A Protein from *Tetrahymena thermophila* That Specifically Binds Parallel-Stranded G4-DNA†

Ted Schierer and Eric Henderson*

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

Received October 8, 1993; Revised Manuscript Received December 8, 1993®

ABSTRACT: G4-DNA is a parallel, four-stranded structure mediated by tetrads of hydrogen-bonded guanines (G-quartets). An abundant protein called *Tetrahymena* G4 binding protein (TGP) that binds to an intermolecular, quadruplex form of d(TTGGGGTTGGGGTTGGGGTTGGGG) under physiological salt conditions has been identified in cellular extracts from the ciliated protozoan *Tetrahymena thermophila*. In binding competition experiments, molecules capable of forming G4 structures compete for binding to TGP, but non-G4-forming molecules and $r(U_2G_4)_4$ do not. TGP binding also requires a single-stranded region adjacent to the G4 structure. During the course of this study, it was determined that Mg^{2+} facilitates the formation of parallel-stranded G4-DNA structures and that high oligonucleotide concentrations are not required to drive formation of these structures. In addition, G4-DNA and TGP/G4-DNA complexes form readily under physiological salt conditions. These data support the proposal that G4-DNA structures exist *in vivo*.

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

Intramolecular G-DNA structures migrate faster than linear forms of the same length in nondenaturing gels (Henderson et al., 1987; Williamson et al., 1989) while intermolecular structures migrate more slowly (Sen & Gilbert, 1990; Acevedo et al., 1991; Sundquist & Klug, 1989). The equilibrium between these structures is cation and concentration dependent, and the formation of the less favored antiparallel G′2-DNA (Sen & Gilbert, 1990) verses parallel-stranded G4-DNA is stabilized by the presence of particular monovalent ($K^+ > Rb^+ > Na^+ > Cs^+ > Li^+$) and divalent cations ($Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$) and high oligonucleotide concentration (Venczel & Sen, 1993). A recent study demonstrated that divalent cations stabilize G-DNA structures at a concentration of 10 mM whereas 1 M monovalent cation concentrations are needed to produce the same effect (Venczel & Sen, 1993). Once formed, intermolecular quadruplex structures are exceedingly stable and recalcitrant to hybridization with complementary C-rich strands (Raghuraman & Cech, 1990; Hardin et al., 1991).

Along with telomeric DNA, a variety of biologically relevant G-rich DNA sequences have been identified, including immunoglobulin switch regions (Sen & Gilbert, 1988), recombination hot spots (Hastie & Allshire, 1989), and gene regulatory regions (Nickol & Felsenfeld, 1983; Pears & Williams, 1988). G4-RNA has also been studied in detail and shown to form structures based upon G-quartets (Cheong & Moore, 1992). G-RNA can mediate dimerization of the HIV genome *in vitro* (Marquet et al., 1991; Sundquist & Heaphy, 1993). These examples support the idea that G-DNA/RNA structures may have important roles in the cell including telomere function [reviewed by Blackburn (1991)], meiotic chromosome pairing (Sen & Gilbert, 1988), HIV genome dimerization (Marquet et al., 1991; Sundquist & Heaphy, 1993), and promoter function (Walsh & Gualberto, 1992).

The exact biological role of G-DNA structures is not clear in any organism. Intramolecular foldback structures formed by $d(T_4G_4)_4$ (Oxy 4) inhibit the activity of *Oxytricha* telomerase (Zahler et al., 1991), an enzyme that adds telomeric repeats to the 3′ end of the chromosome (Grieder & Blackburn, 1985; Zahler & Prescott, 1988). The β subunit of the *Oxytricha* telomeric end binding protein has recently been reported to catalyze the formation of G-quartet structures (Fang & Cech, 1993), suggesting that G-DNA may have a regulatory role in telomere replication. In addition, a variety of proteins have been reported to bind G4-DNA. Chick topoisomerase II (Chung et al., 1992), a yeast nuclease (Liu et al., 1993), MyoD (a transcription factor that regulates myogenesis) (Walsh & Gualberto, 1992), a hepatocyte chromatin protein (QUAD; Weisman-Shomer & Fry, 1993), and macrophage scavenger receptors (Pearson et al., 1993) all bind G4-DNA and suggest potential roles for G-DNA/protein complexes.

We report here the identification and characterization of an abundant G-DNA binding activity from *Tetrahymena thermophila*. This protein, designated *Tetrahymena* G4 binding protein (TGP), binds an intermolecular quadruplex

† This work was supported by Grants GM41023 from the NIH (E.H.) and JFRA 240 from The American Cancer Society (E.H.).

* To whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* February 1, 1994.

0006-2960/94/0433-2240$04.50/0

structure. An extensive comparison of different G4-DNA molecules as substrates for TGP is presented. It is shown that parallel-stranded G4-DNA is preferred over antiparallel quartet structures. The identification of this G4 binding protein lends further support to the proposal that G4-DNA exists *in vivo* and may play an important role in the cell.

MATERIALS AND METHODS

Tetrahymena Strains and Cell Culture. *T. thermophila* strain C3 V was grown using a rotary shaker at 30 °C in 2% PPYS (2% proteose peptone, 0.2% yeast extract, and 0.003% sequestrine) to midlog phase (2.5×10^5 cells/mL).

Extract Preparation. *Tetrahymena* whole cell extracts were prepared according to the procedure of Greider and Blackburn (1987) except that the cells were not mated. Furthermore, double-distilled H_2O was substituted for diethyl pyrocarbonate-treated H_2O and 10 mM Tris-HCl (pH 7.5) was substituted for Dryls for cell washes. Briefly, cells at midlog phase were harvested at 5K rpm for 5 min (Sorvall GSA rotor, 4 °C) and washed twice with a 20–50× volume of cold 10 mM Tris (pH 7.5). Cell pellets were then resuspended in a 5× volume of TMG buffer (10 mM Tris-HCl, pH 7.5, 1 mM $MgCl_2$, 10% glycerol, 10 mM 2-mercaptoethanol), and 1/10 volume of 2% NP-40 was added immediately. Cells were lysed by shaking on a Fisher Genie 2 vortexer (setting 3) for 30 min at 4 °C. The cell lysate was centrifuged at 100000*g* for 60 min at 4 °C. The supernatant, termed S-100 cell extract, was aliquoted and frozen quickly with liquid nitrogen. The final protein concentration was typically 2.5 mg/mL in TMG buffer. The protease inhibitors leupeptin (0.01 mM), pepstatin (0.01 mM), and Pefabloc (0.1 mM) (Boehringer Mannheim) were included in all solutions. Without the addition of leupeptin, TGP still bound to Tet 4 multimers, but the complex migrated below the 517 bp marker, indicating that the protein had been cleaved by a protease (data not shown). Extracts can be thawed 3 times without any loss in TGP binding activity.

DNA Oligonucleotide Synthesis, Purification, and 5′ End Radiolabeling. $d(T_2G_4)_4$ DNA oligonucleotides were gel purified as previously described (Henderson et al., 1987). Briefly, DNA oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer followed by deprotection. After boiling in 1× TBE buffer [89 mM Tris-HCl, 89 mM boric acid, 1.5 mM disodium ethylenediaminetetraacetate (EDTA), pH 8.3] containing 80% formamide for 90 s, the oligonucleotides were separated by electrophoresis through 20% polyacrylamide sequencing gels containing 7 M urea and 1× TBE buffer. The desired species were identified by UV shadowing and excised from the gel. DNA was eluted by shaking in TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) for 12–16 h at room temperature, and the eluted oligonucleotides were desalted over a Sep-Pak C18 column (Waters). Gel-purified Tet 4 oligonucleotides were 5′ end labeled as previously described (Henderson et al., 1987). The labeled Tet 4 was then gel purified (12%, 7 M urea–PAGE) and desalted by C-18 chromatography as above.

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

Electrophoretic Mobility Retardation Assays. For mobility retardation assays, 0.5 pmol of 5′-^{32}P probe were boiled in the presence of 295 mM NaCl, 29 mM KCl, 6 mM $MgCl_2$, 12 mM Tris-HCl (pH 7.5), and 5.3% glycerol and cooled on ice for at least 30 min to allow the probe to form structure. Preformed Tet 1.5 multimers were not boiled. In Figure 2 (lanes 1 and 2), formation of intermolecular Tet 4 structure was prevented by boiling Tet 4 in 10 mM Tris-HCl (pH 7.5) and 4.5% glycerol, but enough Na^+ existed in the Tris buffer to allow formation of the intramolecular foldback monomer (Figure 2, lanes 1 and 2). A 200-fold molar excess of oligo-$d(T)_{24}$ was then added as a nonspecific competitor (5 pmol/μL final concentration) before adding 2 μL of diluted S-100 extract (0.25 μg/μL) to produce a final reaction volume of 20 μL. The final probe concentration was 0.025 pmol/μL in all cases except XYa, XYb, and XYc (0.04 pmol/μL), and final concentrations in the binding reaction were 250 mM NaCl, 25 mM KCl, 5 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.5), 5.5% glycerol, 1 mM 2-mercaptoethanol, and 0.02% NP-40. In competition assays, unlabeled competitors were boiled in the same salt concentrations as the probe and cooled on ice to allow separate structure formation of the competitor and probe before they were combined. Once probe and competitor were combined, a 200-fold molar excess of oligo-$d(T)_{24}$ was then added as a nonspecific competitor (5 pmol/μL final concentration) before adding 0.5 μg of S-100 extract to produce a final reaction volume of 20 μL. The final probe concentration was 0.025 pmol/μL in all cases except with XYa, XYb, and XYc (0.04 pmol/μL). $r(U_2G_4)_4$ binding and competition assays included 20 U of RNasin (Promega). Binding reactions were incubated on ice for 20 min and then loaded onto a 8% nondenaturing polyacrylamide gel in 0.6× TBE. Electrophoresis at 10 V/cm was carried out at room temperature until the bromophenol blue reached 3/4 the length of the gel. The gel was then dried, and the bands were visualized by autoradiography and quantitated using a PhosphorImager (Molecular Dynamics).

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

RESULTS

Characterization of Multimeric and Monomeric Forms of Tet 4. The oligonucleotide $d(T_2G_4)_4$ (Tet 4) was used as a probe to search for proteins that bind to G4-DNA. Its ability to form G4-DNA is demonstrated in Figure 1. Under nondenaturing conditions, slowly migrating Tet 4 species were reproducibly observed in addition to the species migrating to the expected position for Tet 4 monomers (Figure 1A). UV cross-linking experiments demonstrated that the slow migrating structures are due to intermolecular associations of Tet 4. Under denaturing conditions without UV irradiation, both species migrate to the position expected for a 24-nucleotide long molecule (Figure 1B, lanes 1 and 2). Following UV irradiation, at least three cross-linked complexes from the slow migrating band that have sizes much larger than the non-UV-exposed species are observed (Figure 1B, lane 4). Thus, the slow migrating form of Tet 4 is an intermolecular structure.

A control oligo-d(ACTGTCGTACTTGATATGGGGGT) (oligo Y; Sen & Gilbert, 1990) was subjected to the same treatment as Tet 4. Oligo Y was previously shown to form

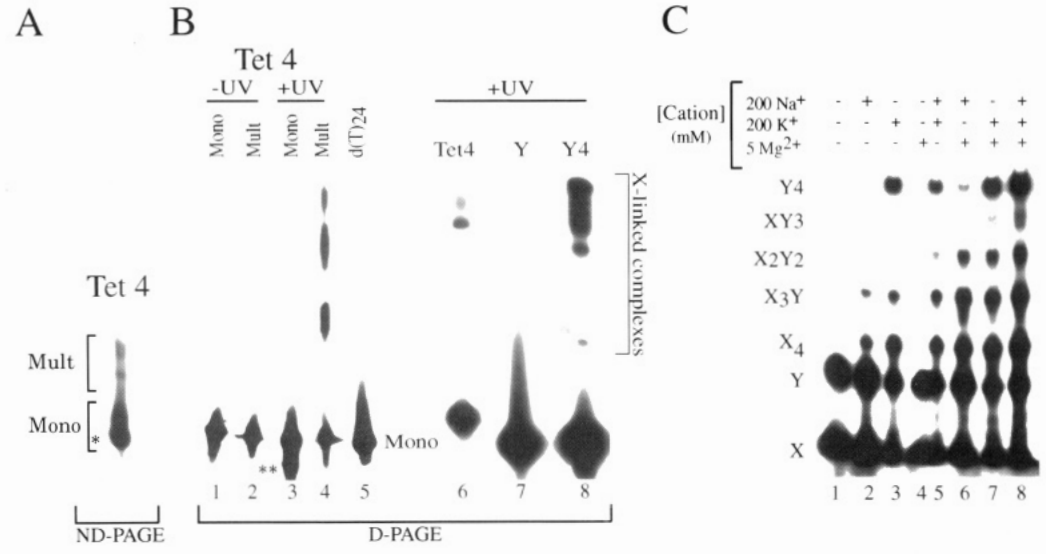

FIGURE 1: Tet 4 forms G4-DNA. (A) ^{32}P-labeled Tet 4 oligonucleotides were analyzed on a nondenaturing gel (ND-PAGE) and separated into monomeric (Mono) and multimeric (Mult) species. (B) Tet 4 species from a nondenaturing gel were cross-linked with short-wave UV (+UV, lanes 3 and 4), excised, and run on a denaturing gel (D-PAGE) next to Tet 4 monomers and multimers that had not been exposed to UV (–UV, lanes 1 and 2). Only the multimeric Tet 4 species formed cross-linked (lanes 4 and 6) species having a size similar to the known tetramer Y4 in lane 8 (see Figure 3A for oligo Y sequence). As expected, Y monomers (lane 7) were unaffected by UV irradiation. $d(T)_{24}$ (lane 5) was used as a marker to indicate the position of unstructured monomers. Tet 4 intramolecular foldback monomers (panel A, *) run faster on denaturing gels after cross-linking (panel B, **). (C) Formation of G4-DNA is greatly facilitated by the presence of Mg^{2+} in addition to Na^+ and K^+. G4-DNA oligos X and Y were incubated in the presence of different combinations of cations indicated at the top in addition to 10 mM Tris-HCl (pH 7.5). The 5 possible X-Y four-stranded molecules are indicated at the left.

four-stranded structures (Sen & Gilbert, 1990). Following UV irradiation, only the Y tetramer (Figure 1B, lane 8) formed cross-linked species. Y monomers (Figure 1B, lane 7) migrated at the position expected for a 23-mer. Slow migrating forms of Tet 4 exposed to UV (Figure 1B, lane 6) ran parallel to cross-linked Y quadruplexes. Thus, we conclude that the Tet 4 multimers very likely consist of complexes of four separate strands.

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

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

Previous studies have shown that high oligonucleotide concentrations facilitate the formation of G4-DNA (Sen & Gilbert, 1988, 1990). Therefore, it was surprising to find

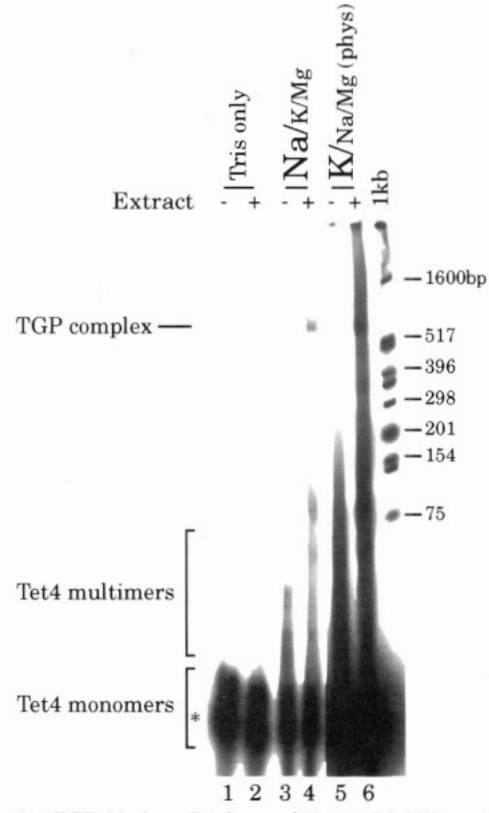

FIGURE 2: TGP binds a G4 form of Tet 4. Mobility retardation assays were used to determine the species of Tet 4 bound by TGP. In the presence of S-100 extract (+ lanes) neither unstructured Tet 4 monomers nor intramolecular foldback monomers (lanes 1 and 2, *) were bound by TGP (lane 2). A TGP complex was formed only when the multimeric species of Tet 4 was available (lanes 4 and 6). Physiological salt conditions (lanes 5 and 6) supported formation of multimeric Tet 4 (lanes 5 and 6) and TGP complex formation (lane 6).

that X and Y formed G4-DNA at concentrations 250-fold more dilute (0.004 μg/μL, Figure 1C) than the conditions used by Sen and Gilbert (1990) (1 μg/μL), indicating that

A

		Compet. Ability	Binding Ability
OXY4	TTTTGGGGTTTTGGGGTTTTGGGGTTTTGGGG (32)	++	+
TET4	TTGGGGTTGGGGTTGGGGTTGGGG (24)	++	+
Y	ACTGTCGTACTTGATATGGGGGT (23)	+	+
$(TET1.5)_n$	$(GGGGTTGGGG)_n$	nd	+
OXY1.5	GGGGTTTTGGGG (12)	-	-
X	TGATATGGGGGC (12)	-	-
Tet1.5	GGGGTTGGGG (10)	nd	-
$r(U_2G_4)_4$	UUGGGGUUGGGGUUGGGGUUGGGG (24)	-	-

nd=not determined

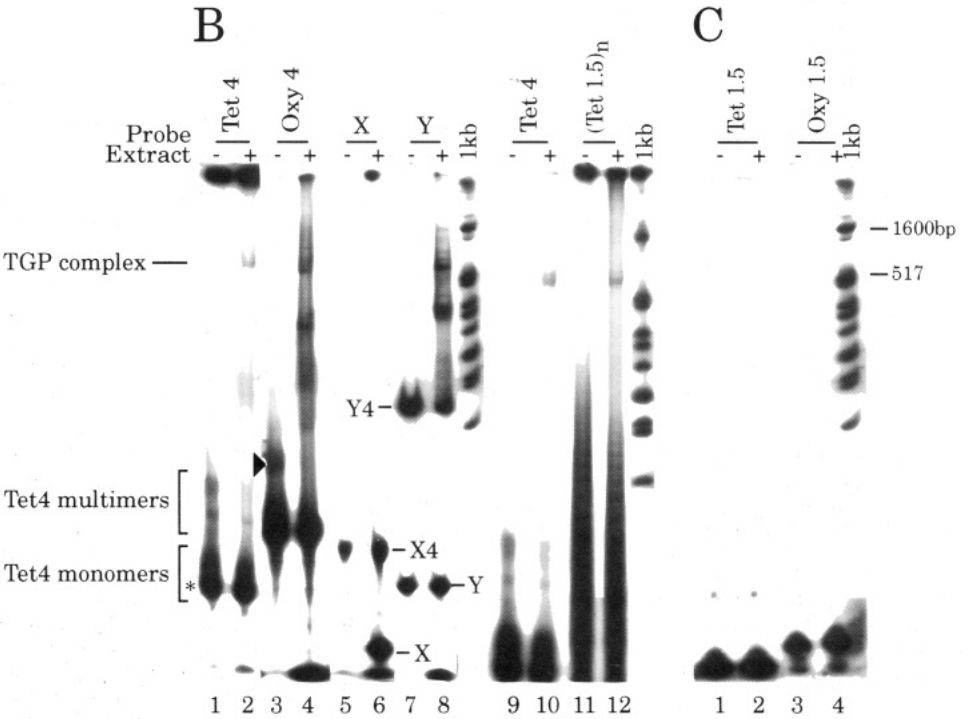

FIGURE 3: TGP binds preferentially to parallel-stranded quadruplex DNA. (A) Summary of binding and competition assays. Several different G-quartet structures were used as competitors for TGP binding. Competition assays using Tet 4 as the probe were consistent with the binding assays shown in panels B and C. Oxy 4 and Tet 4 were the best competitors, indicated by "++". Y4 was a strong competitor (+) while X, Oxy 1.5, and $r(U_2G_4)_4$ were poor competitors (–). (B) In mobility retardation assays, multimeric forms of Tet 4 and Oxy 4 formed TGP complexes (lanes 2 and 4) in addition to the known parallel-stranded quadruplex, Y4 (lane 8). A novel, multimeric, parallel-stranded structure formed by Tet1.5 [$(Tet 1.5)_n$] also formed a TGP complex (lane 12). The TGP complex always migrated between the 517 and 1600 bp markers of the 1 kb ladder (right). "+" or "–" indicates the presence or absence of S-100 extract. (C) The antiparallel hairpin dimer Oxy 1.5 (lanes 3 and 4; Rich et al., 1992) did not form a TGP complex and neither did Tet 1.5 species (lanes 1 and 2).

concentrations over 1 $\mu g/\mu L$ are not necessary to drive formation of G4-DNA.

In addition to the intermolecular Tet 4 species described above, Tet 4 readily folds into an intramolecular quadruplex structure represented by the fastest migrating band (indicated by *, Figure 1A). A G-quartet model for the intramolecular foldback structure was proposed by Williamson (1989) and confirmed by X-ray crystallography (Kang et al., 1992) and NMR (Smith & Feigon, 1992). When cross-linked by UV irradiation, it migrates faster on denaturing gels than unstructured 24-mers (indicated by **, Figure 1B, lane 3; Williamson et al., 1989). This foldback monomer was the dominant species formed by Tet 4 under the salt conditions used in the binding assays described below.

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

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

TGP Binds to Quadruplexes in Parallel-Stranded Arrangements. The structures formed by several G-rich oligos have been well characterized in recent years (Williamson et al., 1989; Sundquist & Klug, 1989; Sen & Gilbert, 1990; Jin et al., 1992; Aboul-ela et al., 1992; Kang et al., 1992; Smith & Feigon, 1992; Gupta et al., 1993). A number of different arrangements (parallel and antiparallel) and strand stoichiometries (monomers, dimers, and tetramers) exist in this structural family. To test the structural requirement for TGP binding, the oligos listed in Figures 3A, 4A, and 5A were used in electrophoretic mobility retardation and binding competition experiments. In Figure 3A, the TGP complex increased in intensity in proportion to the availability of quadruplex DNA when the amount of total protein added to the binding reaction was held constant. This trend is evident for the intermolecular quadruplexes formed by Tet 4, Oxy 4, Y4, and $(Tet 1.5)_n$. Oxy 4 formed a slow migrating species analogous to the Tet

A

		Compet. Ability	Binding Ability
XYa	ACTTGATATGGGGGT (15)	-	-
XYb	CGTACTTGATATGGGGGT (18)	-	-
XYc	TGTCGTACTTGATATGGGGGT (21)	+	+

B

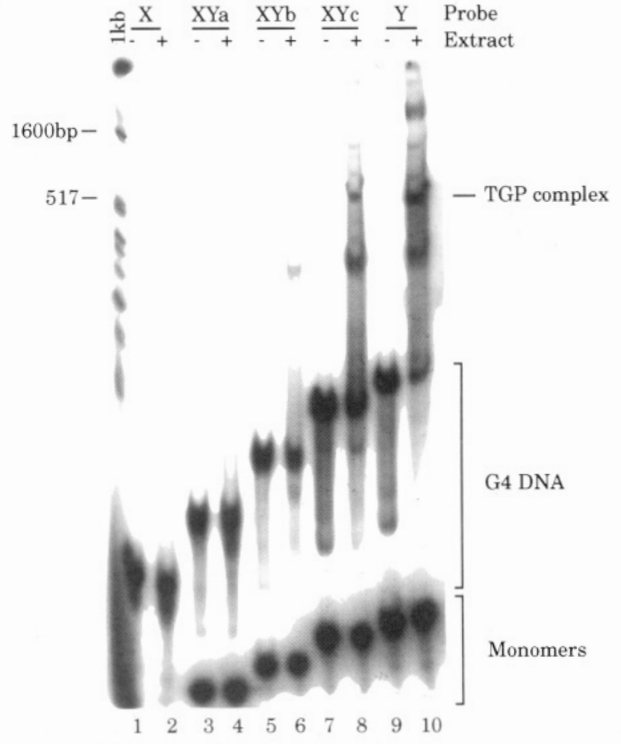

FIGURE 4: TGP has a tail length requirement for optimal binding. (A) Summary of binding and competition assays. Competition experiments with Tet 4 as the probe corroborated the binding assay results. XYa and XYb were both poor competitors (–) while XYc competed as effectively as Y (+). (B) Mobility retardation assays with X (lanes 1 and 2), XYa (lanes 3 and 4), XYb (lanes 5 and 6), XYc (lanes 7 and 8), and Y (lanes 9 and 10) as the probe (indicated at the top) demonstrated that TGP binds only to G4 molecules with a 5′ tail length of at least 15 nucleotides (XYc, lane 8).

4 tetramers (Figure 3B, lane 3, arrowhead). This slow form was shifted preferentially over the monomeric species (Figure 3B, lane 4). Similarly, over 70% of the Y4 molecules were shifted in the presence of protein with as much as 97% shifted in some experiments. In contrast, there was no detectable reduction in Y monomers in the presence of protein (Figure 3B, lane 8). Therefore, Y monomers, like Tet 4 monomers, are not bound by TGP.

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

Competition experiments using Tet 4 as the probe and the oligos listed in Figure 3A as the unlabeled competitor corroborated binding assay results. Four levels of unlabeled competitor were tested (2-fold molar excess over the Tet 4

A

		Compet. Ability	Binding Ability
D5 duplex	TTTTTTTGGCAAAAAACAAAAATAGTAA AAAAAAACCGTTTTTTGTTTTTATCATT (28)	-	nd
CS duplex	TTTTTTTCCCAAAAAAAAAAACAAAAATAGTAA AAAAAAAGGGTTTTTTTTTTTGTTTTTATCATT (33)	-	nd
Telo duplex	AAAACTCGACTTGGGGTTGGGGTAGTGCATCGAC TTTTGAGCTGAACCCCAACCCC	-	nd
1G ss	CTCAAGAACTCA (12)	-	-
4G ss	TGAGTTCTTGAG (12)	-	-
TR1 ss	TTTTGATCCGGAGATGTTTCCCC (23)	-	-
TR2 ss	AAAAAAAAAAACAAAAATAGTAA (23)	-	-
TR3 ss	ACCTTCCGAACTTTTGCAACTTTGAGA (27)	-	-
d(C4A2)4	CCCCAACCCCAACCCCAACCCCAA (24)	na	-
d(T)24	TTTTTTTTTTTTTTTTTTTTTTTT	-	-

nd=not determined
na=not applicable

B

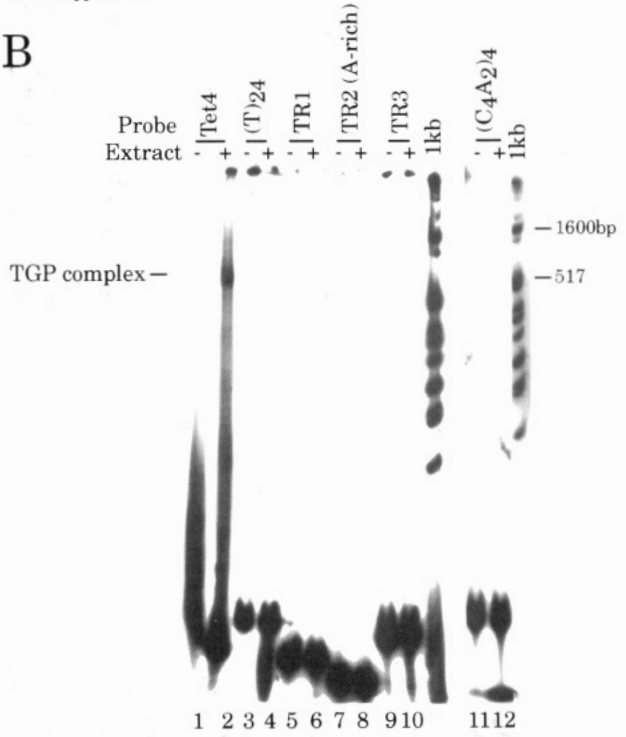

FIGURE 5: Non-G4-forming oligos do not form TGP complexes. (A) Summary of binding and competition assays. Competition assays with Tet 4 as the probe corroborated the binding assay results and also demonstrated that two different A/T-rich duplexes were not bound by TGP since all oligos tested were poor competitors (–). (B) Mobility retardation assays using several different single-stranded oligos as probes were performed to test the ability of A-rich (lanes 7 and 8), T-rich (lanes 3 and 4), and C-rich oligos (lanes 11 and 12) to form a TGP complex. In addition, oligos from the replication origin of *Tetrahymena* rDNA were found not to form TGP complexes (TR1, lanes 5 and 6, and TR2, lanes 9 and 10) even though these oligos bind factors in S-100 extracts under different conditions (A. Umthun, Z. Sibenaller, W. Shaiu, and D. Larson, unpublished results).

probe, 10-fold, 50-fold, and 250-fold). A 200-fold excess of $d(T)_{24}$ was included as a nonspecific competitor. Oxy 4 and Tet 4 were the best competitors, having the ability to compete over 80% of the TGP complex counts at a 50-fold molar excess (Figure 3A). In addition, Y competed much better than X. Sixty percent of the TGP counts were competed by Y at a 50× molar excess whereas X was unable to compete even at a 50× molar excess (Figure 3A).

In contrast to the parallel quadruplexes in Figure 3B, the antiparallel hairpin dimer Oxy 1.5 (Kang et al., 1992; Smith & Feigon, 1992) was not bound by TGP (Figure 3C, lane 4) and neither was the analogous molecule from *Tetrahymena*, Tet 1.5 (Figure 3C, lane 2). Furthermore, Oxy 1.5 was a poor competitor for TGP binding (Figure 3A), indicating that TGP does not bind to antiparallel quadruplex structures.

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

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

The abundance of TGP in S-100 extracts was estimated from the picomoles of G4-DNA shifted. Enough protein exists in 0.5 μg of extract to bind 0.00625 pmol of intermolecular Tet 4. Therefore, there are at least 7.5×10^9 protein molecules per μg of extract (1.6×10^6 copies/cell). The abundance of TGP should facilitate its purification. Attempts to localize TGP activity have thus far been inconclusive.

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

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

Non-G4 Oligos as Substrates for TGP. Several non-G4-forming sequences were also tested for binding to TGP (Figure 5). Oligos TR1, TR2, and TR3 correspond to sequences found in the replication origin of *Tetrahymena* rRNA genes and are bound by factors from *Tetrahymena* (A. Umthun, Z. Sibenaller, W. Shaiu, and D. Larson, unpublished results). T-rich $[d(T)_{24}]$, A-rich (TR2), and C-rich $[d(C_4A_2)_4]$ single-stranded oligos were also tested in binding and competition assays. None were found to bind to or compete for TGP (Figure 5). In addition, the conditions of the assay were such that the complexes normally formed by the oligos TR1, TR2, and TR3 in S-100 extracts were not observed. Oligos which mimic telomeric DNA were also tested. Telomeric duplex oligos (Cardenas et al., 1993) failed to compete (Figure 5A), and telomeric C-strand oligos did not form a TGP complex (Figure 5B, lane 12). Competition experiments using $d(C_4A_2)_4$ were hindered by duplex formation. Finally, two different A/T-rich duplexes from the origin of replication in *Tetrahymena* rDNA were also tested as competitors for TGP and found not to compete at any level (Figure 5A).

DISCUSSION

TGP Binding Specificity. Telomeric and nontelomeric G-rich sequences can form G4-DNA structures [reviewed by Sundquist (1991), Henderson et al. (1987), Williamson et al. (1989), Sundquist and Klug (1989), Sen and Gilbert (1990), Jin et al. (1992), Kang et al. (1992), and Smith and Feigon (1992)]. This study demonstrates that *Tetrahymena* contains an abundant protein, TGP, that binds specifically to G4 structures formed by the telomeric G-strand sequence $d(T_2G_4)_4$ and not to Tet 4 intramolecular foldback monomers or to unstructured Tet 4 monomers. Furthermore, TGP binds parallel-stranded G4-DNA structures having very different primary sequences from Tet 4, demonstrating its structure-specific nature. Taken together, the abundance and structural specificity of TGP suggest that it may have an important function in the cell.

TGP does not bind to G/C-rich or A/T-rich duplex DNA, nor to non-G-rich single-stranded DNA. TGP also fails to bind G-rich RNA monomers $[(U_2G_4)_4]$. However, we were unable to unequivocally test for the ability of G4-RNA to form TGP complexes since $(U_2G_4)_4$ did not readily form G4-RNA under the conditions used in these experiments. Thus, it is still possible that TGP has affinity for G4-RNA.

An unusual feature of TGP is that it may require a single-stranded region in addition to a G-quartet structure region. The single-stranded requirement is most obviously demonstrated with the X and Y oligos. The length of the Y tail is an important feature suggesting that a longer tail somehow affects the G-quartet region in a manner that makes it available for binding by TGP. Single-stranded tails would also be associated with G4-DNA structures formed by Tet 4 and Oxy 4 oligos in which the individual strands are not precisely aligned. The presence of several different intermolecular Tet 4 structures on nondenaturing gels (Figure 1A) suggests that the individual strands of the G4 complexes are arranged in several different ways and likely contain single-stranded tail regions. If single-stranded tails adjacent to a quartet structure are a requirement for TGP binding, it is possible that the absence of a TGP/Oxy 1.5 or Tet 1.5 complex is due to the lack of a single-stranded tail and not due to a specific requirement for parallel DNA. It will be of interest to test antiparallel molecules having long single-stranded tails for TGP binding.

TGP's preference for molecules with a single-stranded tail is similar to the binding properties of the yeast nuclease identified by Liu et al. (1993). The yeast nuclease appears to require G4 structure for cleavage of a single-stranded region 5′ of the G4 structure since single-stranded forms of the oligos capable of forming G4-DNA are not cleaved. Nuclease activity is present in crude extracts containing TGP binding activity. Cleavage products are observed at the bottom of

gels only in the presence of extract (Figure 3B). However, it is not known whether the two activities are related since these extracts contain many proteins. Purification of TGP will allow us to address this question.

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

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

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

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

ACKNOWLEDGMENT

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

REFERENCES

Aboul-ela, F., Murchie, A. I. H., & Lilley, D. M. J. (1992) *Nature 360*, 280–282.
Acevedo, O. L., Dickinson, L. A., Macke, T. J., & Thomas, C. A. (1991) *Nucleic Acids Res. 19*, 3409–3419.
Blackburn, E. H. (1991) *Nature 350*, 569–573.
Blackburn, E. H., & Szostak, J. (1984) *Annu. Rev. Biochem. 53*, 163–194.
Cardenas, M. E., Bianchi, A., & De Lange, T. (1993) *Genes Dev. 7*, 883–894.
Cheong, C., & Moore, P. B. (1992) *Biochemistry 31*, 8406–8414.
Chung, I. K., Mehta, V. B., Spitzner, J. R., & Muller, M. T. (1992) *Nucleic Acids Res. 20*, 1973–1977.
Fang, G., & Cech, T. R. (1993) *Cell 74*, 875–885.
Greider, C. W., & Blackburn, E. H. (1985) *Cell 43*, 405–413.
Greider, C. W., & Blackburn, E. H. (1987) *Cell 51*, 887–898.
Gupta, G., Garcia, A. E., Guo, Q., Lu, M., & Kallenbach, N. R. (1993) *Biochemistry 32*, 7098–7103.
Hardin, C. C., Henderson, E., Watson, T., & Prosser, J. K. (1991) *Biochemistry 30*, 4460–4472.
Hastie, N. D., & Allshire, R. C. (1989) *Trends Genet. 5*, 326–331.
Henderson, E., & Blackburn, E. H. (1989) *Mol. Cell. Biol. 9*, 345–348.
Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. J., & Blackburn, E. H. (1987) *Cell 51*, 899–908.
Jin, R., Breslauer, K. J., Jones, R. A., & Gaffney, B. L. (1990) *Science 250*, 543–546.
Jin, R., Gaffney, B. L., Wang, C., Jones, R. A., & Breslauer, K. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 8832–8836.
Kang, C. H., Zhang, X., Ratliff, R., Moyzis, R., & Rich, A. (1992) *Nature 356*, 126–131.
Klobutcher, L. A., Swanton, M. T., Donini, P., & Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3015–3019.
Liu, Z., Frantz, J. D., Gilbert, W., & Tye, B. K. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3157–3161.
Marquet, R., Baudin, F., Gabus, C., Darlix, J.-L., Mougel, M., Ehresmann, C., & Ehresmann, B. (1991) *Nucleic Acids Res. 19*, 2349–2357.
Nickol, J. M., & Felsenfeld, G. (1983) *Cell 35*, 467–477.
Pears, C. J., & Williams, J. G. (1988) *Nucleic Acids Res. 16*, 8467–8486.
Pearson, A. M., Rich, A., & Krieger, M. (1993) *J. Biol. Chem. 268*, 3546–3554.
Pluta, A. F., Kaine, B. P., & Spear, B. B. (1982) *Nucleic Acids Res. 10*, 8145–8154.
Raghuraman, M. K., & Cech, T. R. (1990) *Nucleic Acids Res. 18*, 4543–4552.
Ryan, D. A., Shapiro, T. A., Rauch, C. A., & Englund, P. T. (1988) *Annu. Rev. Microbiol. 42*, 339–358.
Sen, D., & Gilbert, W. (1988) *Nature 334*, 364–366.
Sen, D., & Gilbert, W. (1990) *Nature 344*, 410–414.
Smith, F. W., & Feigon, J. (1992) *Nature 356*, 164–168.
Sundquist, W. I. (1991) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D. M. J., Eds.) pp 1–24, Springer-Verlag, Berlin.
Sundquist, W. I., & Klug, A. (1989) *Nature 342*, 825–829.
Sundquist, W. I., & Heaphy, S. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3393–3397.
Venczel, E. A., & Sen, D. (1993) *Biochemistry 32*, 6220–6228.
Walsh, K., & Gualberto, A. (1992) *J. Biol. Chem. 267*, 13714–13718.
Weisman-Shomer, P., & Fry, M. (1993) *J. Biol. Chem. 268*, 3306–3312.
Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell 59*, 871–880.
Zahler, A. M., & Prescott, D. M. (1988) *Nucleic Acids Res. 16*, 6953–6972.
Zahler, A. M., Williamson, J. R., Cech, T. R., & Prescott, D. M. (1991) *Nature 350*, 718–720.
Zimmerman, S. B., Cohen, G. H., & Davies, D. R. (1975) *J. Mol. Biol. 92*, 181–192.