

1998

Purification, characterization and molecular cloning of TGP1, a novel G-DNA binding protein from *Tetrahymena thermophila*


Quan Lu
Iowa State University

Ted Schierer
Iowa State University

Sang-Gu Kang
Iowa State University

Eric Henderson
Iowa State University, telomere@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/zool_pubs

 Part of the [Cell and Developmental Biology Commons](#), [Genetics Commons](#), [Genomics Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Lu, Quan; Schierer, Ted; Kang, Sang-Gu; and Henderson, Eric, "Purification, characterization and molecular cloning of TGP1, a novel G-DNA binding protein from *Tetrahymena thermophila*" (1998). *Zoology and Genetics Publications*. 33.
http://lib.dr.iastate.edu/zool_pubs/33

This Article is brought to you for free and open access by the Zoology and Genetics at Iowa State University Digital Repository. It has been accepted for inclusion in Zoology and Genetics Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Purification, characterization and molecular cloning of TGP1, a novel G-DNA binding protein from *Tetrahymena thermophila*

Quan Lu, Ted Schierer, Sang-Gu Kang and Eric Henderson*

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

Received December 23, 1997; Revised and Accepted February 16, 1998

DDBJ/EMBL/GenBank accession no. AF006380

ABSTRACT

G-DNA, a polymorphic family of four-stranded DNA structures, has been proposed to play roles in a variety of biological processes including telomere function, meiotic recombination and gene regulation. Here we report the purification and cloning of TGP1, a G-DNA specific binding protein from *Tetrahymena thermophila*. TGP1 was purified by three-column chromatographies, including a G-DNA affinity column. Two major proteins (~80 and ~40 kDa) were present in the most highly purified column fraction. Renaturation experiments showed that the ~80 kDa protein contains TGP1 activity. Biochemical characterization showed that TGP1 is a G-DNA specific binding protein with a preference for parallel G-DNAs. The TGP1/DNA complex has a dissociation constant (K_d) of $\sim 2.2 \times 10^{-8}$ M and TGP1 can form supershift in gel mobility shift assays. The cDNA coding TGP1 was cloned and sequenced based upon an internal peptide sequence obtained from the ~80 kDa protein. Sequence analyses showed that TGP1 is a basic protein with a pI of 10.58, and contains two extensively hydrophilic and basic domains. Homology searches revealed that TGP1 is a novel protein sharing weak similarities with a number of proteins.

INTRODUCTION

G-quartets are unusual nucleic acid structures first described about three decades ago (reviewed in 1). In a G-quartet, four guanine molecules lie in a plane with each guanine forming G:C (instead of Watson-Crick A:T or G:C) hydrogen bonds with two adjacent guanines (1–3). DNA containing such structures is known as quadruplex, tetraplex or G-DNA (3). Many guanine-rich sequences of biological significance have been found to be capable of forming G-quartet structures *in vitro* under physiological conditions. Such G-rich sequences include most telomeres (4–6), immunoglobulin switch regions (7), a few gene promoters (8–10), fragile X repeats (11) and the dimerization domain in the human immunodeficiency virus (HIV) genome (12,13).

The ability to form G-DNA by these important sequences implies that G-DNA may be biologically relevant. However, direct evidence for the existence of G-DNA *in vivo* is lacking. An alternative approach to study G-DNA is to identify and investigate proteins that

interact with this structure. A number of proteins have been found to be able to bind to G-DNA. These proteins include a hepatocyte chromatin protein QUAD (14), transcription factor MyoD (15), chick topoisomerase II (16), two yeast proteins G4p1 and G4p2 (17,18) and the yeast KEM1 protein (19). In contrast to all other G-DNA binding proteins, KEM1 not only binds to G-DNA, but also acts as a nuclease to cut the single-stranded DNA 5' to the G-quartet domain (19,20). The G-DNA specific nuclease activity of KEM1, together with the studies in which meiotic recombination was found to be severely affected in KEM1 deletion mutants (21,22), suggests that KEM1 and the G-DNA structure play roles in meiotic pairing (20). In addition to G-DNA binding proteins, two proteins have been found to facilitate the formation of G-DNA structure: the *Oxytricha* telomere binding protein β subunit, which acts as a molecular chaperone (23,24), and the multifunctional yeast telomere protein RAPI (25). Since both proteins are telomere-related, their G-DNA promotion activities suggest that G-DNA may have a role in telomere function.

The biological relevance of G-DNA was further suggested by several other studies showing that G-DNA can inhibit the activities of a number of proteins. One of these G-DNA-inhibited proteins is telomerase, the enzyme that synthesizes telomeres (26). It was found that folding of telomere DNA (specifically the primer) into G-DNA inhibited the activity of telomerase *in vitro*, suggesting that formation of G-DNA in the primer could down-regulate telomere elongation *in vivo* (27). Another *in vitro* study showed that an antiparallel G-DNA aptamer can inhibit thrombin activity (28). G-DNA was also shown to inhibit the activity of macrophage scavenger receptors, the glycoproteins which may function in the deposition of lipoprotein cholesterol (29). More interestingly, an oligonucleotide in G-DNA form was found to be a potent inhibitor of HIV-1 integrase (30). These inhibitory activities of G-DNA suggest its potential use as a pharmaceutical agent.

Possible biological functions for G-DNA are diverse, but remain to be definitively demonstrated. Further studies on G-DNA and its related proteins are needed to elucidate the exact role(s) of these structures *in vivo*. The ciliated protozoan, *Tetrahymena thermophila*, has proven to be a useful system for identification of G-DNA binding proteins. This organism has two nuclei, one of which contains a fragmented genome with thousands of telomeres (31). A relatively abundant G-DNA binding activity has been identified in *T. thermophila* (32). The protein, TGP1 (*Tetrahymena* G-DNA binding protein 1), binds to an intermolecular G-DNA form of d(TTGGGG)₄ under physiological conditions, and competition

*To whom correspondence should be addressed. Tel: +1 515 294 2876; Fax: +1 515 294 2876; Email: telomere@iastate.edu

Table 1. Oligonucleotides used in mobility shift and competition assays (oligo GL is from ref. 17)

Y:	ACTGTCGACTTGATATGGGGGT
Tet4:	TTGGGGTTGGGGTTGGGGTTGGGG
rTet4:	UUGGGGUUGGGGUUGGGGUUGGGG
GL:	TATGGGGGAGCTGGGGAAGGTGGGATT
Tet1.5:	GGGGTTGGGG
cY:	ACCCCATATCAAGTACGACACAGT
cTet4:	CCCCAACCCCAACCCCAACCCCA

experiments showed that G-DNAs competed strongly for TGP1 binding, while non-G-DNA and G-RNA oligonucleotides did not (32). In this paper, we report the purification, molecular cloning and further biochemical characterization of TGP1. TGP1 was shown to be an 83 kDa protein with binding activity specific for G-DNA. Complete cDNA sequence of TGP1 was obtained and shown to encode a novel protein with two extensively hydrophilic and basic putative DNA binding domains and weak similarities to a number of other proteins. In light of these data, possible functions for TGP1 are discussed.

MATERIALS AND METHODS

Oligonucleotide synthesis, purification and labeling

Oligonucleotides (oligos) (Table 1) were synthesized on an ABI 394 DNA/RNA synthesizer (ISU DNA facility), and purified as previously described (32). Briefly, oligos were separated on a 20% denaturing polyacrylamide gel (7 M urea), then desired oligo products were excised and purified by C-18 chromatography (Waters). 5'-³²P-labeling of oligos using T4 polynucleotide kinase was carried out according to a standard protocol (33). Radiolabeled oligos were purified by G-25 spin columns (5 prime→3 prime).

Mobility shift experiments and competition assays

Fifty nanograms of ³²P-labeled oligo Y were boiled for 3 min in G-DNA formation buffer (50 mM KCl, 10 mM MgCl₂, 10 mM Tris-HCl, pH 8.0, 5% glycerol), cooled and incubated at room temperature for at least 30 min to make G-DNA. Mobility shift experiments were performed as previously described with minor modifications (32). About 2.5 ng of labeled Y(G4) were mixed with desired amounts of protein extract or TGP1 fractions in binding reaction buffer (10 mM Tris-HCl, pH 7.5, 6% glycerol). One hundred fold (~250 ng) non-specific competitor poly(dI-dC) (Pharmacia) 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. The gel was then vacuum-dried, and exposed to X-ray film or a PhosphorImager screen (Molecular Dynamics).

For competition assays, unlabeled competitors were made as follows. Single-stranded DNA (Y) was made by boiling in ddH₂O for 3 min to prevent G-quartet structure formation, and chilling on ice for 30 min. Double-stranded DNAs (Y-cY, Tet4-cTet4) were formed by annealing of single-stranded oligos with their complementary strands at 95°C for 5 min, 65°C for 10 min and 37°C for 10 min in buffer containing 50 mM NaCl and 10 mM

Tris-HCl, pH 7.5. G-DNAs were formed according to the methods described by Sen and Gilbert (34). In mobility shift assays, unlabeled competitors were diluted as indicated and added to the binding buffer before the addition of TGP1 and labeled Y(G4) probe. About 50 ng of TGP1 and saturating amounts (2.5 ng) of labeled probe Y(G4) were used in each binding reaction. Twenty units of RNasin ribonuclease inhibitor (Promega) were included in reactions where RNA oligo (rTet4) was used. Binding reactions and mobility shift experiments were performed as described above.

SDS-PAGE and silver staining

SDS-PAGE was performed according to the standard protocol (33). After electrophoresis, protein bands were visualized by silver staining following the manufacturer's protocol (BioRad).

UV cross-linking

Cross-linking of G-DNA to TGP1 was performed *in situ* by exposing a wet 8% mobility shift gel on ice to 254 nm UV light with the gel no more than 5 cm from the UV source. The gel was then exposed to film to reveal the position of the TGP1/G-DNA binding complex. The gel piece containing the complex was excised, denatured in SDS sample buffer for 5 min, and polymerized into the stacking gel of a 10% SDS-polyacrylamide gel. After electrophoresis, the gel was dried and exposed to X-ray film. Molecular weight markers were included to identify the positions of cross-linked bands.

Tetrahymena cell culture and extract preparation

Tetrahymena cell culture and total protein extract preparation were performed as previously described (32). Briefly, *T. thermophila* cells (strain C3V) were grown vegetatively to mid-log phase (2.5 × 10⁵ cells/ml) and harvested. Cells were washed twice with 10 mM Tris-Cl, pH 7.5, resuspended in 5 volumes of TMG buffer (10 mM Tris-Cl, pH 7.5, 1 mM MgCl₂, 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), and lysed by addition of one tenth volume of 2% NP-40 (Sigma). The cell lysate was centrifuged at 100 000 g for 70 min. The supernatant (S100) was aliquoted, immediately frozen in liquid nitrogen and stored at -70°C.

Purification of TGP1

All purification steps were conducted at 4°C unless otherwise indicated. *Tetrahymena* S100 protein extract (100 ml, ~10 mg/ml) was quickly thawed and filtered through a 0.4 µm filter (Costar). The filtrate was then loaded onto a SP-Sepharose column (4.91 cm² × 14.5 cm, Pharmacia), which had been equilibrated and packed with at least 2 column volumes of chromatography buffer [CB buffer: 10 mM Tris-HCl, pH 7.7, 1 mM EDTA, 0.01% NP-40, 10% glycerol and 1 mM dithiothreitol (DTT)], at a flow rate of 0.8 ml/min. After the extract had passed through the column, 1.5 column volumes of CB buffer were passed through the column at the same flow rate (0.8 ml/min). The column was then washed with 2 column volumes of CB buffer containing 0.45 M NaCl at a flow rate of 2 ml/min to remove weakly bound proteins. Four column volumes of CB buffer containing 0.8 M NaCl were then applied to elute TGP1 activity. About sixty 3 ml fractions were

collected and assayed for TGP1 activity in mobility shift experiments.

Fractions from the SP-Sepharose column containing TGP1 activity were pooled and dialyzed against CB buffer overnight at 4°C. The fractions were then loaded at a flow rate of 0.5 ml/min onto a DE52 column (3.21 cm² × 4 cm, Whatman), which had been equilibrated with CB buffer. The column was then washed with 1 column volume of CB buffer at a rate of 0.5 ml/min. A linear NaCl gradient (0–0.8 M in CB buffer, 2 column volumes) was applied to elute TGP1 activity at a flow rate of 0.4 ml/min. About thirty 2 ml fractions were collected and assayed in mobility shift experiments.

A G-DNA affinity column (~2 ml) was constructed according to a published procedure (20) with minor modifications. Briefly, 200 µg of 3'-biotinylated oligo Y (Midland) were mixed with 1 mg of oligo Y (5-fold) in the presence of 10 mM MgCl₂ and 100 mM KCl. The mixture was boiled for 3 min and then cooled to room temperature for 30 min to allow G-DNA formation. The biotinylated Y(G4) was then mixed with 2 ml of 50% avidin-agarose (Pierce), which had been equilibrated with CB buffer containing 100 mM KCl. After incubation at room temperature for 1 h, the mixture was poured and packed into a 5 ml disposable column (BioRad). Fractions from the DE52 column containing TGP1 activity were pooled and dialyzed against CB buffer overnight at 4°C. The fractions were then loaded at a flow rate of 0.1 ml/min onto the G-DNA affinity column, which had been equilibrated with CB buffer containing 0.1 M KCl. After the sample passed through, the column was washed with 1 column volume of 0.1 M KCl/CB buffer, and proteins were eluted with a linear KCl gradient (0.1–2.0 M) in CB buffer. About thirty 0.2 ml fractions were collected and tested for TGP1 activity in mobility shift experiments.

Protein renaturation

Renaturation experiments were performed using a procedure similar to that described by Hager and Burgess (35). Briefly, partially purified TGP1 (~10 µg) was displayed on a 10% SDS-PAGE, and the gel was then stained with Coomassie blue. Gel slices containing different protein bands were excised, rinsed with ddH₂O, put into different tubes, and crushed into small pieces. One milliliter of elution buffer (50 mM Tris-HCl, pH 7.5, 0.1% SDS, 0.1 mg/ml BSA, 1 mM DTT, 0.2 mM EDTA and 2.5% glycerol) was then added to each tube. Proteins were eluted from the gel into buffer with agitation for at least 2 h at room temperature. The mixtures were centrifuged for 2 min at 10 000 r.p.m. Supernatants were transferred to fresh tubes, mixed with 4 volumes of 100% acetone (-20°C), stored at -20°C for >2 h and centrifuged for 10 min at 10 000 r.p.m. Precipitates were rinsed with 0.5 ml of 100% methanol (-20°C) twice, air dried and dissolved in 5 ml of 6 M guanidine-HCl in dilution buffer (50 mM Tris-HCl, pH 7.6, 20% glycerol, 0.1 mg/ml BSA, 0.15 M NaCl, 1 mM DTT and 0.1 mM EDTA) for 20 min at room temperature. Solutions were 50-fold diluted with dilution buffer, and proteins were allowed to renature overnight at room temperature. Renatured proteins were tested for TGP1 activity in mobility shift experiments.

K_d determination of TGP1/G-DNA binding complex

Affinity purified TGP1 fractions were pooled, dialyzed against CB buffer and concentrated using Centricon concentrators (Amicon). TGP1 protein concentrations were estimated by SDS-PAGE and Coomassie blue staining. Concentrated TGP1

was then serially-diluted 2-fold and used in mobility shift experiments. Bound and unbound Y(G4) probe was quantitated by a PhosphorImager (Molecular Dynamics). The K_d for TGP1 G-DNA binding reaction was estimated from the plot of percentage of bound Y(G4) versus protein concentration.

Peptide sequencing

Purified TGP1 from affinity column was displayed on a 10% SDS-PAGE and electrotransferred to PVDF membrane according to the manufacturer's protocol (BioRad). The 80 kDa band was excised and N-terminal sequencing was attempted (ISU Protein Facility), but failed due to N-terminal blockage, thus internal peptide sequencing was then performed. About 5 µg of TGP1 (80 kDa) was eluted from 10% SDS-PAGE gel and acetone precipitated (see the Renaturation section for details). The protein was then subjected to CNBr digestion according to the protocol described by Smith (36). After digestion, peptides were resolved on a 16% SDS-PAGE, and transferred to PVDF membrane. One of the peptides (~35 kDa) yielded a clear peptide sequence: ??GRQSAEG?VG?hYIEV?rFgQYi ('?' indicates that no amino acid had been assigned, while letters in lower case indicate tentative assignments).

Molecular cloning of TGP1 cDNA

Total RNA was isolated from mid-log phase *T.thermophila* cells using TRIzol reagent according the manufacturer's protocol (Gibco-BRL). First strand cDNA synthesis from total RNA was done by reverse transcription using the M-MLV reverse transcriptase (Gibco-BRL). Based on the internal peptide sequence and with consideration of *Tetrahymena* genetic codon usage (37), two partially degenerate primers were designed: primer 1 [5'-GGTAGA(C/T)A(A/G)(T/A)(C/G)(T/C)GC(T/C)GAAGG] corresponding to the peptide sequence GRQSAEG, and primer 2 [5'-AATCTGCA(A/G)ACTTC(A/G)AT(A/G)TA] corresponding to the antisense sequence of peptide YIEV?rF. PCR was performed using the first strand cDNAs and primer 1 and primer 2. The PCR product (~55 bp) was cloned into TA cloning vector (Invitrogen), and sequenced (ISU DNA Facility).

Both 5' and 3' rapid amplification of cDNA ends (RACE) were performed according to the recommended protocols (Gibco-BRL). For 3' RACE, a TGP1 gene specific primer (primer 3: 5'-GCTGAAGGAAAAGTCGGTGGTCAC) was designed based on the sequence of the 55 bp PCR product. First strand cDNA synthesis was done using the poly(dT)-anchor oligo (Gibco-BRL). Subsequent PCR was performed over the first strand cDNAs using the anchor and primer 3. PCR products were cloned and sequenced. For 5' RACE, the first strand cDNA synthesis was done using a gene specific primer (primer 4: 5'-TTGACCGAATTTCTTAACCTT-CG), which was based on the DNA sequence obtained in the 3' RACE. After that, the first strand cDNAs were tailed with poly(dC), and purified. PCR was performed over the tailed cDNAs using a poly(dG)-anchor primer and a nested gene specific primer (primer 5: 5'-TAGTGACCACCGACTTTTCC) as primers. PCR products were cloned and sequenced.

Sequence analyses and database searches

GCG (Genetics Computer Group, Inc.) software was used for sequence analyses. BLAST (38) and FASTA (39) database searches were performed at the National Center for Biotechnology

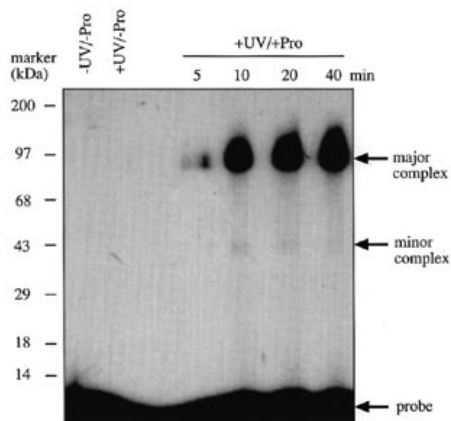


Figure 1. UV cross-linking assay to estimate the molecular weight of TGP1. Radiolabeled Tet4(G4) was cross-linked to TGP1 *in situ*. The band containing the cross-linked complex was excised from a mobility shift assay gel and polymerized into a 10% SDS-PAGE gel for size estimation. The -UV/-Pro and +UV/-Pro lanes contained probe alone without and with UV treatment respectively. The +UV/+Pro lanes contained protein/DNA complexes exposed to UV for different amounts of time (in minutes). The main cross-linked complex had an apparent MW of ~95 kDa. A minor complex ~45 kDa was also observed with longer UV exposure times.

Information (NCBI). Secondary structure analysis was done at the European Molecular Biology Laboratory (EMBL).

RESULTS

Identification of TGP1 by UV cross-linking

To identify proteins responsible for TGP1 activity, UV cross-linking experiments were performed. For these experiments, Tet4(G4) probe and a saturating amount of S100 extract were used. The results are shown in Figure 1. There were no detectable bands in

samples which were not exposed to UV-light and which did not contain both the protein extract and probe. In contrast, a band of ~95 kDa was observed in UV-exposed samples containing both the protein extract and DNA probe. These results suggested that the 95 kDa band was the cross-linked complex between TGP1 and DNA probe. Depending on how many Tet4 monomers cross-linked to TGP1, the molecular weight (MW) of the protein can range from 61 to 87 kDa (the Tet4 monomer has a MW ~8.4 kDa). As UV exposure time increased, a minor cross-linked product ~45 kDa was observed. This 45 kDa band could have resulted from a degradation product of the 95 kDa band, or may represent a different protein. Subsequent experiments (affinity purification and renaturation) suggested that this band represents a different G-DNA binding protein.

Purification of TGP1

A number of chromatography columns were tested on analytical scales (1–10 ml) to determine their usefulness in purifying TGP1. Three columns (SP-Sepharose, DE52 and G-DNA affinity) were chosen for the purification. The first column used, SP-Sepharose, is a cation-exchange column (Fig. 2A). As shown in the mobility shift assay, TGP1 was eluted when the salt concentration reached 0.8 M NaCl. This column resulted in removal of >90% of non-TGP1 proteins (data not shown). The high salt concentration (0.8 M NaCl) needed to elute TGP1 activity suggested that TGP1 is a basic protein, or may contain basic domains. After the SP-Sepharose column, the TGP1 fractions were applied to an anion-exchange (DE52) column (Fig. 2B). TGP1 activity was eluted from the column at NaCl concentrations of 0.15–0.5 M.

To further purify TGP1, a G-DNA affinity column was used. After elution from this column, the fractions were subjected to both mobility shift assay and SDS-PAGE silver staining. As shown in Figure 3A, TGP1 activity was eluted between 0.6 and

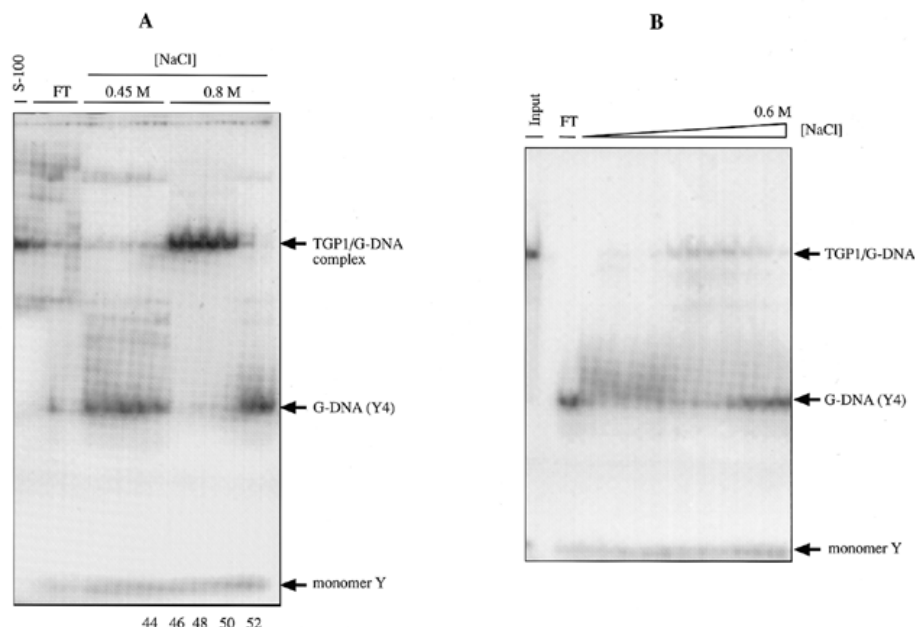


Figure 2. SP-Sepharose and DE52 column purification of TGP1. (A) SP-Sepharose column purification. S100 extracts of vegetatively grown *Tetrahymena* mid-log phase cells (2.5×10^5 /ml) were loaded onto a ~50 ml SP-Sepharose column (Pharmacia). Stepwise NaCl was applied to elute the proteins. Following purification, the fractions were tested for TGP1 activity in mobility shift experiments. Most of the TGP1 activity eluted from the column when 0.8 M NaCl was applied (fractions 48–52). (B) DE52 column purification. The TGP1 fractions from SP-Sepharose column were pooled, dialyzed against CB buffer overnight and loaded onto a ~10 ml DE52 (Whatman) column. Proteins were eluted using a linear salt gradient (0–0.8 M NaCl). TGP1 activity was followed by mobility shift assays.

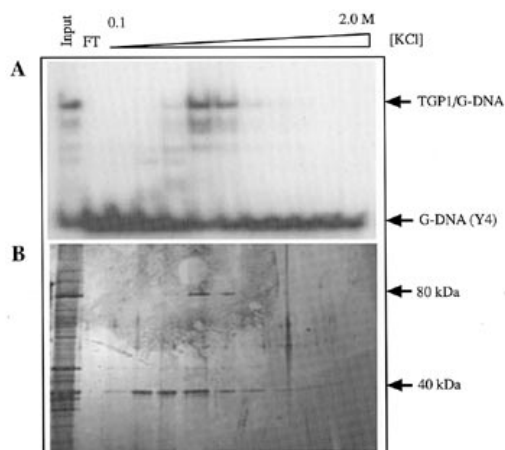


Figure 3. G-DNA affinity column purification of TGP1. The TGP1 fractions from DE52 column were pooled, dialyzed and applied to a 1 ml G-DNA column. (A) A KCl gradient (0–2.0 M) was used to elute proteins from the column. The fractions were tested for TGP1 activity by mobility shift assays. TGP1 activity was eluted at 0.8–1.0 M KCl. The fractions were also tested by SDS-PAGE and silver staining (B) to determine protein components in each fraction. Two major bands (~80 and ~40 kDa) were found in the silver stained gel. The pattern of the 80 kDa band matched that of the TGP1 activity.

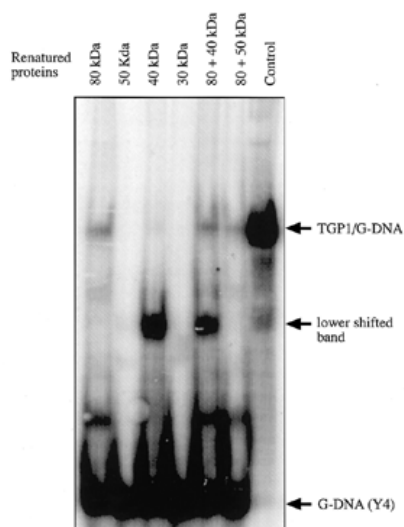


Figure 4. Renaturation of TGP1 activity from SDS-PAGE gel. Proteins (80, 50, 40 and 30 kDa) from SDS-PAGE gel were allowed to renature according to the procedure described in the Materials and Methods. Renatured proteins were tested for TGP1 activity in mobility shift assays (native TGP1 was used in the control lane).

1.0 M KCl with a peak at 0.8 M KCl. In the silver staining (Fig. 3B), two major bands (~80 and ~40 kDa) were present in the TGP1 peak fraction. However, the 80 kDa band pattern in the silver staining correlated best with TGP1 activity shown in the mobility shift assay, suggesting that the 80 kDa band is responsible for TGP1 activity. The 40 kDa protein may correlate with the lower shifted bands in the mobility shift assay.

Renatured ~80 kDa protein contains TGP1 activity

To confirm that the 80 kDa protein was TGP1, renaturation of proteins from SDS-PAGE was performed. The renatured proteins

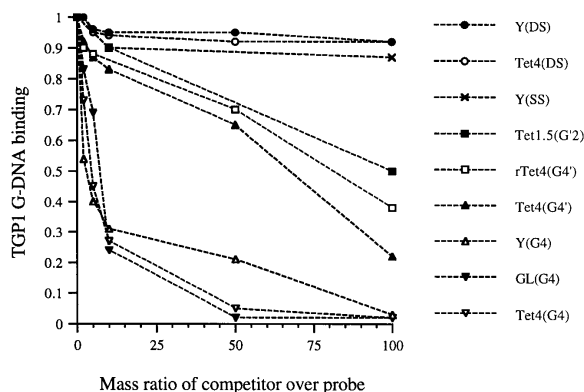


Figure 5. TGP1 G-DNA binding specificity. In competition assays, 2.5 ng of labeled Y(G4) was used as probe. Appropriate amounts of purified TGP1 were used to keep the Y(G4) probe saturated (excessive probe). Unlabeled competitors were diluted (2-, 5-, 10-, 50- and 100-fold) and used in mobility shift assays, as described in the methods section. The nomenclature of different G-quartet forms was according to Sen and Gilbert (34): G4, G'4 and G'2 denote intermolecular parallel, intramolecular foldback, and antiparallel dimer G-quartets, respectively.

were used in mobility shift experiments to assay for TGP1 activity. The results (Fig. 4) showed that the protein renatured from the 80 kDa band had TGP1 activity, while no shifted bands were observed for two control proteins (50 and 30 kDa). These results demonstrated that the 80 kDa band contains the TGP1 protein. The size of TGP1 (80 kDa) is consistent with that predicted by the UV cross-linking experiment. Interestingly, a strong band that shifted to a position lower than the TGP1/G-DNA complex was observed for the renatured 40 kDa protein. To test whether there are any interactions between the 80 and 40 kDa proteins, we mixed both renatured proteins and tested them in the mobility shift assay. No additional shifted bands were observed, suggesting no obvious interactions between these two proteins. However, this result does not rule out the possibility of interactions between these two proteins, because the proteins used in the assay were likely to have been only partially renatured.

TGP1 binds specifically to G-DNA

The DNA binding specificity of TGP1 was studied by competition assays using purified TGP1. Oligonucleotides used in the assays are listed in Table 1. The results are shown in Figure 5. The strongest competitors for TGP1 G-DNA binding are GL(G4), Tet(G4) and Y(G4). These three oligos are in the intermolecular parallel G-DNA (G4) form, suggesting that TGP1 has high binding affinity for such a structure. Non-G-DNAs, including single-stranded DNA (Y) and double-stranded DNAs (Y-cY, Tet4-cTet4), did not compete for TGP1 binding under our assay conditions. Antiparallel G-DNAs (G4' and G'2) showed intermediate competition between those of parallel G-DNAs and non-G-DNAs. G-RNA rTet4(G4) was also tested for competition, and showed much weaker competitive activity than its DNA counterpart Tet4(G4). Hence, our competition results demonstrate that TGP1 is a G-quartet structure-specific binding protein that prefers G-DNAs in the parallel form.

TGP1/G-DNA complex has a K_d of 2.2×10^{-8} M

Purified TGP1 was also used to determine the K_d of the TGP1/G-DNA complex. The result of the mobility shift assay using serially-diluted (2-fold) TGP1 is shown in Figure 6A. A supershift band was observed when high concentrations of TGP1 were used.

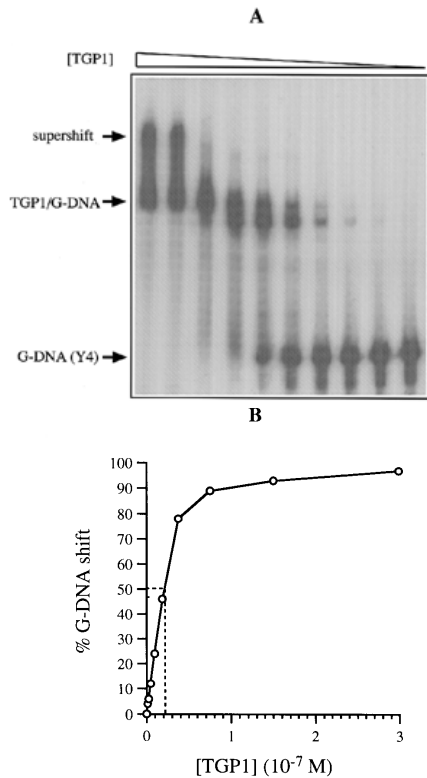


Figure 6. K_d determination of TGP1 G-DNA binding. (A) Mobility shift assays were carried out with 2.5 ng Y(G4) probe and 2-fold serial-diluted TGP1 (the highest TGP1 concentration was $\sim 1 \mu\text{g}$ per reaction). A supershift band was observed when TGP1 of high concentrations was used. (B) K_d estimation. The mobility shift gel (A) was quantitated using a PhosphorImager (Molecular Dynamics). The data were plotted and a K_d of $\sim 2.2 \times 10^{-8}$ M was determined from the graph.

With TGP1 at low concentration, the supershift band disappeared. This supershift could be a dimer of TGP1 binding to a G-DNA substrate. However, because the TGP1 fraction used in the assay contains other proteins (mainly the 40 kDa protein), it is possible that other protein components may be involved in forming such a supershift band. Bands in the mobility shift gel were quantitated and the data were plotted in the graph shown in Figure 6B. The estimated K_d for TGP1/G-DNA binding is $\sim 2.2 \times 10^{-8}$ M, under our assay conditions [2.5 ng Y(G4) probe, 10 mM Tris-HCl, pH 7.5, 4°C]. This K_d value is in a range similar to those of other G-DNA binding proteins such as G4p1 (17) and G4p2 (18).

Molecular cloning of TGP1 cDNA

To clone the cDNA coding the TGP1 protein, peptide sequencing was performed. No sequence data were obtained from the N-terminal sequencing directly from TGP1, possibly due to N-terminal blockage (data not shown). However, a 35 kDa peptide obtained from CNBr digestion of TGP1 yielded an amino acid sequence (??GRQSAEG?VG?hYIEV?rFgQYi).

TGP1 cDNA was cloned based on a PCR strategy. Two partially degenerate primers were designed based on the internal peptide sequence of TGP1. Using these two primers, a ~ 55 bp fragment was amplified from the first strand cDNAs by reverse transcriptase PCR (RT-PCR). The DNA sequence between the two primers encodes a peptide KVGGH which is consistent with the peptide sequence obtained from direct peptide sequencing.

Based on this DNA sequence, the full length sequence of TGP1 cDNA was obtained using 5' and 3' RACE.

Translation of the longest open reading frame (ORF) in the TGP1 full length cDNA sequence produces a 725 amino acid long protein (the TGP1 cDNA and protein sequences had been deposited into GenBank with accession no. AF006380). The sequence of this predicted protein is considered correct and complete for TGP1 for the following reasons: (i) this predicted protein sequence contains the peptide sequence obtained from the direct peptide sequencing; (ii) the predicted molecular weight for TGP1 from the sequence is 83.2 kDa, which is consistent with the SDS-PAGE result (~ 80 kDa); (iii) the amino acid composition of this predicted TGP1 protein matches the composition data obtained from acid hydrolysis of TGP1 (data not shown).

TGP1 is a novel protein with unusual sequence features

Analyses of the predicted TGP1 protein sequence showed that TGP1 has an unbalanced amino acid composition: it is rich in asparagine (13.9%), lysine (9.4%), glutamine (7.2%) and arginine (7.0%), whereas it has only one tryptophan, two cystines and five methionines. The richness of basic amino acids (lysine, arginine and histidine) results in a predicted pI of 10.58, and likely relates to the DNA binding activity of TGP1. Hydrophobicity plot of TGP1 showed that TGP1 contains two extensively hydrophilic and basic regions. One of the two regions is from residue 476 to residue 603. In this 128 amino acid region, there are 109 hydrophilic residues, and among them, 40 asparagines (33%, twice the average percentage in the whole TGP1 sequence), 17 glutamines (14%, twice the average), and more strikingly, 33 basic residues (compared with only five acidic residues). Another hydrophilic and basic region is from residue 251 to 285. The hydrophilic and basic properties of these regions suggest that they may serve as the DNA binding sites (domains) for TGP1. However, none of the known DNA binding motifs were found in the TGP1 sequence (including the hydrophilic and basic regions).

Comparison of the TGP1 protein sequence with available databases using BLAST and FASTA programs showed that TGP1 does not share significant homology or similarity with any other proteins, including many known G-DNA binding proteins, yet is weakly similar to a number of proteins. An entry with one of the best scores ($P = 0.02$) in the BLAST search is the yeast NUF1 protein, which contains coiled-coil structures and is probably a component of the yeast nuclear skeleton (40). The GAP alignment (gap creation penalty: 7; extension penalty: 2) between TGP1 and NUF1 over the whole sequence showed a 22% identity. In addition to NUF1, several other proteins with coiled-coil structures, including the tropomyosin α chain of smooth muscle, paramyosin and myosin regulatory light chain from different species, were returned in the FASTA search. However, a secondary structure prediction program (the PHD program in EMBL) did not predict similar structures for TGP1. Of note, the BLAST search also returned two yeast G-quartet binding proteins, G4p1 and G4p2. TGP1 has a region (152–268 amino acids) similar (19% identity, 40% similarity) to a region in G4p1 (46–162 amino acids) (Fig. 7A). TGP1 also has several short sequences similar to sequences in G4p2, and one of the short sequences is within the longer extensively hydrophilic region in TGP1 (Fig. 7B). However, alignments of TGP1 with either G4p1 or G4p2 over the whole sequence resulted in much lower (probably random) similarity.

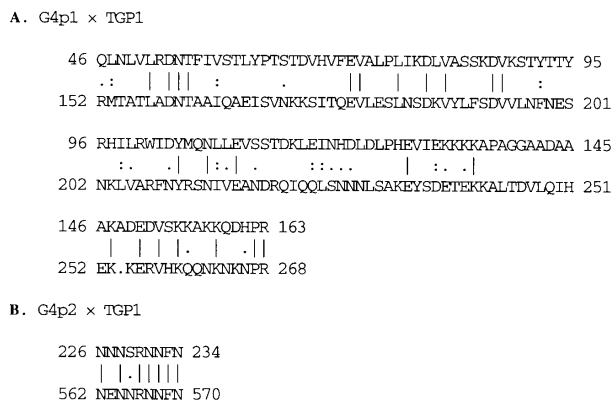


Figure 7. Partial sequence alignments between TGP1 (GenBank accession no. AF006380) and G4p1 (17) (A), and G4p2 (18) (B). The identity (!) and similarities (: and .) between residues are indicated.

DISCUSSION

G-DNAs are novel DNA structures whose proposed biological roles include participation in telomere function, recombination and gene regulation. These putative roles are still unproven and need further exploration. One approach is to identify and study proteins that interact with G-DNA. In this paper, we described the purification, biochemical characterization and molecular cloning of TGP1, the first identified *Tetrahymena* G-DNA binding protein (32).

This study ultimately identified TGP1 as an 83 kDa protein. UV cross-linking experiments suggested that a protein of mw of ~61–87 kDa was responsible for the TGP1 activity. During a three-column purification process, an ~80 kDa protein was found in the most highly purified fractions. Renaturation experiments demonstrated that the ~80 kDa protein contained TGP1 activity. Furthermore, the cloned TGP1 cDNA encodes a protein with a predicted mw of 83 kDa. In addition to the 83 kDa TGP1 protein, we identified an ~40 kDa protein which copurified with TGP1 during the three-column purification. Interestingly, renaturation experiments showed that this 40 kDa protein possessed a G-DNA binding activity distinct from TGP1. Since the N-terminal peptide sequence of the 40 kDa protein was not found in the TGP1 sequence (data not shown), the 40 kDa protein is unlikely to be a degradation product of TGP1, but rather represents an additional G-DNA binding protein.

The binding specificity of TGP1 was evaluated using the purified protein. The results of these experiments are consistent with a previous study (32). The subtle DNA binding specificity of TGP1 distinguishes it from most known G-quartet binding proteins. First, TGP1 showed very low affinity for non-G-quartet nucleic acids (i.e., single- or double-stranded DNAs), whereas some G-DNA binding proteins are able to bind to these structures. Such proteins include the yeast telomere protein RAP1 which binds to the double-stranded yeast telomeric DNA with much higher affinity (25), and a rat hepatocyte protein qTBP42 which can bind to single-stranded DNAs (41). Second, TGP1 has a weaker binding affinity for G-quartet structures formed by RNA, while some G-quartet binding proteins [e.g., a mouse cytoplasmic exoribonuclease mXRN1p (42), two yeast protein G4p1 and G4p2 (17,18)] have high affinity for G-RNA. Third, TGP1 has a higher affinity for parallel-stranded G-DNA than for the antiparallel form. In contrast, most known G-DNA binding proteins do not

distinguish between these two G-DNA forms. The differences in binding specificity suggest that G-DNA/RNA binding proteins can be divided into several subfamilies, each with specificity for different G-quartet forms. TGP1 could therefore belong to a distinct parallel G-DNA-binding protein subfamily.

Sequence analyses showed that none of the known DNA binding motifs were found in TGP1, including a recently identified Myb-like domain characteristic of some telomere binding proteins (43,44). This result is not surprising in light of the fact that TGP1 binds specifically to G-DNA, and has almost no affinity for double- or single-stranded DNAs. It is reasonable to predict that novel DNA binding motifs would exist to account for the specific G-DNA binding activity. Consistent with this hypothesis, two extensively hydrophilic and basic regions, which have no similarity to any known DNA/RNA binding motifs, were identified in TGP1. We propose that the hydrophilic and basic regions comprise novel G-DNA binding domains. We further suggest that these novel domains could be shared by other G-DNA binding proteins to some extent. The weak similarities between sequences of G4p2 and the longer extensively hydrophilic region within TGP1 lend support to this hypothesis. Since no G-DNA specific domains have been identified in any G-DNA binding proteins thus far, further characterization of these putative G-DNA binding domains might provide valuable information on mechanism(s) by which proteins can specifically recognize the G-DNA structure.

Homology/similarity searches revealed that TGP1 is a novel protein with very limited similarity to a number of proteins. The weak similarity between TGP1 and NUF1, the yeast putative nuclear skeleton protein, suggests that TGP1 may be a component of the nuclear skeleton (matrix) in *Tetrahymena* nuclei. Nuclear matrix, the insoluble non-chromatin scaffold structure of nucleus, has been thought to be involved in many nuclear events such as chromosome organization and gene regulation, through associations with chromosomes (45). The relative abundance of TGP1 [$\sim 1.6 \times 10^6$ molecules/cell, (32)] is consistent with a role for TGP1 as a structural protein related to the nuclear matrix. One possibility is that TGP1 may bind to telomeres, where G-DNA structures could form (46), linking the telomeres to the nuclear matrix, thus helping to organize the chromosomes. This speculation is supported by data from other species that link telomere proteins to the nuclear matrix. For example, the yeast telomere protein RAP1 is associated with the nuclear matrix (47), and interacts with the SIR4 protein (48), which is weakly similar to the human nuclear matrix proteins lamin A and C. In humans, the telomere protein TRF was found to be a component of the nuclear matrix, and the TRF/telomeric DNA complexes are associated with the nuclear matrix network (49).

The proposed function for TGP1 in the nuclear matrix is speculative and does not exclude other possible telomere-related functions for TGP1. For example, TGP1 may stabilize telomere-telomere interactions that had been observed in many species (50). Such interactions have been suggested to have a role in chromosome separation during mitosis (51,52). A mutation in *Tetrahymena* telomeric DNA repeat (GGGGTT changed to GGGGTTTT) caused delayed and abnormal separation between the sister chromatids during mitosis (52). The mutant chromatids became elongated up to twice the normal length when they finally separated, indicating stronger physical association exists between the mutant chromatids, especially at the telomeres (52). However, little is known about how telomeres interact with each other. One of the possible mechanisms is that G-rich strands or 3' G-overhangs

(53–58) of telomeres from different chromosomes can stack together by forming a G-DNA structure (46). If that is the case, protein components such as TGP1 could bind to the G-DNA, and stabilize the interaction between sister chromatids. Any alterations in the telomere sequence may change the G-DNA/protein structure, and thus cause problems in chromosome separation.

Preferred TGP1 binding to parallel G-DNA suggests that TGP1 may also be involved in recombination, since parallel G-DNA structures have long been thought to be involved in recombination processes (7). It was proposed that four G-rich DNA strands from sister chromosomes can bind together to form parallel G-DNA stabilized structures (7). Protein factors that interact with such structures could therefore have roles in recombination. The yeast G-DNA specific nuclease KEM1, which recognizes the parallel G-DNA structure and cuts the single-stranded DNA 5' to the G-quartet domain *in vitro*, was proposed to function in such a way in meiotic pairing (20). However, in contrast to KEM1, no nuclease activity was observed for TGP1 (data not shown).

All hypotheses regarding the biological functions for TGP1 and G-DNA need to be rigorously tested by further studies such as immunolocalization and gene disruption, and these experiments are in progress. These studies will further define the function(s) of TGP1, and should also contribute to the understanding of biological role(s) of G-DNA in general.

ACKNOWLEDGEMENTS

The authors want to thank Scott Schaus for critical reading of the manuscript. This work was funded partly by a research grant from the American Cancer Society to E.H. (NP906). Q.L. was supported in part by fellowships from the MCDB program. This is Journal Paper No. J-17788 of the Iowa Agriculture and Home Economics Experiment Station, Ames, IA. Project No. 3064, and supported by Hatch Act and State of Iowa Funds.

REFERENCES

- Guschlbauer, W., Chantot, J. F. and Thiele, D. (1990) *J. Biomol. Struct. Dyn.*, **8**, 491–511.
- Williamson, J. R. (1994) *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703–730.
- Henderson, E. (1995) In Blackburn, E. H. and Greider, C. W. (eds) *Telomeres*. Cold Spring Harbor Laboratory Press, New York, pp. 11–34.
- Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, J. J. and Blackburn, E. H. (1987) *Cell*, **51**, 899–908.
- Sundquist, W. I. and Klug, A. (1989) *Nature*, **342**, 825–829.
- Williamson, J. R., Raghuraman, M. K. and Cech, T. R. (1989) *Cell*, **59**, 871–880.
- Sen, D. and Gilbert, W. (1988) *Nature*, **334**, 364–366.
- Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. and Efstratiadis, A. (1984) *Nucleic Acids Res.*, **12**, 8043–8058.
- Kipatrick, M. W., Torri, A., Kang, D. S., Engler, J. A. and Wells, R. D. (1986) *J. Biol. Chem.*, **261**, 11350–11354.
- Lewis, C. D., Clark, S. P., Felsenfeld, G. and Gould, H. (1988) *Genes Dev.*, **2**, 863–873.
- Fry, M. L. and Loeb, L. A. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4950–4954.
- Awang, G. and Sen, D. (1993) *Biochemistry*, **32**, 11453–11457.
- Sundquist, W. I. and Heaphy, S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3393–3397.
- Weiman-Shomer, P. and Fry, M. (1993) *J. Biol. Chem.*, **268**, 3306–3312.
- Walsh, K. and Gualberto, A. (1992) *J. Biol. Chem.*, **267**, 13714–13718.
- Chung, I. K., Mehta, V. B., Spitzner, J. R. and Muller, M. T. (1992) *Nucleic Acids Res.*, **20**, 1973–1977.
- Frantz, J. D. and Gilbert, W. (1995) *J. Biol. Chem.*, **270**, 20692–20697.
- Frantz, J. D. and Gilbert, W. (1995) *J. Biol. Chem.*, **270**, 9413–9419.
- Liu, Z., Frantz, J. D., Gilbert, W. and Tye, B. -K. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 3157–3161.
- Liu, Z. and Gilbert, W. (1994) *Cell*, **77**, 1083–1092.
- Kim, J., Ljungdahl, P. O. and Fink, G. R. (1990) *Genetics*, **126**, 799–812.
- Tishkoff, D. X., Johnson, A. W. and Kolodner, R. D. (1991) *Mol. Cell. Biol.*, **11**, 2593–2608.
- Fang, G. and Cech, T. R. (1993) *Biochemistry*, **32**, 11646–11657.
- Fang, G. W. and Cech, T. R. (1993) *Cell*, **74**, 875–885.
- Giraldo, R. and Rhodes, D. (1994) *EMBO J.*, **13**, 2411–2420.
- Greider, C. W. (1995) In Blackburn, E. H. and Greider, C. W., (eds) *Telomerase Biochemistry and Regulation. Telomeres*, Cold Spring Harbor Laboratory Press, New York.
- Zahler, A. M., Williamson, J. R., Cech, T. R. and Prescott, D. M. (1991) *Nature*, **350**, 718–720.
- Wang, K. Y., McCurdy, S., Shea, R. G., Swaminathan, S. and Bolton, P. H. (1993) *Biochemistry*, **32**, 1899–1904.
- Pearson, A. M., Rich, A. and Krieger, M. (1993) *J. Biol. Chem.*, **268**, 3546–3554.
- Mazumder, A., Neamati, N., Ojwang, J. O., Sunder, S., Rando, R. F. and Pommier, Y. (1996) *Biochemistry*, **35**, 13762–13771.
- Prescott, D. M. (1994) *Microbiol. Rev.*, **58**, 233–267.
- Schierer, T. and Henderson, E. (1994) *Biochemistry*, **33**, 2240–2246.
- 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 and Sons, New York.
- Sen, D. and Gilbert, W. (1992) *Methods Enzymol.*, **211**, 191–199.
- Hager, D. A. and Burgess, R. R. (1980) *Anal. Biochem.*, **109**, 76–86.
- Smith, B. J. (1993) In Walker, J. M. (ed.), *The Protein Protocols Handbook*. Humana Press Inc., Totowa, NJ, pp. 369–373.
- Martindale, D. W. (1989) *J. Protozool.*, **36**, 29–34.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) *J. Mol. Biol.*, **215**, 403–410.
- Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
- Mirzayan, C., Copeland, C. S. and Snyder, M. (1992) *J. Cell Biol.*, **116**, 1319–1332.
- Sarig, G., Weisman-Shomer, P., Erlitzki, R. and Fry, M. (1997) *J. Biol. Chem.*, **272**, 4474–4482.
- Bashkurov, V. I., Scherthan, H., Solinger, J. A., Buerstedde, J. -M. and Heyer, W. -D. (1997) *J. Cell Biol.*, **136**, 761–773.
- Bilaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M. and Gilson, E. (1996) *Nucleic Acids Res.*, **24**, 1294–1303.
- Cooper, J. P., Nimmo, E. R., Allshire, R. C. and Cech, T. R. (1997) *Nature*, **385**, 744–747.
- Gasser, S. M., Amati, B. B., Cardenas, M. E. and Hofmann, J. F. -X. (1989) *Int. Rev. Cytol.*, **119**, 57–96.
- Sundquist, W. I. (1991) In Eckstein, F. and Lilley, D. M. J. (eds) *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, pp. 1–24.
- Hofmann, J. F. -X., Laroche, T., Brand, A. H. and Gasser, S. M. (1989) *Cell*, **57**, 725–737.
- Moretti, P., Freeman, K., Coodly, L. and Shore, D. (1994) *Genes Dev.*, **8**, 2257–2269.
- Luderus, M. E. E., Steensel, B. V., Chong, L., Sibon, O. C. M., Cremers, F. F. M. and Lange, T. D. (1996) *J. Biol. Chem.*, **135**, 867–881.
- Dernburg, A. F., Sedat, J. W., Cande, W. Z. and Bass, H. W. (1995) In Blackburn, E. H. and Greider, C. W. (eds), *Telomeres*. Cold Spring Harbor Laboratory Press, New York, pp. 295–338.
- Yu, G. L., Bradley, J. D., Attardi, L. D. and Blackburn, E. H. (1990) *Nature*, **344**, 126–132.
- Kirk, K. E., Harmon, B. P., Reichart, I. K., Sedat, J. W. and Blackburn, E. H. (1997) *Science*, **275**, 1478–1481.
- Klobutcher, L. A., Swanton, M. T., Donini, P. and Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3015–3019.
- Pluta, A. F., Kaine, B. P. and Spear, B. B. (1982) *Nucleic Acids Res.*, **10**, 8145–8154.
- Henderson, E. and Blackburn, E. (1989) *Mol. Cell. Biol.*, **9**, 345–348.
- Dionne, I. and Wellinger, R. J. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 13902–13907.
- Makarov, V. L., Hirose, Y. and Langmore, J. P. (1997) *Cell*, **88**, 657–666.
- Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D. and Shay, J. W. (1997) *Genes Dev.*, **11**, 2801–9.