***Oxytricha* telomeric nucleoprotein complexes reconstituted with synthetic DNA**

M.K.Raghuraman¹, Christopher J.Dunn^{2*}, Brian J.Hicke¹ and Thomas R.Cech^{1,2}

¹Department of Molecular, Cellular and Developmental Biology and ²Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309-0215, USA

Received March 1, 1989; Revised and Accepted May 9, 1989

ABSTRACT

The telomere binding protein from macronuclei of *Oxytricha nova* binds macronuclear DNA *in vitro*, protecting the 3'-terminal single-stranded (T₄G₄)₂ tail from chemical and enzymatic probes. We have used synthetic oligodeoxynucleotides to study the binding properties of the telomere protein. It binds at the 3' end of single-stranded oligonucleotides that have the sequence (T₄G₄)_n, where n ≥ 2, reconstituting the methylation protection seen with macronuclear DNA. Three oligonucleotide•protein complexes are resolved in nondenaturing gels, all specific for this sequence. Single-stranded oligonucleotides that have one or more repeats of the sequence C₄A₄ are also recognized, forming a single complex. The dissociation constant for (T₄G₄)₄ is about 19 nM, and for macronuclear DNA is at least 20-fold lower. The basis for this difference is not fully understood, but it is not simply due to the absence of a (C₄A₄)_{2.5}•(G₄T₄)_{2.5} region on the oligonucleotide. Transversions of T's to A's or of G's to C's in the 3' tail portion prevent binding. Changing T's to dU's does not prevent binding, indicating that the hydrophobic 5-methyl group is not required for binding as had been suggested from the salt-stability of the complex. The properties of the DNA-protein complex suggest a revised model for telomere synthesis in *Oxytricha*.

INTRODUCTION

It has long been understood that telomeres, the natural ends of chromosomes, have special properties that are not shared by the ends of broken chromosomes (see ref. 1 for a review). Telomeres stabilize linear chromosomes, preventing both degradation and the end-to-end joining that broken chromosomes undergo. They must also mediate the replication of the chromosomal ends, as normal DNA polymerases would not be able to fill in the gap left by removal of the RNA primer at the 5' end of the chromosome. There is, hence, an obvious need to understand the properties of the nucleic acid and protein components that make up these specialized structures, and the interactions between these components.

One finds a striking similarity in the DNA sequences present in nuclear telomeres of organisms as diverse as yeast, ciliates, trypanosomes, slime molds, plants and humans (2-9): all have tandem repeats of a sequence of the form C₁₋₈(T/A)₁₋₄. The conservation of telomeric sequences suggests that they have properties important for telomere function. While they conform to the general consensus sequence, telomeric DNA from different organisms may differ both in the precise nucleotide sequence and in the organization of the tandem repeats. For instance,

Tetrahymena telomeres are heterogeneous in length, having 30-100 copies of a C₄A₂ repeat, whereas *Oxytricha* telomeric DNA has a defined organization (5):

The sequences internal to the C₄A₄ repeats are unique; no other conserved telomere-associated sequences are known.

There is increasing evidence for another feature common to nuclear telomeres: they occur in the cell as non-nucleosomal complexes with protein (10-14). Isolation and study of the telomeric nucleoprotein complex has been particularly fruitful in *Oxytricha nova*, facilitated by the large number of linear DNA molecules in the cell, about 24×10^6 per macronucleus. Two structural domains have been identified in the *Oxytricha* nucleoprotein complex. A salt-sensitive domain extends from about 45 to 135 base pairs from the end of the chromosome. This domain appears to have DNA wound around the surface of protein, as in a nucleosome, but is smaller than a normal nucleosome (14, 16). It is not known whether histones are present in this domain. The very end of the DNA, including the 16 bases of the 3' single-stranded extension or tail, is bound tightly, but non-covalently, by a protein heterodimer (15-17). This complex is stable in 2 M NaCl (15, 16). The two polypeptides, 55 kDa and 43 kDa in size, can be purified by a scheme devised by Gottschling and Zakian (15). These workers also showed that the purified protein can bind macronuclear DNA *in vitro*, and can form complexes with synthetic DNA as judged by nitrocellulose filter binding. Protection of G's in the 3' tail from chemical modification, seen in whole cells and macronuclei, can also be restored *in vitro* by addition of the protein to purified macronuclear DNA (16).

In this study, we have further explored the use of synthetic oligodeoxynucleotides as substrates for the telomere binding protein. A gel mobility shift assay has revealed that the telomere protein forms three complexes with these oligonucleotides.

MATERIALS AND METHODS

Growth of *Oxytricha nova* and Isolation of Macronuclei

Oxytricha nova strain HO was grown in nonsterile culture with live *Chlorogonium* as the only food source (18). Macronuclei were isolated essentially as described (18). The isolation buffers contained 0.1 mM phenylmethanesulfonyl fluoride (PMSF), 0.1 mM tosylphenylalanine chloromethylketone (TPCK) and 2 mM *p*-chloromercuribenzenesulfonic acid (PCMBs) as protease inhibitors. PCMBs has the additional effect of stabilizing the binding activity of the protein (17).

Oligonucleotide Synthesis

Oligodeoxynucleotides were made on an Applied Biosystems 380B DNA synthesizer and purified by preparative electrophoresis through 10-20% polyacrylamide, 8M urea gels. The purified DNA was resuspended in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA (TE), and quantitated by UV absorbance.

Hybridization of Oligonucleotides

GT46 and CA30 (see Table 1) were hybridized by heating a 1:1 mixture of the oligonucleotides to 90°C in TE and cooling the mixture to room temperature either directly (Fig. 4) or in steps of 5-30 minutes at 65°, 37°, 25° and 0°C (Fig. 2). The degree of hybridization was assayed by electrophoresis through nondenaturing gels, which separates the hybridized from the single-stranded oligonucleotides.

Isolation of Telomere Protein

Telomere protein was isolated essentially as described (15, 16). $15\text{--}20 \times 10^7$ macronuclei were lysed in 6-8 ml of TE plus 2 M NaCl, 0.1 mM TPCK, 0.1 mM PMSF for 2 hours on ice. The supernatant was applied to a Biogel A 15M column equilibrated with TE plus 2 M NaCl. DNA-containing fractions (with the telomere protein bound to the DNA) were collected and dialysed against TE plus 0.1 mM TPCK, 0.1 mM PMSF. CaCl_2 and micrococcal nuclease were added to concentrations of 2 mM and 0.1 units/ml respectively and the solution was concentrated to 5 ml in an Amicon ultrafiltration cell. After an incubation at 37°C for 1 hour to digest the remaining DNA, the protein preparation was concentrated to 1 ml in Centricon-30 microconcentrators. The nuclease was then inactivated by the addition of EGTA to a concentration of 15 mM. The protease inhibitors were diluted in 1-5 ml of TE before being added to the protein solution, as addition of the stock solutions (in methanol or isopropanol) directly to the protein solution seemed to promote precipitation of the protein. The protein was analyzed by electrophoresis through 12% polyacrylamide/SDS gels (19). Protein prepared in this manner is essentially pure as judged by Coomassie Blue staining of SDS/polyacrylamide gels (16, 17, and

Table 1. Sequences of oligonucleotides used.

<u>Oligonucleotide</u>	<u>Sequence</u>
GT32	5' - TTTTGGGGTTTTGGGGTTTTGGGGTTTTGGGG - 3'
GT46	5' - CGGCTTAAGCGGGGTTTTGGGGTTTTGGGGTTTTGGGGTTTTGGGG - 3'
CA30	3' - GCCGAATTCGCCCAAACCCCAAACCCC - 5'
T→U tail	5' - CGGCTTAAGCGGGGTTTTGGGGTTTTGGGGUUUUUGGGUUUUUGGGG - 3'
G→C tail	5' - CGGCTTAAGCGGGGTTTTGGGGTTTTGGGGTTTTCCCTTTTCCCC - 3'
GA32	5' - AAAAGGGGAAAAGGGGAAAAGGGGAAAAGGGG - 3'
TGPermut	5' - GGGGTTTTGGGGTTTTGGGGTTTTGGGGTTTT - 3'
T ₁ G ₂ toCA	5' - TTTTGGGGTTTTGGGGTTTTACACACACGGGG - 3'
XT1A	5' - CTAGGATCTGAGACCTCATGACGAGATCCTTCAGTCTGTGCTGAGTAGACAC - 3'
XT1B	5' - GTGTCTACTCAGCACAGACTGAAAGGATCTCGTCATGAGGTCTCAGATCCTAG - 3'

Fig. 1A). Typically, the final protein concentration was about 2-10 pmol/μl, with at least 60-70% of the protein retaining the 3' terminal single-stranded extension (unpublished observations).

DNA Binding Assays

Methylation protection assays were done essentially as described (16). Purified *Oxytricha* macronuclear DNA that had been labeled at the 3' end with cordycepin, or synthetic oligodeoxynucleotides labeled at the 5' end, were mixed with the desired amount of telomere protein and the volume adjusted to 10 μl with TE. When the labeled DNA was an oligonucleotide, 1 μg of sonicated and heat-denatured calf thymus DNA was included. After 30-60 minutes at 37°C, dimethyl sulfate (DMS) was added to a concentration of 0.1% and the reaction was allowed to proceed for 10 minutes at room temperature. The reaction was stopped by the addition of 1/4 volume of 2.5 M β-mercaptoethanol, 1.5 M sodium acetate, 100 μg/ml carrier DNA and precipitated with three volumes of ethanol. Dried DNA pellets were resuspended in 100 μl of 10% piperidine, heated to 90°C for 30 minutes and lyophilized to remove the piperidine. Dried samples were dissolved in loading buffer (80% formamide, 10 mM NaOH, 1mM EDTA, 0.1% xylene cyanol), heated to 90°C for a minute and loaded on a 10 or 20% polyacrylamide sequencing gel for electrophoresis.

Gel mobility shift assay. Endlabeled oligonucleotides (0.1-1 pmol) were incubated with telomere protein in the presence of 0.5-2 μg sonicated, denatured calf thymus DNA in 10-20 μl of TE for 30-60 minutes. Unless otherwise indicated the incubation temperature was 37-39°C. The samples were then prepared for electrophoresis by the addition of 1/5 volume of 50% glycerol, 0.01% bromophenol blue and loaded on a 10% polyacrylamide (55:1 acrylamide::bis-acrylamide) gel that had been pre-run at 5-15 mA for 1-2 hours. Electrophoresis was for 2-6 hours at 4°C and 5-15 mA. The buffer used was 44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA (0.5x TBE). Quantitation of radioactive bands (Fig. 4A) was by scanning of a dried gel in an Ambis radioanalytic imaging system.

Nitrocellulose filter binding assay. The protein-DNA complex was formed at 37°C as described above. The samples were diluted to 100 μl with ice-cold TE immediately before being applied to nitrocellulose filter discs (24 mm circles, 0.45 μm pore size, Schleicher & Schuell) that had been soaked for at least 10 minutes in TE. Filters were then washed three times with 1 ml of cold TE and the amount of bound DNA was quantitated in a liquid scintillation counter.

Derivation of the Binding Constant

The dissociation constant (K_D) for a system at equilibrium can be described by the equation

$$K_D = \frac{[\text{Protein}]_{\text{free}} \cdot [\text{DNA}]_{\text{free}}}{[\text{Complex}]}$$

for a complex with a 1:1 molar ratio of protein to DNA. The concentration of added protein is kept constant and the amount of complex formed is measured with increasing amounts of added DNA. When the DNA is in excess, the protein is saturated and the amount of complex formed reaches a plateau. At half saturation, the amount of free protein and the amount of complex are

Figure 1. Specific binding of single-stranded oligonucleotides containing repeats of G₄T₄ or C₄A₄ by telomere protein. (A) Coomassie Blue-stained SDS/polyacrylamide gel showing the isolated telomere protein. Indicated on the left are the positions and molecular weights (in kilodaltons) of marker proteins run in the same gel. (B,C,D): 5'-[³²P]-labeled oligonucleotides were incubated with telomere protein and 1 µg of sonicated, denatured calf thymus DNA as described below. The complexes were separated from free DNA by electrophoresis through nondenaturing gels. (B) A gel shift assay with GT32 as substrate; the three complexes are indicated. GT32 was incubated with telomere protein (lane 1) or without protein (lane 4); lane 2, the protein was treated for 5 minutes at 37°C with SDS (0.1%) and proteinase K (100 µg/ml), after which the DNA was added and incubation continued for another 30 minutes; lane 3, the protein and DNA were mixed and incubated for 30 minutes, and then treated with SDS and proteinase K for 5 minutes at 37°C. (C) A gel shift assay comparing GT32 and the control, non-telomeric sequence oligonucleotides XT1A and XT1B as substrates. (D) Binding of CA30, which contains repeats of the sequence C₄A₄.

equal; these two terms cancel out of the equation, and the K_D is equal to the concentration of free DNA.

RESULTS

Reconstitution of Telomeric Nucleoprotein Complex with Synthetic Oligodeoxynucleotides

We used a version of the gel shift assay developed by Schneider et al. (20) to monitor binding of various DNA oligonucleotides by the telomere binding protein, the assumption being that DNA bound to protein would move slower than free DNA in nondenaturing gels. As shown in Fig. 1B and Fig. 6, the oligonucleotide GT32, which has the telomere-like sequence (T₄G₄)₄,

forms three DNA-protein complexes. We call these complexes 1, 2 and 3 in order of increasing mobility. Complex 1 remains in the well of the gel. We tried a variety of conditions and gel systems, including agarose gels, to see if complex 1 could be made to enter the gel; all attempts were uniformly unsuccessful.

Formation of all three complexes requires the continued presence of protein, as treatment of the protein with SDS and proteinase K either before or after incubation with DNA led to no complex formation (Fig. 1B). Further, Coomassie Blue staining of a gel containing protein-DNA complexes showed that the three complexes each contain protein; the DNA by itself gave no staining under these conditions (data not shown). All three complexes are specific for the telomeric DNA sequence: no complex is formed when the oligonucleotides XT1A and XT1B, of non-telomeric sequence, are used (Fig. 1C). All three complexes are also seen when we use GT46•CA30, which mimics the structure of natural telomeric DNA; as expected from previous work (15), oligonucleotides that are double-stranded through their entire length are not recognized (data not shown).

In a separate study (21), we show that complexes 1 and 2 have the same protein-DNA contacts, while complex 3 has a different set of protein-DNA contacts and is less stable than complexes 1 and 2. Complex 2 has almost the same properties as complex 1, and arises from slightly proteolyzed protein (21).

The telomere protein also binds single-stranded DNAs containing the C₄A₄ sequence of *Oxytricha* telomeric DNA (Fig. 1D). One complex predominates, comigrating with complex 3, although a slower-moving complex can be detected upon prolonged exposure of the film. The smallest oligonucleotide recognized has the sequence C₄A₂ (data not shown).

An interesting property of these complexes is that within limits, the electrophoretic mobility of a given complex seems to be independent of the length of the DNA in the complex. For instance, GT32 (32 nucleotides long) and T→U tail (46 nucleotides long) form complexes that have the same mobility (Fig. 6). This observation holds for oligonucleotides as short as sixteen and as long as seventy nucleotides (data not shown). The mobility of the (C₄A₄)_n complexes is also independent of the length of the added DNA, being the same with n = 1, 2, 4 or 8 (data not shown).

Competition experiments confirmed that the added GT-rich oligonucleotides were being specifically recognized by the telomere-binding protein. The ability of unlabeled oligonucleotides to compete with labeled macronuclear DNA was tested using the methylation protection assay, where formation of the complex results in a characteristic pattern of protection of G's in the 3' single-stranded tail upon treatment with dimethyl sulfate (16). Labeled macronuclear DNA was added to protein that had been pre-incubated with various unlabeled oligonucleotides, and then tested for methylation protection. When the protein is pre-incubated with oligonucleotides of the proper telomeric DNA sequence, the characteristic protection pattern is not seen, while oligonucleotides with non-telomere sequences leave the protection pattern unaffected (Fig. 2).

Figure 2. Competition between synthetic oligonucleotides and macronuclear DNA.

Telomere protein was incubated at room temperature with 100 pmol of GT46 (single-stranded, with G_4T_4 repeats), GT46•CA30 (GT46 annealed to a C_4A_4 -containing oligonucleotide to form a natural end), or non-specific oligonucleotides XT1A or XT1B. Thirty minutes later, 3'-labeled macronuclear DNA (0.7 pmol) was added to each sample and after another 15 minutes at room temperature, tested for protein binding by the methylation protection assay. The numbering of G's at the left is from the 3' end; protected bases are indicated at the right with asterisks. The two leftmost lanes are controls, with no protein added and with no competitor.

Single-stranded GT46, which has ten bases of non-telomere sequence preceding 36 bases of $(T_4G_4)_n$ (Table 1) and GT46•CA30, which forms a structure that mimics natural *Oxytricha* telomeres, are both competitors of macronuclear DNA. A comparison of the amount of GT46 or GT46•CA30 required to prevent protection suggests that single-stranded GT46 competes as well as GT46•CA30 (data not shown). In the converse experiment, labeled GT32 DNA was incubated with the protein in the presence of up to a 4-fold molar excess of unlabeled *Oxytricha* macronuclear DNA. The amount of oligonucleotide bound, monitored by the gel shift assay, decreases with increasing amounts of competitor (data not shown). Calf thymus DNA in comparable amounts has no effect on the amount of labeled DNA bound. In fact, complex formation with labeled oligonucleotides is routinely done in the presence of a 10- to 20-fold molar excess of calf thymus DNA ends.

Figure 3. Methylation protection of macronuclear DNA and GT32 oligonucleotide by telomere protein. 3'-labeled macronuclear DNA (left) or 5'-labeled GT32 (right) was subjected to methylation, cleavage and electrophoresis, after incubation in the presence or absence of telomere protein. The G's are numbered from the 3' end. Protected bases are indicated with asterisks.

Footprinting of Reconstituted Complexes with DMS

Although almost all preparations of the *Oxytricha* telomere protein have DNA-binding activity to the extent that they cause retention of macronuclear DNA on nitrocellulose filters, some preparations can bind inaccurately in that they do not confer the characteristic methylation

protection pattern on the G's in the 3' tail (17). Hence, it was important to ascertain that the reconstituted telomere protein•oligonucleotide complexes were authentic as judged by the methylation protection assay. To this end, macronuclear DNA or GT32 oligonucleotide DNA was reconstituted into complexes with the telomere protein, treated with DMS, and cleaved at the modified positions. The resulting DNA fragments were separated by electrophoresis through a denaturing polyacrylamide gel (Fig. 3). Note that the ^{32}P label was at opposite ends of the two DNAs: at the 3' end of macronuclear DNA and at the 5' end of the GT32 oligonucleotide. Bound protein protects the same bases in the oligonucleotide as in macronuclear DNA (Fig. 3; protected bases marked with asterisks), although the protein:DNA ratio has to be 10-20-fold higher for the oligonucleotide to be protected. It would therefore appear that the nature of the protein-DNA interaction is the same whether the DNA is macronuclear, with its characteristic termini, or a synthetic, single-stranded oligonucleotide that is a simple repeat of the T_4G_4 sequence. Further, the oligonucleotide•protein complex is stable in 2 M NaCl, as are the native complex and complexes reconstituted with macronuclear DNA (data not shown).

The protein shows absolute discrimination between the 5' and the 3' halves of the substrate: although the oligonucleotide used in this experiment has four repeats of the sequence T_4G_4 , all of the stable interactions that can be detected by this assay are limited to the 3'-terminal sixteen nucleotides. This is the region of the DNA that would correspond to the 3' single-stranded extension of macronuclear DNA. Sequences 5' to the protected bases must contribute to the binding, however, because the efficiency of binding of $(\text{T}_4\text{G}_4)_2$ is low compared to GT32 (data not shown).

DNA Structure and Efficiency of Protein Binding

Our observations suggested that the native telomeric DNA structure (i.e., 20 base pairs of $\text{G}_4\text{T}_4\text{•C}_4\text{A}_4$ duplex and sixteen bases of T_4G_4 DNA as a 3' extension) is not required for binding by the telomere protein. For instance, the single-stranded oligonucleotide GT46 is as good a competitor for macronuclear DNA as is GT46•CA30 , which mimics a natural telomeric DNA structure (Fig. 2). Also, the protein protects the same bases from DMS in both single-stranded GT32 and in macronuclear DNA (Fig. 3). However, as mentioned earlier, 5' sequences do contribute to the efficiency of binding. Do the 5' sequences have to be in the form of a duplex, as suggested by Gottschling and Zakian (15), or is the structure of the 5' sequence irrelevant? We addressed this question in two ways.

Free GT46 (single-stranded) is separable from free GT46•CA30 (duplex, with a 3' tail) in the nondenaturing gel system used in the gel shift assays. A mixture of GT46 and GT46•CA30 was incubated with telomere protein and the depletion of the two forms of substrate with increasing amounts of added protein was measured (Fig. 4A). In a complementary experiment, the amount of protein•DNA complex formed with increasing amounts of added protein was assayed by nitrocellulose filter-binding for each form of DNA substrate (Fig. 4B). In both sets of experiments, the ^{32}P label was always at the 5' end of GT46. By either assay, it would appear

Figure 4. DNA structure and efficiency of protein binding. (A) Gel shift assay. An approximately equimolar mixture of 5'-[32 P]-GT46 (o) and 5'-[32 P]-GT46•CA30 (●) was incubated with different amounts of protein and separated by electrophoresis through a nondenaturing gel. The amount of free DNA of each form remaining with increasing amounts of added protein is plotted. (B) Nitrocellulose filter-binding assay. 5'-[32 P]-GT46 (o) and 5'-[32 P]-GT46•CA30 (●) were separately incubated with different amounts of protein. The amount of complex formed, measured by the nitrocellulose filter-binding assay, is plotted as a function of the amount of added protein.

that the protein does not prefer one form of DNA over the other, as both substrates are used equally well and both give rise to comparable amounts of complex. It should be mentioned that in these experiments, hybridization of GT46 to CA30 was incomplete; unlabeled, single-stranded CA30 was also present in the mixture of GT46 and GT46•CA30.

The Equilibrium Binding Constant

The dissociation constant (K_D) of a DNA-binding protein is often determined by measuring the concentration of protein required to bind half of the DNA in the assay, the K_D

being the concentration of free protein at that point. The feasibility of this approach depends on a knowledge of the protein concentration. Since a sizeable fraction of the telomere protein preparation retains a portion of the 3' tail fragment of DNA (17; unpublished observations), measuring the concentration of protein available for binding is not easy. We therefore used a variation of this method, keeping the protein concentration constant for a given experiment and varying the concentration of DNA (22). Assuming a 1:1 ratio of protein to DNA in the complex (see below), the K_D would be the concentration of free DNA at half saturation of the protein (as described in Materials and Methods). This method has the advantage that no prior knowledge of the protein concentration is required, and the concentration of the input DNA is very accurately known. The binding constant obtained by this method is a reflection of the affinity of individual binding sites for the substrate, and is independent of the number of binding sites per protein heterodimer. The potential drawback with this approach is that release of the tail fragment during the experiment would affect the K_D measurements. However, since the tail fragment is retained even after several hours of treatment at 37°C with micrococcal nuclease, it seems unlikely that it would be released during the sixty-minute incubation used for the experiment. We therefore assume that protein retaining the 3' tail fragment is inactive in binding exogenous DNA. Before doing these experiments, we confirmed that the formation of complex was not kinetically limited, i.e., that increasing the incubation period beyond the usual 60 minutes did not increase the amount of complex formed.

Formation of complexes was measured using the nitrocellulose filter-binding assay. Binding curves obtained using 5'-[³²P]-GT32 and three different concentrations of protein are shown (Fig. 5). The dotted lines in Figure 5 are theoretical curves calculated using the equation described in Materials and Methods. The concentration of binding sites was estimated from the concentration of the complex at the plateau for the lowest protein concentration used ($\square$); an estimate of the K_D derived from each set of measurements was used to generate the theoretical curve for that set of measurements. The K_D values obtained are 12×10^{-9} , 15×10^{-9} and 29×10^{-9} M, giving an average of $19 \pm 7 \times 10^{-9}$ M for GT32. This is an overall affinity measurement and does not distinguish between the contribution of individual kinds of complexes or potential interaction between complexes.

The stoichiometry of the protein and DNA in the complex has not been conclusively established. However, both methylation protection (16) and protein-protein crosslinking studies (17) support a model of one protein heterodimer per DNA end. Our estimation of the K_D assumes a protein:DNA ratio of 1. The best fit of the theoretical curves to the observed data is obtained when we use a protein:DNA ratio of 1 for the theoretical calculations, but we cannot rule out a protein:DNA ratio of 2.

The concentration of the complex when the protein is saturated (i.e., at the plateau in Fig. 5) is a measure of the concentration of binding sites available. We estimate, from this value, that the concentration of active binding sites is about 1 pmol/ μ l in the preparation of protein used

Figure 5. Measuring the equilibrium binding constant. Various amounts of labeled GT32 were incubated with a fixed amount of telomere protein and the resulting complex quantitated by nitrocellulose filter-binding. The concentration of the complex is plotted as a function of the total concentration of GT32 present. Measurements were taken at three different protein concentrations (1x, $\square$; 3x, $\circ$; and 5x, Δ). Each point on the graph is the mean of three samples; the vertical bars, where present, indicate the minimum and maximum value obtained for each point. Where vertical bars are not shown, the variation was less than half the size of the symbol used. The dotted lines represent theoretical curves, generated as described in the text.

in these experiments. This is consistent with our estimates of total protein concentration and the fraction of protein that, because it retains the 3' tail fragment, is expected to be inactive in binding exogenous DNA (see Materials and Methods).

Competition, protein titration and methylation protection experiments all indicate that the value of the K_D for *Oxytricha* macronuclear DNA is less than 0.05 times that obtained with GT32.

Sequence and Length Requirements for Binding

Sequence requirements for binding were analyzed using additional oligonucleotides, listed in Table 1. The only oligonucleotides bound by the protein are GT32, T $\rightarrow$ U tail and TGPermut (Fig. 6, Table 2). TGPermut forms only complex 3. Preliminary results (not shown) indicate that the complex with T $\rightarrow$ U tail, like that with GT32, is salt-stable.

As mentioned earlier, the size of the DNA seems to have no effect on the electrophoretic mobility of the DNA-protein complex in a nondenaturing gel. This property of the complexes gave us a convenient way of determining the minimum length of DNA required for binding.

Figure 6. Sequences required for binding. Various oligonucleotides with specific alterations of the G_4T_4 motif were tested by the gel shift assay for their performance as substrate for the telomere-binding protein.

Using base-specific chemical cleavage methods (23), a complete size distribution or "ladder" of DNA was generated from 5'-[^{32}P]-GT32. After treatment with T4 polynucleotide kinase to remove the 3'-phosphate groups, the ladder of DNA was allowed to form complexes, which were then isolated from a nondenaturing gel and their DNA content analyzed by electrophoresis in a denaturing gel (Fig. 7). Complex 1 is formed predominantly from integral multiples of the T_4G_4 repeat, the smallest DNA bound being about 16 nucleotides long. Complex 3 binds to DNA of length 25-28 nucleotides (i.e., DNAs that have T tracts at their 3' ends), in addition to binding the full-length oligonucleotide, as might be predicted from its ability to bind the TGPermut oligonucleotide.

DISCUSSION

The study of *Oxytricha* macronuclear telomeres is providing the first detailed information about nonviral telomeric structures. Previous work has demonstrated the presence of two polypeptides that are bound to the ends of macronuclear DNA in a salt-resistant complex (15). These polypeptides, which function as a heterodimer, can be added back to purified macronuclear DNA, reconstituting the methylation protection seen in whole cells (16, 17). One of the limitations of this system is that at present, there is no way of introducing exogenous sequences

Table 2. Formation of complexes between oligonucleotides and telomeric protein.

<u>Oligonucleotide</u>	<u>Complex 1</u>	<u>Complex 2</u>	<u>Complex 3</u>
GT32	Yes	Yes	Yes
T→U tail	Yes	Yes	Yes
TGPermut	No	No	Yes
GA32	No	No	No
T ₁ G ₂ toCA	No	No	No
G→C tail	No	No	No

into *Oxytricha*, which makes it impossible to alter telomeric DNA sequences and study the effects of these changes *in vivo*. Such information could be especially useful, now that other telomere-related activities, such as the telomere terminal transferase, are coming to light (24, 25). These activities are being studied mainly *in vitro*, with defined, synthetic DNAs as substrate. Understanding the interplay between two fundamental telomere functions, elongation by the terminal transferase and capping of the chromosomal end by the binding protein, would be facilitated by understanding how the binding protein interacts with synthetic DNA. In this paper, we have described our use of synthetic DNA oligonucleotides as substrates for reconstitution of the telomeric nucleoprotein complex *in vitro*. We find that the telomere protein does form complexes with single-stranded oligonucleotides that bear repeats of the G₄T₄ sequence motif.

We have used gel electrophoresis under nondenaturing conditions to separate complexes from free DNA. Under these conditions, three complexes are resolved, one barely entering the gel. Since both subunits of the protein are extremely basic (pI > 8.5; ref. 17), one might expect that the only reason the complexes move toward the anode at all is because of the bound DNA. We find that the mobility is unaffected by the size of the DNA. This is in keeping with previous observations (26, 27) that the electrophoretic mobility of a complex is determined mainly by the size and charge of the protein, although there are instances where the size of the DNA (28) or the conformation of bound DNA (29, 30) can also be a factor.

All three complexes are specific as judged by the following criteria: the requirement for the proper DNA sequence, resistance of the complex to competition by macronuclear DNA, and the detection of specific DNA-protein contacts by methylation protection (unpublished results) and interference (21) of individual complexes. When multiple complexes are observed in other systems, it is often the case that crude extracts are used as the source of protein, resulting in the non-specific binding of protein to the probe DNA (e.g., ref. 31). Such artefacts are not likely with the telomere protein, as the protein preparation that we use is essentially homogeneous, containing only the two telomere protein subunits (Fig. 1A) and the added micrococcal nuclease.

Figure 7. The minimum length required for binding. 5'-[^{32}P]-GT32 was treated with DMS or hydrazine and cleaved at the modified bases. The G and T sequence ladders so produced were pooled and treated with T4 polynucleotide kinase to remove the 3'-phosphates. This complete ladder was then used as substrate in a gel shift assay, and gel slices containing complexes 1 and 3 were recovered. The gel slices, after being soaked in sequencing gel loading buffer, were layered on a 10% polyacrylamide sequencing gel to allow the DNA fragments in the complexes to run into the sequencing gel. The lane marked Starting material is the ladder used as substrate, also loaded on the sequencing gel as a size marker. The sizes of the fragments, in nucleotides, are indicated. As a control to check for degradation of the DNA during the experiment, full-length GT32 was used in a duplicate experiment.

The micrococcal nuclease by itself has no detectable DNA binding activity under these conditions (32). Given the extremely tight association between the two subunits of the protein (17), it is also unlikely that the difference between the three complexes is a reflection of differences in the subunit composition of the complexes. The same DNA-protein contacts are seen with the oligonucleotide as with macronuclear DNA, further confirming accurate reconstitution with the oligonucleotide. Also, these interactions are resistant to 2 M NaCl, as observed both with telomeric complexes in macronuclei and with complexes reconstituted with macronuclear DNA (16).

One would not predict, *a priori*, that the telomere protein would recognize C₄A₄ repeats in single-stranded DNA, the (C₄A₄)_{2.5} in macronuclear telomeres presumably being base-paired to (G₄T₄)_{2.5} *in vivo*. It is conceivable that this property of the protein is related to telomere replication, when the protein might be required to recognize C₄A₄ primers. An alternative and perhaps trivial explanation is that the binding of C₄A₄ sequences is mediated not by direct contacts between the protein and the DNA, but by hybridization of the C₄A₄ strand to the G₄T₄ tail that is retained by the protein (17). That the C₄A₄ oligonucleotides form only one of the three complexes seen with the G₄T₄ oligonucleotides argues against this idea, but does not rule it out. Because of this concern, detailed studies of C₄A₄ binding have been postponed until protein that is free of DNA can be obtained.

The structure of the DNA upstream from the 3'-terminal (T₄G₄)₂ seems to have little effect on the efficiency of binding. A single-stranded oligonucleotide is bound as well as a synthetic DNA construct that mimics the structure of natural *Oxytricha* telomeric DNA (i.e., 20 base pairs of G₄T₄•C₄A₄ and 16 bases of 3' G₄T₄ repeats). This was contrary to expectation, previous work (15) having suggested that both the duplex and the single-stranded portions were required. A possible explanation for this discrepancy is that the quantitation in the previous work was done with the radioactive label in the CA-rich strand, and could therefore have been affected by the affinity of the protein for single-stranded C₄A₄ sequences. It has been noted (15, 32) that what appears to be a three-stranded structure forms when excess C₄A₄ strand is annealed to the G₄T₄ strand. In our hands, this structure, which would have a region of single-stranded C₄A₄ sequence, is bound better than the 1:1 GT₄₆•CA₃₀ structure, supporting the notion that binding to single-stranded CA-rich sequence could have affected the previous report.

In the experiments of Fig. 4, hybridization of GT₄₆ to CA₃₀ was incomplete; hence, unlabeled, single-stranded CA₃₀ was also present in the mixture of GT₄₆ and GT₄₆•CA₃₀. It is conceivable that the presence of increasing amounts of protein in the binding assay drove the hybridization of GT₄₆ to CA₃₀ closer to completion. There are thus at least two ways of explaining the observed depletion of single-stranded GT₄₆ upon addition of protein (Fig. 4A). First, the protein could be binding single-stranded GT₄₆ directly. Second, increasingly efficient hybridization of GT₄₆ to CA₃₀ due to added protein would also be manifested as a depletion of single-stranded GT₄₆ in the gel shift assay even if single-stranded GT₄₆ was never directly

bound by the protein. The nitrocellulose filter-binding studies (Fig. 4B) rule out the second explanation, since complex-formation is as efficient with GT46 alone, with no CA30 DNA present at all, as with GT46•CA30.

The protein has an apparent dissociation constant of about 19 nM for GT32. As mentioned before, we assume that the 3' tail fragment, which is retained by the protein even after extensive treatment with micrococcal nuclease, is not released during the course of the binding experiments. This binding constant may therefore be considered as providing a frame of reference; it may subject to revision if DNA-free protein becomes available.

Binding to macronuclear DNA is at least 20-fold tighter. As discussed above, we do not believe that base-pairing of the upstream $(G_4T_4)_{2.5}$ to $(C_4A_4)_{2.5}$ is the reason for the tighter binding to macronuclear DNA. The affinity could be affected by the length of the DNA substrate: GT64 (5'-[T₄G₄]₈-3') is about twice as good a substrate as GT32, and the affinity for GT46 is intermediate in value (data not shown). In addition, the source of the DNA could affect its suitability as substrate: chemically synthesized DNA has been reported to have modified bases (e.g., ref. 33), albeit at a very low frequency. Alternatively, the macronuclear DNA could have modifications that the synthetic DNA would lack.

The affinity of the protein for macronuclear DNA ($K_D < 1$ nM) is comparable to that of other DNA-binding proteins that have specific recognition sequences. For example, the *Escherichia coli* *trp* repressor•DNA complex and the *E. coli* *lac* repressor•DNA complex have K_D values of 5×10^{-10} M and 1×10^{-13} M respectively (27, 22). Single-stranded DNA-binding proteins that show no sequence specificity tend to have slightly lower affinities for their substrates: for instance, 5×10^{-8} to 5×10^{-10} M for bacteriophage T4 gene 32 protein (34), and 10^{-8} to 10^{-10} M for *E. coli* SSB (35).

The single-stranded DNA-binding proteins typically tend to show cooperative behavior, although this may be salt-dependent (36). When they bind cooperatively, contiguous sections of the DNA are occupied, more DNA being occupied with increasing amounts of added protein. If the telomere protein does bind in a cooperative manner, it is not in the same way as the gene 32 protein or SSB, in that it occupies only 16 nucleotides even when it is at saturating concentrations.

We have tested a small subset of the sequence changes that are possible in the binding site (Table 2). Where possible, these changes were designed to avoid sequences that could form Watson-Crick base pairs. The G₄ and T₄ sequences both seem to be important for binding. The G→C tail oligonucleotide, which has C₄T₄ repeats in the 3' tail region, is not bound at all even though it has G₄T₄ sequences internally, again demonstrating the specificity of the protein for G₄T₄ sequences at the 3' terminus. Changing the T's in the 3' tail to U's (T→U tail) does not prevent binding. The obvious model, that it is the 5-methyl groups that are responsible for the non-ionic component of the interaction, must therefore be revised. Since this oligonucleotide does bind less well than does GT46, however, the methyl groups must be making some contribution

to the binding interaction. Other candidates for non-ionic interactions include the stacking of aromatic amino acid side chains with the bases, and possible hydrogen bonding between the bases and the protein, as suggested by Price and Cech (17). The oligonucleotide T₁G₂toCA is not recognized by the protein. Perhaps the (CA)₄ sequence is an inhibitor of binding, for modification of these positions in GT32 does allow complex 3 to form (21). Alternatively, one could postulate that surfaces of the G and T bases not affected by the chemical modification are required for binding, in which case changing these G's and T's to alternating C's and A's could prevent binding.

The shortest oligonucleotide that will serve as substrate is about 16 nucleotides long, 5'-(T₄G₄)₂-3' (Fig. 7). This corresponds to the length of the 3' single-stranded tail in macronuclear DNA. Zahler and Prescott (25) have proposed a model for *Oxytricha* telomere replication: after the addition of multiple repeats of G₄T₄ sequence by the telomere terminal transferase, binding of the telomere protein and exonucleolytic trimming of the overbuilt telomere to the protein binding site could result in a 3' tail of defined length. In view of our observations that the protein has no detectable affinity for internal G₄T₄ sites and that the minimal binding site is 16 nucleotides long, a modified model for telomere replication can be presented: that there is competition between the terminal transferase to add G₄T₄ sequences, and the binding protein to bind at the end of the lengthening 3' tail and presumably prevent further elongation by the terminal transferase. The binding event would occur only after the addition of two T₄G₄ repeats, that being the minimum binding site. This model would also explain the observation that a small fraction of the macronuclear DNA population has more than the canonical 36 bases of G₄T₄ sequence (16): if the length of the telomere is regulated by a balance in the rates of addition by the transferase and binding by the binding protein, some fraction of the replicating molecules could have extra G₄T₄ repeats added on before the 3' end is occupied by the binding protein.

How does the protein discriminate between the 3' end and internal sites in the DNA? An obvious feature of the 3' end that is not found anywhere else in the DNA is the 3' hydroxyl group. However, the hydroxyl group *per se* is not required, demonstrated by the efficient binding of macronuclear DNA labeled at the 3' end with cordycepin (3'-dA). Preliminary results indicate that the presence of a phosphate group at the 3' end does strongly inhibit binding. Since the presence of the bulky cordycepin does not prevent binding, one can infer that steric hindrance is not the entire reason behind the inhibition by the 3' phosphate. A more plausible explanation is that the strong negative charge associated with the terminal phosphate, perhaps in combination with a steric effect, is responsible for the poor binding of 3'-phosphorylated DNA. The same mechanism might also allow the protein to ignore internal binding sites.

ACKNOWLEDGMENTS

We are grateful to Cheryl Grosshans for oligonucleotide synthesis, Gregg MacDonald for growing *Oxytricha* and *Chlorogonium*, and Carolyn Price for helpful discussions. This work was supported in part by NIH grant GM28039.

*Present address: Department of Chemistry, Northern Arizona University, Flagstaff, AZ 86011, USA

REFERENCES

1. Blackburn, E.H. and Szostak, J.W. (1984) *Annu. Rev. Biochem.* **53**, 163-194.
2. Szostak, J.W. and Blackburn, E.H. (1982) *Cell* **29**, 245-255.
3. Blackburn, E.H. and Gall, J.G. (1978) *J. Mol. Biol.* **120**, 33-53.
4. Yao, M.-C. and Yao, C.-H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7436-7439.
5. Klobutcher, L.A., Swanton, M.T., Donini, P. and Prescott, D.M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3015-3019.
6. Blackburn, E.H. and Challoner, P.B. (1984) *Cell* **36**, 447-457.
7. Emery, H.S. and Weiner, A.M. (1981) *Cell* **26**, 411-419.
8. Richards, E.J. and Ausubel, F.M. (1988) *Cell* **53**, 127-136.
9. Moyzis, R.K., Buckingham, J.M., Cram, L. S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.-R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6622-6626.
10. Blackburn, E.H. and Chiou, S.-S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2263-2267.
11. Budarf, M.L. and Blackburn, E.H. (1986) *J. Biol. Chem.* **261**, 363-369.
12. Cheung, M.K., Drival, D.T., Littau, V.C. and Johnson, E.M. (1981) *J. Cell Biol.* **91**, 309-314.
13. Berman, J., Tachibana, C.Y. and Tye, B.-K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3713-3717.
14. Gottschling, D.E. and Cech, T.R. (1984) *Cell* **38**, 501-510.
15. Gottschling, D.E. and Zakian, V.A. (1986) *Cell* **47**, 195-205.
16. Price, C.M. and Cech, T.R. (1987) *Genes and Dev.* **1**, 783-793.
17. Price, C.M. and Cech, T.R. (1989) *Biochemistry* **28**, 769-774.
18. Swanton, M.T., Greslin, A.F. and Prescott, D.M. (1980) *Chromosoma* **77**, 203-215.
19. Laemmli, U. K. (1970) *Nature* **227**, 680-685.
20. Schneider, R., Gander, I., Muller, U., Mertz, R. and Winnacker, E.L. (1986) *Nucleic Acids Res.* **14**, 1303-1317.
21. Raghuraman, M.K. and Cech, T.R., manuscript in preparation.
22. Riggs, A.D., Suzuki, H. and Bourgeois, S. (1970) *J. Mol. Biol.* **48**, 67-83.
23. Maxam, A.M. and Gilbert, W. (1980) In Grossman, L. and Moldave, K. (eds), *Methods in Enzymology*, Academic Press, New York, Vol. 65, pp. 499-560.
24. Greider, C.W. and Blackburn, E.H. (1985) *Cell* **43**, 405-413.
25. Zahler, A.M. and Prescott, D.M. (1988) *Nucleic Acids Res.* **16**, 6953-6972.
26. Bading, H. (1988) *Nucleic Acids Res.* **16**, 5241-5248.
27. Carey, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 975-979.
28. Gustafson, T.A., Miwa, T., Boxer, L.M. and Kedes, L. (1988) *Mol. Cell. Biol.* **8**, 4110-4119.
29. Straney, D.C. and Crothers, D.M. (1985) *Cell* **43**, 449-459.
30. Liu-Johnson, H.-L., Gartenberg, M.R. and Crothers, D.M. (1986) *Cell* **47**, 995-1005.
31. Aguilera, R.J., Akira, S., Okazaki, K. and Sakano, H. (1987) *Cell* **51**, 909-917.
32. Dunn, C.J. (1988) M.S. Thesis, University of Colorado, Boulder.
33. Eadie, J.S. and Davidson, D.S. (1987) *Nucleic Acids Res.* **15**, 8333-8349.
34. von Hippel, P.H., Kowalczykowski, S.C., Lonberg, N., Newport, J.W., Paul, L.S., Stormo, G.D. and Gold, L. (1982) *J. Mol. Biol.* **162**, 795-818.
35. Lohman, T.M. and Overman, L.B. (1985) *J. Biol. Chem.* **260**, 3594-3603.
36. Kowalczykowski, S.C., Lonberg, N., Newport, J.W. and von Hippel, P.H. (1981) *J. Mol. Biol.* **145**, 75-104.