2000

Two Tetrahymena G-DNA-binding proteins, TGP1 and TGP3, share novel motifs and may play a role in micronuclear division

Quan Lu
Iowa State University

Eric Henderson
Iowa State University, telomere@iastate.edu

Nucleic Acids Research, 2000, Vol. 28, No. 15 **2993–3001**

Two *Tetrahymena* G-DNA-binding proteins, TGP1 and TGP3, share novel motifs and may play a role in micronuclear division

Quan Lu and Eric Henderson*

Department of Zoology and Genetics, Molecular, Cellular and Developmental Biology Program, Iowa State University, Ames, IA 50011, USA

Received February 22, 2000; Revised May 10, 2000; Accepted June 13, 2000 DDBJ/EMBL/GenBank accession no. AF136448

ABSTRACT

G-DNA is a four-stranded DNA structure with diverse putative biological roles. We have previously purified and cloned a novel G-DNA-binding protein TGP1 from the ciliate *Tetrahymena thermophila*. Here we report the molecular cloning of TGP3, an additional G-DNA-binding protein from the same organism. The TGP3 cDNA encodes a 365 amino acid protein that is homologous to TGP1 (34% identity and 44% similarity). The proteins share a sequence pattern that contains two novel repetitive and homologous motifs flanking an extensively hydrophilic and basic region. A nuclear fractionation experiment showed that TGP1 and TGP3 activities are localized predominantly in the nuclear fraction. To further investigate the biological roles of the proteins *in vivo*, we have generated separate macronuclear gene knockout (KO) strains (TGP1KO and TGP3KO) for each of the two genes. Southern blot analysis demonstrated that the macronuclear copies of each gene were completely disrupted. Mobility shift assays showed that the corresponding G-DNA-binding activity for each protein was abolished in the KO strains. Growth analysis showed that both KO strains grew at near wild-type rates, indicating that neither of the genes is essential for cell growth. Nevertheless, nuclear staining analysis revealed that both TGP1KO and TGP3KO cells have an increased occurrence (more than 2-fold) of extra micronuclei, implying faulty control of micronuclear division in the KO cells.

INTRODUCTION

G-DNA, also known as G4, G-quartet or G-tetraplex DNA, is a family of four-stranded DNA structures characterized by a novel motif called the G-quartet (1,2). In a G-quartet, four guanine molecules (one from each of the four DNA strands) lie in a plane with each guanine forming G:G hydrogen bonds with its two neighbors. Although the existence of G-DNA *in vivo* remains undetermined, many G-rich sequences of biological relevance can readily form the G-DNA structure *in vitro* under near physiological conditions. These sequences include most telomeric DNA (3–5), an immunoglobulin switch region (6), microsatellite sequences (7) and a few gene promoters (8–10). Once formed, G-DNA is usually more stable than double-stranded DNA, evoking the suggestion that this degree of stability may preclude the dynamic properties required for biological activity. However, recent identification of several proteins with G-DNA unwinding activities (11–13) argues in favor of a functional role for G-DNA *in vivo*.

Evidence is accumulating which implies that G-DNA plays roles in diverse biological processes. First, G-DNA has implied function(s) at telomeres, the terminal structure of the eukaryotic chromosome (14). Two telomere-binding proteins, the yeast RAP1 protein (15) and *Oxytricha* TEBP β subunit (16), were found to facilitate the formation of G-DNA *in vitro*. G-DNA formation at telomeres may regulate the activity of telomerase, the enzyme responsible for the regulation of telomere length, as suggested by a study showing that telomerase was inhibited by G-DNA primers (17). Second, G-DNA may have roles in recombination. The yeast KEM1 protein, which is essential for meiotic recombination, is a G-DNA-specific nuclease (18,19). LR1, a B cell-specific protein that binds to the immunoglobulin switch region (which is G-rich), was found to exhibit G-DNA-binding activity (20). Third, G-DNA may play a role in gene regulation. This is supported by studies identifying G-DNA-binding activities in proteins known to be involved in the control of gene expression, including the transcription factor MyoD (21) and eukaryotic topoisomerase II (22). Finally, G-DNA may play other roles *in vivo*, as novel G-DNA-binding proteins, including a hepatocyte chromatin protein QUAD (23) and two yeast proteins G4p1 and G4p2 (24,25), have been identified.

We have studied G-DNA and G-DNA-binding proteins in the ciliate *Tetrahymena thermophila*. *Tetrahymena* is a potentially rich source of G-DNA since its macronuclear genome is fragmented and thus contains a large proportion of telomere DNA (G-rich T_2G_4 repeats) (26). Three G-DNA-binding activities (TGP1, TGP2 and TGP3) had been identified from *Tetrahymena* total protein extracts (27–29). While TGP2 contains dyhydrolipoamide dehydrogenase activity (29),

*To whom correspondence should be addressed. Tel: +1 515 296 6128; Fax: +1 515 296 6570; Email: eric@bioforcelab.com

TGP1 was cloned and shown to be a novel protein (27). In this paper, we report the molecular cloning of a third *Tetrahymena* G-DNA-binding protein, TGP3. TGP3 is homologous to TGP1 and both proteins share an intriguing sequence pattern that may constitute a novel putative G-DNA-binding domain. To investigate the biological functions of these two novel proteins, we have generated macronuclear gene disruption strains for each of the genes separately. Both gene knockout (KO) strains (TGP1KO and TGP3KO) grow at near normal rates, indicating that neither gene is essential for cell growth and survival. However, nuclear staining analysis revealed that in both KO cells the percentage of cells containing multiple (more than two) micronuclei is much higher than that in the wild-type cells. These data suggest that the control of micronuclear division in the KO cells is faulty.

MATERIALS AND METHODS

Tetrahymena cell culture and extract preparation

Cells were cultured in 2% PPYS medium (2% proteose peptone, 0.2% yeast extract, 10 μM $FeCl_3$) at 30°C with constant shaking (125 r.p.m.). Cell cultures were not generally allowed to exceed a concentration of 5×10^5 cells/ml. Cells were transferred to fresh medium at 1:1000 dilution. Stocks were maintained at room temperature in low aeration in either 1% PPYS for 1 month or soybean medium (1 soybean autoclaved in 10 ml ddH_2O) for 6 months. *Tetrahymena* total protein extract was prepared as previously described (28). Briefly, *T.thermophila* cells (strain C3V) were grown to mid log phase (2.5×10^5 cells/ml) and harvested. Cells were washed twice with 10 mM Tris–HCl, pH 7.5, resuspended in 5 vol of TMG buffer (10 mM Tris–HCl, pH 7.5, 1 mM $MgCl_2$, 10% glycerol and 10 mM β-mercaptoethanol) with protease inhibitors (0.01 mM leupeptin, 0.01 mM pepstatin and 0.1 mM Pefabloc; all from Boehringer-Mannheim, Indianapolis, IN) and lysed by addition of a 1/10 vol of 2% NP-40 (Sigma, St Louis, MO). The cell lysate was centrifuged at 100 000 *g* for 70 min. The supernatant (S100) was immediately frozen in liquid nitrogen and stored at –70°C.

Protein purification and peptide sequencing

TGP3 was purified using essentially the same protocol previously described for TGP1 purification (27). Three chromatography columns were used for TGP3 purification, including a SP-Sepharose column, a DE52 column and a G-DNA affinity column. Partially purified TGP3 (~5 μg) was resolved by 10% SDS–PAGE and transferred to PVDF membrane. The protein band was excised and used for N-terminal peptide sequencing (ISU Protein Facility). For internal peptide sequencing ~5 μg TGP3 protein transferred to PVDF membrane was digested with CNBr according to the method of Smith (30). Digested peptides were separated by 16% SDS–PAGE and transferred to PVDF membrane. Each peptide band was excised and subjected to N-terminal peptide sequencing.

TGP3 cDNA and gene cloning

Mid log phase *T.thermophila* (strain CU428) cells were used for both RNA and DNA extractions. Total genomic DNA was prepared as previously described (31). RNA extraction was performed using the TriZol reagent (Life Technologies, Rockville, MD) according to a protocol recommended by the manufacturer. Rapid amplification of cDNA ends (RACE) was performed using RACE kits from Gibco BRL. For 3′-RACE, first strand cDNA was synthesized using a poly(dT)-anchor primer (Gibco BRL). Subsequent PCR was performed with the first strand cDNAs using an anchor and a partially degenerate primer TGP3U [5′-TGCAGAAC(T/C)AACAA(T/C)TA(T/C)AGAAAA] corresponding to the peptide cRTNNYRK. The predominant PCR product was cloned and sequenced (ISU DNA Facility). For 5'-RACE, two gene-specific primers, 3R1 (5′-TGTTAGTGTTGTTGTTGTTGC) and 3R2 (5′-AGCTTAGTGGAATCTCTTAGGC), were synthesized based on sequences obtained from the 3′-RACE. Primer 3R2 was used in first round PCR and 3R1 was used as a nested primer in second round PCR. The predominant PCR product was cloned and sequenced. The genomic copy of the TGP3 gene was cloned by PCR with *Tetrahymena* genomic DNA using two gene-specific primers termed 3U (5′-TAACAACTAAGTCTCTCCTC) and 3R (5′-ATTCACTCATTGCTTAGTGGC). A 1.7 kb PCR product was purified and sequenced.

Sequence analyses

The GCG (Genetics Computer Group, Madison, WI) software package was used for sequence analyses. BLAST searches were performed over the World Wide Web at the National Center of Biotechnology Information. The program Boxshade was used to process multiple sequence alignments.

Nuclear fractionation

Tetrahymena nuclear fractionation was performed using a protocol slightly modified from Higashinakagawa (32). Essentially, $\sim 1 \times 10^7$ mid log phase cells were harvested and resuspended in 10 vol of ice-cold 0.25 M sucrose and 10 mM $MgCl_2$. Cells were lysed by the addition of NP-40 (final concentration 0.2%) followed by vigorous pipetting (usually 50 times) on ice. The extent of cell lysis was checked by methyl green staining and microscopy to ensure that >80% of the cells were lysed. Solid sucrose was added to the lysed cell solution at a final concentration of 2.1 M. The mixture was stirred vigorously at the maximal setting (Vortex Genie 2; Fisher, Pittsburgh, PA) at 4°C for 40 min and centrifuged at 50 000 r.p.m. (Ti70.1 rotor; Beckman-Coulter, Fullerton, CA) for 2 h. The supernatant was dialyzed against a dialysis buffer (10 mM Tris–HCl, 7.5, 1 mM $MgCl_2$, 5% glycerol) with protease inhibitors (0.01 mM leupeptin, 0.01 mM pepstatin, 0.1 mM Pefabloc) at 4°C overnight and saved as *Tetrahymena* cytoplasmic extract. The pellet containing mostly nuclei (confirmed by methyl green staining) was rinsed twice, resuspended in the dialysis buffer and lysed by the addition of a 1/10 vol of 2% NP-40 and maximum stirring at 4°C for 30 min. The lysed nuclear mixture was centrifuged at 50 000 r.p.m. (Ti70.1 rotor) and the supernatant saved as *Tetrahymena* nuclear extract.

Targeted disruption of somatic TGP1 and TGP3 genes

p4T2-1 (a gift from D.M. Gorovsky, University of Rochester) is a plasmid containing the neomycin resistance (*neo*) gene cassette, whose expression confers paromomycin resistance in *Tetrahymena*. pGEXT-TGP1 and pGEXT-TGP3 are pGEXT-easy vectors (Promega, Madison, WI) containing cloned TGP1 and TGP3 genes (full genomic copies with all exons and introns

included), respectively. Unique cloning sites (*Xho*I and *Bam*HI for TGP1; *Xho*I and *Xma*I for TGP3) were created in an exon of the cloned TGP gene by an inside-out PCR amplification of pGEXT-TGP vectors. Primers used were: 1KOU (*Xho*I) (5′-CCGCTCGAGGCTAAGGTAGCTGTCATTC) and 1KOR (*Bam*HI) (5′-CGGGATCCCACTGCTGCTATCCAAGCTG) for pGEX-TGP1; 3KOU (*Xho*I) (5′-AACTCGAGATAATTC-CTCCTCTTCCTG) and 3KOR (*Xma*I) (5′-TTCCCGGGTTA-TCTGTTTTAACAGCGGC) for pGEX-TGP3. Amplified PCR products and the p4T2-1 plasmid were digested with appropriate restriction enzymes indicated in the primers and ligated to create gene disruption vectors. The disruption vectors (pTGP1KO and pTGP3KO) thus contain TGP genes that are disrupted by the *neo* gene cassette.

A gold particle-mediated biolistic gun protocol was used for *Tetrahymena* transformation (33). Disruption vectors were linearized by *Apa*I digestion, purified, dissolved in ddH_2O at ~2 µg/µl and used to coat gold particles as follows. Five microliters of DNA (2 µg/µl) was mixed with 25 µl of 1.0 µm gold particles (Bio-Rad, Hercules, CA) (pretreated and stored at –20°C in 50% glycerol at a concentration of 60 mg/ml), 10 µl of 0.1 M spermidine (Sigma) and 25 µl of 2.5 M $CaCl_2$. The mixture was then vortexed (setting 4, Vortex Genie 2) at 4°C for 10 min. The gold particles were washed once with 70% ethanol and once with 100% ethanol and then resuspended in 20 µl of 100% ethanol. Ten microliters of the resuspended DNA-coated gold particles was loaded onto a flying disk (Bio-Rad) and air dried. About 1×10^7 mid log phase *Tetrahymena* cells were starved in 10 mM Tris–HCl, pH 7.5, at 30°C with shaking for ~15 h, washed, resuspended in 1 ml of 10 mM HEPES buffer, pH 7.5, spread onto a moist filter paper (Whatman, Rockland, MA) and bombarded with DNA-coated gold particles at 900 p.s.i. using a biolistic PDS-1000/He particle delivery system (Bio-Rad). Bombarded cells were immediately resuspended in 50 ml of 2% PPYS medium, cultured at 30°C for 4 h to allow cell recovery and plated into 96-well microtiter plates at 150 µl/well. Paromomycin (Sigma) was added to the wells to a final concentration of 120 µg/ml. After incubation in a humid chamber for at least 3 days, wells with actively growing cells were counted and replicated to plates with fresh 2% PPYS medium containing 200 µg/ml paromomycin.

Southern blot analysis

About 15 µg *Tetrahymena* genomic DNA was digested with restriction enzyme(s) at 37°C overnight, purified by phenol/chloroform extraction and ethanol precipitation and separated on 0.8% agarose gel. The gels were depurinated, denatured and blotted onto MagnaGraph nylon membrane. Membrane blots were probed with ^{32}P-labeled TGP1 or TGP3 gene fragment (Random-Primer Labeling Kit; Promega). Hybridization was performed according to a standard protocol (34). The blots were exposed to X-ray films or to a Phosphorimager screen (Molecular Dynamics, Sunnyvale, CA).

Growth analysis

Mid log phase cells were used to inoculate 50 ml of 2% PPYS medium at an initial cell density of 200 cells/ml. Cells were cultured at 30°C with constant shaking (125 r.p.m.). At various time points during the culturing, 100 µl of the cell culture was taken out to determine the culture density using a Coulter counter (Beckman-Coulter). The data were then plotted and the doubling time was estimated by curve fitting (Microsoft Excel; Microsoft, Seattle, WA). The analysis (cell culturing and counting) was repeated at least twice to minimize sample variations.

Mobility shift assay

Mobility shift assays were performed as previously described (27). Fifty nanograms of ^{32}P-labeled oligo Y (ACTGTCG-TACTTGATATGGGGGT) was boiled for 3 min in G-DNA formation buffer (50 mM KCl, 10 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.0, 5% glycerol), cooled and incubated at room temperature for at least 30 min. Usually, >90% of oligo Y is in G-DNA form after the boiling–cooling procedure. About 2.5 ng labeled Y(G4) was mixed with protein extract in a binding reaction buffer (10 mM Tris–HCl, pH 7.5, 6% glycerol). One hundred-fold (~250 ng) non-specific competitor poly(dI-dC)·poly(dI-dC) (Pharmacia, Peapack, NJ) was added to each binding reaction. The total volume of each reaction was 20 µl. After incubation on ice for 20 min, the reaction mixtures were loaded onto a 6% polyacrylamide gel. Electrophoresis was carried out in 0.6× TBE at room temperature. Gels were vacuum dried and exposed to X-ray film or to a Phosphorimager screen (Molecular Dynamics).

Nuclear staining

Tetrahymena cells in mid log phase were washed, resuspended in 10 mM Tris–HCl, pH 7.5, at a concentration of 10^5 cells/ml, fixed by addition of a one thousandth volume of 37% formaldehyde at room temperature for 1 min and stained by addition of a 1/10 vol of Hoechst DNA dye (Sigma) solution (10 mg/ml) for 10 min. Nuclei were visualized using an BH-2 fluorescence microscope (Olympus, Melville, NY). Images were captured and processed in the program NIH Image.

RESULTS

Molecular cloning of TGP3

We have previously purified and cloned a novel G-DNA-binding protein, TGP1, from *T.thermophila* (27). During TGP1 purification, a 40 kDa protein was found to co-purify with TGP1 (~82 kDa) and account for an additional G-DNA-binding activity, and this was referred to as TGP3. Because of its apparent co-purification with TGP1, TGP3 was purified using essentially the same procedure as that for TGP1. The highest purified fractions contained two predominant proteins (80 kDa TGP1 and 40 kDa TGP3) as revealed by silver stained SDS–PAGE (27).

To clone the TGP3 gene, direct N-terminal peptide sequencing was first performed with purified TGP3 protein. However, no clear sequence data were obtained, probably due to N-terminal blockage of the purified protein. Internal peptide sequencing was thus attempted. Purified TGP3 protein was digested with CNBr, resulting in three major peptides (data not shown). Each of the peptides was subjected to N-terminal sequencing and one of the peptides yielded a clear sequence (RTNNYRKQNNNQQRKNN). Based on this sequence, along with consideration of *Tetrahymena* genetic codon usage (35), a long partially degenerate primer (TGP3U) was designed and used directly in a 3'-RACE [the other primer was the

A

MAEQVTNNQVSPQGEQKPKVQKRPRTVQLESKDPVKIADLQPRNNNSNFV
GKVIEVQVLEKGNNKQGNPRKFLKGLIGDDTGVVRFDLAVKNDVVFKVDD
VVSFDKAMNKVNKDGHHYIEVKRFGKYEITNGSIAAVKTDNNISTKIIPP
LPEGEKKQKIIKKSNNKQNNNGKVNNNNNNNGDNNTNNNGNKMNNRTNNY
RKQNNNQQRKNNNNNNSNNNNTNNSQNYKNYKQKPKRFHQALLTPIKNLK
PGQNGQTVKGKVFDVQSYQKTIKNREATFFKGRVADDTANINFDFMMKEK
TISEGDMVIFTNIMNKVSETGHHYIVVGKYGNYSVLNEKQININEQENKS
SIEYAISSSNNATKQ

B

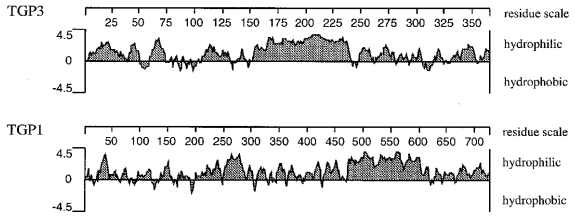

Figure 1. (**A**) Deduced TGP3 protein sequence. The peptide sequence obtained from internal peptide sequencing is underlined. Putative nuclear localization sequences (NLS) are dot underlined. The cDNA sequence has been deposited in GenBank (accession no. AF136448). (**B**) Hydropathy analysis of TGP1 and TGP3 proteins using the algorithm of Kyte and Doolittle (51). The hydropathy profiles of the two proteins are very similar: both proteins are very hydrophilic and contain a long and continuous hydrophilic region (residues 470–610 for TGP1 and 154–240 for TGP3). The long and continuous hydrophilic regions are also rich in asparagine and basic amino acid residues (data not shown).

poly(dT)-anchor primer provided in the RACE kit]. A PCR product of ~600 bp was amplified. Full-length TGP3 cDNA was obtained by doing a 5'-RACE using the sequence information generated from the 3'-RACE. The deduced amino acid sequence (Fig. 1A) contains the peptide obtained from the direct protein sequencing and the predicted molecular weight (40.2 kDa) of the deduced protein is consistent with the TGP3 protein size observed on SDS–PAGE (27). These data suggest that the cloned TGP3 cDNA sequence is most likely complete and correct. In addition to the cDNA cloning, TGP3 genomic DNA was also cloned based on a PCR strategy. Two primers (one from the 5'-end of the TGP3 cDNA and the other from the 3'-end) were used in PCR amplification of *Tetrahymena* genomic DNA. A predominant 1.7 kb band was amplified and sequenced (data not shown). The genomic DNA contains six exons and five introns. The sequence of the exons matches exactly the cloned TGP3 cDNA sequence, confirming the authenticity of the cDNA sequence obtained. All the introns (ranging from 42 to 187 bp) are AT-rich and contain consensus eukaryotic splicing sites.

TGP3 is homologous to TGP1

TGP3 protein has several notable sequence features. First, it is rich in basic residues and has a predicted pI of 10.67. Second, it contains a long and continuously hydrophilic region (residues 154–240; Fig. 1B), which is also rich in asparagine and basic residues. Third, it contains two putative nuclear localization signals (NLS) (dot underlined in Fig. 1A). Notably, all three features are also shared by TGP1 (27). Namely, TGP1 has a pI of ~10.5, has a hydrophobicity plot strikingly similar to that of TGP3 (Fig. 1B) and contains a putative NLS.

BLAST searches revealed that TGP1 and TGP3 are homologous to each other. Sequence alignment showed that TGP3 is homologous to the 3'-end of TGP1 (Fig. 2A). The sequence identity between the two proteins is 34% and the level of similarity is 44%. The alignment creates a long artificial gap in the TGP3 sequence. Immediately following this gap are the long and continuous hydrophilic regions. The alignment between these regions is of very low complexity (most identical residues are asparagines) and, thus, probably only reflects the hydrophilic and asparagine-rich nature of the regions. In contrast, the alignment between sequences flanking the gap and the hydrophilic regions is of high complexity and the identical residues are often in clusters. For the purpose of clarity in the following section, we designate the sequence upstream of the hydrophilic region (and the gap) region A and the downstream sequence region B.

Novel repetitive and homologous motifs in TGP sequences

When we used region A of the TGP3 sequence in a BLAST search, we found that it is homologous not only to region A of TGP1 (which is expected from the sequence alignment shown in Fig. 2A) but also to region B of TGP1. Using region B in a BLAST search, we found a similar situation. This cross-homology between the A and B regions indicates that the four regions (two from each protein) are homologous to each other. Subsequent multiple sequence alignment revealed that similarities exist among the four regions (Fig. 2B). There are 18 identical amino acid residues among all the four regions and 22 additional sites identical among at least three of the four regions. Even though these identical sites constitute only ~22% of the whole sequence in the alignment, most of these sites tend to be in clusters, suggesting that the similarities are functionally significant. A pattern-initiated BLAST search (36) revealed no other proteins sharing such motifs. In summary, sequence analysis and comparison showed that TGP1 and TGP3 share a similar sequence pattern that contains two novel repetitive and homologous motifs flanking an extensively hydrophilic and basic region.

TGP1 and TGP3 activities are localized predominantly to the nuclei

To study the *in vivo* roles of these two novel G-DNA-binding proteins, we first determined the subcellular localization of the proteins. Using a standard protocol we fractionated *Tetrahymena* cells into cytoplasmic and nuclear extracts, which were then tested for TGP G-DNA-binding activities in mobility shift assays (Fig. 3). The cytoplasmic extract contained TGP2 activity but almost no TGP1 and TGP3 activities. In contrast, the nuclear extract contained both TGP1 and TGP3 activities but no TGP2 activity. TGP2 has been identified as a dihydrolipoamide dehydrogenase (29), which is localized mainly in the mitochondrial membrane and, thus, it is expected to be found in the cytoplasmic fraction. The fractionation data demonstrate that, in contrast to TGP2, TGP1 and TGP3 activities are localized mainly to the nuclei. This result is consistent with the identification of NLSs and the DNA-binding nature of TGP1 and TGP3 proteins.

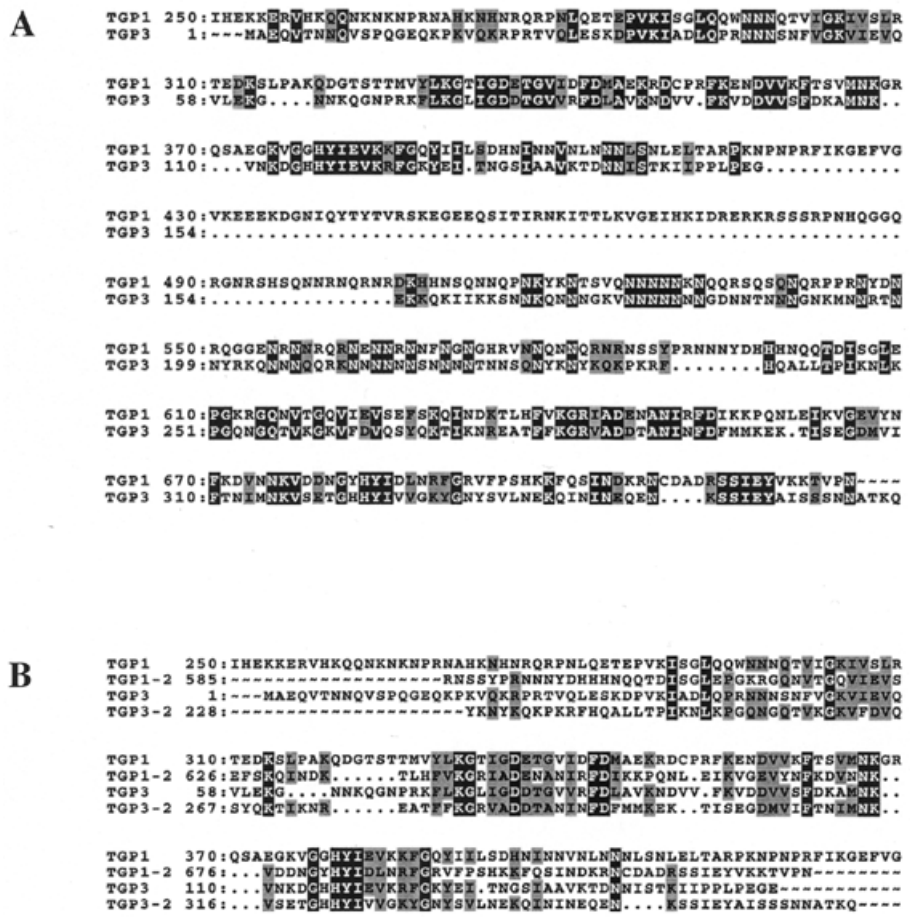

Figure 2. TGP1 and TGP3 are homologous and share conserved motifs. (**A**) Sequence alignment of TGP3 and TGP1. (**B**) Sequence alignment of four domains (two from TGP1 and two from TGP3) that flank the long and continuous hydrophilic regions. Identical and similar amino acid residues are indicated by shading in dark and gray, respectively. For the alignment of the four domains (B), only residues that are conserved in at least three of the four domains are shaded.

Macronuclear transformation of KO vectors

To directly study the biological function(s) of the proteins, the TGP1 and TGP3 genes were separately disrupted in the macronucleus. Gene disruption vectors (pTGP1KO and pTGP3KO for TGP1 and TGP3, respectively) were constructed by inserting a *neo* gene cassette (whose expression confers paromomycin resistance in *Tetrahymena* cells; 37) into an exon (to ensure that the *neo* cassette would not be excised during transcription of TGP genes) of the target gene (Fig. 4A and B). The KO vectors were separately transformed into starved *Tetrahymena* CU428 cells using a gold particle-mediated transformation protocol (33). Several dozen transformants were obtained for both TGP1 and TGP3 disruption vectors. Two transformants from each group (1KOA and 1KOB for TGP1; 3KOA and 3KOB for TGP3) were randomly selected for further analyses.

Complete macronuclear gene disruption in KO cells

The *Tetrahymena* macronucleus is not diploid (as is the micronucleus) but rather contains an average of ~50 copies for each gene (26). In a typical transformation, gene disruption initially occurs through homologous recombination in only very few (most likely one) copies of the target gene. However, the macronucleus divides amitotically (distributing gene copies randomly to daughter cells), resulting in a phenomenon known as phenotypic assortment. Under selection pressure, the initial disrupted gene copies tend to be sorted to homogeneity. Thus, unless there is counter-selective pressure to maintain copies of the wild-type gene, complete gene disruption will be achieved. On the other hand, a gene that is essential for viability can only be partially disrupted.

Selected KO transformants (1KOA, 1KOB, 3KOA and 3KOB) were cultured in 500 μg/ml paromomycin medium for at least 200 generations to allow complete phenotypic assortment. Genomic DNA was then isolated from the KO cells and used in Southern blot analysis to determine the extent of gene disruption (Fig. 4C and D). For the analysis of TGP1 gene disruption (Fig. 4C), a TGP1 gene-specific probe was hybridized to genomic DNA digested by *Eco*RI or *Eco*RV. In the wild-type genomic DNA, a single band was detected with either *Eco*RI (2.4 kb) or *Eco*RV (8.0 kb) digestion. In DNA from the two TGP1KO strains (1KOA and 1KOB), the hybridized band shifted to 4.0 (*Eco*RI) or 4.2 kb (*Eco*RV), indicating insertion of the *neo* cassette into the TGP1 gene locus. No hybridized band of wild-type size was detected in the TGP1KO cells, suggesting that the TGP1 gene was completely disrupted. Similarly, for the analysis of TGP3 gene disruption (Fig. 4D), a TGP3 gene-specific probe was hybridized to

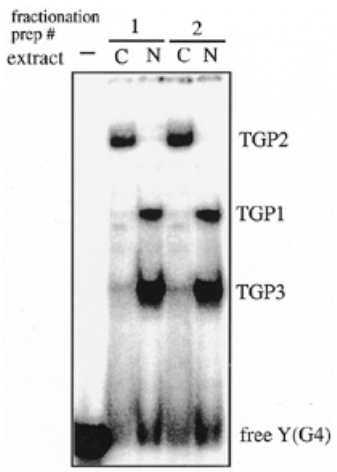

Figure 3. TGP1 and TGP3 activities are localized predominantly to the nuclei. *Tetrahymena* cells were fractionated into cytoplasmic (C) and nuclear (N) fractions according to a standard method (25). Protein extracts were made from the fractions and tested for G-DNA-binding activities in mobility shift assay. The first lane is a control with no protein extract. While TGP2, which is a dihydrolipoamide dehydrogenase, was found mainly in the cytoplasmic fraction, the TGP1 and TGP3 activities were found predominantly in the nuclear fraction. Two separate fractionation experiments were done (prep #1 and #2) to show reproducibility of the assay.

*Eco*RI- or *Eco*RV-digested genomic DNA. The hybridized band shifted from ~6 kb in wild-type cells to ~7 kb in the TGP3 KO cells (*Eco*RI digestion) or from ~12 to ~7 kb (*Eco*RV digestion), indicating that the *neo* cassette had been inserted into the TGP3 gene locus. No wild-type band was detected in TGP3KO genomic DNA, suggesting complete disruption of the TGP3 gene.

To further confirm complete disruption of the TGP genes in the macronucleus, the KO cells were cultured in medium containing no paromomycin. If a gene has not been completely disrupted (meaning wild-type copies of the gene still exist in the macronuclear genome), culturing without drug selection will result in a reversion of the wild-type gene (through phenotypic assortment). However, if a gene has been completely disrupted, no such reversion will occur. After growth in paromomycin-free medium for 2 weeks, DNA from the KO cells was checked by Southern blot analysis. No reversion of the wild-type gene was observed in either TGP1KO or TGP3KO cells (data not shown), demonstrating that each of the genes had been completely disrupted in the macronucleus.

The G-DNA-binding activities are abolished in the KO cells

Since the *Tetrahymena* macronucleus is transcriptionally active and responsible for all protein expression in the cell, complete gene disruption in this nucleus should result in the elimination of protein expression of the targeted gene. To determine whether the expression of TGP proteins had been abolished in the KO cells, we assayed the G-DNA-binding activities in total protein extracts made from the wild-type, TGP1KO and TGP3KO cells (Fig. 5). While the extract from wild-type cells contained three G-DNA-binding activities

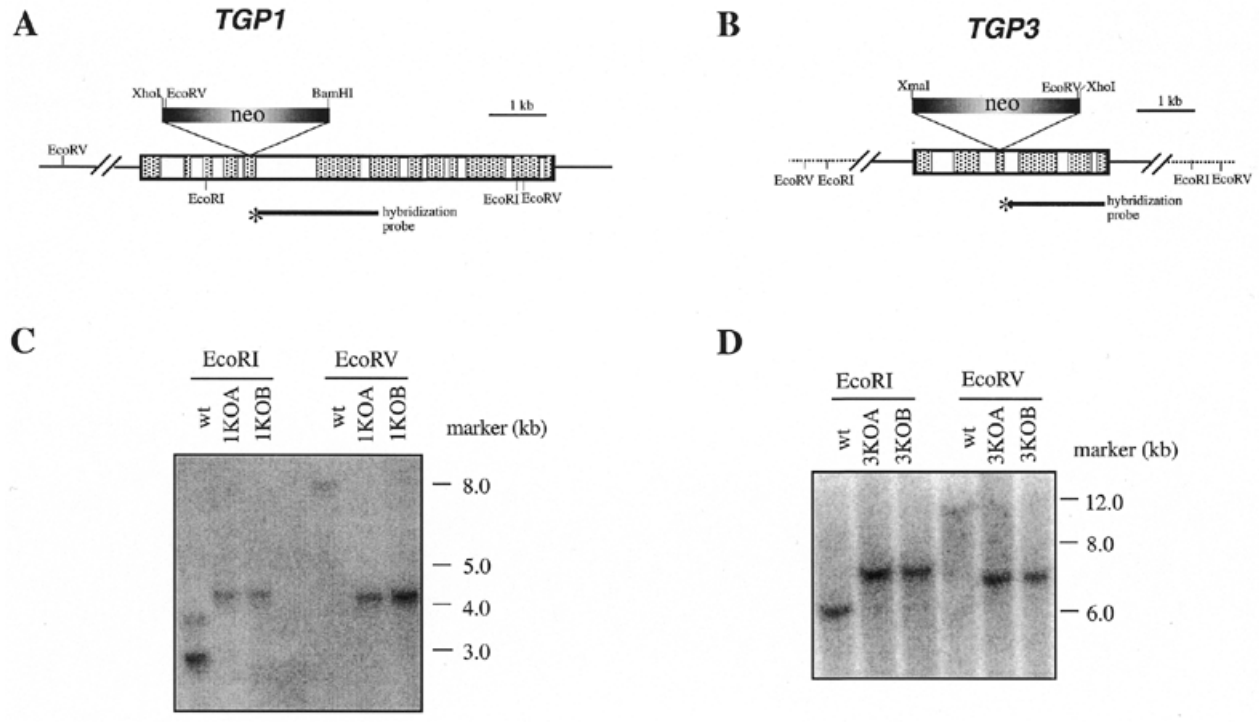

Figure 4. The TGP1 (or TGP3) gene in the macronucleus was completely disrupted. (**A** and **B**) Diagrams of macronuclear TGP1 and TGP3 genomic organizations. Exon–intron structures of the genes are shown with the exons indicated by the dotted pattern. Gene disruption was done by inserting the *neo* gene cassette into an exon (the fifth exon for TGP1 and the third for TGP3) of the target TGP gene. A 1 kb scale bar is shown. (**C** and **D**) Southern blot analysis of TGP1 (TGP3) KO strains. wt, wild-type; 1KOA and 1KOB, two of the TGP1 KO stains; 3KOA and 3KOB, two of the TGP3 KO stains. Genomic DNA was extracted, digested with *Eco*RI or *Eco*RV and hybridized with TGP gene-specific probes. Probes used in the hybridization are indicated in (A) and (B).

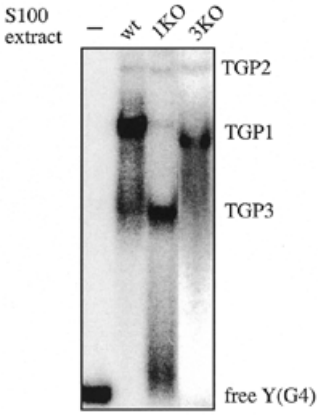

Figure 5. Mobility shift assay of total protein extracts from wild-type, TGP1KO and TGP3KO cells. A roughly equal amount of protein extract (from ~10^3 cells) was used in each lane (except the first lane, which is a control with no protein extract).

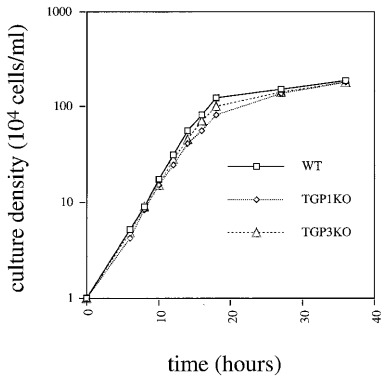

Figure 6. Growth curves of wild-type, TGP1KO and TGP3KO cells. Cells were cultured at 30°C in 2% PPYS medium with constant shaking (125 r.p.m.). Culture density was determined by cell counting at different time points and plotted as a growth curve on a logarithmic scale. Cell doubling time was determined by curve fitting. KO cells have a slightly slower doubling time (TGP1, 2.75 h; TGP3KO, 2.6 h; wild-type, 2.45 h).

(TGP1, TGP2 and TGP3), TGP1KO cells exhibited only two G-DNA-binding activities (TGP2 and TGP3) but had no TGP1 activity, indicating that TGP1 activity had been abolished in the TGP1KO cells. Similarly, the TGP3 activity was lost in TGP3KO cells. The result showed that complete disruption of the targeted TGP gene in the macronucleus results in abolition of the corresponding G-DNA-binding activity.

In addition to loss of the particular G-DNA-binding activity, the mobility shift assay showed a change in the other G-DNA-binding activity in the KO cells (Fig. 5). For example, in TGP1KO cells where TGP1 activity was abolished, there was an increase in TGP3 activity compared with that in the wild-type cells. This is intriguing, since TGP1 and TGP3 share significant homology with each other. Therefore, one might expect that there is certain level of redundancy between the two proteins. However, in TGP3KO cells, TGP1 activity did not increase as substantially as TGP3 did in TGP1KO cells. Instead, the TGP1 activity shifted to a lower position, suggesting that the loss of TGP3 may somehow affect the architecture of the G-DNA–TGP1 complex.

KO cells grow at near normal rate

To determine if the disruption of TGP genes affects cell growth, we assayed the growth rates of the KO cells. Wild-type and KO cells were cultured in parallel under optimal growth conditions (30°C, 125 r.p.m. shaking, 2% PPYS medium). Cell numbers were usually counted every 2 h during the culturing and the data were used to construct the growth curves shown in Figure 6. Cell doubling rates were determined by curve fitting. The doubling rates of TGP1KO and TGP3KO cells were about 2.75 and 2.6 h, respectively. Both doubling rates are only slightly slower than that of the wild-type cells (2.45 h). In addition to cell growth rate, we also examined the cell morphology and swimming rate, but found no apparent defects in these aspects in the KO cells (data not shown). These data indicate that, while there might be some subtle growth changes in the KO cells, disruption of either the TGP1 or TGP3 gene has no profound effects on cell growth and behavior, suggesting that neither TGP1 nor TGP3 is essential.

Extra micronuclei in KO cells

The TGP1 and TGP3 proteins are specific DNA-binding proteins and predominantly localize to the nuclei, as shown by the nuclear fractionation experiment. To determine if disruption of the TGP genes affects nuclear structure, we examined the nuclei in the KO cells. Nuclei of TGP1KO, TGP3KO and wild-type cells were stained with Hoechst DNA dye and imaged using fluorescence microscopy. The normal *Tetrahymena* nuclear structure is shown in Figure 7A. Initially, a cell contains one micronucleus and one macronucleus. The micronucleus begins to divide, forming two micronuclei, followed by macronuclear elongation and separation. The nuclei are then distributed into daughter cells. In normal cells, there should be no more than two micronuclei. However, in both TGP1KO and TGP3KO cells, we observed large numbers of cells containing more than two micronuclei (sometimes up to six) (Fig. 7B). To obtain statistically significant data, a large number (~3000) of cells from each group (wild-type, TGP1KO and TGP3KO cell lines) were stained with Hoechst DNA dye and the number of micronuclei in the cell was determined. In both TGP1KO and TGP3KO cells, the percentages of multiple micronuclei were significantly higher than that in the wild-type cells (Fig. 7C). For example, cells containing three micronuclei account for 2.5% of TGP1KO cells and 3.5% of TGP3KO cells, while only 1% of wild-type cells have three micronuclei. In the case of cells containing four or more micronuclei, the percentages of such cells in TGP1KO and TGP3KO strains are also higher than that in wild-type cells. In addition, percentages of cells containing two nuclei are higher in KO cells than those in wild-type cells. In summary, our data revealed an increased occurrence of multiple micronuclei in both TGP1KO and TGP3KO cells.

DISCUSSION

The functional significance of G-DNA in biological systems remains elusive and requires further exploration. We previously cloned a novel *Tetrahymena* G-quartet DNA-binding protein, TGP1 (27). In this paper we report the molecular

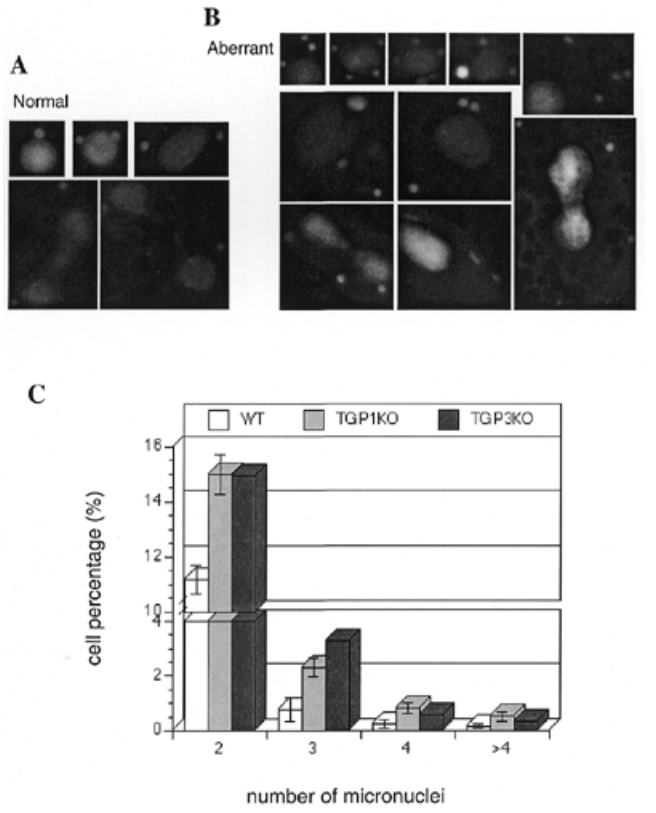

Figure 7. Increased occurrence of extra micronuclei in both TGP1KO and TGP3KO strains. (**A** and **B**) Normal and aberrant nuclear structures. *Tetrahymena* nuclei were stained with Hoechst DNA dye. Images of stained nuclei were captured and processed using the program NIH Image. A *Tetrahymena* cell usually contains no more than two micronuclei (A). In the KO cell lines, there are a large number of cells containing more than two micronuclei (B). (**C**) Percentages of cells containing multiple micronuclei in wild-type and KO cells. The experiment was repeated three times over a 1 month period. About 3000 cells from each cell line were counted according to the number of micronuclei in the cell.

cloning of an additional *Tetrahymena* G-DNA-binding protein, TGP3, that is homologous to TGP1. Furthermore, we have investigated the biological role(s) of the novel proteins by functional analyses, including gene disruption.

Novel motifs shared by TGP1 and TGP3

To our knowledge, the homology between TGP1 and TGP3 represents the first such finding among all known G-quartet-binding proteins. The two proteins share an intriguing and novel sequence pattern. This sequence pattern is composed of two repetitive motifs flanking an extensively hydrophilic and basic region. We speculate that this sequence pattern may constitute a novel G-quartet-specific binding domain or motif. A model consistent with the observed sequence arrangement is shown in Figure 8. In this model the two flanking repetitive regions specifically recognize and bind to G-quartet DNA, while the hydrophilic and basic region, which is of low sequence complexity, increases the binding strength through electrostatic interaction with the DNA substrate. Future detailed domain analysis by serial deletion will help identify a minimal domain accounting for G-DNA binding.

We found no significant homologs of TGPs in available protein databases nor did we find proteins with the general

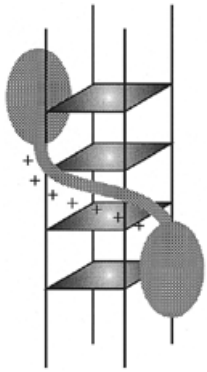

Figure 8. Model for TGP–G-DNA binding. The two homologous motifs in the TGP proteins specifically recognize and bind to the G-DNA, while the extensively hydrophilic and basic region between the homologous motifs strengthens the protein–G-DNA binding through electrostatic interaction with the G-DNA.

sequence pattern shared by TGP1 and TGP3. However, it is possible that TGP-like proteins exist in other organisms but their relationship to TGP1 and TGP3 in *Tetrahymena* is masked by changes in primary sequence. This is the case for several ciliate proteins. For example, the *Oxytricha* protein TEBP has no apparent sequence homologs but contains structural motifs called oligonucleotide/oligosaccharide folds that are shared by many other proteins (38). It may be most productive to search for TGP homologs in other ciliates such as *Oxytricha* and *Euplotes*. Identification of TGP homologs from these organisms will help define shared higher order features such as the putative G-DNA-binding domain and facilitate the search for homologs in more distantly related organisms.

Biological role(s) of TGP proteins

We showed by a nuclear fractionation experiment that both TGP1 and TGP3 are localized predominantly to the nuclei. This is consistent with the presence of NLSs in both proteins. We began to address the possible biological role(s) of these proteins by disrupting the genes encoding them in the macronucleus. We observed an increased occurrence of extra micronuclei in both TGP1KO and TGP3KO cells. One plausible explanation for this extra micronuclear phenotype is that the micronucleus in the KO cells divides prematurely and becomes partially uncoupled from cellular division. However, the mechanism by which a G-DNA-binding protein could cause this is not clear. Recently, telomere–telomere association has been suggested to be involved in the process of mitotic nuclear division (39,40). A study by Kirk *et al*. (41) showed that a mutation in *Tetrahymena* telomeric DNA repeats ($T_2G_4 \rightarrow T_4G_4$) caused severely delayed micronuclear division. Since most telomeres contain a single-stranded G-rich overhang (42–47) that can adopt the G-DNA structure *in vitro*, one possibility is that the TGP proteins are involved in G-DNA-mediated telomere association. In this model, removal of the proteins could cause weakened association at the telomeres and thus allow sister chromatids to separate prematurely, producing the extra micronuclear phenotype observed in the TGPKO cells. However, we have not demonstrated any TGP–telomere association at this time and it is equally possible that the TGP proteins interact with G-DNA domains other than those proposed to occur at telomeres. Additional studies using, for example, recently

developed G-DNA-specific dyes (48) or antibodies (49,50) along with immunolocalization data are needed to determine the sub-nuclear location of TGP1 and TGP3.

Other than the faulty control of micronuclear division, we did not detect any significant morphological or physiological changes in the KO cells. Both TGP1 and TGP3 KO cells grow at near normal rate and have a normal morphology, suggesting that neither TGP1 nor TGP3 is an essential gene. Since TGP1 and TGP3 share significant homology with each other, one might expect that the proteins may have similar functions *in vivo* and, furthermore, there could be a certain level of functional redundancy between the proteins. Indeed, we found that TGP3 activity increases in TGP1 KO cells, suggesting that TGP3 may compensate for the loss of TGP1 by increasing its expression. This possible functional redundancy between TGP1 and TGP3 may explain the relatively mild phenotype we observed in the KO cells. Future emphasis will be on creating a TGP1/TGP3 double KO in anticipation of a more severe and revealing phenotype.

In summary, we have cloned a new *Tetrahymena* G-DNA-binding protein TGP3. This protein shares a putative G-DNA-binding domain with a previously cloned *Tetrahymena* G-DNA-binding protein TGP1. Functional analyses suggest that these two novel proteins localize in the nuclei and may play a role in micronuclear division. Future studies, including protein domain analysis, immunolocalization and double gene disruption, will be needed to further elucidate the biological functions of the two novel proteins and the G-DNA structure with which they interact.

ACKNOWLEDGEMENTS

The authors want to thank Scott Schaus for critical reading of the manuscript. Q.L. was supported in part by fellowships from the MCDB program.

REFERENCES

1. Guschlbauer,W., Chantot,J.F. and Thiele,D. (1990) *J. Biomol. Struct. Dyn.*, **8**, 491–511.
2. Henderson,E. (1995) In Blackburn,E.H. (ed.), *Telomeres*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 11–34.
3. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco,J.J. and Blackburn,E.H. (1987) *Cell*, **51**, 899–908.
4. Sundquist,W.I. and Klug,A. (1989) *Nature*, **342**, 825–829.
5. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) *Cell*, **59**, 871–880.
6. Sen,D. and Gilbert,W. (1988) *Nature*, **334**, 364–366.
7. Weitzmann,M.N., Woodford,K.J. and Usdin,K. (1998) *J. Biol. Chem.*, **273**, 30742–30749.
8. Evans,T., Schon,E., Gora-Maslak,G., Patterson,J. and Efstratiadis,A. (1984) *Nucleic Acids Res.*, **12**, 8043–8058.
9. Kipatrick,M.W., Torri,A., Kang,D.S., Engler,J.A. and Wells,R.D. (1986) *J. Biol. Chem.*, **261**, 11350–11354.
10. Lewis,C.D., Clark,S.P., Felsenfeld,G. and Gould,H. (1988) *Genes Dev.*, **2**, 863–873.
11. Baran,N., Pucshansky,L., Marco,Y., Benjamin,S. and Manor,H. (1997) *Nucleic Acids Res.*, **25**, 297–303.
12. Harrington,C., Lan,Y. and Akman,S.A. (1997) *J. Biol. Chem.*, **272**, 24631–24636.
13. Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) *J. Biol. Chem.*, **273**, 27587–27592.
14. Zakian,V.A. (1995) *Science*, **270**, 1601–1607.
15. Giraldo,R. and Rhodes,D. (1994) *EMBO J.*, **13**, 2411–2420.
16. Fang,G. and Cech,T.R. (1993) *Cell*, **74**, 875–885.
17. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) *Nature*, **350**, 718–720.
18. Liu,Z., Frantz,J.D., Gilbert,W. and Tye,B.-K. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3157–3161.
19. Liu,Z. and Gilbert,W. (1994) *Cell*, **77**, 1083–1092.
20. Dempsey,L.A., Sun,H., Hanakahi,L.A. and Maizels,N. (1999) *J. Biol. Chem.*, **274**, 1066–1071.
21. Chung,I.K., Mehta,V.B., Spitzner,J.R. and Muller,M.T. (1992) *Nucleic Acids Res.*, **20**, 1973–1977.
22. Walsh,K. and Gualberto,A. (1992) *J. Biol. Chem.*, **267**, 13714–13718.
23. Weiman-Shomer,P. and Fry,M. (1993) *J. Biol. Chem.*, **268**, 3306–3312.
24. Frantz,J.D. and Gilbert,W. (1995) *J. Biol. Chem.*, **270**, 20692–20697.
25. Frantz,J.D. and Gilbert,W. (1995) *J. Biol. Chem.*, **270**, 9413–9419.
26. Prescott,D.M. (1994) *Microbiol. Rev.*, **58**, 233–267.
27. Lu,Q., Schierer,T., Kang,S.G. and Henderson,E. (1998) *Nucleic Acids Res.*, **26**, 1613–1620.
28. Schierer,T. and Henderson,E. (1994) *Biochemistry*, **33**, 2240–2246.
29. Kee,K., Niu,L. and Henderson,E. (1998) *Biochemistry*, **37**, 4224–4234.
30. Smith,B.J. (1993) In Walker,J.M. (ed.), *The Protein Protocols Handbook*. Humana Press, Totowa, NJ, pp. 369–373.
31. Larson,D., Spangler,E.A. and Blackburn,E.H. (1987) *Cell*, **50**, 477–483.
32. Higashinakagawa,T., Narushima-Iio,M., Saiga,H., Kondo,S. and Mita,T. (1992) *Chromosoma*, **101**, 413–419.
33. Cassidy-Hanley,D., Bowen,J., Lee,J.H., Cole,E., VerPlank,L.A., Gaertig,J., Gorovsky,M.A. and Bruns,P.J. (1997) *Genetics*, **146**, 135–147.
34. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.M., Seidman,J.G., Smith,J.A. and Struhl,K. (1992) *Current Protocols in Molecular Biology*. John Wiley & Sons, New York, NY.
35. Martindale,D.W. (1989) *J. Protozool.*, **36**, 29–34.
36. Zhang,Z., Schaffer,A.A., Miller,W., Madden,T.L., Lipman,D.J., Koonin,E.V. and Altschul,S.F. (1998) *Nucleic Acids Res.*, **26**, 3986–3990.
37. Kahn,R.W., Andersen,B.H. and Brunk,C.F. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 9295–9299.
38. Horvath,M.P., Schweiker,V.L., Bevilacqua,J.M., Ruggles,J.A. and Schultz,S.C. (1998) *Cell*, **95**, 963–974.
39. Hawley,R.S. (1997) *Science*, **276**, 1215.
40. Hawley,R.S. (1997) *Science*, **275**, 1441–1443.
41. Kirk,K.E., Harmon,B.P., Reichart,I.K., Sedat,J.W. and Blackburn,E.H. (1997) *Science*, **275**, 1478–1481.
42. Klobutcher,L.A., Swanton,M.T., Donini,P. and Prescott,D.M. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 3015–3019.
43. Pluta,A.F., Kaine,B.P. and Spear,B.B. (1982) *Nucleic Acids Res.*, **10**, 8145–8154.
44. Henderson,E. and Blackburn,E. (1989) *Mol. Cell. Biol.*, **9**, 345–348.
45. Dionne,I. and Wellinger,R.J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 13902–13907.
46. Makarov,V.L., Hirose,Y. and Langmore,J.P. (1997) *Cell*, **88**, 657–666.
47. Wright,W.E., Tesmer,V.M., Huffman,K.E., Levene,S.D. and Shay,J.W. (1997) *Genes Dev.*, **11**, 2801–2809.
48. Arthanari,H., Basu,S., Kawano,T.L. and Bolton,P.H. (1998) *Nucleic Acids Res.*, **26**, 3724–3728.
49. Brown,J.C., Brown,B.A., Li,Y. and Hardin,C.C. (1998) *Biochemistry*, **37**, 16338–16348.
50. Brown,B.A., Li,Y., Brown,J.C., Hardin,C.C., Roberts,J.F., Pelsue,S.C. and Shultz,L.D. (1998) *Biochemistry*, **37**, 16325–16337.
51. Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.