as this will strongly support the involvement of the KEM 1 protein in meiosis.

G-DNA binding proteins can be divided into 3 categories: 1) Proteins identified on the basis of their ability to bind G-DNA 2) Previously characterized proteins tested for their ability to bind G-DNA 3) Proteins which promote the formation of G-DNA structures. TGP (Tetrahymena G4 binding protein), described in chapters 1 and 2, falls into the first category. Other proteins from the first category include QUAD, a hepatocyte chromatin protein (Weisman-Shomer & Fry, 1993) and the KEM 1 protein (Liu et al., 1993). The genes for TGP and QUAD have not yet been identified. Characterization of their biological function may be more directly addressed once the purified proteins are further characterized, the genes cloned and antibodies to the purified proteins are generated. A clue to the biological role of TGP may be provided by its preference for binding G-DNA molecules having a long single stranded tail adjacent to the G-quartet region (Schierer & Henderson, 1994). The single strand binding property allows TGP to fit into the model described for the KEM 1 protein's involvement in meiotic recombination described above.

Several proteins also fall into the second category including topoisomerase II, Mf3 (an avian protein), RAP 1 and macrophage scavenger receptors. Topoisomerase II purified from chicken was shown to bind G4-DNA structures formed by the immunoglobulin switch region sequences (Chung et al., 1992). In addition, topoisomerase II has been implicated in recombination due to its high reactivity toward an alternating purine-pyrimidine sequence which functions as a recombination hot spot in the β-globin gene (Spitzner et al., 1989). Thus, topoisomerase II, RAP 1, and the KEM 1 protein all have potential roles in meiotic recombination and each of them binds to G4-DNA suggesting that G-DNA structures have a role in meiotic recombination as well.

RAP 1 is a multifunctional yeast protein found to bind to G-DNA and promote its formation. RAP 1 promotes the formation of parallel-stranded G4-
DNA rather than antiparallel G-DNA even though it binds to preformed structures of both types (Giraldo et al., 1994). One model suggests that RAP 1 promotes formation of G4-DNA by acting as a surface which brings together the G-strands thereby increasing their local concentration (Giraldo et al., 1994). Several G-DNA binding proteins have basic domains which may serve as the surface which brings the G-strands together (Fang & Cech, 1993). RAP 1 has also been found to stimulate meiotic recombination at the HIS4 locus in yeast (White et al., 1991). A mutation in the RAP 1 binding site which eliminates RAP 1 binding reduced the frequency of meiotic crossing over (White et al., 1991).

RAP 1 in conjunction with the KEM 1 protein could carry out formation of G4-DNA and recombination of G-rich sequences during meiosis. Since KEM 1 only binds G4-DNA and does not promote its formation, RAP 1 would be responsible for the initial formation of G4-DNA upon which KEM 1 acts (KEM 1 model already discussed above). Another possible function of RAP 1 involves mediating telomere-telomere associations. In support of this possibility, it was found that yeast telomeres acquire single-stranded overhangs in late S phase and these overhangs mediate telomere-telomere pairing in vivo (Wellinger et al., 1993). Telomere-telomere association could then provide a topological constraint for modulating chromatin structure and gene expression. Further constraints could be provided by anchoring telomeres in the nuclear matrix. In support of this, telomeric sequences have been shown to be bound to nuclear matrix components (de Lange, 1992).

The third category of G-DNA binding proteins currently consists of RAP 1 (discussed above) and the β subunit of the Oxytricha end-binding protein. The work of Fang and Cech (1993a) demonstrated that the β subunit of the Oxytricha telomeric end-binding complex promotes the formation of both parallel and antiparallel G-quartet structures. The rate constant for G-quartet structure formation in the presence of the β subunit was 10^5 fold higher than the rate constant for G-quartet formation in the absence of the β subunit (Fang & Cech, 1993b). Thus, the β subunit acts as a molecular chaperone for nucleic acid structure since it facilitates the folding of a DNA structure without remaining bound to the substrate. This chaperone function differs from RAP 1 since RAP 1 remains bound to G-DNA structures. In the model proposed by Liu and Gilbert (Liu & Gilbert, 1994), the β subunit could facilitate formation of the G4-DNA
structure during meiotic recombination since the KEM 1 protein has not been shown to perform this function. The biological relevance of the G-DNA binding capability of the remaining proteins mentioned above (Mf3 and macrophage scavenger receptors) is not clear.

Non-telomeric G-DNA of biological importance

Several non-telomeric sequences of biological importance containing blocks of guanines have been examined for their ability to form G-quartet structures. These include immunoglobulin switch regions and FMR 1 (fragile X gene). Sequences from the immunoglobulin switch regions were shown to form a four-stranded, G-quartet structure (Sen & Gilbert, 1988). Such a structure may have a role in the recombination of these sequences during differentiation of B lymphocytes to plasma cells. As mentioned above, proteins linked to recombination events bind to G-DNA structures (RAP 1 and topoisomerase II).

The (CGG)\textsubscript{n} repeats of the fragile X locus have also been found to form G-quartet structures. The expansion of these repeats to more than 200 compared with 30-175 repeats in normal individual and carriers is correlated with the fragile X mental retardation phenotype (Hansen et al., 1993). The FMR 1 gene is not expressed in fragile X patients (Pieretti et al., 1991) and its replication is delayed to the G2/M transition of the cell cycle compared with S phase replication in normal males (Hansen et al., 1993). In addition, the 5' region of the gene is methylated in fragile X patients and methylation of cytosine residues of the (CGG)\textsubscript{n} repeats has been shown to stabilize the G-quartet structures formed by these repeats (Fry & Loeb, 1994; Hardin et al., 1993). The correlation of repeat expansion and FMR 1 methylation with stabilization of G-quartet structures formed by these sequences suggests that G-quartet structures may help regulate delayed replication and transcriptional silencing observed at the FMR 1 locus.

Another medically important application of G-DNA research stems from the discovery of a G-DNA aptamer which inhibits thrombin (Bock et al., 1992). The aptamer, d(GGTTGCTGCTGGTGG), was shown to form an antiparallel quadruplex structure (Macaya et al., 1993; Wang et al., 1993). Inhibition of thrombin is desired when anticoagulation is needed such as during cardiovascular surgery and to treat vascular diseases such as myocardial infarction and cerebral infarction (Wang et al., 1993). Heparin is the most widely used anticoagulant, but
its long lifetime in the blood makes it difficult to reverse its anticoagulation activity. The G-DNA aptamer will likely have a shorter lifetime in the blood and less severe side effects than heparin (Wang et al., 1993).

G-rich sequences capable of forming G-DNA structures have been found in promoter elements (Pears & Williams, 1988; Nickol & Felsenfeld, 1983; Murchie & Lilley, 1992), but it is not clear if their ability to form structure has biological significance. In addition, codon 12 of the human ras gene has G-rich regions which were shown to be capable of forming G-quartet structures (Smith et al., 1989). The authors suggest that G-quartet structures may serve as a signal for DNA methylation since methyltransferase is stimulated by structural changes near a d(pCG) site (Smith et al., 1989).

**G-quartet structures are formed by HIV genomic RNA**

RNA sequences with blocks of contiguous guanines have been shown to form four-stranded, G-quartet structures. The sequence UG₄U was shown to be a parallel-stranded tetramer by NMR analysis (Cheong & Moore, 1992). The sequence of the RNA genome of HIV-1 contains regions of contiguous guanines similar to those previously shown to mediate G-quartet structures. Mature retroviral virions contain two copies of their single-stranded RNA genomes which associate within an RNA-gag protein complex. This dimerization of the two copies of genomic RNA is an important step in the HIV-1 life cycle that appears to negatively regulate translation and positively regulate encapsidation (Marquet et al., 1991).

The mechanism which mediates RNA genome dimerization is not known, but several lines of evidence suggest that the dimer is mediated by G-quartets. The RNA genomes dimerize in vitro in the absence of protein leading to the speculation that dimerization occurs primarily through direct RNA-RNA interactions (Marquet et al., 1991). It was further shown that cation stabilization of the RNA dimer follows the same order as that obtained for G-quartet structures (K⁺ > Na⁺ > Li⁺ > Cs⁺; (Sundquist & Heaphy, 1993). Finally, DMS treatment of the dimer revealed that two blocks of guanines in the region essential for dimerization were protected from methylation (Awang & Sen, 1993).
Telomeric G-DNA superstructures

In addition to the G-quartet structures already mentioned, telomeric oligonucleotides are able to form large superstructures. The formation of these superstructures is dependent on the availability of free 3' or 5' guanine. If thymine is located on the 3' or 5' end of the oligonucleotides mentioned above, superstructures do not form (Sen & Gilbert, 1992). The oligonucleotides dT9G3 or dG3T9 formed higher order structures in multiples of four strands on nondenaturing gels demonstrating that G-quartet-based tetramers assemble end to end to form these structures (Sen & Gilbert, 1992). The arrangement of individual strands in G-DNA superstructures is out-of-register. Methylation protection of only the center G in T12G3 confirmed the out-of-register alignment (Sen & Gilbert, 1992). Similar to the study by Sen and Gilbert (1992), superstructures were also obtained with dT4G4 but not with dT4G4T (Lu et al., 1992).

Marsh and Henderson (1994) extended superstructure studies by Sen and Gilbert to show the large growth potential of G-DNA superstructures. By imaging superstructures formed by d(G4T2G4) with the atomic force microscope (AFM), it was shown that superstructures up to one micrometer in length can be formed (Marsh et al., 1995). Growth of G-wires was affected by the cationic environment, incubation temperature, DNA concentration, and duration of the incubation. The longest G-wires obtained were incubated in 50 mM Na+, 50 mM Tris-HCl pH 7.5, and 10 mM Mg2+ at 37 °C for 12-24 hours (Marsh et al., 1995).

The biological relevance of G-DNA superstructures is not clear. Their existence in the cell is possible since telomeric overhangs in vivo have terminal guanines (Henderson & Blackburn, 1989) instead of terminal thymines, and superstructures can assemble in physiological salt conditions (Marsh & Henderson, 1994). The significance of G-DNA superstructures may become more clearly established in a non-biological context by their use in nanotechnology applications. The large size of G-wires and the ability of nucleic acids to be modified with functional moieties enhance the potential of G-wires for use as scaffolds for nanostructure design and construction (Marsh et al., 1995).
**Atomic Force Microscopy of G-DNA/protein complexes**

The atomic force microscope provides images of the topological features of biological samples by translating the movement of a cantilever probe in contact with the sample into a visual image of the surface. The first AFM images of G-DNA structures were obtained by Marsh et al. (1995). The telomeric oligonucleotide $G_4T_2G_4$ was used for these studies because it forms large, multimeric superstructures termed G-wires. The sample height of G-wires is two to three fold higher than double-stranded plasmid DNA, but the difference in diameter between G-wires and B-DNA predicted from X-ray crystallographic studies is much less (1.4 fold). Thus, G-wires appear to be more resistant to compression during AFM sample preparation and imaging than double-stranded DNA (Marsh et al., 1995).

Due to their size stability in AFM studies, G-wires provide a good model system for studying DNA-protein interactions in the AFM. Mobility shift assays demonstrated that G-wires are bound by TGP (Schierer & Henderson, 1994). AFM images of G-DNA/protein complexes were obtained after partial purification of TGP. These provide the first images of G-DNA/protein complexes. Further AFM studies using the TGP/G-wire system in solution may provide images of real time binding and release of G-DNA/protein complexes, as well as subunit stoichiometry, mode of binding and the effect of protein binding on DNA structure.
CHAPTER 1: A PROTEIN FROM TETRAHYMENA THERMOPHILA
THAT SPECIFICALLY BINDS PARALLEL-STRANDED G4-DNA

A paper reprinted with permission from Biochemistry 33, 2240-2246. Copyright
1994 American Chemical Society

Ted Schierer and Eric Henderson*
Department of Zoology and Genetics
Iowa State University
Ames, IA 50011
USA

*To whom correspondence should be addressed

ABSTRACT

G4-DNA is a parallel, four-stranded structure mediated by tetrads of
hydrogen-bonded guanines (G-quartets). An abundant protein called Tetrahymena
G4 binding protein (TGP) that binds to an intermolecular, quadruplex form of
d(TTGGGGTTGGGTGGGGTTGGGG) under physiological salt conditions has
been identified in cellular extracts from the ciliated protozoan Tetrahymena
thermophila. In binding competition experiments, molecules capable of forming
G4 structures compete for binding to TGP, but non-G4 forming molecules and
r(U2G4)4 do not. TGP binding also requires a single-stranded region adjacent to
the G4 structure. During the course of this study, it was determined that Mg2+
facilitates the formation of parallel-stranded G4-DNA structures and that high
oligonucleotide concentrations are not required to drive formation of these
structures. In addition, G4-DNA and TGP/G4-DNA complexes form readily under
physiological salt conditions. These data support the proposal that G4-DNA
structures exist in vivo.
INTRODUCTION

Most telomeric DNA consists of simple repetitive sequences containing blocks of G/C base pairs with an asymmetric distribution of guanine on one strand (G-strand) and cytosine on the other (C-strand) (reviewed by Blackburn & Szostak, 1984). The G-strand extends approximately 12-16 nucleotides beyond the end of G/C duplex, forming a 3' overhang in organisms where this has been studied (Klobutcher et al., 1981; Pluta et al., 1982; Henderson & Blackburn, 1989). Synthetic oligonucleotides containing telomeric G-strand sequences are able to form unusual structures mediated by a cyclic hydrogen-bonded arrays of guanines (G-quartets) (reviewed by Sundquist, 1991; Williamson et al., 1989; Sundquist & Klug, 1989; Kang et al., 1992; Smith and Feigon, 1992). These structures include antiparallel, intramolecular quadruplex (G'2-DNA) (Henderson et al., 1987; Williamson et al., 1989; Sen & Gilbert, 1990; Jin et al., 1990; Kang et al., 1992) and parallel-stranded intermolecular quadruplex (G4-DNA) structures (Zimmerman et al., 1975; Sen & Gilbert, 1990; Jin et al., 1992; Aboul-ela et al., 1992; Gupta et al., 1993).

Intramolecular G-DNA structures migrate faster than linear forms of the same length in non-denaturing gels (Henderson et al., 1987; Williamson et al., 1989) while intermolecular structures migrate more slowly (Sen & Gilbert, 1990; Acevedo et al., 1991; Sundquist & Klug, 1989). The equilibrium between these structures is cation and concentration dependent, and the formation of the less favored antiparallel G'2-DNA (Sen & Gilbert, 1990) verses parallel-stranded G4-DNA is stabilized by the presence of particular monovalent (K+ > Rb+ > Na+ > Cs+ > Li+) and divalent cations (Sr2+ > Ba2+ > Ca2+ > Mg2+) and high oligonucleotide concentration (Venczel & Sen, 1993). A recent study demonstrated that divalent cations stabilize G-DNA structures at a concentration of 10 mM whereas 1 M monovalent cation concentrations are needed to produce the same effect (Venczel & Sen, 1993). Once formed, intermolecular quadruplex structures are exceedingly stable and recalcitrant to hybridization with complementary C-rich strands (Raghuraman & Cech, 1990; Hardin et al., 1991).
Along with telomeric DNA, a variety of biologically relevant G-rich DNA sequences have been identified, including immunoglobulin switch regions (Sen & Gilbert, 1988), recombination hot spots (Hastie & Allshire, 1989) and gene regulatory regions (Nickol & Felsenfeld, 1983; Pears & Williams, 1988). G4-RNA has also been studied in detail and shown to form structures based upon G-quartets (Cheong & Moore, 1992). G-RNA can mediate dimerization of the HIV genome in vitro (Marquet et al., 1991; Sundquist & Heaphy, 1993). These examples support the idea that G-DNA/RNA structures may have important roles in the cell including telomere function (reviewed by Blackburn, 1991), meiotic chromosome pairing (Sen & Gilbert, 1988), HIV genome dimerization (Marquet et al., 1991; Sundquist & Heaphy, 1993), and promoter function (Walsh & Gualberto, 1992).

The exact biological role of G-DNA structures is not clear in any organism. Intramolecular foldback structures formed by d(T4G4)4 (Oxy 4) inhibit the activity of Oxytricha telomerase (Zahler et al., 1991), an enzyme that adds telomeric repeats to the 3' end of the chromosome (Grieder & Blackburn, 1985; Zahler & Prescott, 1988). The subunit of the Oxytricha telomeric end binding protein has recently been reported to catalyze the formation of G-quartet structures (Fang & Cech, 1993) suggesting that G-DNA may have a regulatory role in telomere replication. In addition, a variety of proteins have been reported to bind G4-DNA. Chick topoisomerase II (Chung et al., 1992), a yeast nuclease (Liu et al., 1993), MyoD (a transcription factor that regulates myogenesis) (Walsh & Gualberto, 1992), a hepatocyte chromatin protein (QUAD, Weisman-Shomer & Fry, 1993) and macrophage scavenger receptors (Pearson et al., 1993) all bind G4-DNA and suggest potential roles for G-DNA/protein complexes.

We report here the identification and characterization of an abundant G-DNA binding activity from Tetrahymena thermophila. This protein, designated Tetrahymena G4 binding protein (TGP) binds an intermolecular quadruplex structure. An extensive comparison of different G4-DNA molecules as substrates for TGP is presented. It is shown that parallel-stranded G4-DNA is preferred over antiparallel quartet structures. The identification of this G4 binding protein lends further support to the proposal that G4-DNA exists in vivo and may play an important role in the cell.
MATERIALS AND METHODS

*Tetrahymena strains and cell culture.* *Tetrahymena thermophila* strain C3 V was grown using a rotary shaker at 30°C in 2% PPYS (2% proteose peptone, 0.2% Yeast extract, and 0.003% sequestrine) to mid-log phase (2.5 X 10⁵ cells/ml).

*Extract preparation.* *Tetrahymena* whole cell extracts were prepared according to the procedure of Greider and Blackburn (1987) except that the cells were not mated. Furthermore, double distilled H₂O was substituted for diethyl pyrocarbonate-treated H₂O and 10 mM Tris-HCl (pH 7.5) was substituted for Dryls for cell washes. Briefly, cells at mid-log phase were harvested by centrifugation at 5K rpm for 5 minutes (Sorvall GSA rotor, 4°C) and washed twice with a 20-50x volume of cold 10 mM Tris-HCl (pH 7.5). Cell pellets were then resuspended in a 5x volume of cold 10 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, 10% glycerol, 10 mM 2-mercaptoethanol) and 1/10 volume of 2% NP-40 was added immediately. Cells were lysed by shaking on a Fisher Genie 2 vortexer (setting 4) for 30' at 4°C. The cell lysate was centrifuged at 100,000 x g for 60 minutes at 4°C. The supernatant, termed S-100 cell extract, was aliquoted and frozen quickly with liquid nitrogen. The final protein concentration was typically 2.5 mg/ml in TMG buffer. The protease inhibitors leupeptin (.01 mM), pepstatin (.01 mM), and Pefabloc (.1 mM) (Boehringer Mannheim) were included in all solutions. Without the addition of leupeptin, TGP still bound to Tet 4 multimers, but the complex migrated below the 517 bp marker indicating that the protein had been cleaved by a protease (data not shown). Extracts can be thawed 3 times without any loss in TGP binding activity.

*DNA oligonucleotide synthesis, purification and 5’ end radiolabeling.* d(T₂G₄)₄ DNA oligonucleotides were gel purified as previously described (Henderson et al., 1987). Briefly, DNA oligonucleotides were synthesized on an Applied Biosystems DNA Synthesizer followed by deprotection. After boiling in 1X TBE buffer (89 mM Tris-HCl; 89 mM boric acid; 1.5 mM disodium ethylenediamine tetraacetate (EDTA); pH 8.3) containing 80% formamide for 90 seconds, the oligonucleotides were separated by electrophoresis through 20% polyacrylamide sequencing gels containing 7M urea and 1x TBE buffer. The desired species were identified by UV shadowing and excised from the gel. DNA
was eluted by shaking in TE buffer (10 mM Tris-HCl, pH 7.5, 1mM EDTA) for 12-16 hours at room temperature, and the eluted oligonucleotides were desalted over a Sep-Pak C18 column (Waters). Gel purified Tet 4 oligonucleotides were 5' end labeled as previously described (Henderson et al., 1987). The labeled Tet 4 was then gel purified (12%, 7M urea PAGE) and desalted by C18 chromatography as above.

**Electrophoretic quadruplex assays.** Oligos X and Y at a concentration of 1 pmol/ml were boiled in the indicated salts in addition to 10 mM Tris-HCl (pH 7.5) and 4.5% glycerol and cooled 10' on ice to allow structure formation before loading an 8% polyacrylamide gel. Gels were run in 0.6x TBE at room temperature at 10 V/cm.

**Electrophoretic mobility retardation assays.** For mobility retardation assays, 0.5 picomoles of 5' ³²P-probe were boiled in the presence of 295 mM NaCl, 29 mM KCl, and 6 mM MgCl₂, 12 mM Tris-HCl (pH 7.5) and 5.3% glycerol and cooled on ice for at least 30' to allow the probe to form structure. Preformed Tet 1.5 multimers were not boiled. In Figure 2 (lanes 1, 2), formation of intermolecular Tet 4 structure was prevented by boiling Tet 4 in 10 mM Tris-HCl (pH 7.5) and 4.5% glycerol, but enough Na⁺ exists in the Tris buffer to allow formation of the intramolecular foldback monomer (Figure 2, lanes 1, 2). A 200 fold molar excess of oligonucleotide d(T)₂₄ was then added as a nonspecific competitor (5 pmol/ml final concentration) before adding 2 ml of diluted S-100 extract (0.25 mg/ml) to produce a final reaction volume of 20 ml. The final probe concentration was 0.025 pmol/ml in all cases except XYa, XYb and XYc (0.04pmol/ml), and final concentrations in the binding reaction were 250 mM NaCl, 25 mM KCl, 5 mM MgCl₂, 10 mM Tris-HCl (pH 7.5), 5.5% glycerol, 1 mM 2-mercaptoethanol and 0.02% NP-40. In competition assays, unlabeled competitors were boiled in the same salt concentrations as the probe and cooled on ice to allow separate structure formation of the competitor and probe before they were combined. Once probe and competitor were combined, a 200-fold molar excess of oligonucleotide d(T)₂₄ was then added as a nonspecific competitor (5 pmol/ml final concentration) before adding 0.5 mg of S-100 extract to produce a final reaction volume of 20 ml. The final probe concentration was 0.025 pmol/ml in all cases except with XYa, XYb and XYc (0.04pmol/ml). r(U₂G₄)₄ binding and competition assays included 20U of RNasin (Promega).
Binding reactions were incubated on ice for 20 minutes and then loaded onto a 8% non-denaturing polyacrylamide gel in 0.6x TBE. Electrophoresis at 10 volts/cm was carried out at room temperature until the bromophenol blue reached 3/4 the length of the gel. The gel was then dried and the bands were visualized by autoradiography and quantitated using a PhosphorImager (Molecular Dynamics).

UV cross-linking. Cross-linking of the multimeric Tet 4 species was performed in situ by exposing a wet 8% non-denaturing polyacrylamide gel containing the multimeric species to 254 nm UV light for 30 minutes on ice with the gel no more than 5 cm from the UV source. The gel was exposed to film to reveal the positions of the putative multimeric and monomeric species. Gel pieces containing putative multimeric or monomeric Tet 4 were then excised, placed into 1 ml TE buffer (pH 7.5) and shaken overnight to elute the DNA which was then purified on a C-18 (Waters) column. The purified DNA was mixed with an equal volume of 80% formamide/1x TBE and boiled 5 minutes before analysis by 7M urea-12% PAGE.

RESULTS

Characterization of multimeric and monomeric forms of Tet 4. The oligonucleotide d(T2G4)4 (Tet 4) was used as a probe to search for proteins that bind to G4-DNA. Its ability to form G4-DNA is demonstrated in Figure 1. Under non-denaturing conditions, slowly migrating Tet 4 species were reproducibly observed in addition to the species migrating to the expected position for Tet 4 monomers (Figure 1A). UV cross-linking experiments demonstrated that the slow migrating structures are due to intermolecular associations of Tet 4. Under denaturing conditions without UV irradiation, both species migrate to the position expected for a 24 nucleotide long molecule (Figure 1B, lanes 1, 2). Following UV irradiation, at least three cross-linked complexes from the slow migrating band that have sizes much larger than the non-UV exposed species are observed (Figure 1B, lane 4). Thus, the slow migrating form of Tet 4 is an intermolecular structure.

A control oligonucleotide d(ACTGTCGTACCTGATATGGGGGT) (oligo Y, Sen & Gilbert, 1990) was subjected to the same treatment as Tet 4. Oligo Y was
Figure 1. Tet4 forms G4-DNA. (A) 32P-labeled Tet4 oligonucleotides were analyzed on a non-denaturing gel (ND-PAGE) and separated into monomeric (Mono) and multimeric (Mult) species. (B) Tet4 species from a non-denaturing gel were crosslinked with short-wave UV (=UV lanes 3 and 4), excised and run on a denaturing gel (D-PAGE) next to Tet4 monomers and multimers that had not been exposed to UV (-UV lanes 1 and 2). Only the multimeric Tet4 species formed cross-linked (lanes 4 and 6) species having a size similar to the known tetramer Y4 in lane 8 (see Figure 3A for oligo Y sequence). As expected, Y monomers (lane 7) were unaffected by UV irradiation. d(T)24 (lane 5) was used as a marker to indicate the position of unstructured monomers. Tet4 intramolecular foldback monomers (Figure 1A, *) run faster on denaturing gels after crosslinking (Figure 1B, **). (C) Formation of G4-DNA is greatly facilitated by the presence of Mg2+ in addition to Na+ and K+. G4-DNA oligos X and Y were incubated in the presence of different combinations of cations indicated at the top in addition to 10 mM Tris-HCl (pH 7.5). The 5 possible X-Y four-stranded molecules are indicated at the left.
previously shown to form four-stranded structures (Sen & Gilbert, 1990).
Following UV irradiation, only the Y tetramer (Figure 1B, lane 8) formed cross-linked species. Y monomers (Figure 1B, lane 7) migrated at the position expected for a 23-mer. Slow migrating forms of Tet 4 exposed to UV (Figure 1B, lane 6) ran parallel to cross-linked Y quadruplexes. Thus, we conclude that the Tet 4 multimers very likely consist of complexes of four separate strands.

Formation of G-quartet structures has been shown to dependent on monovalent cation and DNA concentrations (Williamson et al., 1989; Sundquist & Klug, 1989; Sen & Gilbert, 1990; Hardin et al., 1991; Venczel & Sen, 1993). We found that the addition of 5 mM Mg\(^{2+}\) facilitated the formation of G4-DNA when 200 mM Na\(^+\) and 200 mM K\(^+\) were present (Figure 1C, lanes 6, 7), but Mg\(^{2+}\) alone did not facilitate G4-DNA formation (Figure 1C, lane 4). Monovalent cations alone (Figure 1C, lanes 2, 3) were not as effective at promoting G4 formation as Na\(^+\) plus Mg\(^{2+}\) (Figure 1C lane 6) and K\(^+\) plus Mg\(^{2+}\) (Figure 1C, lane 7). The role of Mg\(^{2+}\) as a facilitator has been observed previously with the Oxy 4 intramolecular, antiparallel quartet structure (Zahler et al., 1991). Zahler (1991) found the T_m of Oxy 4 in K\(^+\)/Mg\(^{2+}\) to be 40 °C higher than Oxy 4 in Mg\(^{2+}\) and 30 °C higher than Oxy 4 in Na\(^+\)/Mg\(^{2+}\). This study shows that Mg\(^{2+}\) can also facilitate the formation of parallel-stranded G4-DNA. However, thermodynamic studies of Mg\(^{2+}\) as a facilitator of G4-DNA formation were not performed. Thus, it is possible that Mg\(^{2+}\) facilitates interconversion between forms, but that it actually destabilizes any given form.

The salt conditions optimal for G4 formation determined in DNA assays (250 mM Na\(^+\), 25 mM K\(^+\), and 5 mM Mg\(^{2+}\)) were used in subsequent binding and competition assays. In addition, physiological salt conditions (100 mM K\(^+\), 10 mM Na\(^+\), and 5 mM Mg\(^{2+}\)) supported the formation of intermolecular forms of Tet 4 indicating that formation of these structures in vivo is possible (Figure 2, lane 6).

Previous studies have shown that high oligonucleotide concentrations facilitate the formation of G4-DNA (Sen & Gilbert, 1988, 1990). Therefore, it was surprising to find that X and Y formed G4-DNA at concentrations 250-fold more dilute (0.004 mg/ml, Figure 1C) than the conditions used by Sen and Gilbert (1990) (1 mg/ml) indicating that concentrations over 1 mg/ml are not necessary to drive formation of G4-DNA.
In addition to the intermolecular Tet 4 species described above, Tet 4 readily folds into an intramolecular quadruplex structure represented by the fastest migrating band (indicated by * Figure 1A). A G-quartet model for the intramolecular foldback structure was proposed by Williamson (1989) and confirmed by X-ray crystallography (Kang et al., 1992) and NMR (Smith & Feigon, 1992). When cross-linked by UV irradiation, it migrates faster on denaturing gels than unstructured 24-mers (indicated by ** Figure 1B, lane 3; Williamson et al., 1989). This foldback monomer was the dominant species formed by Tet 4 under the salt conditions used in the binding assays described below.

Identification of a Tetrahymena DNA-binding protein specific for G4-DNA. Electrophoretic mobility retardation assays were used to identify a protein in Tetrahymena extracts (TGP) that bound preferentially to the intermolecular form of Tet 4 despite the fact that it makes up only 10% of the total Tet 4 DNA species (Figure 2, lane 4). Specificity for multimeric Tet 4 was demonstrated by the lack of TGP complexes in lanes 1 and 2 (Figure 2) where multimeric Tet 4 is absent but the intramolecular form is present (see Materials and Methods). Comparison of lanes 1 and 2 with lanes 3 and 4 in Figure 2 demonstrates that formation of Tet 4 multimers is salt dependent and that TGP binds to the intermolecular form of Tet 4 and not to the unstructured or intramolecular foldback monomers.

TGP complex formation was completely inhibited by pre-incubation of S-100 extract above 50 °C or by extensive protease treatment indicating that TGP is a protein (data not shown). TGP could still bind G4-DNA after limited proteolytic cleavage suggesting that its G4 binding domain may be separable from other domains of the protein (data not shown).

TGP binds to quadruplexes in parallel-stranded arrangements. The structures formed by several G-rich oligonucleotides have been well characterized in recent years (Williamson et al., 1989; Sundquist & Klug, 1989; Sen & Gilbert, 1990; Jin et al., 1992; Aboul-ela et al., 1992; Kang et al., 1992; Smith & Feigon, 1992; Gupta et al., 1993). A number of different arrangements (parallel and antiparallel) and strand stoichiometries (monomers, dimers and tetramers) exist in this structural family. To test the structural requirement for TGP binding, the oligonucleotides listed in Figures 3A, 4A, and 5A were used in
Figure 2. TGP binds a G4 form of Tet 4. Mobility retardation assays were used to determine the species of Tet 4 bound by TGP. In the presence of S-100 extract (+ lanes) neither unstructured Tet 4 monomers nor intramolecular foldback monomers (lanes 1 and 2, *) were bound by TGP (lane 2). A TGP complex was formed only when the multimeric species of Tet 4 was available (lanes 4 and 6). Physiological salt conditions (lanes 5 and 6) supported formation of multimeric Tet 4 (lanes 5, 6) and TGP complex formation (lane 6).
electrophoretic mobility retardation and binding competition experiments. In Figure 3A, the TGP complex increased in intensity in proportion to the availability of quadruplex DNA when the amount of total protein added to the binding reaction was held constant. This trend is evident for the intermolecular quadruplexes formed by Tet 4, Oxy 4, Y4, and (Tet 1.5)n. Oxy 4 formed a slow migrating species analogous to the Tet 4 tetramers (Figure 3B, lane 3, arrowhead). This slow form was shifted preferentially over the monomeric species (Figure 3B, lane 4). Similarly, over 70% of the Y4 molecules were shifted in the presence of protein with as much as 97% shifted in some experiments. In contrast, there was no detectable reduction in Y monomers in the presence of protein (Figure 3B, lane 8). Therefore, Y monomers, like Tet 4 monomers are not bound by TGP.

Tet 1.5 formed intramolecular foldback structures at low concentration (Figure 3C, lanes 1, 2) and a ladder of multimers at high concentration (Figure 3B, lanes 11, 12). The multimers are a previously unreported structure currently under study in our laboratory and their proposed structure is a parallel-stranded, G4 arrangement (T. Marsh and E. Henderson, in preparation). As shown in lane 12 (Figure 3B), a TGP complex was formed in the presence of Tet 1.5 multimers, but not in the presence of Tet 1.5 intramolecular foldback structures (Figure 3C, lanes 1, 2). X4 is an exception to TGP's preference for G4-DNA seen with Tet 4, Oxy 4, Y4, and Tet 1.5 and its further examination is described below.

Competition experiments using Tet 4 as the probe and the oligonucleotides listed in Figure 3A as the unlabeled competitor corroborated binding assay results. Four levels of unlabeled competitor were tested (2-fold molar excess over the Tet 4 probe, 10-fold, 50-fold, and 250-fold). A 200-fold excess of d(T)24 was included as a nonspecific competitor. Oxy 4 and Tet 4 were the best competitors having the ability to compete over 80% of the TGP complex counts at a 50 fold molar excess (Figure 3A). In addition, Y competed much better than X. 60% of the TGP counts were competed by Y at a 50x molar excess whereas X was unable to compete even at a 50x molar excess (Figure 3A).

In contrast to the parallel quadruplexes in Figure 3B, the antiparallel hairpin dimer Oxy 1.5 (Kang et al., 1992; Smith & Feigon, 1992) was not bound
Figure 3. TGP binds preferentially to parallel-stranded quadruplex DNA. (A) Summary of binding and competition assays. Several different G-quartet structures were used as competitors for TGP binding. Competition assays using Tet 4 as the probe were consistent with the binding assays shown in B and C. Oxy 4 and Tet 4 were the best competitors indicated by "++". Y4 was a strong competitor (+) while X, Oxy 1.5, and r(U2G4)4 were poor competitors (-). (B) In mobility retardation assays, multimeric forms of Tet 4 and Oxy 4 formed TGP complexes (lanes 2 and 4) in addition to the known parallel-stranded quadruplex, Y4 (lane 8). A novel, multimeric, parallel-stranded structure formed by Tet 1.5 ((Tet 1.5)n) also formed a TGP complex (lane 12). The TGP complex always migrated between the 517 and 1600 bp markers of the 1 kb ladder (right). "+" or "-" indicates the presence or absence of S-100 extract (C) The antiparallel hairpin dimer Oxy 1.5 (lanes 3 and 4, Kang et al. 1992) did not form a TGP complex and neither did Tet 1.5 species (lanes 1 and 2).
by TGP (Figure 3C, lane 4) and neither was the analogous molecule from *Tetrahymena*, Tet 1.5 (Figure 3C, lane 2). Furthermore, Oxy 1.5 was a poor competitor for TGP binding (Figure 3A) indicating that TGP does not bind to antiparallel quadruplex structures.

The only common sequence element of the G4 oligonucleotides tested for binding to TGP was GGGG. Thus, TGP has no apparent sequence requirements other than blocks of four or more contiguous guanines. Oligos that contain guanines separated by 1 or more nucleotides were not bound by TGP (Figure 5A, 4G and 1G oligonucleotides). These data suggest that TGP is a G4-structure specific protein and not a primary sequence specific protein.

The Tet 4 probe consistently formed only 1 major TGP complex (Figure 3A, lane 2). The complexes formed by the G4-DNA’s (Oxy 4, Y4, (Tet 1.5)n) that migrated with the Tet 4 TGP complex in Figure 3A (lane 2) are likely to contain the same proteins since Oxy 4 and Y4 both compete successfully for binding to TGP when Tet 4 is the probe (Figure 3A). However, until TGP is purified to homogeneity, it remains a formal possibility that the proteins binding the different G4-DNA complexes are different proteins. The identity of the bands binding to Oxy 4 and Y4 which did not migrate with the Tet 4 TGP complex is under investigation (Figure 3A, lanes 4, 8). It is not known whether these bands represent different G4 binding proteins or if some of them represent different numbers of TGP proteins bound to a single G4 molecule.

The abundance of TGP in S-100 extracts was estimated from the picomoles of G4-DNA shifted. Enough protein exists in 0.5 ug of extract to bind 0.00625 pmol of intermolecular Tet 4. Therefore, there are at least 7.5 x 10⁹ protein molecules per ug of extract (1.6 x 10⁶ copies/cell). The abundance of TGP should facilitate its purification. Attempts to localize TGP activity have thus far been inconclusive.

*TGP binds preferentially to G4 molecules with long single-stranded tails.* Surprisingly, the tetrameric forms of oligonucleotide X did not form a TGP complex (Figure 3B, lane 6) suggesting that TGP requires more than just G-quartet structure for binding. In agreement with the lack of an X4-TGP complex, there was no detectable loss of X4 species in the presence of extract (Figure 3B, compare X4 bands from lanes 5 and 6). In contrast, 70-97% of Y4 counts were
reproducibly shifted in Y4 binding assays under the same conditions, and a prominent TGP complex was always present.

The only difference between X and Y is 11 additional nucleotides of random sequence 5' to the G-quartet structure region on oligo Y. To test if the length of this tail region facilitates binding by TGP, oligonucleotides with tails intermediate in length between X and Y were synthesized and used in competition experiments. Oligos XYa, XYb, and XYc become progressively more like Y in length with XYa being the shortest (Figure 4A). XYc competed for binding to TGP much better than XYb or XYa at all levels of competitor (Figure 4A). XYb and XYa were indistinguishable in their ability to compete for TGP. Binding assays corroborated the competition results. Only Y and XYc formed a strong TGP complex (Figure 4B, lanes 8 and 10). In contrast, XYb and XYa formed very weak TGP complexes having less than 2% of the counts present in the Y TGP complex (Figure 4B, lanes 6 and 4 respectively). Thus, a 5' tail of at least 15 nucleotides in addition to G-quartet structure is needed for optimal binding to these G4 structures. However, TGP is not a single strand binding protein since it did not bind monomeric forms of any of the oligonucleotides tested in this study.

Non-G4 oligonucleotides as substrates for TGP. Several non-G4 forming sequences were also tested for binding to TGP (Figure 5). Oligos TR1, TR2, and TR3 correspond to sequences found in the replication origin of Tetrahymena rRNA genes and are bound by factors from Tetrahymena (A. Umthun Z. Sibenaller, W. Shaiu, and D. Larson, unpublished results). T-rich (d(T)24), A-rich (TR2), and C-rich (d(C4A2)4) single stranded oligonucleotides were also tested in binding and competition assays. None were found to bind to or compete for TGP (Figure 5). In addition, the conditions of the assay were such that the complexes normally formed by the oligonucleotides TR1, TR2 and TR3 in S-100 extracts were not observed. Oligos which mimic telomeric DNA were also tested. Telomeric duplex oligonucleotides (Cardenas et al., 1993) failed to compete (Figure 5A), and telomeric C-strand oligonucleotides did not form a TGP complex (Figure 5B, lane 12). Competition experiments using d(C4A2)4 were hindered by duplex formation. Finally, two different A/T rich duplexes from the origin of replication in Tetrahymena rDNA were also tested as competitors for TGP and found not to compete at any level (Figure 5A).
Figure 4. TGP has a tail length requirement for optimal binding. (A) Summary of binding and competition assays. Competition experiments with Tet 4 as the probe corroborated the binding assay results. XYa and XYb were both poor competitors (-) while XYc competed as effectively as Y (+).

(B) Mobility retardation assays with X (lanes 1, 2), XYa (lanes 3, 4), XYb (lanes 5, 6), XYc (lanes 7, 8), and Y (lanes 9, 10) as the probe (indicated at the top) demonstrated that TGP binds only to G4 molecules with a 5' tail length of at least 15 nucleotides (XYc, lane 8).
Figure 5. Non-G4 forming oligos do not form TGP complexes. (A) Summary of binding and competition assays. Competition assays with Tet 4 as the probe corroborated the binding assay results and also demonstrated that two different A/T-rich duplexes were not bound by TGP since all oligos tested were poor competitors (-). (B) Mobility retardation assays using several different single stranded oligos as probes were performed to test the ability of A-rich (lanes 7-8), T-rich (lanes 3-4), and C-rich oligos (lanes 11, 12) to form a TGP complex. In addition, oligos from the replication origin of Tetrahymena rDNA were found not to form TGP complexes (TR1, lanes 5-6 and TR3, lanes 9-10) even though these oligos bind factors in S-100 extract under different conditions (A. Umthun, Z. Sibenaller, W. Shaiu, and D. Larson, unpublished results).
DISCUSSION

TGP binding specificity. Telomeric and non-telomeric G-rich sequences can form G4-DNA structures (reviewed by Sundquist, 1991; Henderson et al., 1987; Williamson et al., 1989; Sundquist & Klug, 1989; Sen & Gilbert, 1990; Jin et al., 1992; Kang et al., 1992; Smith & Feigon, 1992). This study demonstrates that Tetrahymena contains an abundant protein, TGP, that binds specifically to G4 structures formed by the telomeric G-strand sequence d(T2G4)4 and not to Tet 4 intramolecular foldback monomers or to unstructured Tet 4 monomers. Furthermore, TGP binds parallel-stranded G4-DNA structures having very different primary sequences from Tet 4, demonstrating its structure-specific nature. Taken together, the abundance and structural specificity of TGP suggest that it may have an important function in the cell.

TGP does not bind to G/C rich or A/T rich duplex DNA, nor to non-G-rich single-stranded DNA. TGP also fails to bind G-rich RNA monomers ((U2G4)4). However, we were unable to unequivocally test for the ability of G4-RNA to form TGP complexes since (U2G4)4 did not readily form G4-RNA under the conditions used in these experiments. Thus, it is still possible that TGP has affinity for G4-RNA.

An unusual feature of TGP is that it may require a single-stranded region in addition to a G-quartet structure region. The single-stranded requirement is most obviously demonstrated with the X and Y oligos. The length of the Y tail is an important feature suggesting that a longer tail somehow affects the G-quartet region in a manner that makes it available for binding by TGP. Single-stranded tails would also be associated with G4-DNA structures formed by Tet 4 and Oxy 4 oligonucleotides in which the individual strands are not precisely aligned. The presence of several different intermolecular Tet 4 structures on non-denaturing gels (Figure 1A) suggests that the individual strands of the G4 complexes are arranged in several different ways and likely contain single-stranded tail regions. If single-stranded tails adjacent to a quartet structure are a requirement for TGP binding, it is possible that the absence of a TGP/Oxy 1.5 or Tet 1.5 complex is due to the lack of a single-stranded tail and not due to a specific requirement for parallel DNA. It will be of interest to test antiparallel molecules having long single-stranded tails for TGP binding.
TGP’s requirement of a single stranded tail is similar to the binding properties of the yeast nuclease identified by Liu et al. (1993). The yeast nuclease appears to require G4 structure for cleavage of a single-stranded region 5' of the G4 structure since single-stranded forms of the oligonucleotides capable of forming G4-DNA are not cleaved. Nuclease activity is present in crude extracts containing TGP binding activity. Cleavage products are observed at the bottom of gels only in the presence of extract (Figure 3b). However, it is not known whether the two activities are related since these extracts contain many proteins. Purification of TGP will allow us to address this question.

Since chick topoisomerase II has been shown to bind G4-DNA (Chung et al., 1992), phosphocellulose purified TGP was tested for topoisomerase II activity. TGP was incapable of decatenating K-DNA (data not shown, Ryan et al., 1988) using conditions which support K-DNA decatenation by purified human topoisomerase. Thus, it is seems unlikely that TGP is *Tetrahymena* topoisomerase II. It is also unlikely that TGP is a telomeric end binding protein. A *Tetrahymena* end-binding protein migrates to a different position in mobility retardation assays than TGP (H. Sheng and E. Henderson, unpublished results). In addition, oligonucleotides which mimic the 3' end of the telomere (Cardenas et al., 1993) do not compete for binding to TGP (data not shown).

*Biological Relevance of TGP.* Several proteins have been reported to bind G-quartet structures, including the transcription factor MyoD (Walsh & Gualberto, 1992), macrophage scavenger receptors (Pearson et al., 1993), chick topoisomerase II (Chung et al., 1992), and a novel yeast factor (Liu et al., 1993). Recently, Fang and Cech (1993) demonstrated that the β subunit of *Oxytricha* catalyzes G-quartet formation. However, a specific biological role for any G-quartet binding protein has yet to be demonstrated *in vivo.*

It has been suggested that G-quartet structures present an array of phosphates that are favorable for nonspecific binding by basic proteins (J. Williamson, personal communication). In support of this idea, human and yeast topoisomerase II sequences and macrophage scavenger receptors contain clusters of basic residues which could serve as a nonspecific binding domain for G4-DNA (Pearson et al., 1993). However, given TGP’s specificity for intermolecular, parallel-stranded quartet structures verses intramolecular,
antiparallel quartet structures, it is likely that its biological role involves specific interactions with G4-DNA.

G4 DNA has been implicated in biological events involving contact or exchange between DNA domains (e.g., meiotic chromosome pairing (Sen & Gilbert, 1988), regulation of gene activity (Nickol & Felsenfeld, 1983; Pears & Williams, 1988), and gene rearrangement (Hastie & Allshire, 1989)). Furthermore, one G4 binding protein, topoisomerase II, is involved in passing a DNA duplex through another duplex. In all of these cases one can envision a state where two duplexes are aligned and operated upon. In addition, during processes like recombination, it is likely that single stranded regions will exist adjacent to transient four stranded domains. Thus, a protein like TGP, that binds to both four stranded and adjacent single stranded domains, could be involved in this type of reaction.

ACKNOWLEDGMENTS

The authors wish to thank B. Mixan and P. Qiu for preliminary studies, T. Marsh for kindly providing Tet 1.5 multimers, A. Umthun for TR and duplex oligonucleotides, S. Ahmed for the C-strand oligonucleotide, J. Zhao for K-DNA and D. Larson, S. Ahmed and T. Marsh for critical reading of the manuscript.

REFERENCES

CHAPTER 2: PURIFICATION AND CHARACTERIZATION OF TGP, A PROTEIN FROM TETRAHYMENA THAT BINDS TO PARALLEL-STRANDED G4-DNA

A manuscript in preparation for the journal *Biochemistry*

Ted Schierer, Zhen Hou, Luming Niu, and Eric Henderson

Department of Zoology and Genetics
2112 Molecular Biology Building
Iowa State University
Ames, Iowa 50011

ABSTRACT

G4-DNA is a parallel-stranded, alternative DNA conformation. In order to better understand the biological role of G4-DNA structures, a protein (TGP) which binds G4-DNA was identified and initially characterized (Schierer & Henderson, 1994). Its further characterization and purification has been pursued. The main component of TGP is an 83 kDa protein that binds to G4-DNA possibly as a homomultimer. Chromatography matrices most useful for TGP purification include the cation exchangers S-Sepharose and Biorex 70.

INTRODUCTION

The G-quartet, a cyclic, hydrogen-bonded array of guanines, was first proposed by Gellert et al. (Gellert et al., 1962) in order to explain the remarkable stability of the gels formed by guanylic acid (GMP). Interest in the study of G-quartet structures has increased dramatically in recent years since it was discovered that biologically relevant DNA sequences such as telomeres (Henderson et al., 1987; Williamson et al., 1989; Sundquist & Klug, 1989) and immunoglobulin switch regions (Sen & Gilbert, 1988) form unusual DNA structures.
Subsequent studies showed important features of telomeric sequences which influence the structure including the number of contiguous guanines and the number of thymines between blocks of guanines. It was determined that G-quartets cannot exist in isolation from other quartets and therefore at least two contiguous guanines are necessary for the formation of a G-quartet structures in sequences containing at least two blocks of guanines (Jin et al., 1990). Sequences having only one block of guanines need at least three contiguous G's to form G-quartet structures (Sen & Gilbert, 1992). In addition, the number of T's between the blocks of guanines determines whether the structure will be an antiparallel hairpin dimer or a parallel-stranded tetramer (Balagurumoorthy et al., 1992; Guo et al., 1993). At least two T's are necessary to allow formation of a hairpin foldback dimer in sequences with two telomeric repeats. If only one T is present between the G-blocks, the structure formed is a parallel-stranded tetramer (Balagurumoorthy et al., 1992; Guo et al., 1993).

The crystal structures of both parallel and antiparallel G-quartet structures have been determined (Laughlan et al., 1994; Kang et al., 1992). All parallel-stranded G-quartet structures studied to date are right-handed helices with entirely anti glycosidic conformations (Gupta et al., 1993; Aboul-ela et al., 1992; Wang & Patel, 1993; Wang & Patel, 1992). Antiparallel G-DNA structures are characterized by alternating syn and anti glycosidic conformations along each strand in the quadruplex (Kang et al., 1992; Smith & Feigon, 1992).

Two approaches have been taken to address the biological significance of G-DNA structures. One approach involves identifying biologically relevant sequences capable of forming G-DNA structures. Several different G-rich DNA or RNA sequences having potential biological roles have been tested for their ability to form G-DNA or G-RNA structures. DMS treatment of HIV genomic RNA dimer revealed that two blocks of guanines in the region essential for dimerization were protected from methylation (Awang & Sen, 1993). Dimerization of the two copies of HIV genomic RNA is an important step in the HIV-1 life cycle that appears to negatively regulate translation and positively regulate encapsidation (Marquet et al., 1991). Sequences from the immunoglobulin switch regions were shown to form a four-stranded, G-quartet structure (Sen & Gilbert, 1988). Such a structure may have a role in the recombination of these sequences during differentiation of B lymphocytes to plasma cells. The fragile X (CGG)n repeats
form a G-quartet structure which is stabilized by methylation of the cytosine residues of the (CGG)$_n$ repeats (Fry & Loeb, 1994; Hardin et al., 1993). The correlation of repeat expansion and FMR 1 methylation with stabilization of G-quartet structures formed by these sequences suggests that G-quartet structures may help regulate delayed replication and transcriptional silencing observed at the FMR 1 locus (Pieretti et al., 1991; Hansen et al., 1993). Another medically important application of G-DNA research stems from the discovery of a G-DNA aptamer which inhibits thrombin (Bock et al., 1992). The aptamer, d(GGTTGGTGTTGGG), was shown to form an antiparallel quadruplex structure (Macaya et al., 1993; Wang et al., 1993).

The second approach involves the identification and characterization of proteins which bind to G-DNA structures. Several different proteins have been identified which bind to both parallel and antiparallel G-DNA structures. Among these, the strongest functional evidence comes from a yeast nuclease which binds to a parallel-stranded G-DNA structure (Liu et al., 1993). A homozygous deletion of the gene for this protein (KEM 1) blocks meiosis at the 4N stage. FACS analysis of KEM 1 deletion mutants revealed that the mutants underwent premeiotic DNA synthesis but arrested before spore wall formation (Tishkoff et al., 1991). The KEM 1 protein also has strand exchange activity (Kolodner et al., 1987). Taken together, these results suggest a role for KEM 1 in meiotic recombination.

In support of the proposal that G4-DNA structures have a role in meiotic recombination, telomeric sequences have been found to strongly stimulate meiotic recombination (White et al., 1993). Furthermore, RAP 1, which has been shown to bind telomeres in vivo (Conrad et al., 1990) and G-DNA structures (Giraldo & Rhodes, 1994), stimulates meiotic recombination (White et al., 1991). A mutation in the RAP 1 binding site which eliminates RAP 1 binding reduced the frequency of meiotic crossing over (White et al., 1991). Another enzyme implicated in recombination, topoisomerase II, was shown to bind G4-DNA structures (Chung et al., 1992). Topoisomerase II has been implicated in recombination due to its high reactivity toward an alternating purine-pyrimidine sequence which functions as a recombination hot spot in the β-globin gene (Spitzner et al., 1989). Thus, topoisomerase II, RAP 1, and the KEM 1 protein all have potential roles in meiotic
recombination and each of them binds to G4-DNA suggesting that G-DNA structures have a role in meiotic recombination as well.

Continued studies of G4-DNA binding proteins are needed to elucidate the exact biological function of both the proteins and the G-DNA structures to which they bind. *Tetrahymena* may provide a unique opportunity to study parallel-stranded G-DNA/protein complexes. *Tetrahymena* has an abundance of G-rich DNA. Furthermore, it exhibits nuclear dimorphism: Only the micronucleus undergoes meiosis, and the macronucleus divides by fission. Thus, a protein which only binds to parallel-stranded G-DNA in meiotic chromosome pairs may be localized exclusively in the micronucleus. In a previous study, we identified an activity from *Tetrahymena* (TGP) that binds to parallel-stranded G4-DNA but not to anti-parallel G-DNA (Schierer & Henderson, 1994). In this study, we report the identification of the main protein involved in TGP/G4-DNA complexes and its partial purification.

**MATERIALS AND METHODS**

*Tetrahymena strains and cell culture.* *Tetrahymena thermophila* strain C3 V was grown using a rotary shaker at 30°C in 2% PPYS (2% proteose peptone, 0.2% Yeast extract, and 0.003% sequestrine) to mid-log phase (2.5 X 10^5 cells/ml).

*Extract preparation.* *Tetrahymena* whole cell extracts were prepared according to the procedure of Greider and Blackburn (1987) except that the cells were not mated. Furthermore, double distilled H2O was substituted for diethyl pyrocarbonate-treated H2O and 10 mM Tris-HCl (pH 7.5) was substituted for Dryls for cell washes. Briefly, cells at mid-log phase were harvested at 5K rpm for 5 minutes (Sorvall GSA rotor, 4 °C) and washed twice with a 10x volume of cold 10 mM Tris-HCl (pH 7.5). Cell pellets were then resuspended in a 1x volume of TMG buffer (10 mM Tris-HCl pH 7.5, 1 mM MgCl2, 10% glycerol, 10 mM 2-mercaptoethanol) and 1/10 volume of 2% NP-40 was added immediately. Cells were lysed by shaking on a Fisher Genie 2 vortexer (setting 4) for 30’ at 4 °C. The cell lysate was centrifuged at 100,000 x g for 60 minutes at 4 °C. The supernatant, termed S-100 cell extract, was aliquoted and frozen quickly with liquid nitrogen. The final protein concentration was typically 10-20 mg/ml in
TMG buffer. The protease inhibitors leupeptin (0.01 mM), pepstatin (0.01 mM), and Pefabloc (0.1 mM) (Boehringer Mannheim) were included in all solutions.

**DNA oligonucleotide synthesis, purification and 5' end radiolabeling.** Oligo Y (dACTGTCGTACTTGATATGGGGGT; Sen & Gilbert, 1990) was gel purified as previously described (Henderson et al., 1987). Briefly, DNA oligonucleotides were synthesized on an Applied Biosystems DNA Synthesizer followed by deprotection. After boiling in 50% formamide for 4', the oligonucleotides were separated by electrophoresis through 20% polyacrylamide sequencing gels containing 7M urea and 1x TBE buffer. The desired species were identified by UV shadowing and excised from the gel. DNA was eluted by shaking in TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) for 24-48 hours at room temperature, and the eluted oligonucleotides were desalted over a Sep-Pak C18 column (Waters). Gel purified Tet 4 oligonucleotides were 5' end labeled as previously described (Henderson et al., 1987). The labeled Y was then gel purified (12%, 7M urea PAGE) and desalted by C18 chromatography as above.

**Electrophoretic mobility retardation assays and unit definition.** Oligo Y was chosen as the probe for detecting TGP because of its ease of structure formation and because it is only capable of forming one type of G-quartet structure, a parallel-stranded tetrameric quadruplex (Sen & Gilbert, 1990; Laughlan et al., 1994). For mobility retardation assays, 0.5 picomoles of 5' 32P-probe were boiled in the presence of 5 mM KCl, and 5 mM MgCl2, 10 mM Tris-HCl (pH 7.5) and 5% glycerol and slow cooled at room temperature for at least 30' to allow the probe to form structure before transferring to ice. A 200 fold molar excess of oligonucleotide d(T)24 was then added as a nonspecific competitor (5 pmol/ml final concentration) before adding protein to produce a final reaction volume of 20 ml. The final probe concentration was 0.025 pmol/ml, and final concentrations of other components in the binding reaction were 5 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl (pH 7.5), 5% glycerol, 1 mM 2-mercaptoethanol and 0.02% NP-40. Under these reaction conditions, one unit of TGP activity is defined as the binding of 0.5 fmol of Y4 (4-stranded form of oligo Y) as determined by mobility retardation assays and quantitation by phosphorimager analysis (Molecular Dynamics). Binding reactions were incubated on ice for 20 minutes and then loaded onto a 6% non denaturing polyacrylamide gel in 0.6x TBE. Electrophoresis at 10 volts/cm was carried out
at room temperature until the bromophenol blue reached 3/4 the length of the gel. The gel was then dried and the bands were visualized by autoradiography and quantitated using a PhosphorImager (Molecular Dynamics).

**SDS-PAGE and molecular weight estimation.** 6 and 10% separating gels and stacking gels were prepared according to *Current Protocols in Molecular Biology* (Smith, 1987; p. 10.2.1-10.2.9). SDS sample buffer and 1x SDS electrophoresis buffer were prepared according to *Guide to Protein Purification* (Garfin, 1990; p. 430). Protein samples were boiled in a 1-4x volume of sample buffer for 5 minutes and loaded. Electrophoresis was carried out at 16-24 mA constant current at room temperature until the dye front reached the bottom of the gel. Molecular weights of unknown proteins were estimated using a calibration curve prepared by plotting log MW verses Rf for the molecular weight standards (Garfin, 1990; p. 440). The unknown molecular weights were calculated from the Rf using the equation for the calibration curve.

**UV cross-linking.** Cross-linking of Tet 4 species to TGP was performed in situ by exposing a wet 8% EMRA to 254 nm UV light for the indicated number of minutes on ice with the gel no more than 5 cm from the UV source. The gel was exposed to film overnight to reveal the positions of the TGP/Tet 4 complex. Gel pieces containing the TGP/Tet 4 complex were then excised, denatured by boiling 5' in SDS sample buffer, and polymerized into the stacking gel of a 10% SDS-PAGE next to prestained molecular weight standards for molecular weight estimation. DTT was substituted for 2-mercaptoethanol in the sample buffer, since BME inhibited gel polymerization. After electrophoresis at 30 mA, the gel was dried and exposed to X-ray film to reveal the position of cross-linked species. The film was aligned on the dried gel to mark the position of the molecular weight standards and the molecular weight was estimated from the molecular weight standards' calibration curve.

**Preparative Mobility Retardation Assay.** In order to estimate the size the proteins involved in TGP/G4-DNA complexes without the aid of a cross-linked probe, a concentrated preparation of TGP was prepared from the high activity fractions of an S-Sepharose column and used in a mobility retardation assay so that a sufficient amount of protein would be present for staining. 2 ml of high activity S-Sepharose fractions were concentrated to a final volume of 100 ml using Centricon 10 concentrators (Amicon). 10 ml of this concentrated fraction
was used in a standard mobility retardation assay (see above) and the band containing TGP was visualized by autoradiography of the wet gel. The desired gel piece was then excised, denatured as above (see UV cross-linking) and polymerized into the stacking gel of a 6% SDS-PAGE for molecular weight estimation. After electrophoresis at 20 mA constant current, the gel was silver stained using Biorad silver stain according to the manufacturer's instructions. The molecular weight of TGP was estimated from the molecular weight standards' calibration curve.

**Ion Exchange Chromatography.** A Pharmacia S-Sepharose column (4.91 cm$^2 \times 14.5$ cm) was poured and then equilibrated and packed at 3.6 ml/min with 2 column volumes of chromatography buffer (CB: 10 mM Tris/HCl (pH 7.7), 1 mM EDTA, 0.01% NP-40, 10% glycerol, and 1 mM DTT). 10 ml of S100 extract was equilibrated with 0.01 mM leupeptin, 0.1 mM pefabloc and 1 mM EDTA and loaded onto the column at a rate of 0.82 ml/min. After the extract had passed into the column, it was followed with 70 ml of CB at the same flow rate (0.82 ml/min). The loaded column was then washed with 100 ml of CB containing 0.5 M NaCl at a flow rate of 2 ml/min to remove weakly binding proteins from the column. A 200 ml, 0.5-1.5 M NaCl/CB linear gradient was then applied to elute TGP. Fractions were aliquoted and frozen in liquid N$2$. Column fractions were assayed for TGP activity with mobility retardation assays.

A Biorad Biorex 70 column (4.91 cm$^2 \times 50$ cm) was equilibrated according to the manufacturer and then poured and packed with 500 ml of CB. 17 ml of S100 extract was equilibrated with 0.01 mM leupeptin, 0.1 mM pefabloc and 1 mM EDTA and loaded onto the column at a rate of 0.82 ml/min. After the extract had passed into the column, it was followed with 80 ml of CB at the same flow rate (0.82 ml/min) and 40 ml of CB at 1.5 ml/min. The column was then washed with 160 ml of CB containing 450 mM NaCl to remove weakly binding proteins from the column. TGP was then eluted using a 462 ml, 0.45 M-1 M NaCl/CB linear gradient. Fractions were aliquoted and frozen in liquid N$2$. Column fractions were assayed for TGP activity with mobility retardation assays.

**Affinity Chromatography** A 2 ml Affi-blue column (Biorad) was poured and packed according to the manufacturer's instructions. The packed column was equilibrated with 10 column volumes of CB buffer. The load for the Affi-blue column was prepared by combining, concentrating and dialyzing the high-Y4-
binding-activity fractions from a 70 ml S-Sepharose column. The fractions were first dialysed into TMG (10 mM Tris-HCl (pH 7.5), 1 mM MgCl₂, and 10% glycerol) and then concentrated 10 fold by ultrafiltration using YM membranes (Amicon). Approximately 1/10 of the total volume of the the S-Sepharose high activity fractions was loaded onto the 2 ml Affi-blue column at a flow rate of 0.13 ml/min followed by 3 ml of CB buffer. A 0-2M NaCl/CB salt gradient was then applied to fractionate proteins. 20, 1 ml fractions were collected. The protease inhibitors leupeptin (5 μM) and pefabloc (50 μM) from Boehringer Mannheim were included in all wash buffers. Fractions were assayed for Y4-binding activity before being frozen in liquid nitrogen.

A 1 ml Y4-affinity column was constructed by binding preformed, biotinylated Y4 complexes to avidin agarose beads (Pierce). Biotinylated Y4 complexes were formed by adding a 6 fold excess of unbiotinylated oligo Y to biotinylated oligo Y in the presence of 100 mM KCl and 5 mM MgCl₂ to ensure that each complex capable of binding the avidin-agarose beads contained only one biotinylated molecule. The preformed complexes were then added to an equilibrated and packed column of avidin agarose and allowed to bind overnight at 4 °C. Excess oligo Y was washed from the column with 5 column volumes of CB buffer containing 100 mM KCl; 5 mM MgCl₂. The load for the Y4-affinity column was prepared by combining, concentrating and dialysing high activity fractions (fractions 19-23) of the Affi-blue column shown in Figure 3A. These fractions were dialysed into 10 mM Tris-HCl, 1 mM EDTA and 10% glycerol and then concentrated 10 fold by ultrafiltration using YM membranes (Amicon) to a final volume of 200 μl. All of the concentrated Affi-blue material was loaded onto the Y4-affinity column at a flow rate of 0.15 ml/min and followed with 3 ml of CB buffer containing 50 mM KCl, 10 μM leupeptin, and 100 μM pefabloc. The column was then washed with a salt gradient (50 mM-2M NaCl/CB; 5 μM leupeptin, 50 μM pefabloc) to fractionate proteins. 20, 1 ml fractions were collected during the gradient.

**RESULTS**

*Identification of TGP* UV cross-linking analysis was used to identify the protein bands responsible for the shifts seen in mobility shift assays used in the
initial characterization of TGP (Schierer & Henderson, 1994). The TGP band from a mobility retardation assay such as that found in Figure 2A was excised after different UV exposure times in order to tag the protein for visualization on an SDS-PAGE. The excised band was then polymerized into the stacking gel of a 10% SDS-PAGE for estimation of molecular weight. Increasing exposure times result in more extensive cross-linking of the radiolabeled probe to the protein. A cross-linked complex of approximately 97 kDa is visualized after exposure of the protein gel to X-ray film (Figure 1A). In all cases, the molecular weight of the unknown protein bands was estimated from a calibration curve of the molecular weight standards (log MW vs Rf (relative mobility); Garfin, 1990). Because one to four molecules of Tet 4 could be cross-linked to the protein or cross-linked to each other, the size estimation for the protein can range from 89 kDa for one molecule of d(T2G4)4 to 63.8 kDa for four molecules of d(T2G4)4 since each molecule has a molecular weight of 8.4 kDa. Since over 75% of the total Tet 4 molecules remain monomeric or form cross-linked dimers after 20 minutes exposure to UV (Schierer & Henderson, 1994), the estimated TGP/Tet 4 molecular weight can be narrowed to between 89 and 80.6 kDa. A doublet of the main cross-linked species is visible in the 5 minute exposure lane. The position of the two bands in the doublet can be accounted for by the molecular weight of one molecule of Tet 4 suggesting that the upper band has two molecules of Tet 4 cross-linked to the protein and the lower band has one molecule of Tet 4 attached.

A minor DNA/protein cross-linked products was observed around 43 kDa using both Tet 4 and Y4 (data not shown) as the DNA probe. The minor products could represent a degradation product of TGP that still retains the binding domain or it could represent a minor component of the TGP/G4-DNA complex. High activity fractions from S-Sepharose contain a minor band of approximately 43 kDa in addition to the dominant 83 kDa band (Figure 2) suggesting that the minor band is part of the TGP/G4-DNA complex.

A complementary experiment to the cross-linking analysis was performed to more accurately estimate the molecular weight of TGP. As before, the TGP band from a mobility retardation assay was excised and polymerized into the stacking gel of an SDS-PAGE. A 6% SDS-PAGE was used since this percentage provides optimal resolution for proteins in this range. Cross-linking was not performed to visualize the protein. Rather, as much protein as possible was used
Figure 1. UV cross-linking assay to estimate the size of TGP. Radiolabeled Tet 4 molecules were cross-linked to TGP in situ. The band containing the cross-linked complex was excised and polymerized into a 10% SDS-PAGE for size estimation. The main cross-linked complex was resolved at 97 KD and a minor complex at 42 KD. The -UV/-Pro and +UV/-Pro, lane contained probe alone without and with UV treatment. The +UV/+Pro lanes contained TGP/probe complexes exposed to UV for different numbers of minutes. In B, a band containing the TGP/probe complex from a preparative mobility retardation assay was excised and polymerized into the stacking gel of a 6% SDS-PAGE for size estimation (TGP/MRA lane). The SDS gel resolved a protein band of 85 KD. Lane "M" contains molecular weight standards.
in the TGP/G4-DNA binding reaction to provide enough material to visualize on the protein gel without the aid of the radiolabeled probe. Because of the large amount of protein used for the mobility retardation assay, this experiment is termed a preparative mobility retardation assay. In this experiment a single band of approximately 85 kDa was visualized (Figure 1B).

In order to determine if bands corresponding to the molecular weights estimated by UV cross-linking and preparative mobility retardation assays were present in high activity fractions from Biorex 70 and S-Sepharose columns, these column fractions were concentrated and separated by 6% SDS-PAGE. In both Coomassie and silver stain gels, one main band is present which corresponds to the estimated 85 kDa molecular weight. The position of the main band from three different 6% SDS-PAGE was averaged to obtain an approximate molecular weight of 83 kDa. The 83 kDa band is the most abundant protein in these fractions since it is the only band visualized in unconcentrated fractions (Figure 2C, lane 5). In addition, the 83 kDa band is the main band present in the high activity fractions from the two most effective columns used for TGP purification (Biorex 70 and S-Sepharose). The 83 kDa band is the only band identified so far that is consistently present in the high activity fractions of columns used for TGP purification (Table I). Based upon the UV cross-linking analysis, the preparative mobility retardation assay, and protein gels from high activity chromatography fractions, we conclude that the 83 kDa protein is the main component of the TGP/G4-DNA complexes.

**Purification of TGP** Several different columns were tested on an analytical scale (1-10 ml column volumes) to determine their usefulness in purifying TGP from other cellular components (Table I). TGP activity was assayed by testing column fractions for the ability to bind to the four-stranded form of oligo Y (Y4) in standard mobility retardation assays (Figure 2A and B; see also Materials and Methods). The columns having the highest salt elution (Phosphocellulose, Biorex 70, and S-Sepharose) were all cation exchange columns indicating that TGP has a basic binding domain as would be expected for a G4-DNA binding protein. Each of these columns performs well as the first step in the purification scheme as TGP requires at least 500 mM NaCl/CB for elution of bound protein from the column. S-Sepharose has the narrowest activity peak of the cation exchange columns and thus the highest concentration of the 83 kDa band. Comparing the specific
Figure 2. S-Sepharose (A) and Biorex 70 (B) fractionation of TGP. Whole cell S100 extracts were loaded onto each column and TGP activity was eluted using CB buffer containing a high concentration of NaCl. TGP activity from column fractions was assayed using mobility shift assays (A and B) with oligo Y as the probe. Peak TGP activity elutes from both columns above 500 mM NaCl. In part A, lanes 1 and 3 are DNA (Y4) alone; lane 4 represents the eluent collected during loading of the column; lane 5 represents the eluent collected during a 0.5 M NaCl wash step and lanes 6-16 represent column fractions collected during a 0.5 M-1.5 M NaCl gradient. In part B, lane 2 represents the eluent collected during loading of the column; lanes 3-5 represent the eluent collected during a 0.45 M NaCl wash step and lanes 6-16 represent column fractions collected during a 0.45 M-1.5 M NaCl gradient. Part C. Coomassie stained 6% SDS-PAGE analysis of high salt fractions from S-Sepharose and Biorex 70. The 83 KD band is the most abundant protein in the high activity fractions of both columns. In Part C, 20 λ of a Biorex 70 peak activity fraction (BR 70) are qualitatively compared to 10 and 20 λ of an S-Sepharose peak activity fraction ([10], [20]; [un]=unconcentrated, 20λ; [ ] = concentrated), and to S100 (2.5 μg). M= molecular weight standard
Figure 3. Affi-blue and Y4 affinity fractionation of TGP. (A) Fractions containing peak TGP activity from an S-sepharose column were desalted by dialysis, concentrated by ultrafiltration, and loaded onto an Affi-blue column for further fractionation. The activity of the material loaded onto the Affi-blue column is represented by the "SS-load +" lane. "-" lanes represent DNA alone incubated in the CB buffer or S100 extract buffer. "Flow through" fractions were collected from the point at which the column was loaded until one column volume of CB buffer had flowed through. "Wash fractions-Affi-blue" represent the elution of TGP activity using CB buffer containing a high concentration of NaCl (0-1.1 M linear gradient). The NaCl concentration in high activity fractions ranges from 290 mM in fraction 19 to 1.1 M in fraction 23. (B) Fractions 19-22 from the affi-blue column in (A) were desalted by dialysis, concentrated by ultrafiltration and loaded onto a Y4-affinity column. The activity of the material loaded onto the Y4-affinity column is represented by the "AB-load +" lane. "-" lanes represent DNA alone incubated in the CB buffer or S100 extract buffer. "Flow through" fractions were collected from the point at which the column was loaded until one column volume of CB buffer containing 50 mM KCl had flowed through. "Wash fractions-Y4-affinity" represent the elution of TGP activity using CB buffer containing a high concentration of NaCl (0-1.3 M linear gradient). The NaCl concentration in high activity fractions ranges from 670 mM in fraction 13 to 1.3 M in fraction 23.
activity (U/mg) of the starting material (S100 extract) to the specific activity of
the high salt/high activity fractions from the S-Sepharose column, the purification
of the 83 kDa protein by this column is at least 90 fold. In addition, TGP accounts
approximately 50% of the total protein in the high activity fractions as determined
by SDS-PAGE analysis.

<table>
<thead>
<tr>
<th>Matrix Name</th>
<th>Matrix type</th>
<th>[Na+] needed for TGP elution (mM)</th>
<th>Separation from bulk protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biorex 70</td>
<td>Cation exchange</td>
<td>&gt;550</td>
<td>&gt;95%</td>
</tr>
<tr>
<td>S-Sepharose</td>
<td>Cation exchange</td>
<td>&gt;500</td>
<td>&gt;95%</td>
</tr>
<tr>
<td>Phosphocellulose</td>
<td>Cation exchange</td>
<td>&gt;650</td>
<td>&gt;95%</td>
</tr>
<tr>
<td>Y4</td>
<td>Affinity</td>
<td>&gt;400</td>
<td>nd</td>
</tr>
<tr>
<td>Sephacryl S200HR</td>
<td>Size exclusion</td>
<td>na</td>
<td>&gt;49%</td>
</tr>
<tr>
<td>DE52</td>
<td>Affinity</td>
<td>&gt;100</td>
<td>nd</td>
</tr>
<tr>
<td>Heparin agarose</td>
<td>Affinity</td>
<td>&gt; 50</td>
<td>nd</td>
</tr>
<tr>
<td>Q-Sepharose</td>
<td>Affinity</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Phenyl agarose</td>
<td>Hydrophobic</td>
<td>No activity</td>
<td></td>
</tr>
</tbody>
</table>

nd=not determined; na=not applicable

Biorex 70 and S-Sepharose columns were chosen for scale up after the
initial column screen (100-200 ml column volumes). Phosphocellulose (Whatman)
was not chosen for scale up since it is only usable for 1 week after precyling.
Scaled up S-sepharose and Biorex 70 columns performed as effectively as the
analytical scale columns in terms of the elution salt concentration.

Several shifted bands in the mobility retardation assays coelute with the
main TGP band except when testing fractions with very high activity peaks
which contain the highest quantity of the 83 kDa band (Figure 2A, fractions 6 and
7). In addition, when concentrated fractions 6 and 7 are diluted over a gel filtration
column, the minor bands reappear in the mobility retardation assays (data not
shown). This suggests that the presence of these minor bands is dependent on the
concentration of the 83 kDa band and represent (n-1) mers of the fully formed
TGP complex.

TGP was further purified by Affi-blue and Y4-affinity chromatography. The
high Y4-binding-activity fractions from an S-Sepharose column were desalted,
concentrated and loaded onto an Affi-blue column. The high activity fractions
from the Affi-blue column (fractions 19-23) were then desalted, concentrated and
loaded onto a Y4-affinity for an additional 1.75 fold purification over the S-Sepharose column (Figure 3). The total fold purification of the 83 kDa protein is about 150 fold over the S100 starting material. Two bands (83 and 50 kDa) coelute with the highest Y4 binding activity (fraction 15; Figure 3B) and are the most abundant proteins in the high activity fractions (Figure 4). The 83 kDa protein makes up approximately 0.5-1.5% of the total cell protein and the 50 kDa protein makes up approximately 0.1-0.4% of the total cell protein.

**DISCUSSION**

We have reported the identification of the major component of TGP and its partial purification. UV cross-linking and preparative mobility shift assay analysis in conjunction with SDS-PAGE analysis of high TGP-activity fractions reveal a dominant 83 kDa protein. Proteins in addition to the 83 kDa protein may be part of the complex, but these components, if any, are far less abundant than the 83 kDa protein. Three minor shifted bands below the main TGP band consistently coelute with the main TGP band except in the highest activity fractions where the concentration of the 83 kDa protein is the highest (Figure 2A, compare fractions 4 and 5 to fractions 6 and 7). This suggests that the presence of these minor bands is dependent on the concentration of the 83 kDa band and represent (n-1)mers of the TGP complex. The reappearance of the minor shifted bands after gel filtration may be due to dilution of the 83 kDa protein below a critical concentration necessary for complex formation or it may be due to the presence of a different protein involved in the TGP complex. Since there are four consistent bands, the TGP complex may be a homotetramer of the 83 kDa subunit. The minor bands could also represent a component of the complex other than the 83 kDa subunit which has not yet been identified. A 50 kDa protein consistently coelutes with TGP activity and is present in the most highly purified fractions of the Y4-affinity column. This may represent an additional subunit of TGP or a different G4 binding protein. The protein/G4-DNA complex in fractions 17-19 (Figure 3B) shifts into the well. The 50 kDa protein is present in these fractions (Figure 4) but the 83 kDa protein is not indicating that the 50 kDa protein may aggregate under non-denaturing conditions.
Figure 4. 8% SDS-PAGE/silver stain analysis of column fractions. 83 and 50 kDa proteins are consistently present with the peak of TGP1 Y4-binding-activity from different chromatographic separations. "M" contains molecular weight standards. "S100 " contains 5 μl of a 1:50 dilution of S100 extract. "[SS]" contains 10 μl of concentrated, peak-Y4-binding-activity fractions from an S-Sepharose column. "[f14-15]" and "[f18-19]" contain 10 μl of concentrated, peak-Y4-binding-activity fractions from a Y4 affinity column run after an S-Sepharose column (SS/Y4). "[SS/AB]" contains 10 μl of concentrated, peak-Y4-binding-activity fractions from an Affi-blue column run after an S-Sepharose column. "[un] f15 (or f17)" and "[f15] (or 17)" contain 10 μl of un-concentrated ([un]) or concentrated peak-Y4-binding-activity fractions from a Y4 affinity column run after S-Sepharose and Affi-blue columns (SS/AB/Y4).
TGP bound tightly to all cation exchange columns tested (S-Sepharose, Biorex 70, and Phosphocellulose) indicating that TGP has a basic domain. Basic domains have been identified in other G-DNA binding proteins such as topoisomerase II, macrophage scavenger receptors and the β subunit of the Oxytricha telomere end-binding protein. It is the basic domain that has been proposed to provide a surface which brings four strands of G-rich DNA into close proximity so that the formation of a G-quartet structure is facilitated by increasing the local DNA concentration (Fang & Cech, 1993).

In attempts to study the biological role of G-quartet DNA structures in vivo, the study of proteins that bind these structures has proved to be a useful approach. A model for the biological function of one of these proteins (KEM 1) has been proposed by Liu and Gilbert (Liu & Gilbert, 1994). They propose that a parallel-stranded G4-DNA structure is formed by the G-rich strands of the bivalent during meiotic recombination and bound by the KEM 1 protein. The nuclease activity of KEM 1 then cleaves strands to be recombined and the strand exchange activity follows to transfer of chromosomal segments between homologues. The proposed model may be the mechanism responsible for recombination of G-rich sequences such as telomeres and immunoglobulin switch regions. In support of the proposal that G4-DNA structures have a role in meiotic recombination, telomeric sequences have been found to strongly stimulate meiotic recombination (White et al., 1993). Furthermore, RAP 1, which has been shown to bind telomeres in vivo (Conrad et al., 1990) and G-DNA structures (Giraldo & Rhodes, 1994), stimulates meiotic recombination (White et al., 1991). A clue to the biological role of TGP may be provided by its preference for binding G-DNA molecules having a long single stranded tail adjacent to the G-quartet region (Schierer & Henderson, 1994). The single strand binding property allows TGP to fit into the model described for the KEM 1 protein's involvement in meiotic recombination described above.

Once antibodies to the KEM 1 protein are made, cytological studies can be performed to test whether this protein is associated with the synaptonemal complex during meiosis. Furthermore, antibodies to this protein may be useful in disrupting meiotic recombination in wild type cells. Evidence such as this will strongly support the involvement of the KEM 1 protein in meiosis. Another useful
approach would involve making antibodies to the G4-DNA structures themselves as has been done with Z-DNA, but this approach has yet to be accomplished.

Our results lend further evidence that G4-DNA binding proteins are conserved throughout divergent organisms and may provide an opportunity to study the localization of G-DNA binding proteins. As explained in the introduction, a protein which only binds to parallel-stranded G-DNA may be localized exclusively in the micronucleus since the macronucleus does not undergo meiosis. This hypothesis can be tested once labeled antibodies against TGP are made.

REFERENCES

Biochemistry 32, 1899-904.
6621-6625.
88, 9755-9759.
GENERAL CONCLUSIONS

Chapter 1 described the identification of a protein from the ciliated protozoan *Tetrahymena thermophila*, TGP, that binds to parallel-stranded G4-DNA. TGP was identified using mobility retardation assays to test *Tetrahymena* extracts for the presence of proteins that bind to G-DNA structures but not to duplex or single-stranded DNA. In binding competition experiments, molecules capable of forming G4 structures competed for binding to TGP, but non-G4 forming molecules did not. TGP binding also requires a single-stranded region adjacent to the G4 structure. It was also determined that Mg$^{2+}$ facilitates the formation of parallel-stranded G4-DNA structures and that high oligonucleotide concentrations are not required to drive formation of these structures. In addition, G4-DNA and TGP/G4-DNA complexes form readily under physiological salt conditions. These data support the proposal that G4-DNA structures exist in vivo. In Chapter 2, the further characterization and purification of TGP has been pursued. The main component of TGP is an 83 kDa protein that possibly binds to G4-DNA as a homomultimer. Chromatography matrices most useful for TGP purification include the cation exchangers S-Sepharose and Biorex 70.

In relation to the overall goal of determining a biological role for G-quartet structures and the proteins which bind them in the cell, TGP has some important characteristics in common with other G-DNA binding proteins. It binds preferentially to G4-DNA molecules having a single-stranded region adjacent to the G-quartet structure portion of the molecule. At least three other G-DNA binding proteins have been shown to possess this characteristic including the *Oxytricha* β subunit (Fang & Cech, 1993), topoisomerase II (Chung et al., 1992) and the KEM 1 protein (Liu & Gilbert, 1994). Topoisomerase II and KEM 1 both cleave the single stranded region 5' to the G-quartet region. In the model proposed for the function of the KEM 1 protein, the cleaved single strands are transferred to the recipient homologue during meiotic recombination (Liu & Gilbert, 1994). The single strand binding property of TGP indicates that it could also be implicated in meiotic recombination. The possible involvement of TGP in recombination can be tested in vitro with strand exchange assays. These assays have been used to
characterize the mechanism by which bacterial Rec A and eukaryotic Sep 1 proteins promote homologous pairing and strand exchange (Tishkoff et al., 1991).

Another important characteristic of G-DNA binding proteins, in cases where it is known, is the presence of a basic domain thought to provide a surface which brings G-rich strands into close proximity to facilitate G-DNA structure formation. The proposed function of the basic domain in vivo is to act as a molecular chaperone for nucleic acid structure formation (Fang & Cech, 1993). Fang and Cech (1993) showed that the *Oxytricha* β subunit strongly promotes the rate of formation of G-DNA structures. It is likely that TGP has a basic domain such as that possessed by the *Oxytricha* β subunit since TGP binds almost exclusively to cation exchange columns.

In summary, TGP has properties consistent with a role in meiotic recombination such as that implicated for other G-DNA binding proteins. Future research involving TGP can be modeled after the *Oxytricha* β subunit. After purification of this protein along with the α subunit in 1987 (Price & Cech, 1987), it required 4 more years to clone and express the genes for these proteins in *E. coli* (Gray et al., 1991). Two more years were required for kinetic characterization of its ability to promote the formation of G-DNA structures (Fang & Cech, 1993). Characterization of TGP such as has been done with the *Oxytricha* β subunit will likely reveal properties that provide insight into its biological function.
REFERENCES


ACKNOWLEDGEMENTS

Graduate school has been one of the most important experiences of my life. It has caused me to mature, and it has opened up my eyes to exciting opportunities I did not know about before. I will never forget the lessons I have learned here nor the people who helped me through. Although I'm not extremely vocal about it, I feel that I was greatly helped by the direction and spiritual strength provided by my personal relationship with Jesus. I also want to express my sincere thanks to Eric for his honesty and for being a very reasonable and enjoyable PI and to Drena, Al and Basil for their helpful comments and encouragement. In addition, I received a great deal of encouragement and assistance from my lab mates and family.

All the great people in the Henderson/Larson labs made life much more pleasant than it would have been in their absence. Thanks to SA and TCM for their encouragement and for introducing me to the concert scene; Luming and Du for giving me advise about women; Hong, Zhen, and Jamie for their helpfulness, and Audra for all the spiritual discussions. Thanks also to Zita, Ang, Lori, Meilin, Ping, and Joong Kook for all the fun times in lab. I also want to thank my sister and parents for always believing in me when I had difficulty believing in myself.
APPENDIX A: IDENTIFICATION AND CHARACTERIZATION OF A PUTATIVE TELOMERE END-BINDING PROTEIN FROM TETRAHYMENA THERMOPHILA


Hong Sheng, Zhen Hou, Ted Schierer, Drena L. Dobbs, and Eric Henderson*
Department of Zoology and Genetics, Iowa State University,
Ames, Iowa 50011 USA

*To whom correspondence should be addressed
Identification and Characterization of a Putative Telomere End-Binding Protein from *Tetrahymena thermophila*

HONG SHENG, ZHEN HOU, TED SCHIERER, DRENA L. DOBBS, AND ERIC HENDERSON*

Department of Zoology and Genetics, Iowa State University, Ames, Iowa 50011

Received 23 June 1994/Returned for modification 28 July 1994/Accepted 27 October 1994

Telomeric DNA of *Tetrahymena thermophila* consists of a long stretch of (TTGGGG), double-stranded repeats with a single-stranded (TTGGGG), 3' overhang at the end of the chromosome. We have identified and characterized a protein that specifically binds to a synthetic telomeric substrate consisting of duplex DNA and the 3' telomeric repeat overhang. This protein is called TEP (telomere end-binding protein). A change from G to A in the third position of the TTGGGG overhang repeat converts the substrate to a human telomere analog and reduces the binding affinity approximately threefold. Changing two G's to C's in the TTGGGG repeats totally abolishes binding. However, permutation of the *Tetrahymena* repeat sequence has only a minor effect on binding. A duplex structure adjacent to the 3' overhang is required for binding, although the duplex need not contain telomeric repeats. TEP does not bind to G-quartet DNA, which is formed by many G-rich sequences. TEP has a greatly reduced affinity for RNA substrates. The copiency of TEP is at least 10^4 per cell, and it is present under different conditions of cell growth and development, although its level varies. UV cross-linking experiments show that TEP has an apparent molecular mass of ~65 kDa. Unlike other telomere end-binding proteins, TEP is sensitive to high salt concentrations.

Telomeres are the natural ends of eukaryotic chromosomes. They protect chromosomes from nuclease degradation and from end-to-end ligation, ensure complete replication of chromosomes, and are involved in chromosome organization and nuclear architecture (2, 22, 34, 47). Telomeres typically contain an array of short (5- to 8-bp) sequence repeats which are G rich in the strand that extends to the 3' end of the chromosome (2, 22, 34, 47). In those cases studied in molecular detail, it has been shown that the G-rich strand forms a 3' single-stranded overhang of 12 to 16 nucleotides at the chromosomal terminus (20, 25, 32). Most telomeric sequences fit the consensus Cx-w(T/A)x-t (2, 22, 34, 47).

It is important to characterize proteins that bind to telomeres because they are intimately involved in telomere-mediated chromosome stabilization. Moreover, telomere-binding proteins must interact with telomerase, an enzyme involved in telomere replication and maintenance, whose activity is implicated in both cancer and aging (10). Telomeric DNA is associated with two types of proteins in vivo. Internal telomere-binding proteins interact with the duplex region of telomeric repeats. These include PPT, identified in *Physarum polycephalum*, and RAPI, identified in the yeast *Saccharomyces cerevisiae* (3, 4, 8, 9, 27). PPT is a 10-kDa heat-stable protein that binds specifically to the 3' overhang telomeric sequence and the duplex (8). RAPI is a multifunctional protein that, in addition to binding telomeric repeats in yeast cells, binds to the upstream activating sequences of many genes and to silencer elements. Underexpression of RAPI reduces telomere length, whereas overproduction increases both telomere length and heterogeneity (7, 28, 41). Another internal duplex telomere-binding protein has been characterized in extracts of mammalian cells and may bind along the length of mammalian telomeres (48).

A second type of telomere-binding protein binds specifically to the duplex and 3' overhang structures at the telomeric terminus. Telomere end-binding proteins have been isolated from *Euplotes crassus* and *Oxytricha nova* and have recently been identified in *Xenopus* egg extracts (5, 15, 16, 33, 35). The proteins from the ciliate species bind specifically to the T4G4 or T4G2 repeats at the 3' overhang and protect the telomeric DNA from chemical modification and Bal 31 nuclease digestion (15, 16, 33, 35, 36). These protein-telomeric DNA complexes are resistant to high concentrations of salt (e.g., 2 M NaCl or 6 M CsCl) (33, 35). The *Oxytricha* telomere end-binding protein is a 98-kDa heterodimer containing subunits of 56 kDa (α subunit) and 41 kDa (β subunit). Although both subunits are required for maximal binding activity, the binding domain is located entirely in the α subunit. The DNA binding activity is stabilized when the β subunit is present. The *Euplotes* end-binding protein has a single subunit of 51 kDa (33), which is homologous to the α subunit of the *Oxytricha* protein (42).

To further characterize the interaction between telomeric DNA and telomere-binding proteins, we have identified a protein (telomere end-binding protein [TEP]) from *Tetrahymena thermophila* that binds specifically to the 3' overhang telomeric repeats of synthetic telomeres. Our results show that both the (TTGGGG), telomeric overhang sequence and the duplex structure adjacent to it are necessary for TEP binding activity. This protein is distinct from previously identified telomere end-binding proteins in that its binding is salt sensitive. Purification of TEP will provide an excellent opportunity to further investigate the interaction between a telomere end-binding protein and telomerase (17) in *T. thermophila*, the organism in which telomerase is best characterized.

**MATERIALS AND METHODS**

Preparation of cell extracts. Mating-cell extracts were prepared as described previously (18), with several modifications. Briefly, strains *C* and *Cm* were grown to mid-log phase (3 × 10^6 to 5 × 10^6 cells per ml) at 30°C. Cells were then washed twice with 10 mM Tris-Cl (pH 7.5) and resuspended in an equal volume of the wash buffer. After starvation for 24 h, cells from the two cultures were mixed together and incubated for 9 h without shaking. Pairing efficiency was greater than 90%. Cells were harvested by centrifugation and resuspended in 5
TABLE 1. Oligonucleotide probes and competitors

<table>
<thead>
<tr>
<th>Oligonucleotide*</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST (S) 4</td>
<td>AAAACCTAAGCTGTCATCGACGGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>ST (D) 4</td>
<td>AAAACCTAAGCTGTCATCGACGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>SI (S) 4</td>
<td>TTGGACGTATCGACGGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>SI (D) 4</td>
<td>AAAACCTAAGCTGTCATCGACGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>NS (S)</td>
<td>AAAACCTAAGCTGTCATCGACGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>NS (D)</td>
<td>AAAACCTAAGCTGTCATCGACGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>ST (D) GC</td>
<td>AAAACCTAAGCTGTCATCGACGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>ST (D) H</td>
<td>AAAACCTAAGCTGTCATCGACGGGTTGGGGTTGGGGT</td>
</tr>
<tr>
<td>ST (D) R</td>
<td>GCAACCACGACUGACACGACGACGACGACGACGACC</td>
</tr>
<tr>
<td>TelLOOP</td>
<td>ATTGGGTTGGGTTGGGTTGGGTTGGGTTGGGTTGGG</td>
</tr>
</tbody>
</table>

* ST, specific tail; SI, specific internal; (D), double stranded; (S), single stranded; NS, nonspecific sequence; H, human telomere sequence (TTAGGG); GC, mutated version (TTGGG) of the telomere sequence; R, RNA version (UUGGGG) of the telomere sequence; 0 to 4, number of Gs at the end of the telomeric sequences.

For UV cross-linking in situ, radiolabeled ST/D probe (0.8 pmol) was incubated with 5 µl of radiolabeled oligonucleotides [ST(D), ST(S), and NS(D)]. The mixtures were irradiated with 254-nm light for 30 min and separated by sodium dodecyl sulfate (SDS)-10% polyacrylamide gel electrophoresis (PAGE). Gels were dried under vacuum at 80°C, and protein-DNA complexes were visualized by autoradiography.

For UV cross-linking in solution, the reaction mixtures were incubated on ice for 30 min before loading. Gel electrophoresis and autoradiography were performed as described for the EMRA procedure.

**Estimation of TEP abundance.** In a series of binding assays, various amounts of crude cell extract (5106) were incubated with a constant amount of ST/D probe (Table 1). The mixtures were subsequently loaded onto a gel as described above. Radioactivity in specific bands on the gel was quantitated with a Phosphorlmager (Molecular Dynamics). The number of DNA molecules bound by TEP was calculated from the quantitated bound/total ratio and the total number of probe molecules used in each reaction. Assuming that each TEP molecule binds one DNA molecule, the number of bound TEP molecules should be the same as the bound DNA molecules. The amount of TEP in Telomera cells was then estimated from the number of TEP molecules and the number of cells that gave rise to the amount of SI100 extract present in each binding assay.
Identification of TEP. EMRAs were used to identify complexes between cellular proteins and a synthetic DNA probe that mimics the structure found at natural telomeres. This probe, ST(D)4 (see Table 1 for names and description of all probes used in this study), has a duplex portion of random sequence and a 3' extension consisting of two telomeric repeats, (TTGGGG)_2. EMRA analysis of S100 extract from mating cells by using ST(D)4 as the probe revealed a specific DNA-protein complex that migrated more slowly than the free probe (Fig. 1, lanes 1 to 4). A 10-fold and a 30-fold molar excess of unlabeled ST(D)4 effectively competed with the labeled probe for complex formation (Fig. 1, lanes 5 and 6). In contrast, a 30-fold and a 100-fold molar excess of a nonspecific competitor with a similar structure but an altered sequence in the 3' overhang (NS(D); Table 1) did not compete with ST(D)4 (Fig. 1, lanes 7 and 8). Quantitation of the results from several experiments showed that a 10-fold molar excess of unlabeled ST(D)4 competed for more than 90% of the binding activity (Fig. 1, lane 5; see also Fig. 3, lane 2), whereas a 100-fold molar excess of unlabeled NS(D) competed for less than 20% (Fig. 1, lane 8; see also Fig. 3, lane 4). When radiolabeled NS(D) was used as the binding substrate under conditions identical to those used for ST(D)4, no shifted band was observed (Fig. 1, lanes 9 to 13). ST(D)4 and NS(D) have exactly the same sequence in their double-stranded portions, differing only in the 3' overhang, which is (TTGGGG)_2 in ST(D)4 but a random sequence in NS(D). Other experiments indicated that binding did not occur if the telomeric sequence was located internally in the duplex region of the probe (see below). Taken together, these results suggest that the binding observed with ST(D)4 is dependent upon the telomeric repeat sequence in the 3' overhang, as would be expected for a telomere end-binding protein. TEP activity can be attributed to a protein(s), since proteinase K or heat treatment (75°C) of S100 extracts could abolish its binding with ST(D)4 (data not shown). In addition, complex formation between TEP and ST(D)4 was not altered in the presence of RNase A (data not shown), suggesting that RNA is not a component of TEP.

In addition to the specific DNA-TEP complex, two faster-migrating complexes (arrowheads in Fig. 1) were also detected. Neither their appearance nor their susceptibility to competition ...
TEP binding activity is sensitive to high salt concentrations. Previous studies have shown that telomere end-binding proteins found in O. nova, E. crassus, and Xenopus egg extracts are resistant to high salt concentrations (e.g., 2 M NaCl) (4, 33, 35). In contrast, the Tetrahymena DNA-TEP complex is salt sensitive (Fig. 2). Optimal binding for TEP occurred at 0 to 50 mM LiCl, and the binding activity decreased at higher salt concentrations. More than 90% of the binding activity was lost when the salt concentration reached 450 mM (Fig. 2). This result explains why high-salt extraction protocols that were successfully used to purify telomere end-binding proteins in other systems were not successful with T. thermophila (data not shown).

TEP specifically recognizes the 3' overhang of DNA substrates. If TEP is a telomere end-binding protein with properties similar to those found in other species, it would be expected to show specificity with regard to both sequence and arrangement of the two repeats in the 3' overhang. To investigate the sequence specificity, a DNA substrate with an overhang containing the human telomere repeat sequence TTAGGG [ST(D)4; Table 1] was used in competition experiments with radiolabeled ST(D)4 as the probe. Figure 3 shows that the human telomeric sequence competed somewhat less effectively than its Tetrahymena counterpart: ST(D)H required at least a 30-fold molar excess to achieve the same level of competition as was observed with a 10-fold molar excess of unlabeled ST(D)4 (Fig. 3; compare lane 2 with lanes 5 to 7). When the 3' overhang of the radiolabeled probe was changed from (TTAGGG)2 to (TTGCGG)2, no shifted band was detected (Fig. 1, lanes 19 to 21). Thus, TEP can distinguish telomeric from nontelomeric sequences but has only slightly reduced affinity for telomeric sequences from phylogenetically distant species.

To determine whether permutation of the Tetrahymena G-strand sequence alters the binding efficiency of TEP, a series of oligonucleotides with the same length of 3' overhang but different arrangements of G's within the (TTGGGG)2 sequence were tested in competition experiments (Fig. 3). The relative ability of these oligonucleotides to compete for TEP was ST(D)4 ≈ ST(D)3 > ST(D)2 > ST(D)1 > ST(D)0 > NS(D). Therefore, the permutation of the telomere repeat is not critical as long as G's are present at the 3' end.

TEP does not bind G-DNA. G-rich telomeric and nontelomeric oligonucleotides are capable of forming unusual structures that are extremely stable (G-DNA, a four-stranded arrangement stabilized by G tetraplexes [39]). Therefore, it was of interest to determine whether the double-stranded DNA substrates used in this study could form such structures under the assay conditions used. Various combinations of oligonucleotides were allowed to interact with one another under the binding assay conditions (except that SI100 was not present in the mixture) and examined on 12% nondenaturing polyacrylamide gels previously shown to reveal structural variability in telomeric oligonucleotides (21, 46). Under these conditions, DNA bands were observed only in positions expected for Watson-Crick duplexes; no aberrantly migrating species were detected (data not shown). Furthermore, when Li\(^+\), known to decrease the stability of G-quartet structures, was used as the monovalent cation in binding assays, complex formation was not impaired (Fig. 2 and 4). Finally, when ST(S)4 was induced to form an intermolecular G-quartet DNA complex in the presence of K\(^+\), TEP did not bind it (data not shown). Taken together, these results demonstrate that TEP does not require a G-DNA structure for binding.

TEP requires a duplex structure adjacent to the 3' overhang. Complexes formed between S100 proteins and G-rich single-stranded substrates [e.g., ST(S)4 and SI(S)4] were nonspecific in nature and therefore probably unrelated to TEP.
FIG. 3. Effect of alteration and permutation of G-strand overhang sequence on TEP binding. Radiolabeled ST(D)4 (0.06 pmol) was incubated with SI00 extract (2.3 μg) from mating cells in the presence of different competitors. Lane 1, no competitor; lane 2, ST(D)4; lanes 3 and 4, NS(D); lanes 5 to 7, ST(D)H; lanes 8 to 10, ST(D)0; lanes 11 and 12, ST(D)3; lanes 13 and 14, ST(D)2; lanes 15 and 16, ST(D)1. All of these oligonucleotides have the same duplex sequence (Table I). The molar excess of the competitor is indicated above each lane. The arrow marks the DNA-TEP complex. The arrowhead indicates a nonspecific complex. The numbers below the lanes indicate the residual DNA-TEP complex (compared with lane 1) in the presence of a 10-fold molar excess of indicated competitors. One microgram of unlabeled poly[d(I-C)] as a nonspecific competitor was present in all experiments.

(TEP prefers a free 3' end. To test whether the presence of a free 3' end is important for TEP binding, several DNAs in which two single-stranded TTGGGG repeats occupy internal positions were used as competitors. The competitors contained duplexes at either end or both ends. As shown in Fig. 5, none of these three molecules competed as efficiently as ST(D)4, which contains the two single-stranded TTGGGG repeats at the 3' end. These results indicate a strong preference of TEP for substrates with single-stranded TTGGGG repeats at a free 3' end, as would be expected for a telomere end-binding protein (15, 16, 33, 35, 42).

To investigate the sequence requirements in the duplex domain for TEP binding, an oligonucleotide containing the sequence (TTGGGG)_2 in the duplex portion near a random sequence 3' overhang, SI(D)4, was used as a probe in the binding assay. As shown in Fig. 1, lanes 14 to 18, and Fig. 4, lanes 11 to 14, no complex formation was observed under the conditions used. Furthermore, when an oligonucleotide containing both telomeric duplex and overhang regions was used in competition experiments, it competed only marginally better than ST(D)4 (data not shown). This oligonucleotide (Te-LOOP; Table I) contained a tetraloop sequence that minimized formation of slipped structures which would lack the desired overhang and confuse interpretation of competition data. Thus, it appears that TEP binding specificity is dictated by the G-strand overhang and an adjacent duplex, but that the duplex sequence requirements are quite relaxed and telomeric repeats are not required.

TEP does not efficiently bind to RNA oligonucleotides. To investigate the affinity of TEP for RNA analog of telomeric sequence, an RNA oligonucleotide [ST(S)R] with the same sequence as that of ST(S)4 (except one nucleotide at the 5' end) was generated by in vitro transcription (see Materials and Methods). The integrity of the RNA was verified electro-
FIG. 4. Both the (TTGGGG)$_2$ overhang and the adjacent duplex are required for TEP binding. (A) Various probes were incubated in the presence (+) or absence (-) of 5 μg of Sl100 protein from mating cells. LiCl was present at 150 mM in all binding assays. Lanes 1 to 6, ST(D)$^4$; lanes 7 to 10, NS(D); lanes 11 to 14, SI(S)$^4$; lanes 15 to 18, SI(S)$^4$. For the competition experiments, unlabeled ST(D)$^4$ (lanes 3, 4, 9, 13, and 17) or NS(D) (lanes 5, 6, 10, 14, and 18) was used as a competitor. The molar excess of the competitor is indicated above each lane. Binding assays in lanes 1, 2, 7, 8, 11, 12, 15, and 16 did not contain any competitor. The arrow marks the position of the DNA-TEP complex. (B) Radiolabeled ST(D)$^4$ (lanes 1 to 4) or ST(S)$^4$ (lanes 5 to 10) was incubated with or without 6 μg of Sl100 extract from starved cells and examined by EMRA (as described in Materials and Methods except that KCl or LiCl was omitted and 1 mM EDTA was present). Lane 1, 2, 5, and 6, no competitors; lanes 3 and 9, specific competitors; lanes 4 and 10, nonspecific competitors; lanes 7 and 8, single-stranded specific and nonspecific competitors. The molar excess of each competitor is indicated below the competitor designation. The position of the DNA-TEP complex is aligned with that in panel A. One microgram of unlabeled poly(d(I-C)) as a nonspecific competitor was present in all experiments.

phoretically before it was used in the binding assays (data not shown). When ST(S)$^R$ was incubated with Sl100 extract in the presence of a 50-fold molar excess of nonspecific single-stranded RNA, only one radioactive complex was detected (Fig. 6A). It is unlikely that TEP was responsible for this RNA-protein complex, because the latter moved much faster than TEP-DNA complex on the same gel (Fig. 6A, lanes 1 and 3). Moreover, binding competition assays revealed that the RNA-protein interaction is nonspecific in nature, since ST(D)$^4$ (specific) and NS(D) (nonspecific) competitors had equivalent effects on the complex formation (Fig. 6A, lanes 3 to 5). Furthermore, when a (UUGGGG)$_4$ RNA oligonucleotide [(U$_G$)$_4$];

FIG. 5. TEP prefers (TTGGGG)$_2$ repeats at a free 3' end. (A) Radiolabeled ST(D)$^4$ (0.06 pmol) was incubated with Sl100 extract (5 μg) in the presence of a 30-fold molar excess of different competitors. The competitor used is indicated above each lane. (B) The sequences and arrangements of the competitors used in the experiment shown in panel A are listed. One microgram of unlabeled poly(d(I-C)) as a nonspecific competitor was present in all experiments.
FIG. 6. TEP has higher affinity for DNA than for RNA substrates. (A) Radiolabeled ST(D)4 (lane 1) or ST(S)R (lanes 2 to 5) was used as the probe in EMRA in the presence and absence of 6 μg of S100 extract from starved cells. A 50-fold molar excess of nonspecific RNA competitor transcribed from pBluescript SK II (see Materials and Methods) was present in all reactions. Lanes 1 to 3, no DNA competitors; lanes 4 and 5, specific and nonspecific DNA competitors at a 30-fold molar excess. The DNA-TEP complex is indicated by an arrow. (B) Radiolabeled ST(D)4 was incubated with 5 μg of S100 protein from mating cells in the presence or absence of different competitors. Lane 1, no competitor; lanes 2 and 3, 10- and 50-fold molar excesses of unlabeled ST(D)4; lanes 4 and 5, 50- and 100-fold molar excesses of unlabeled NS(D); lanes 6 and 7, 50- and 150-fold molar excesses of unlabeled (UUGGGG)4. The DNA-TEP complex is aligned with the DNA-TEP complex in panel A. One microgram of unlabeled poly[d(I-C)] as a nonspecific competitor was present in all experiments.

Table 1) was used to compete with radiolabeled ST(D)4 for TEP binding, a 50-fold and a 150-fold molar excess of (U2G4)4 caused only 40 and 62% reduction in TEP binding to ST(D)4, whereas a 10-fold molar excess of ST(D)4 abolished 90% of the DNA-TEP complexes under the same conditions (Fig. 6B). Thus, the RNA analog of the G-strand telomeric sequence has a much lower affinity for TEP than the DNA duplex/overhang structure.

**TEP activity varies as a function of cell growth and development.** To investigate the possibility that TEP activity varies as a function of cell growth and/or development, protein extracts were made from mid-log-phase cells, stationary-phase cells, starved cells, and mating cells. TEP was detected in every case, although the relative activity levels differed (Fig. 7). The activity was highest in starved cell extracts and stationary-phase cell extracts and lower in mid-log-phase cell extracts and mating cell extracts. This trend was reproducible, although the absolute activity levels varied in different trials. In an attempt to ensure that most of the TEP was released from the cellular DNA during extract preparation, and thus available for extraction, the salt concentration was adjusted to 300 mM before the cells were lysed in a control experiment. No significant change in the amount of TEP in S100 extracts was detected in binding assays (data not shown).

**TEP has a copy number of at least 2 × 10^4 per cell and a molecular mass of approximately 65 kDa.** The abundance of TEP in *Tetrahymena* cells was estimated by quantitation of EMRA gels like the one shown in Fig. 1 (see Materials and Methods); on the basis of the calculation, there are at least 2 × 10^4 TEP molecules per mating cell. This value roughly corresponds to the number of telomeres per cell in *T. thermophila* (~4 × 10^4).

The molecular mass of TEP was estimated by UV cross-linking experiments both in solution and in situ (see Materials and Methods). UV cross-linking in solution gave rise to only one prominent protein that was specifically UV cross-linked to ST(D)4 (Fig. 8, lanes 1 to 4). An ST(D)4-protein complex of similar gel migration rate was also evident by UV cross-linking in situ (data not shown). No specific DNA-protein complex was observed when NS(D) or ST(S)4 was used as the probe (Fig. 8, lanes 5 to 14). This result is perfectly consistent with those obtained from EMRAs (e.g., Fig. 4), strongly suggesting that TEP is responsible for the specific DNA-protein complex in lanes 2 and 4 of Fig. 8. Prestained protein size markers run on the same gel were used to estimate the approximate molecular mass of TEP. Previous work has shown that under the conditions used, protein-DNA complexes usually migrate with the same electrophoretic mobility as the protein alone (44), although this is not always true (23). The apparent molecular mass of TEP is ~65 kDa.

**DISCUSSION**

Identification and characterization of telomere-binding proteins from a wide variety of eukaryotes will facilitate our understanding of how telomeres function. In this paper, we re-
port the identification of a relatively salt-sensitive protein from *T. thermophila* which specifically binds to the (TTGGGG)$_2$ sequence present in the 3' overhang of synthetic telomeres. This protein, TEP, binds to synthetic telomeres having two repeats of the *Tetrahymena* G-strand telomeric sequence as long as this sequence is adjacent to duplex DNA. However, TEP does not exhibit a strong requirement for a telomere sequence in the duplex region. This is consistent with TEP being an end-binding factor, with other proteins being responsible for binding to the exclusively duplex region. TEP showed roughly equal binding affinity to all possible permutations of the *Tetrahymena* G-strand sequence. This could be due to insensitivity to subtle differences in our assay or reflect the presence of ragged ends at natural telomeres in *T. thermophila*. Finally, it is curious that TEP binding is decreased only about threefold upon changing the *Tetrahymena* sequence to that found in human and other telomeres (TTAGGG), suggesting that the A residue in the altered sequence is not critical for complex formation. Thus, TEP is a good candidate for a telomere end-binding protein analogous to those characterized in other eukaryotes (11, 15, 16, 33, 35, 42) but with somewhat relaxed sequence and structural requirements for its substrate.

Unusual DNA structures stabilized by G tetrads are formed by many telomeric G-strand repeat sequences and other G-rich sequences. This form of DNA, G-DNA, was originally characterized with guanine derivatives (1, 19) and later characterized in telomeric G-strand sequences (21, 38-40, 45, 46). G-DNA's role at telomeres or elsewhere in the chromosome remains unclear. Nonetheless, several proteins have been identified in various organisms, including *T. thermophila*, that specifically bind to G-quartet DNA, suggesting that it does have a physiological role (6, 12, 13, 26, 37, 43). Recently, the β subunit of the telomere end-binding protein from *O. nova* and RAP1 in *S. cerevisiae* have been shown to facilitate G-DNA formation in vitro (13, 14). In contrast to these proteins, TEP does not bind to G-DNA.

Unlike other telomere end-binding proteins, TEP binding is sensitive to salt concentration. The salt-resistant nature of telomere-binding proteins in other ciliates facilitated their identification and purification (5, 15, 33, 35). As more is learned about TEP, it will be of interest to determine what differences between its mode of binding and that of its putative homologs in other species give rise to its salt-sensitive characteristic.

Several vertebrate DNA-binding factors with affinity for single-stranded telomeric TTAGGG repeats were recently identified in nuclear extracts of murine and HeLa cells (24, 29, 30). Subsequently these factors were shown to be components of heterogeneous nuclear ribonucleoproteins. These factors bind more tightly to single-stranded RNA oligonucleotides having r(UUAGGG) repeats than to DNA of the same sequence. We tested the RNA version of the DNA G-strand sequence used in this study and found binding to be greatly reduced. Additional studies using (UUGGGG)$_4$ as a binding substrate corroborated this observation. Therefore, telomeric DNA, rather than RNA containing telomeric sequences, is likely to be the natural binding substrate of TEP.

TEP was detected in cells grown under a variety of physiological conditions including log phase, stationary phase, star-
FIG. 8. UV cross-linking indicates a molecular mass for TEP of ~65 kDa. Radiolabeled ST(D)4 (lanes 1 to 4), NS(D) (lanes 5 to 8), or ST(S)4 (lanes 9 to 14) was incubated alone (lanes 1, 5, and 9) or allowed to bind TEP (6 μg of S100 extract from starved cells in lanes 2 to 4, 6 to 8, and 10 to 14) in the presence and absence of a 30-fold molar excess of various types of oligonucleotide competitors as indicated above the lanes. The reaction mixtures were irradiated with UV light and separated by SDS-PAGE. Protein size markers are indicated at the left in kilodaltons. The specific DNA-TEP complex revealed by SDS-PAGE is marked by an arrow. One microgram of unlabeled poly[d(I-C)] as a nonspecific competitor was present in all experiments.

vation, and conjugation. It was somewhat surprising that starved cells and stationary-phase cells had the highest levels of TEP since it might be predicted that TEP would be up-regulated following mating or during log-phase growth, when new telomeres are being generated at a rapid rate. A possible explanation for this observation is that the proportion of TEP in starved cell and stationary-phase cell extracts is higher because of a reduction in the concentration of other cellular proteins. Therefore, at a given total protein concentration in starved or stationary-phase cell extracts, the TEP activity would appear to be elevated relative to that in nonstarved and log-phase cells. Additional studies will be necessary to understand the regulation of TEP expression during cell growth and development in T. thermophila.

Purification and further characterization of TEP should provide insight into the mechanism by which telomeres function. Moreover, since telomerase, the enzyme responsible for telomere replication and maintenance, has been best characterized in T. thermophila, the interplay of this fascinating enzyme with other telomere-binding factors can be further investigated in this well-studied system.

ACKNOWLEDGMENTS

We thank A. Miller for suggestions on experiments involving RNA, and we thank S. Ahmed and A. Myers for critically reviewing the manuscript.

This work was supported by Public Health Service grant GM-41023 from the National Institutes of Health.

REFERENCES

binder to and promotes the formation of DNA quadruplexes in telomeric DNA. EMBO J. 13:2411-2420.


