Effect of monovalent cation-induced telomeric DNA structure on the binding of *Oxytricha* telomeric protein

M.K.Raghuraman^{1*} 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, USA

Received April 2, 1990; Revised and Accepted June 22, 1990

ABSTRACT

Oligonucleotides bearing 4 repeats of telomeric deoxyguanosine-rich sequence undergo a monovalent cation-induced transition to a folded conformation with G-G base pairs, modeled as a 'G-quartet' structure. We have now deduced the rates of folding and unfolding of d(TTTTGGGG)₄, which has four repeats of the *Oxytricha* telomeric DNA sequence. The estimated average values of ΔG for the folded form at 37°C are -2.2 kcal/mol and -4.7 kcal/mol in 50 mM Na⁺ and K⁺, respectively. The fully folded DNA is not recognized by the *Oxytricha* telomere-binding protein; the substrate for protein binding has properties consistent with its being partly or fully unfolded. In confirmation of this conclusion, prevention of DNA folding by methylation enables the protein to bind as rapidly in the presence of monovalent cations as in their absence. The slow unfolding ($t_{1/2}$ = 4 hr and 18 hr at 37°C in Na⁺ and K⁺, respectively) of the DNA suggests that such structures would be long-lived if they formed *in vivo*, unless they can be actively unfolded. The inability of the telomere-binding protein to bind the stable, folded form of the 4-repeat telomeric sequence is a problem that may be circumvented *in vivo* by avoiding four single-stranded repeats.

INTRODUCTION

Telomeres, the nucleoprotein structures found at the ends of linear chromosomes, serve a variety of functions. These include stabilization and replication of the ends, and the organization of chromosomes within the nucleus. The DNA component of all nuclear telomeres consists of tandem repeats of a short, deoxyguanosine (dG) rich sequence that fits the consensus d(T/A)₁₋₄dG₁₋₈ (1). The nucleotide sequence of the telomeric repeat is very similar in organisms as diverse as *Tetrahymena* (d(T₂G₄)), *Oxytricha* and *Stylonychia* (d(T₄G₄)), yeast (d(TG₁₋₃)), plants (d(T₃AG₃)) and humans (d(T₂AG₃)) (2-7). In all examples studied, the complementary dC-rich strand is recessed at the 5' end, leaving a 3' single-stranded extension consisting of two repeats of the dG-rich sequence, regardless of

whether the internal, double-stranded portion of the telomeric sequence is homogeneous or heterogeneous in length (3, 8).

Single-stranded telomeric G-rich repeats have special properties. Lipps *et al.* (9) and Oka and Thomas (10) demonstrated that the ends of deproteinized macronuclear DNA from *Oxytricha* aggregate in the presence of certain monovalent cations. Henderson *et al.* (11) found that synthetic oligonucleotides bearing telomeric G-rich repeats migrate aberrantly fast during electrophoresis in nondenaturing gels. Evidence from nuclear magnetic resonance studies indicated that these oligonucleotides could form G-G base pairs, and that some of the guanines were in the *syn* conformation.

The 'G-quartet' model (12, 13) accounts for the observations mentioned above. This model is related to earlier observations of parallel four-stranded helices of poly(G) (e.g., references 14-16), but differs in that adjacent stretches of the polynucleotide backbone are anti-parallel. In the model, the telomeric DNA folds into stacks of square planar quartets of deoxyguanosine residues, each dG forming Hoogsteen base pairs with its neighbors in the plane of the quartet. The structure could be formed either by a single molecule with four repeats of the G-rich repeat (12), or by the interaction of two molecules each bearing two repeats of the telomeric DNA sequence (12, 13). In the presence of monovalent cations, the N7 positions of the guanine bases are protected from methylation by dimethyl sulfate (DMS; 12, 13, 17). These results are consistent with the G-quartet model, but inconsistent with hairpin structures (11), in which only half of the dG's involved in G-G base pairs would be protected. The most direct evidence for the tertiary structure comes from the characterization of UV-induced cross-links connecting nucleotides separated by two G-rich repeats in the primary sequence (12). Assuming that the *Tetrahymena* and *Oxytricha* telomeric four-repeat molecules form related structures, the location of the cross-links is consistent with the G-quartet model and inconsistent with models proposed previously (11).

The remarkable conservation of telomeric DNA sequence across kingdoms means that most telomeric sequences can be accommodated in the G-quartet model. Indeed, oligonucleotides bearing telomeric G-rich sequences from diverse organisms can form compact structures *in vitro* (11). *Tetrahymena* and *Oxytricha*

* To whom correspondence should be addressed

+ Present address: Department of Genetics, SK-50, University of Washington, Seattle, WA 98195, USA

telomeric DNA support the maintenance of linear minichromosomes in yeast even though they differ from each other and from the yeast telomeric DNA sequence (18, 19), suggesting that telomeric DNAs may also have a common structure *in vivo*. A further implication might be that the structure of telomeric DNA, rather than the nucleotide sequence *per se*, is required for some telomere functions.

In *Oxytricha* and *Euplotes*, the ends of the macronuclear DNA are bound by protein *in vivo* (9, 20–23). The *Oxytricha* protein, a heterodimer, can be isolated and reconstituted with macronuclear DNA or telomeric oligonucleotides, giving complexes that resemble the native complex in their methylation protection pattern (21, 22, 24–26). Other telomere-related activities include a telomere terminal transferase (telomerase), which uses an internal RNA template to add telomeric repeats to the 3' end of telomeric DNA primers (27–34), and a primase-like activity that uses telomeric DNA as template to make RNA containing the sequence C₄A₄ (35).

All three activities (the telomere-binding protein, the telomerase and the primase) have been assayed *in vitro* with synthetic, G-rich DNA, often in the absence of the complementary strand. It is important to assess the relevance of G-DNA structure (36) for each of these activities. We now present evidence that the stable, fully folded telomeric DNA structure is not recognized by the *Oxytricha* telomeric DNA-binding protein. We have used as substrate the molecule d(T₄G₄)₄, which has the advantage that unlike its *Tetrahymena* counterpart, it can be readily switched between unfolded and folded conformations by adjusting the ionic conditions.

Although we think that the G-quartet model correctly describes the folded structure of telomeric DNA, the conclusions drawn in this paper are independent of the validity of the model. We therefore use 'G-DNA' as a neutral term to describe the stable folded form of the G-rich strand of telomeric DNA comprising four repeats of the telomeric sequence.

MATERIALS AND METHODS

Preparation of DNA

Oligonucleotide synthesis was as previously described (26). The DNA sequences were: Oxy-4, d(T₄G₄)₄; TS18, d(TAATACGACTCACTATAG); CG₁₀-AU, d(A₃TA₃TA₃TA₃TAT₄A₅TATA₃CCGCGC₄CTATAGTGAGTCGTATTA). DNA was labeled at the 5' end with γ -[³²P]-ATP and T4 polynucleotide kinase. The enzyme was then inactivated by heating to 90°C, and the DNA was de-salted by chromatography in Sephadex G-25 columns (Pharmacia) in water. Alternatively, the DNA was adsorbed on Sep-Pak columns (Waters), washed with water and eluted in 50% acetonitrile. De-salted DNA was lyophilized and resuspended in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA (TE).

Isolation of Telomere Protein

Growth of *Oxytricha nova* and isolation of macronuclei was as described previously (37, 24). Macronuclei (20–50 × 10⁷) were pelleted from lysis buffer by centrifugation at 400 × g and resuspended in 20–30 ml of 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.1 mM phenylmethanesulfonyl fluoride (PMSF), 0.1 mM tosylphenylalanine chloromethylketone (TPCK). Solid CsCl was added at 1 g CsCl/ml of nuclei suspension. The debris was removed by centrifugation at 10,000 × g for 15 min; the clear lysate was transferred to Quickseal (Beckman) tubes and

centrifuged in a vertical rotor exactly as though plasmid DNA were being prepared, except that the temperature was lowered to 15°C and the speed was maximized to reduce preparation time. Fractions of 0.5 ml were collected from the top of the CsCl gradients, and the fractions containing the DNA were identified by measuring the UV absorbance at 260 nm. The telomere protein remained bound to the DNA under these conditions, and was free of any other protein as judged by Coomassie Blue staining of sodium dodecyl sulfate (SDS)/polyacrylamide gels. Subsequent treatment of the protein was as previously described (24, 25): the protein was dialysed, concentrated and treated with micrococcal nuclease to remove the DNA. After inactivation of the nuclease with 15 mM ethylene glycol bis(β-aminoethyl ether) N,N,N',N'-tetraacetic acid (EGTA), a final step was added: the protein preparation was extensively dialysed against 5% glycerol in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.1 mM TPCK, 0.1 mM PMSF. The concentration of active protein was ~2 μM. Protein prepared by this method was indistinguishable in polypeptide content and DNA-binding activity from protein prepared by previously published methods.

Protein Binding Assay

The protein-DNA complex was formed as described in the Figure legends. Nitrocellulose filter-binding was done essentially as described previously (25): samples were diluted into 50 or 100 volumes of ice-cold TE and immediately applied to the filters, which were then washed three times with 1 ml of cold TE. Dilution with TE, which reduces the salt-induced, non-specific sticking of DNA to the filters, does not cause immediate dissociation of the complexes (24). Nor does it promote further binding of protein, as binding is slow at 0°C and at these dilutions of protein and DNA (data not shown).

Gel Mobility Shift Assay

Gel mobility shift assays were done as previously described (25). Where noted, 50 mM NaCl or KCl was included in the gel and in the buffer. Electrophoresis was at 4°C at 15 mA when salt was not included, and at 60–70 mA when salt was included. Gels were pre-run for 3 hr before the samples were loaded. Quantitation of radioactivity in bands was by scanning of dried gels in an Ambis radioanalytic imaging system.

Methylation/Interference of Folding

Isolation of methylated DNA that could or could not fold was as previously described (12).

RESULTS AND DISCUSSION

The Fully Folded G-DNA Structure is Not a Substrate for Telomere Protein

UV Cross-linking and Telomeric Protein Binding

Earlier work had shown that a specific intramolecular cross-link is formed upon irradiation of folded Oxy-4 DNA with ultraviolet light (12). The presence of bound telomere protein prevented the formation of this cross-link (Fig. 1A). Oxy-4 that had been incubated in TE buffer and Oxy-4 that had been released from telomeric complexes by treatment with SDS were both cross-linked (Fig. 1A). The increased cross-linking in the SDS-treated samples (compared to the samples in TE) was presumably because of the Na⁺ in the detergent.

Cross-linked and linear forms of Oxy-4 were recovered from a denaturing gel and assayed for protein binding. Oxy-4 that had

Figure 1. (A) Intramolecular UV cross-linking of DNA inhibited by bound protein. Telomeric complexes were formed with 5' [32 P]-Oxy-4 DNA and then irradiated for various lengths of time either directly or after the addition of 1% SDS. Oxy-4 that had been incubated in TE buffer was also irradiated as a control. Cross-linked (xlink) and non-cross-linked (Oxy-4) material were resolved by electrophoresis in a denaturing, 10% polyacrylamide gel, an autoradiogram of which is shown. (B) Binding of UV-irradiated DNA by telomere protein. Cross-linked (X) and linear forms of Oxy-4 (I, irradiated but not cross-linked; O, non-irradiated) were recovered from a gel such as the one in (A), incubated with telomere protein and then loaded on a 10% nondenaturing polyacrylamide gel for electrophoresis. The positions in the gel of the telomeric complexes 1 and 3 (26) and of the unfolded and folded or cross-linked forms of free DNA are indicated.

been irradiated but not cross-linked was bound by the telomere protein, while the cross-linked DNA, which co-migrated with the folded form of Oxy-4 in the nondenaturing gel, was not bound (Fig. 1B).

These observations suggest that the DNA in the telomeric DNA-protein complex is not in the folded form. Further, the methylation protection pattern of the telomeric DNA *in vivo* and after reconstitution into protein-DNA complexes *in vitro* does not encompass the pattern of protection seen for intra- or inter-molecular telomeric DNA structures (22, 26, 12, 13). However, none of these observations rules out the possibility that the protein binds to folded G-DNA and then perturbs it so as to change the methylation protection pattern and prevent cross-linking, or that the cross-link alters the folded structure, preventing the protein from binding. We therefore tested the effect of telomeric DNA structure on protein binding by comparing the kinetics of folding and unfolding of the DNA with the kinetics of complex formation.

Rate of G-DNA Unfolding

The G-quartet model (12, 13) predicts that the Watson-Crick base-pairing face of the G residues in the fully folded form of Oxy-4 should not be available for hybridization to an oligonucleotide of complementary sequence. If one were to add a C₄A₄-containing oligonucleotide to folded Oxy-4 (i.e., in Na⁺

or K⁺), any hybridization would be limited to Oxy-4 molecules that had partly or fully unfolded, or to Oxy-4 molecules that had been 'invaded' by the C₄A₄ strand. Duplex-formation occurring by invasion of the folded G-DNA would vary in rate with the concentration of the C₄A₄ oligonucleotide. If, on the other hand, the G-DNA structure had to unfold before it became available for hybridization, unfolding of the structure would be the rate-limiting step for duplex-formation (provided that the concentration of the C₄A₄ strand was high enough to trap any Oxy-4 molecule that became available for hybridization). The rate of duplex-formation would then be an apparent first order process, and would be a measure of the rate of unfolding of the G-DNA structure.

5' [32 P]-Oxy-4 was allowed to fold at 37°C in buffer containing 50 mM NaCl or KCl, and then incubated at the same temperature with a 20- or 200-fold excess of unlabeled oligonucleotide of the sequence d(C₄A₄)₈. Samples were removed at various times and the duplex and folded forms of Oxy-4 were resolved by electrophoresis at 4°C in non-denaturing gels containing 50 mM monovalent cation (Fig. 2B, C).

In both Na⁺ and K⁺, there was an initial 'burst' of hybridization comprising about 30% (in Na⁺) or 10% (in K⁺) of the Oxy-4 present, followed by a slow increase in the extent of hybridization (Fig. 2D). In TE buffer, Oxy-4 DNA was

Figure 2. Rate of G-DNA unfolding. (A) Schematic illustration of the design of the experiments in Figures 2 and 3. (B, C) 5'-[³²P]-Oxy-4 at 50 nM was boiled for 5 min in 10 mM Tris, pH 8.0, 1 mM EDTA, 5% glycerol and either 50 mM NaCl or 50 mM KCl. The DNA was then transferred to 37°C and incubated for 1 hr. An aliquot of the DNA was removed from each sample (= 0 min). The remainder of the sample was mixed with an equal volume of 1 μM or 10 μM unlabeled, pre-warmed d(C₄A₄)₈ in the same buffer, and incubation was continued at 37°C. At various times after the addition of d(C₄A₄)₈, aliquots were removed, chilled to 0°C and loaded on gels containing the same salt as the sample. Early time points had the longest electrophoresis times, resulting in the apparent change in electrophoretic mobility between early and late time points. Autoradiograms of typical gels run in 0.5× TBE + 50 mM NaCl (B) or 0.5× TBE + 50 mM KCl (C) are shown. The positions of the folded and hybridized forms of Oxy-4 are marked. The arrow in (C) indicates a band that showed no change in abundance over the time course (see text). The radioactivity in the duplex and folded forms of Oxy-4 was quantitated. The plot (D) shows the result with 0.5 μM d(C₄A₄)₈ as the 'trap' (○, NaCl; △, KCl). The lines represent theoretical curves fit to the data points. Because the data could not be adequately fit to a single first order kinetic component after the initial burst, we have chosen to portray the slow-hybridizing phase as having two first order kinetic components (see text). The average half-lives (see text) represent the average of all components after the initial burst.

completely converted to the duplex form within 5 min (data not shown), providing an end-point of 100% hybridization for this experiment. The observed half-life of the folded form in Na⁺ (representing the average of at least two first order components) was 3 hr at 20-fold excess of d(C₄A₄)₈ (Fig. 2D) and 5 hr at 200-fold excess (data not shown), giving a mean value of $t_{1/2} \approx 4$ hr. In K⁺, the half-lives were 11 hr and 25 hr at 20-fold and 200-fold excess of d(C₄A₄)₈, respectively, giving a mean value of $t_{1/2} \approx 18$ hr. Since the slow phase could not be fit with

a single exponential (see Fig. 2 legend), the apparent half-lives are not related to first order rate constants in any simple way.

Hybridizations done at 4°C resulted in little or no increase in the amount of duplex formed after the initial burst (data not shown). Therefore, although the samples took up to 5 min to enter the gel at 4°C, the extents of hybridization in Fig. 2 represent the Oxy-4 molecules that unfolded at 37°C. Only one duplex species was resolved (Fig. 2B, C), suggesting that the electrophoretic mobility of the duplex was unaffected by the

Figure 3. Rate of protein binding in the presence of monovalent cations. The experiment described in the Fig. 2 legend was modified: telomere protein (75 nM or 750 nM) was substituted for $d(C_4A_4)_8$; sheared, denatured calf thymus DNA (0.6 $\mu\text{g}/\mu\text{l}$) and BSA (100 $\mu\text{g}/\text{ml}$) were included in the incubation, and complex formation was assayed by binding to nitrocellulose filters. The solution conditions were as described in Fig. 2. Binding was also tested in TE buffer as a control. The time course of protein binding in Na^+ (○, ●), K^+ (△, ▲) and in TE (□) is shown. Open and filled symbols represent samples with 75 nM and 750 nM protein, respectively. The lines through the points are theoretical curves fit to the mean observed rates of binding. The broken lines are theoretical for the mean observed rates of hybridization to $d(C_4A_4)_8$ in Na^+ (— — —) and K^+ (- - -).

'register' of binding of the 32-nucleotide Oxy-4 to the 64-nucleotide $d(C_4A_4)_8$.

The slow hybridization was not due to a general effect of monovalent cations, as increasing the ionic strength generally favors hybridization (38–40), and hybridization is not differentially affected by different monovalent cations (41, 42). Accordingly, hybridization of the complementary, non-telomeric oligonucleotides TS18 and $\text{CG}_{10}\text{-AU}$ (see Materials and Methods) was extremely rapid under the conditions described in the Fig. 2 legend, going to 75% completion within 5 min. No difference in hybridization of these oligonucleotides was observed in buffers containing Na^+ compared to K^+ (data not shown). Nor is slow hybridization of telomeric oligonucleotides in monovalent cations a property of the primary sequence of the oligonucleotides, as methylated Oxy-4 that was incapable of forming a stable folded structure was rapidly hybridized in Na^+ and K^+ , whereas methylated Oxy-4 that could fold was slow to hybridize in Na^+ and K^+ , but hybridized rapidly in TE (data not shown).

Because there was no increase in the observed rate of hybridization with a 10-fold increase in the concentration of the complementary strand, we conclude that it was the availability of hybridizable Oxy-4 DNA that was rate-limiting. The simplest explanation for the observed rates of hybridization is that the fully folded G-DNA structure cannot be hybridized; any molecule that unfolds sufficiently becomes available for hybridization and, in the presence of excess complementary strand, is trapped in the duplex form before it can re-fold. There is no evidence for invasion of the G-DNA structure by the $d(C_4A_4)_8$ strand; whatever the mechanism of formation of a stable duplex, the slow step reflects an intrinsic property of the G-DNA structure. An alternative explanation, that the hybridization rate decelerates because it represents an approach to equilibrium, can be eliminated by the following argument: if there were a substantial

amount of unhybridized, folded G-DNA at equilibrium in the presence of 0.5 μM $d(C_4A_4)_8$, then the addition of 10-fold more $d(C_4A_4)_8$ would have significantly perturbed the equilibrium and resulted in an apparent increased unfolding rate.

Thus, the rate of accumulation of the duplex measures of the rate of unfolding of the G-DNA structure. Similarly, Davies and Rich (43) and Fresco and Massoulié (44) have previously proposed multi-stranded structures for polyinosinic acid and polyriboguanilic acid, based in part on their poor hybridization to polyribocytidylic acid in the presence of Na^+ and K^+ . Secondary structure of single-stranded DNA and RNA is also known to reduce accessibility to oligonucleotide hybridization probes (45–47).

In addition to differences in the size of the initial burst of hybridization and the subsequent rate of oligonucleotide binding, Na^+ and K^+ forms of Oxy-4 differed in the presence of a band that was seen only in gels containing K^+ (see arrow, Fig. 2B). This species had an electrophoretic mobility equal to or slightly greater than that of unfolded DNA. It was seen even in the absence of $d(C_4A_4)_8$, and its relative abundance (about 5% of the total Oxy-4) did not change over a 3–4 hr incubation with $d(C_4A_4)_8$. Thus, this species does not consist of unfolded Oxy-4 DNA. It may be an alternative conformation or a stable intermediate peculiar to K^+ (48), or a dimer of Oxy-4.

Telomeric Complex Formation Compared to the Rate of G-DNA Unfolding

We reasoned that if the telomere protein could bind to the folded form of Oxy-4, the rate of binding would be limited by the protein concentration, and at some protein concentration, it would exceed the rate of hybridization to $d(C_4A_4)_8$. If the G-DNA structure were not a substrate for the protein, binding would be limited by the rate of unfolding of the DNA, and would therefore be an apparent first order process, as described for hybridization to $d(C_4A_4)_8$.

The experiment described above was modified, the *Oxytricha* telomere protein being substituted for $d(C_4A_4)_8$ and nitrocellulose filter-binding being used to monitor the formation of the protein-DNA complex. The rate of protein binding was greatest in TE buffer and slowest in K^+ , showing a striking similarity to the kinetics of G-DNA unfolding (Fig. 3). As with hybridization to $d(C_4A_4)_8$, there was an initial burst in the binding of Oxy-4 by the telomere protein in both Na^+ and K^+ , followed by a slow increase in the amount of complex formed (Fig. 3). The burst of protein binding was smaller than that of hybridization. In both Na^+ and K^+ , a 10-fold increase in the protein concentration had a slight (< 2-fold) effect on the size of the initial burst, but did not affect the rate of subsequent binding. Addition of more Oxy-4 (in Na^+ or K^+) at the end of the 3 hr of incubation resulted in another burst of binding, indicating that the protein was still active and that the DNA substrate was limiting (data not shown).

Hence, the fully folded form of Oxy-4 is not a substrate for the protein, nor does the protein cause the unfolding of the G-DNA structure. If the alternative model (that the folded structure is the substrate for the protein) were correct, binding would have been fastest in K^+ and slowest in TE buffer. It is not clear whether the protein and the $d(C_4A_4)_8$ oligonucleotide recognize the same form of Oxy-4. Since the rate of protein binding never exceeds the rate of hybridization, it appears that the protein binds no earlier in the unfolding pathway of Oxy-4 than does $d(C_4A_4)_8$, assuming that there is a single pathway.

Figure 4. Rate of G-DNA folding: hybridization compared to protein binding. (A) Schematic representation of the experimental design. (B) 5'-[³²P]-Oxy-4 DNA was boiled for 5 min in TE buffer containing 5% glycerol, and then incubated for 1 hr at 37°C. NaCl or KCl was added to a final concentration of 50 mM. The concentration of Oxy-4 DNA was 50 nM after the addition of salt. At various times after the addition of salt, aliquots were mixed with an equal volume of 30 μ M d(C₄A₄)₈ in the same buffer. Hybridization was allowed to proceed for 20 seconds at 37°C, after which the samples were chilled and loaded on nondenaturing gels that contained the same salt. The 0 min time points come from aliquots of Oxy-4 that were removed and hybridized just prior to the addition of the monovalent cation. The fraction of Oxy-4 that hybridized was taken to be a measure of the fraction that had not yet folded. Semi-logarithmic plot shows the decrease in the extent of hybridization (relative to the 0 min time point) as a function of time after addition of NaCl ($\circ$, $\bullet$) or KCl (Δ , $\blacktriangle$). The data have been corrected by subtraction of the values obtained at 27 hr ($\sim$ 30% in Na⁺ and 12% in K⁺), to show only the population that could fold. Open and filled symbols represent the data from two independent experiments. The lines represent theoretical fits to the data points, with two components of different half-lives. (C) Telomere protein (0.75 μ M after mixing with the DNA) was substituted for d(C₄A₄)₈, and complex formation in Na⁺ ($\circ$) and K⁺ (Δ) was assayed by nitrocellulose filter binding. The end point (at 2 days) was $<1\%$ in both monovalent cations. The lines are theoretical curves fit to the data. The broken lines represent folding of Oxy-4 measured by hybridization in Na⁺ (---) and K⁺ (---), and are taken from (B).

Rate of G-DNA Folding

5'-[³²P]-Oxy-4 was equilibrated to 37°C for 1 hr in TE, and folding was initiated by the addition of NaCl or KCl. At various times after the addition of salt, samples were removed and mixed with a 600-fold excess of d(C₄A₄)₈. After 20 seconds of hybridization, the samples were chilled to 0°C and loaded on non-denaturing gels that contained the appropriate monovalent cation. Figure 4B shows the relative amount of Oxy-4 that could be trapped by d(C₄A₄)₈ as a function of the time of incubation of Oxy-4 in Na⁺ or K⁺. The curve has at least two phases in a semi-logarithmic plot. The $t_{1/2}$ of the fast population is about 6.5 min in Na⁺ and ≤ 0.5 min in K⁺, giving rate constants of 0.0017 s⁻¹ and ≥ 0.023 s⁻¹ respectively.

Telomeric Complex Formation as a Function of DNA Folding

5'-[³²P]-Oxy-4 was allowed to fold as described above; at various times after the addition of monovalent cation, samples were removed and incubated for 30 seconds with telomere protein (the '10 $\times$ ' concentration in Fig. 3 was used). Complex formation was assayed by nitrocellulose filter-binding.

Addition of monovalent cation to Oxy-4 DNA resulted in its becoming resistant to protein binding (Fig. 4C). The binding was best at early times after initiation of folding and became progressively poorer. As with hybridization, the decrease in the fraction of DNA being trapped showed at least two phases. In both Na⁺ and K⁺, the protein was apparently poorer at trapping the Oxy-4 DNA than was d(C₄A₄)₈.

These data provide additional evidence that the fully folded G-DNA structure is not a substrate for protein binding. It is certainly not an obligate substrate, because binding would then have been poorest at early times and best at late times after addition of monovalent cation.

The DNA folding experiments provide an internal control for the effects of the monovalent cation on hybridization and protein binding. In both hybridization and protein binding, the difference between the early and late time points was the time spent by the Oxy-4 DNA in the presence of monovalent cation; samples at early and late times were hybridized or bound to protein in identical ionic conditions and for the same amount of time. Differences between early and late time points must therefore be due to the effect of the monovalent cation on the Oxy-4 DNA itself, and not on rates of hybridization or protein binding.

The Oxy-4 DNA becomes resistant to protein binding at rates slightly greater than the rates of DNA folding as measured by hybridization. The difference could be due to a number of factors, including limitations of the experimental method. Both assays (hybridization and protein binding) involved incubation with Oxy-4 DNA for 20–30 seconds at 37°C. If the trapping of available Oxy-4 DNA is faster with d(C₄A₄)₈ than with protein, the Oxy-4 would have up to an additional 20–30 seconds to fold in the protein binding assay compared to the hybridization assay, giving apparent decay rates that are greater with the protein binding assay. Alternatively or in addition, the d(C₄A₄)₈ DNA may recognize a larger subset of the stages along the folding pathway than does the protein. For example, if the d(C₄A₄)₈ could hybridize to the unfolded form of Oxy-4 as well as to some intermediate in the folding pathway while the protein could only bind to the unfolded form, there would be a difference in the rate of decrease of substrate for hybridization compared to complex formation. This difference should hold true for both directions in the folding pathway, i.e., the apparent rate of *unfolding* as measured by hybridization should be greater than the rate of protein binding. Such a difference, although slight, is observed (Fig. 3).

Figure 5. Telomeric protein binding with DNA methylated to prevent folding. (A) Time course of complex formation with methylated Oxy-4, based on a nitrocellulose filter-binding assay. Methylated, 5'-[³²P]-Oxy-4 that had folded in Na⁺ (filled symbols) or had not folded (open symbols) was recovered from a nondenaturing gel and assayed for protein binding in TE buffer (■, □) or in TE containing 50 mM NaCl (●, ○) or KCl (▲, △). (B) Time course showing the decrease of free DNA upon protein binding, measured by the gel mobility shift assay. Methylated, 5'-[³²P]-Oxy-4 was recovered from Na⁺ or K⁺ gels and assayed for binding in the same salt as was used in its isolation. Unmodified 5'-[³²P]-Oxy-4 was also tested in Na⁺ and K⁺. □ and ■, binding of unmodified DNA in Na⁺ and in K⁺, respectively; ● and ○, binding in Na⁺ of methylated Oxy-4 that had and had not folded, respectively; ▲ and △, binding in K⁺ of methylated Oxy-4 that had and had not folded, respectively. (C) Time course showing the decrease of free methylated and unmodified DNA upon protein binding in TE. Unmodified Oxy-4 DNA (□) and folded (filled symbols) and non-folded (open symbols) forms of methylated Oxy-4 that had been recovered from Na⁺ (●, ○) and K⁺ (▲, △) gels were incubated with telomere protein in TE buffer. The free and bound forms of the DNA were resolved by electrophoresis in nondenaturing polyacrylamide gels. The lines through the points in (A), (B) and (C) are intended to aid in visualization of the data, and do not represent theoretical curves. To retain clarity, only one line is drawn when sets of data points overlap.

Protein Binding with DNA Methylated to Prevent Folding

Another way to control for the effect of the monovalent cation on the protein or the protein-DNA interaction is to methylate the DNA and isolate a sub-population that is incapable of forming a stable folded structure. Any effect of monovalent cations on the protein binding of this sub-population would be solely due to effects on the protein or protein-DNA interaction.

5'-[³²P]-Oxy-4 was treated with DMS at 65°C in TE, where the DNA is unstructured as judged by the uniformity of the methylation pattern, then incubated at 0°C in Na⁺ or K⁺ to allow folding. Unfolded and fully folded forms were isolated from nondenaturing gels. The same extent of methylation resulted in

much greater interference of folding in Na⁺ than in K⁺ (data not shown).

The methylated DNA was tested for protein binding in two ways. First, unfolded and folded forms of methylated Oxy-4 that had been recovered from a Na⁺ gel were separately incubated with telomere protein in TE, TE + NaCl, or TE + KCl. Complex formation (as a function of time of incubation with protein) was assayed by binding to nitrocellulose filters (Fig. 5A). Second, unfolded and folded forms of methylated Oxy-4 were recovered from Na⁺ and K⁺ gels, and were used as substrates for the protein in gel mobility shift assays. Protein binding of each DNA species was tested in the monovalent cation used for its isolation (Fig. 5B), as well as in TE buffer (Fig. 5C).

Unmodified DNA was also tested, as a control for the effect of methylation on protein binding (Fig. 5B, C).

The results were comparable: methylated DNA that was incapable of folding was rapidly bound in all conditions, while DNA that could form the G-DNA structure was bound slowly in Na^+ and K^+ , but bound rapidly in TE buffer. Note that the 'unfolded' DNA used in the filter-binding assay was recovered from a Na^+ -containing gel. In the presence of K^+ , some of this DNA could re-fold, resulting in apparent slower binding of the unfolded DNA in K^+ compared to Na^+ . In the gel mobility shift assay, where the unfolded and folded forms of free Oxy-4 were resolved, the depletion of the unfolded DNA in K^+ and Na^+ coincided exactly.

These experiments confirm that the poor binding by the protein in Na^+ and K^+ is due to the structure of the DNA substrate and not due to effects of the ionic strength on the protein or the protein-DNA interaction. The enhanced binding of the non-folding, methylated DNA in Na^+ and K^+ cannot be ascribed to interaction of the protein with the methyl group itself, because in TE buffer, unmodified Oxy-4 is bound as well as is methylated Oxy-4. The methylated, non-folding DNA is a slightly *poorer* substrate in TE, presumably because it is the most heavily modified (Fig. 5C). Nor is the slow binding of folded, methylated Oxy-4 a result of the chemical modification, as this DNA is as good a substrate as unmodified DNA in TE buffer. Also, the pattern of methylation/interference of protein binding is distinct from that of G-DNA folding (26, 12).

The increased rate of binding of methylated, non-folded Oxy-4 DNA in Na^+ (Fig. 5A) suggests that, were it not for its effect on DNA structure, the cation would facilitate rather than inhibit the protein-DNA interaction. This is likely to be a general effect of the ionic strength, rather than a specific Na^+ effect, because Cs^+ and Li^+ also increase the rate of complex formation with methylated DNA that cannot fold (data not shown).

Binding of telomeric DNA by the telomere protein results in the formation of two kinds of complexes *in vitro* (25, 26). The

formation of the different complexes was not affected by the presence of Na^+ or K^+ , as the relative abundance of the two complexes was the same in TE, Na^+ and K^+ (data not shown).

More than One Folded Form of Oxy-4

We interpret the data in Fig. 2–4 as reflecting two populations of Oxy-4 DNA molecules. One population, giving rise to the burst (Fig. 2), behaves as though it were unfolded, or folded into some non-G-quartet structure that is rapidly hybridized by the complementary strand. In the DNA folding experiments, this population is represented by the fraction (about 30% in Na^+ and 12% in K^+) that was in a form that could hybridize even after 17 and 27 hours of incubation in monovalent cation (see legend to Fig. 4). The other population gives rise to the slow phase (Fig. 2) and represents Oxy-4 DNA molecules folded into forms that are hybridized only upon unfolding partly or fully. The unfolding kinetics are best explained by there being at least two conformations in this population. Thus, we conclude that Oxy-4 DNA consists of multiple conformations of the G-quartet form that are slow-hybridizing, and at least one alternative, folded structure that is fast-hybridizing. This is consistent with the observation of two electrophoretically distinct species of Oxy-4 in Na^+ gels (12).

At the end-point of the incubation in the DNA folding experiments (Fig. 4), there was less Oxy-4 DNA available for protein binding than for hybridization. After two days at 37°C , only $\cong 1\%$ of the Oxy-4 DNA could be bound by protein (compared to 10–30% being available for hybridization). There may be alternative, fairly stable conformations that can be invaded by the $\text{d}(\text{C}_4\text{A}_4)_8$; such conformations may not necessarily be substrates for the protein.

Candidates for alternative structures include variations on the G-quartet theme (e.g., molecules that fold 'out of register', leaving smaller stacks of quartets and larger loops connecting the corners of the stacks), and duplex, hairpin structures or intermediates analogous to those previously proposed for G-DNA

Figure 6. A model depicting the events in the replication of macronuclear DNA ends (25, 27–30, 35). Only one end of the DNA is shown. Of all the structures shown in this model, only the deoxyribonucleoprotein complex in the upper left corner is known to occur *in vivo* (22). While telomerase activity (34) and replication occur *in vivo*, the *in vivo* substrates and products of these reactions are not proven.

(11, 48). Because the Watson-Crick base-pairing faces of some of the dG's in hairpin structures would still be accessible, one might expect such structures to hybridize more readily than the fully folded G-quartet form.

The Equilibrium Constant of the G-DNA Structure

A more detailed analysis of the kinetics of G-DNA folding and unfolding leads to the conclusion that the fast-folding population corresponds to the slow-unfolding one (49). The conclusion is based on the correlation between the sizes of the fast-folding and slow-unfolding populations, and on the argument that Oxy-4 molecules that remain unhybridized in the DNA folding experiments (Fig. 4) could not represent Oxy-4 molecules that give rise to the initial burst of hybridization (Fig. 2D) because the molecules in the burst are rapidly hybridized even at 4°C with just a 20-fold excess of d(C₄A₄)₈. Thus, the estimated average equilibrium constants (K_{eq}) of the fully folded conformations are $K_{eq}(\text{Na}^+) = 0.0017 \text{ s}^{-1}/4.8 \times 10^{-5} \text{ s}^{-1} \approx 35$ and $K_{eq}(\text{K}^+) = 0.023 \text{ s}^{-1}/1.1 \times 10^{-5} \text{ s}^{-1} \approx 2100$, giving values of $\Delta G = -2.2 \text{ kcal/mol}$ in 50 mM Na⁺ and $\Delta G = -4.7 \text{ kcal/mol}$ in 50 mM K⁺ at 37°C. Since the unfolding rate is the average of at least two kinetic components, and the folding rates are rough estimates (especially in K⁺), these values should be considered to be provisional estimates, subject to refinement. Because the faster kinetic component of the slow-hybridizing phase (see Fig. 2D legend) is small compared to the slowest kinetic component, the average half-life of the slow phase is roughly the same as the half-life of the slowest component. The values of K_{eq} obtained using the $t_{1/2}$ of the slowest component are therefore similar to the values given above.

The difference in the values of the equilibrium constant predicts that the Na⁺ form should be more readily destabilized. This is, in fact, what is observed: a given extent of methylation interferes with structure-formation much more strongly in Na⁺ than in K⁺. Also, the T_m of the related Tet-4 molecule, d(T₂G₄)₄, is reported to be about 40°C higher in K⁺ than in Na⁺ (I. Tinoco, Jr., and C. C. Hardin, personal communication). The lack of methylation interference in K⁺ does not argue against the G-quartet model; it just means that the low level of methylation used is not sufficient to prevent formation of the structure. The G residues are still *protected* from methylation in the presence of K⁺ (12, 13).

DNA Structure and K_D of the Protein-DNA Complex

The previously determined value of the dissociation constant (K_D) of the protein for single-stranded telomeric DNA needs to be revised. It seems likely that the 20-fold difference in the values of the K_D for macronuclear DNA ($K_D \leq 1 \text{ nM}$) and Oxy-4 DNA ($K_D = 19 \text{ nM}$) reported by Raghuraman *et al.* (25) is at least in part due to the effect of monovalent cations on the structure of Oxy-4, as no particular care was taken to desalt the DNA and protein preparations in those experiments. The presence of the C₄A₄ strand would presumably prevent the formation of the intramolecular G-quartet structure and increase the apparent extent of protein binding. Thus, both macronuclear DNA (25) and a synthetic oligonucleotide that was annealed to C₄A₄ DNA to give a 3' G-rich extension (21) would give the appearance of being better substrates for the protein than is the G-strand DNA by itself. In fact, when care is taken to reduce the concentration of salt in the samples, the K_D for the oligonucleotide-protein complex decreases about 10-fold (data not shown). It should be noted that DNA structure could similarly affect the outcome of

any experiment that required accessibility of telomeric DNA to binding or hybridization.

Implications for Telomere Function

A simple model for macronuclear DNA replication (Fig. 6) is useful in assessing the relevance of G-DNA structure to biological function. In this model, replication of the DNA end begins with the dissociation of the telomere protein. This is presumably required whether or not replication is initiated at the ends, and could be effected by a relatively simple regulatory event (such as phosphorylation). After replication and primer removal, one end of each daughter molecule retains the original 3' tail, and can be directly bound by the telomere protein. The other end is presumably blunt-ended and might be directly extended by telomere terminal transferase to give a 3' extension. Alternatively, it might undergo another round of replication to leave a 3' protruding end, which would then be extended by the terminal transferase. Regulation of the length of the newly synthesized telomere could either be effected by binding of the telomere protein (25) or by nucleolytic processing of over-elongated telomeres to give telomeres of the correct length (30). These ends are then bound by the telomere protein.

In such a model, G-DNA structure is not *required* for any of the steps in macronuclear DNA replication. Indeed, it is far from certain that this structure normally forms *in vivo*. In order for non-replicating *Oxytricha* macronuclear DNA to form the intramolecular G-quartet structure, twelve base pairs would need to be melted, assuming that the d(C₄A₄) repeats are completely base-paired and that there are only two d(T₄G₄) repeats in the 3' extension. Other circumstances where the G-quartet structure might form include over-extension by telomerase, and inter-molecular cohesion in the absence of bound protein. Given the measured half-life of the folded structure *in vitro*, it could potentially persist for more than one cell generation if it formed *in vivo*. While such structures might function adequately as telomeres insofar as providing stabilization of the ends, the cell might require their removal to ensure that other telomeric functions are not impaired.

Note that the only structure shown in Fig. 6 that is known to occur *in vivo* is the telomeric nucleoprotein complex. Hence, although G-DNA structure prevents binding of the *Oxytricha* telomeric protein (this work) and inhibits elongation by telomerase *in vitro* (50), it would be premature to make firm conclusions about the relevance of G-DNA structure to telomere function *in vivo*. The conservation of telomeric G-rich sequences in evolution remains the strongest argument that such a function exists.

ACKNOWLEDGEMENTS

We would like to thank Cheryl Grosshans for oligonucleotide synthesis, Gregg MacDonald and Wanda Leong for growing *Oxytricha* and *Chlorogonium*, and Jamie Williamson and Dan Herschlag for stimulating discussions. This work was supported by NIH grant GM28039. T.R.C. is an American Cancer Society Professor and Investigator, Howard Hughes Medical Institute.

REFERENCES

1. Blackburn, E. H., and Szostak, J. W. (1984) *Annu. Rev. Biochem.* **53**, 163–194.
2. Blackburn, E. H., and Gall, J. G. (1978) *J. Mol. Biol.* **120**, 33–53.
3. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3015–3019.

4. Oka, Y., Shiota, S., Nakai, S., Nishida, Y., and Okubo, S. (1980) *Gene* 10, 301–306.
5. Shampay, J., Szostak, J. W., and Blackburn, E. H. (1984) *Nature* 310, 154–157.
6. Richards, E. J., and Ausubel, F. M. (1988) *Cell* 53, 127–136.
7. 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. U.S.A.* 85, 6622–6626.
8. Henderson, E. R., and Blackburn, E. H. (1989) *Mol. Cell. Biol.* 9, 345–348.
9. Lipps, H. J., Grussem, W., and Prescott, D. M. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 2495–2499.
10. Oka, Y., and Thomas, C. A., Jr. (1987) *Nucleic Acids Res.* 15, 8877–8898.
11. Henderson, E., Hardin, C. C., Wolk, S. K., Tinoco, I. Jr., and Blackburn, E. H. (1987) *Cell* 51, 899–908.
12. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) *Cell* 59, 871–880.
13. Sundquist, W. I., and Klug, A. (1989) *Nature* 342, 825–829.
14. Gellert, M., Lipsett, M. N., and Davies, D. R. (1962) *Proc. Natl. Acad. Sci. U.S.A.* 48, 2013–2018.
15. Zimmerman, S. B., Cohen, G. H., and Davies, D. R. (1975) *J. Mol. Biol.* 92, 181–192.
16. Sen, D., and Gilbert, W. (1988) *Nature* 334, 364–366.
17. Henderson, E. R., Moore, M., and Malcolm, B. A. (1990) *Biochemistry* 29, 732–737.
18. Szostak, J. W., and Blackburn, E. H. (1982) *Cell* 29, 245–255.
19. Pluta, A. F., Dani, G. M., Spear, B. M., and Zakian, V. A. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 1475–1479.
20. Gottschling, D. E., and Cech, T. R. (1984) *Cell* 38, 501–510.
21. Gottschling, D. E., and Zakian, V. A. (1986) *Cell* 47, 195–205.
22. Price, C. M., and Cech, T. R. (1987) *Genes and Dev.* 1, 783–793.
23. Price, C. M. (1990) *Mol. Cell Biol.* 10, 3421–3431.
24. Price, C. M., and Cech, T. R. (1989) *Biochemistry* 28, 769–774.
25. Raghuraman, M. K., Dunn, C. J., Hicke, B. J., and Cech, T. R. (1989) *Nucleic Acids Res.* 17, 4235–4253.
26. Raghuraman, M. K., and Cech, T. R. (1989) *Cell* 59, 719–728.
27. Greider, C. W., and Blackburn, E. H. (1985) *Cell* 43, 405–413.
28. Greider, C. W., and Blackburn, E. H. (1987) *Cell* 51, 887–898.
29. Greider, C. W., and Blackburn, E. H. (1989) *Nature* 337, 331–337.
30. Zahler, A. M., and Prescott, D. M. (1988) *Nucleic Acids Res.* 16, 6953–6972.
31. Shippen-Lentz, D., and Blackburn, E. H. (1989) *Mol. Cell. Biol.* 9, 2761–2764.
32. Shippen-Lentz, D., and Blackburn, E. H. (1990) *Science* 247, 546–552.
33. Morin, G. B. (1989) *Cell* 59, 521–529.
34. Yu, G.-L., Bradley, J. D., Attardi, L. D., and Blackburn, E. H. (1990) *Nature* 344, 126–132.
35. Zahler, A. M., and Prescott, D. M. (1989) *Nucleic Acids Res.* 17, 6299–6317.
36. Cech, T. R. (1988) *Nature* 332, 777–778.
37. Swanton, M. T., Greslin, A. F., and Prescott, D. M. (1980) *Chromosoma* 77, 203–215.
38. Wetmur, J. G., and Davidson, N. (1968) *J. Mol. Biol.* 31, 349–370.
39. Britten, R. J., Graham, D. E., and Neufeld, B. R. (1974) *Methods Enz.* 29E, 363–406.
40. Van Ness, J., and Hahn, W. E. (1982) *Nucleic Acids Res.* 10, 8061–8077.
41. Marmur, J., and Lane, D. (1960) *Proc. Natl. Acad. Sci. U.S.A.* 46, 453–461.
42. Gruenwedel, D. W., Hsu, C., and Lu, D. S. (1971) *Biopolymers* 10, 47–68.
43. Davies, D. R., and Rich, A. (1958) *J. Am. Chem. Soc.* 80, 1003–1004.
44. Fresco, J. R., and Massoulié, J. (1963) *J. Am. Chem. Soc.* 85, 1352–1353.
45. Wu, R. (1972) *Nature New Biol.* 236, 198–200.
46. Agarwal, K. L., Brunstedt, J., and Noyes, B. E. (1981) *J. Biol. Chem.* 256, 1023–1028.
47. Gamper, H. B., Cimino, G. D., and Hearst, J. E. (1987) *J. Mol. Biol.* 197, 349–362.
48. Sen, D., and Gilbert, W. (1990) *Nature* 344, 410–414.
49. Raghuraman, M. K. (1990) Ph. D. thesis, University of Colorado, Boulder.
50. Zahler, A. M., Williamson, J. R., Cech, T. R., and Prescott, D. M., manuscript in preparation.