Parameters That Influence the Extent of Site Occupancy by a Candidate Telomere End-binding Protein*

Antonio Gualberto‡§¶, Jason Lowry‡∥, Irma M. Santoro‡**, and Kenneth Walsh‡∥‡‡

From the ‡Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106, the ∥Division of Cardiovascular Research, St. Elizabeth's Medical Center, Tufts University School of Medicine, Boston, Massachusetts 02135, and the §Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599

The MF3 protein specifically recognizes telomeric and non-telomeric DNA probes that can form G·G base-paired structures (Gualberto, A., Patrick, R. M., and Walsh, K. (1992) Genes & Dev. 6, 815–824). Here we further characterize the nucleic acid recognition properties of MF3 and present a mathematical analysis that evaluates the potential extent of telomere site occupancy by this factor. The substitution of dI at dG positions in telomeric DNA probes revealed that a single dG at any position within the internal repeat was sufficient for high affinity binding to MF3. The RNA analogs of high affinity DNA sites were not bound specifically by MF3, but the substitution of dU for dT in a DNA probe had little or no effect on binding. These data demonstrate that ribose ring structure is a critical feature of nucleoprotein complex formation, and this ribose specificity may enable MF3 to occupy sites of unusual DNA structure while minimizing interactions with cellular RNAs. Collectively, the nucleic acid binding properties of MF3 suggest that it may occupy a significant fraction of sites at telomere ends or other G-rich regions of altered DNA structure in vivo.

Telomeres consist of long double-stranded stretches of G-rich repeat sequences including $5'$-T_2AG_3-$3'$, which is found in vertebrates, and $5'$-T_4G_4-$3'$, which is found in ciliates. In some organisms, 12–16-nucleotide single-stranded extensions of the G-rich strand have been documented at the telomere end (1, 2). Single-stranded oligonucleotide probes of telomeric repeats can form hairpin duplex, triplex, and quadruple helical structures through G·G base pairing, and specific protein complexes with these structures have been reported (3–9). These unusual nucleoprotein complexes may regulate telomere length or have a capping function that prevents chromosome degradation or recombination. Some ciliates have $>10^7$ telomere ends, and the proteins that bind to these ends can be selectively released from 2 M salt-extracted macronuclear DNA by nuclease treatment (10). Molecular clones for telomere end-binding proteins

have been isolated from Oxytricha nova and Stylonychia mytilis (11–13). However, the identification of telomere end-binding proteins in higher eucaryotes is a more formidable task due to the relatively low abundance of these ends. Mammalian, avian, and Xenopus cell extracts contain single-stranded DNA-binding activities that recognize telomeric oligonucleotide probes (3, 14–17). The detection of these binding activities in cell extracts from diverse species indicates that the telomeric nucleoprotein structures may be highly conserved throughout evolution. However, it is difficult to know with certainty whether the telomeric probe-binding activities detected in higher eucaryotes represent true telomeric proteins or whether they have entirely different functions within the cell. Recently, Ishikawa et al. (17) and McKay and Cooke (14) reported that single-stranded telomeric probe-binding activities detected in HeLa cell and mouse liver extracts represent abundant heterogeneous nuclear ribonucleoproteins that may function in pre-mRNA splicing and in chromosome capping. In view of these findings, it may be desirable to elucidate the nucleic acid recognition properties of a putative telomere-binding protein to determine whether these properties are consistent with the hypothesis that the factor can occupy sites of altered DNA structure under physiological conditions. Based upon these measurements, an initial assessment can be made about whether a telomeric probe-binding activity can potentially occupy sites at chromosome ends in vivo.

Previously, we identified a potent single-stranded DNA-binding activity, referred to as MF3, that specifically recognized telomeric and non-telomeric sequences (18). High affinity binding occurred with probes containing two or more single-stranded stretches of guanine residues that were capable of G·G base pairing (3). Here we report that a single guanine N-2 group is essential for factor binding, and this critical amino group can occur at any of the three guanine positions within the internal repeat. This nucleic acid specificity indicates that MF3 could bind to G-rich regions of altered DNA structure, such as telomere ends, or to abundant cellular RNAs, which form a multitude of structures containing non-canonical base pairs. To test the latter possibility, the RNA analogs of high affinity DNA-binding sites were tested for their ability to bind to MF3. The RNA probes did not form specific nucleoprotein complexes with MF3. The lower affinity for RNA appears to result from the presence of the 2'-hydroxyl on the ribose ring and not from differences in pyrimidine structure (i.e. uracil versus thymidine). A mathematical analysis is presented that accounts for nonspecific binding to nucleic acids, including RNA, and evaluates the ability of MF3 to occupy specific sites of G-rich regions of altered DNA structure in vivo. These calculations indicate that the nucleic acid recognition properties of MF3 are consistent with the hypothesis that it can function by

binding to telomere ends or other sites of unusual DNA structure within the genome.

MATERIALS AND METHODS

Preparation of Factor—To remove bulk protein and endogenous nucleic acids from the telomeric probe-binding activity, MF3 was partially purified from a number of sources including the breast and thigh muscle of day 12–14 embryonic chickens, nuclear extracts of adult turkey erythrocytes as described previously (3), or nuclear extracts of Jurkat cells. Embryonic tissue extracts were typically made in standard buffer consisting of 10% glycerol, 25 mM Tris, pH 7.5, 1 mM EDTA, 2 mM dithiothreitol, 100 mM NaCl, and a 1 μg ml^{-1} concentration of the protease inhibitors leupeptin, aprotinin, chymostatin, and pepstatin (Boehringer Mannheim). Tissues were disrupted with ultrasonic vibration at 4 °C using a Branson sonifier at settings 3–5 with a ½-inch tip, and the homogenates were clarified by centrifugation at 25,000 $\times$ g for 30 min. Cell extracts were applied to a heparin-Sepharose column (Pharmacia Biotech Inc.), and the telomere-binding activity was eluted with a step gradient from 550 to 800 mM NaCl. Nuclear extracts of turkey erythrocytes were prepared as described by Evans *et al.* (19). The activity was partially purified by sequential chromatography through columns of DEAE-Sepharose (Sigma) and heparin-Sepharose. The nuclear extract was applied to a DEAE-Sepharose column, and the telomere-binding activity was eluted by a step gradient from 150 to 400 mM NaCl. These active fractions were applied to a heparin-Sepharose column, and MF3 was eluted with a step gradient from 500 to 850 mM NaCl. Typically, the MF3 activity could be detected following the initial fractionation step, but not in the crude cell extracts. For some analyses, the pool of MF3 activity from heparin-Sepharose chromatography was diluted with standard buffer and applied to a column of single-stranded DNA-cellulose (Pharmacia) that was equilibrated in standard buffer with 600 mM NaCl. MF3 activity was eluted with a step gradient to 800 mM NaCl, but the MF3 activity in these fractions was unstable to storage. The MF3 activity in Jurkat cell nuclear extracts was partially purified by heparin-Sepharose chromatography.

DNA Oligonucleotide Synthesis and in Vitro Transcription—The DNA oligonucleotides were prepared with an Applied Biosystems 391EP DNA synthesizer using the phosphoramidite method and purified by denaturing polyacrylamide gel electrophoresis and Sep-Pak C_{18} cartridges (Waters Associates). Phosphoramidites were from Applied Biosystems Inc., except for dU, which was obtained from Peninsula Laboratories, Inc. (Belmont, CA), and dI, which was obtained from Glen Labs. The MRE[1] probe contains the promoter sequences from positions −73 to −100 from the chicken skeletal α-actin gene. Single-stranded telomeric probes contained repeats of 5'-TTAGGG-3', or derivatives thereof, downstream of a portion of the noncoding strand of the MRE sequence (5'-CGGCCGTCGCCATATTTG-3') that was shown previously not to be bound by MF3 (18). The complete noncoding strand of the MRE, designated MREnc, is bound by MF3, and it has the sequence 5'-CGGCCGTCGCCATATTTGGGTGTCGGGC-3'. The RNA analog of the noncoding strand of the skeletal actin MRE was synthesized by *in vitro* transcription from a short synthetic DNA template (20). The transcription reaction utilized T7 RNA polymerase (U. S. Biochemical Corp.) and an 18-base pair double-stranded promoter region upstream from a 28-nucleotide 5'-overhang corresponding to the coding strand of the MRE (except for position +1, which was changed from C to a G to increase the efficiency of transcription). Transcription reactions were incubated for 3 h at 37 °C in a buffer containing 40 mM Tris, pH 8.3, 20 mM MgCl₂, 5 mM dithiothreitol, 1 mM spermidine, 8 mg ml^{-1} polyethylene glycol, a 4 mM concentration of each nucleotide, and 2.5 units μl^{-1} T7 RNA polymerase. The concentration of each DNA strand was 0.25 μM. The reactions were terminated by extractions with phenol and chloroform, and the nucleic acids were precipitated with ethanol. The transcript was isolated on a preparative 15% polyacrylamide gel with 50% urea. The concentration of RNA was determined by measuring the absorbance at 260 nm and confirmed by visual inspection on an analytical gel stained with ethidium bromide. The MREnc RNA product was sensitive to treatment with RNase or alkali, but it was stable when incubated with DNase. The VET2.0 RNA probe was made by a similar procedure, but the template for the transcription reaction was 5'-ATTATGCTGAGTGATATCCCCGGCAGCGGTATAAATCCCAATCCC-3'. Transfer RNA was from brewers' yeast (Boehringer Mannheim).

DNA Binding Assay—Electrophoretic mobility shift assays were typically performed with 0.02–0.6 ng of ^{32}P-labeled DNA fragments and

between 0.5 and 2.0 μl of heparin-Sepharose column fractions (with protein concentrations from 0.1 to 1.0 mg ml^{-1}). Binding reactions contained 10 mM Tris, pH 7.5, 30 mM KCl, 1 mM EDTA, 1 mM dithiothreitol, 8% glycerol, and 0.1–1.0 μg of poly(dI-dC)·poly(dI-dC) in a total volume of 10 or 15 μl. Binding reactions were initiated by the addition of protein, and following a 5–10 min incubation at room temperature, binding mixtures were loaded on a 5% polyacrylamide gel. Electrophoresis was performed at 20 V cm^{-1} in 22 mM Tris borate buffer with 0.5 mM EDTA. Gels were dried and exposed to film overnight at −70 °C with an intensifying screen. Probes used in the binding assays were labeled with polynucleotide kinase and [γ-^{32}P]ATP (>4500 Ci/mmol; Amersham Corp.). Typically, the labeled probes were purified with Elutip-d columns (Schleicher & Schuell) according to the directions of the manufacturer. Competition experiments were carried out with a fixed concentration of probe and varying levels of nonlabeled competitor. Binding constants for the complexes between MF3 and the VET2.0, VET3.0, and VET4.0 probes were determined by incubating constant amounts of partially purified fraction from embryonic chicken skeletal muscle or adult turkey erythrocytes with increasing amounts of specific DNA sequence for 5–10 min at 25 °C prior to gel electrophoresis under standard conditions (3). Following electrophoresis and autoradiography, the regions corresponding to the protein-DNA complex and the free DNA were excised from the gel and quantified by liquid scintillation counting. In other experiments, relative affinities were determined by comparing the test competitor with the nonlabeled probe in side-by-side experiments.

Chemical Protection Assays—DNA oligonucleotide probes, end-labeled with ^{32}P, were methylated for 30 s in a solution consisting of 1 part dimethyl sulfate in 20 parts 10 mM Tris, 1 mM EDTA, and 500 mM NaCl at 37 °C. Reactions were stopped by the addition of 9 M 2-mercaptoethanol, and the probes were precipitated by the addition of ethanol. Cleavage of the methylated residues was performed by incubating the probes in 100 μl of 1 M piperidine at 90 °C for 30 min. The samples were subsequently dried under vacuum and dissolved in deionized formamide prior to electrophoresis on a denaturing 20% polyacrylamide gel. The intensities of the bands corresponding to the methylated guanosine or inosine bands in the two telomeric repeats were analyzed with a PhosphorImager (Molecular Dynamics, Inc.) equipped with Image-Quant 3.3 software.

Theoretical—Specific site occupancy levels were calculated with Theorist software (version 1.51) on a Macintosh Quadra 800 computer. Plots were imported to Canvas (version 3.0.6) for graphical presentation.

RESULTS

A Single dG Exocyclic Amino Group Is Essential for Factor Binding—A previous study of MF3 binding to single-stranded probes containing two telomeric repeats revealed that the substitution of dI at all dG positions in the internal repeat completely blocked factor binding (3). Since the dI base, hypoxanthine, lacks the N-2 amino group, these data demonstrated that 1 or more of the guanine residues of the internal repeat contribute a critical exocyclic amino group necessary for the formation of the nucleoprotein complex. To further explore the involvement of exocyclic amino groups in specific factor recognition, a series of dI-substituted derivatives of a vertebrate telomeric probe (VET2.0) were synthesized and assayed for their ability to bind to MF3 (Fig. 1). All derivative probes had dI substituted for dG in the terminal repeat (positions 4–6) and different combinations of dI and dG in the internal repeat (positions 1–3). An electrophoretic mobility shift assay revealed that nucleoprotein complexes were formed between MF3 and the parental probe, VET2.0 (end structure = GGGTTAGGG), and with VET2.1 (GGGTTAIII) and VET2.2 (IGGTTAIII), while VET2.3 (IIGTTAIII) was a slightly weaker binder (Fig. 2A). However, VET2.6 (IIITTAIII), which has dI substituted at all of the telomeric dG positions, was incapable of significant factor binding. Competition experiments were performed to analyze the relative affinities of these probes for the factor (Fig. 2B). VET2.0, VET2.1, VET2.2, and VET2.3 were comparable in their ability to compete for complex formation, but the VET2.6 derivative, with no dG residues, was an ineffective competitor. The dissociation constant for the VET2.0 nucleoprotein com-

---

[1] The abbreviations used are: MRE, muscle regulatory element; XTEF, *Xenopus* telomere end-binding factor.

Parent Probe:

```
                              123     456
VET 2.0   5'- CGGCCGTCGCCATATTTG  TTAGGG  TTAGGG -3'
```

Inosine-substituted Probes:

		Internal Repeat	Terminal Repeat

VET 2.1   5'- CGGCCGTCGCCATATTTG **TTAGGG TTAIII** -3'

VET 2.2   5'- CGGCCGTCGCCATATTTG **TTAIGG TTAIII** -3'

VET 2.3   5'- CGGCCGTCGCCATATTTG **TTAIIG TTAIII** -3'

VET 2.4   5'- CGGCCGTCGCCATATTTG **TTAIGI TTAIII** -3'

VET 2.5   5'- CGGCCGTCGCCATATTTG **TTAGII TTAIII** -3'

VET 2.6   5'- CGGCCGTCGCCATATTTG **TTAIII TTAIII** -3'

FIG. 1. **Single-stranded telomeric probe sequence and deoxyguanosine position designation.** Two vertebrate telomeric repeats occur 3' of an 18-nucleotide sequence that is inert in terms of its ability to bind to factor (18). Derivative probes have deoxyinosine substituted for deoxyguanosine in the internal or terminal telomeric repeats. VET3.0 has three and VET4.0 has four wild-type telomeric repeats (not shown) that occur downstream of the same 18-nucleotide single-stranded fragment contained in the VET2.0 probe and its dI-substituted derivatives.

plex is 1.5 nM (3). Similar dissociation constant values were obtained using DNA probes with either three or four telomeric repeats, referred to as VET3.0 and VET4.0, respectively (Table I). In contrast, the competition data in Fig. 2B indicate that the VET2.6 nucleoprotein dissociation constant is $>10^{-7}$ M. These data demonstrate that high affinity factor binding can occur with probes containing two telomeric repeats and a single dG exocyclic amino group.

In a previous study (3), it was found that the exocyclic N-2 amino groups of the internal telomeric repeat, but not the terminal telomeric repeat, were required for the formation of a structure that was partially protected from methylation by dimethyl sulfate. Since these same groups were also required for specific factor recognition, these data provided correlative evidence that MF3 recognizes a nucleic acid structure that is more complex than a random coil, and they indicated that the nucleoprotein complex may have a specific G·G base pair configuration at the core-binding site. To extend these analyses to the series of dI-substituted derivatives analyzed in this study, methylation protection studies were performed on the VET2.1, VET2.2, VET2.3, and VET2.6 probes (Fig. 3). The methylation protection of N-7 groups in the terminal repeat was most extensive with the VET2.1 probe, which has three exocyclic N-2 groups in the internal repeat. The VET2.2 probe, with two internal N-2 groups, and VET2.3, with one internal N-2 group, gave protection patterns that were intermediate between the VET2.1 and VET2.6 protection patterns. Collectively, these data provide further evidence that the VET2 series of probes can form a chemically protected structure that requires the presence of at least one exocyclic N-2 group within the internal telomeric repeat. Furthermore, these data indicate that the stability of the chemically protected structure is dependent upon the total number of internal exocyclic amino groups in that the extent of protection is more pronounced with three N-2 groups than with two or one.

Next we performed binding experiments to determine whether the position of the dG exocyclic amino group in the internal telomeric repeat is critical for factor binding. For this experiment, the derivatives VET2.1 (end structure = GGGTTAIII), VET2.3 (IIGTTAIII), VET2.4 (IGITTAIII), and VET2.5 (GIITTAIII) were compared for their ability to compete for nucleoprotein complex formation in an electrophoretic mobility
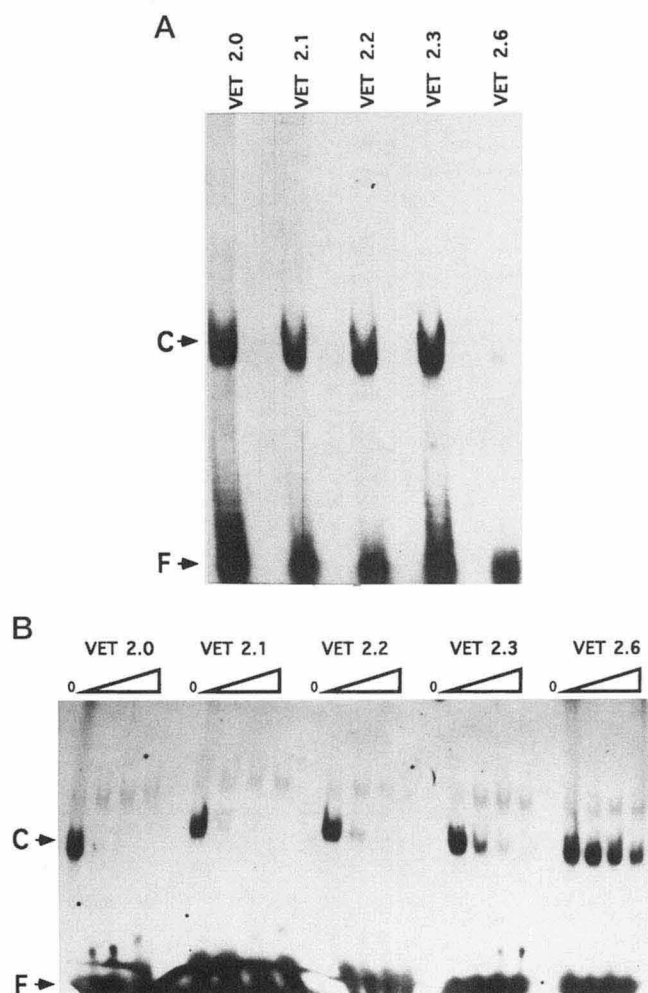

FIG. 2. **A single guanine exocyclic amino group occurring in the internal telomeric repeat is required for factor binding.** *A,* the VET2.0 parental probe and deoxyinosine-substituted probes were tested for their ability to form nucleoprotein complexes with factor in an electrophoretic mobility shift assay. For each assay, ~10,000 cpm (0.1 pmol) of each probe was used with 0.2 μg of poly(dI-dC)·poly(dI-dC). *B,* shown is the competition for nucleoprotein complex formation with parental and deoxyinosine-substituted telomeric probes. Complexes were formed between factor and radiolabeled VET2.0 in the absence or presence of the designated competitor. Nonlabeled competitor oligonucleotides were present at 0, 10, 30, or 100 ng (from *left* to *right*). The factor source for these experiments was turkey erythrocytes. The nucleoprotein complex (*C*) and the free probes (*F*) are indicated.

shift assay (Fig. 4). All of these dI-substituted probes were capable of significant competition, but the VET2.5 and VET2.4 probes (which have a single dG at positions 1 and 2, respectively) were slightly weaker competitors than the VET2.3 probe, which has a single dG at position 3. However, all the derivative probes with a single dG in the internal repeat were dramatically better at binding MF3 than was the probe that had a dI substitution at every dG position (VET2.6). Collectively, the ranking of probe affinity for MF3 is as follows: VET2.0, VET2.1, VET2.2 > VET2.3 > VET2.4, VET2.5 ≫ VET2.6. These data demonstrate that the guanine that contributes the critical N-2 group can occur at any dG position within the internal repeat, although position 3 was optimal for binding to MF3.

Specificity for DNA Involves Ribose Ring Structure—A concern with single-stranded DNA-binding activities is the possibility that the physiological role of these proteins is to bind to RNA, the predominant single-stranded nucleic acid within the cell. To address the issue of DNA *versus* RNA specificity for

Vertebrate Telomeric Probe-binding Activity

TABLE I
Dissociation constants for MF3 binding to DNA and RNA

Nucleic acid	K_D
	M
ss VET2.0 DNA[a,b]	1.5×10^{-9}
ss VET3.0 DNA[a]	1.6×10^{-9}
ss VET4.0 DNA[a]	1.6×10^{-9}
ss MREnc DNA[a,b]	0.4×10^{-9}
ss dU-MREnc DNA[c]	1×10^{-9}
ss MREnc RNA[c]	5×10^{-6}
tRNA[c]	2×10^{-6}
ss control DNA[c]	1×10^{-5}
ds control DNA[c]	3×10^{-4}
Poly(dI-dC) · poly(dI-dC)[c]	1×10^{-3}
Poly(dA-dT) · poly(dA-dT)[c]	8×10^{-4}

[a] Measured directly (see "Materials and Methods"; ss, single-stranded; nc, noncoding; ds, double-stranded).

[b] Reported in Ref. 3.

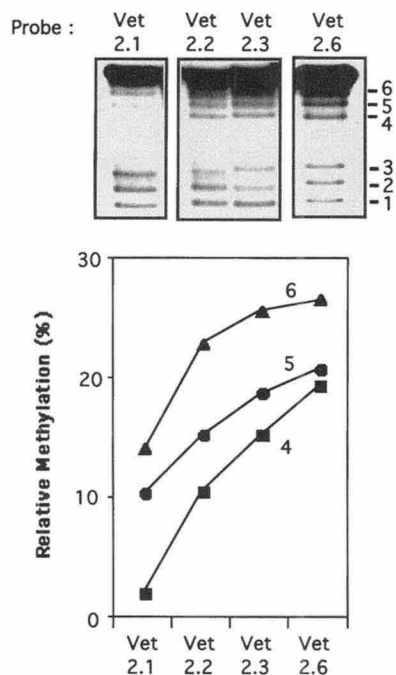[c] Estimated from competition experiments (see "Materials and Methods").

FIG. 3. **Chemical protection analyses of the deoxyinosine-substituted telomeric probes in the absence of protein.** Probes were compared regarding their sensitivity to methylation and cleavage following treatment with dimethyl sulfate. Methylation patterns of the telomeric repeats are shown above, where the guanosine/inosine residues are denoted as described for Fig. 1. The relative methylation of each probe at positions 4–6 is plotted below as the fraction of total methylation (sum of positions 1–6). These band intensity patterns indicate trends in methylation protection in the terminal repeat (positions 4–6). The ranking of degree of chemical protection at the terminal repeat is VET2.1 > VET2.2 > VET2.3 > VET2.6. Furthermore, for every probe analyzed, the greatest protection occurred at position 4, while position 6 was least protected.

MF3, the RNA analogs of two high affinity DNA-binding sites were tested for their abilities to form nucleoprotein complexes with MF3. The RNA analog of VET2.0 was synthesized using T7 RNA polymerase and a template that initiated transcription from a short double-stranded promoter situated upstream from the complementary strand of the VET2.0 probe. Unlike the VET2.0 DNA probe, the VET2.0 RNA probe was unable to form a nucleoprotein complex with MF3 (Fig. 5). No degradation of the RNA or DNA probe was detected when aliquots of the
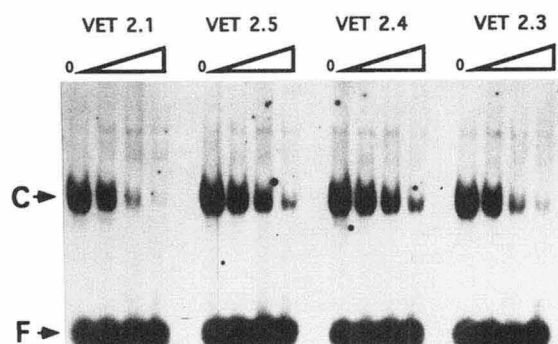

FIG. 4. **The exocyclic amino group required for factor binding can occur at any guanine position in the internal telomeric repeat.** The indicated probes were tested for their ability to compete for complex formation in an electrophoretic mobility shift assay. Nucleoprotein complexes were formed between the factor and the radiolabeled MREnc probe in the presence or absence of nonlabeled inosine-containing oligonucleotides. The MREnc sequence has a 4-fold higher affinity for factor than the VET2.0 probe that was used in Fig. 2B (3). For each assay, ~10,000 cpm (0.1 pmol) of MREnc probe was used with 0.2 μg of poly(dI-dC)·poly(dI-dC), and the factor source was turkey erythrocytes. Nonlabeled competitor oligonucleotides were present at 0, 10, 30, or 100 ng (from *left* to *right*). The nucleoprotein complex (C) and the free probes (F) are indicated.
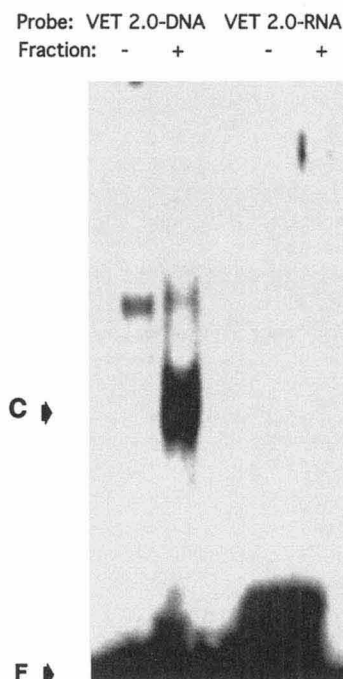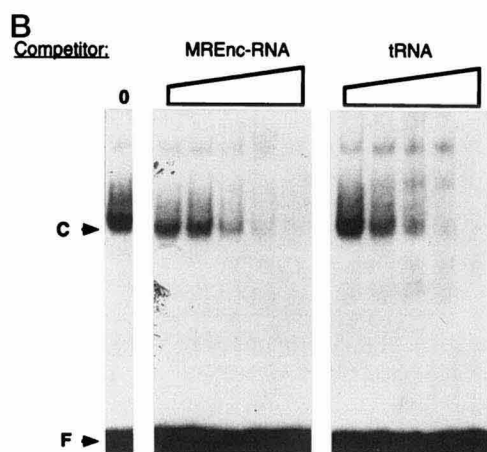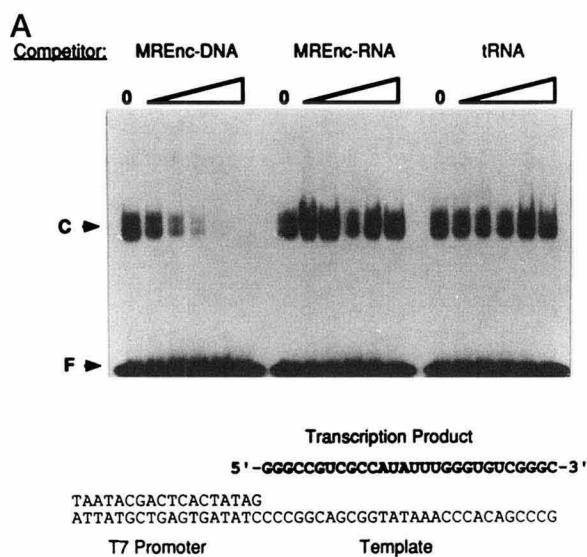

FIG. 5. **MF3 forms specific nucleoprotein complexes with a VET2.0 DNA probe, but not with the corresponding VET2.0 RNA probe.** The RNA analog of the VET2.0 DNA probe was synthesized using T7 RNA polymerase and a synthetic DNA template that had a double-stranded T7 promoter region upstream from the template sequence. Electrophoretic mobility shift assays were performed in the presence or absence of MF3 from Jurkat cells that was purified by heparin-Sepharose chromatography. Binding reactions contained 30,000 cpm of the DNA or RNA probe. Aliquots of each binding mixture were removed and analyzed by autoradiography following electrophoresis on a denaturing 20% polyacrylamide gel to confirm probe integrity.

binding reactions were subjected to electrophoresis on a denaturing 20% polyacrylamide gel (data not shown).
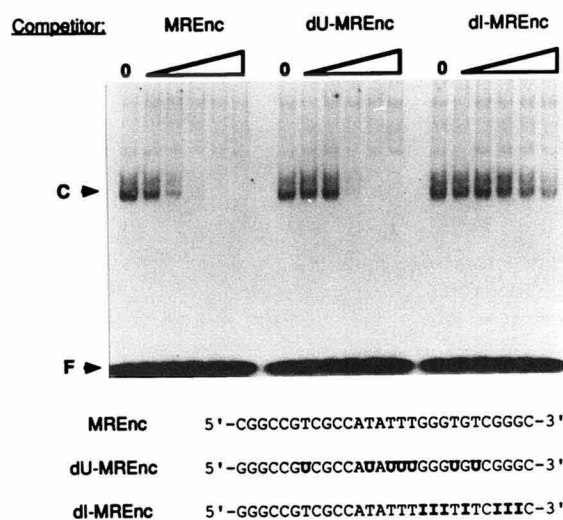
Competition experiments were also performed with the RNA analog of another DNA site that is referred to as MREnc (Fig. 6A). This probe corresponds to the noncoding strand of the chicken skeletal actin promoter from positions −73 to −100, where MF3 specifically recognizes the region encompassing the two G-rich repeats in the 3'-portion of this molecule (18). This

A

Competitor: MREnc-DNA   MREnc-RNA   tRNA

C ►

F ►

Transcription Product

5'-GGGCCGUCGCCAUAUUUGGGUGUCGGGC-3'

TAATACGACTCACTATAG
ATTATGCTGAGTGATATCCCCGGCAGCGGTATAAACCCACAGCCCG

T7 Promoter   Template

B

Competitor: MREnc-RNA   tRNA

0

C ►

F ►

FIG. 6. **Comparison of DNA and RNA in the competition of the specific nucleoprotein complex.** Complexes were formed between factor and the ^{32}P-labeled MREnc probe in the presence or absence of nonlabeled competitors. Competitors were the noncoding DNA strand of MRE (MREnc DNA), the RNA analog of this MRE sequence (MREnc RNA), and tRNA. The RNA analog of the skeletal actin MREnc probe was synthesized using T7 RNA polymerase and a synthetic DNA template that had a double-stranded T7 promoter region upstream from the template sequence. Factor was from embryonic chicken skeletal muscle, and electrophoretic mobility shift assays were performed with 25 pg of ^{32}P-labeled probe and 100 ng of poly(dI-dC)·poly(dI-dC). *A*, for each competition set, the first lane contained no competitor (0), and subsequent lanes contained 0.3, 1.0, 3.0, 10, and 30 ng of competitor in this order. *B*, binding reactions were performed in the absence of competitor (0) or with competition sets containing increasing amounts of MREnc RNA or tRNA. Competition sets contained 30 ng, 100 ng, 300 ng, 1 μg, or 3 μg of the designated RNA competitor in this order. The nucleoprotein complex (*C*) and the free probes (*F*) are indicated.

RNA analog, MREnc RNA, was compared with deoxyribonucleic MREnc (MREnc DNA) for binding to MF3 in quantitative competition experiments (Fig. 6A). The MREnc DNA sequence was a potent competitor, but the RNA analog of this sequence and tRNA were relatively ineffective in their ability to compete for complex formation. Based on quantitative experiments with higher levels of competitor, the affinity of MF3 for RNA was 3–4 orders of magnitude lower than for the specific DNA site (Fig. 6B). The binding to MREnc RNA was nonspecific because this probe was no more effective than tRNA in its ability to compete for the formation of the specific nucleoprotein complex. We note that the affinity of MF3 for nonspecific sites in RNA was higher than for nonspecific sites in duplexed DNA, which have dissociation constants of ~10^{-4} M (Table I).

Competitor: MREnc   dU-MREnc   dI-MREnc

C ►

F ►

MREnc   5'-CGGCCGTCGCCATATTTGGGTGTCGGGC-3'

dU-MREnc   5'-GGGCCGUCGCCAUAUUUGGGUGUCGGGC-3'

dI-MREnc   5'-GGGCCGTCGCCATATTTIIITITCIIIC-3'

FIG. 7. **Nucleotide structure requirements for high affinity binding to DNA.** Complexes were formed between the ^{32}P-labeled MREnc probe and factor in the presence or absence of nonlabeled competitor DNAs. DNA competitors were the noncoding strand of the MRE (MREnc) and its analogs, dU-MREnc and dI-MREnc. The analog dU-MREnc has deoxyuridine substituted for deoxythymidine throughout the probe. The analog dI-MREnc has deoxyinosine substituted for deoxyguanosine in the 3'-portion of the sequence. The dU-MRE and dI-MRE analogs have G instead of C at the most 5'-position, so they correspond to the MREnc RNA sequence used as a competitor in Fig. 4. For each competition set, the first lane contained no competitor (0), and subsequent lanes contained 0.3, 1.0, 3.0, 10, and 30 ng of competitor in this order. Factor was partially purified from embryonic chicken skeletal muscle, and electrophoretic mobility shift assays were performed with 25 pg of ^{32}P-labeled probe and 100 ng of poly(dI-dC)·poly(dI-dC).

Further experiments were performed to determine the structural requirements for high affinity binding to specific sequences in single-stranded deoxyribonucleic acid. The ability of a protein to distinguish a single-stranded sequence of DNA from the same sequence of RNA can result from the presence of thymidine in place of uridine or from the absence of the 2'-hydroxyl group from the pyranose ring. dU was substituted in place of dT over the entire noncoding strand of the MRE to test whether the methyl group at position 5 of the pyrimidine base had a role in determining MF3 specificity (Fig. 7). The modified sequence is referred to as dU-MREnc. Quantitative competition assays revealed that the substitution of dU for dT had a minimal (<3-fold) effect on the affinity of MF3 for this sequence (Fig. 7). From these data, it appears that the specific recognition of DNA *versus* RNA results from the difference in ribose ring structure between these nucleic acids. In RNA, the 2'-hydroxyl groups impose severe stereochemical constraints on the overall conformation of the nucleic acid (21), and these data are consistent with the hypothesis that MF3 recognizes a specific DNA structure that may be more complex than a random coil of single-stranded nucleic acid (3). Finally, the substitution of dI for dG in the 3'-portion of the MREnc probe diminished factor binding by a factor of 50 (Fig. 7), indicating that MF3 recognizes similar structural features in the MREnc and telomeric DNA probes.

Estimates of Site Occupancy—We wondered whether the nucleic acid recognition parameters of MF3 described above are consistent with the hypothesis that MF3 could occupy a significant fraction of telomere ends or other sites of altered DNA conformation *in vivo*. To address this issue, mathematical analyses were performed to evaluate the possibility of MF3 binding to a specific telomeric site in the presence of physiologically relevant levels of RNA and duplexed DNA that would act as nonspecific competitors. For these calculations, we considered a situation in which a protein is in equilibrium with a specific

binding site in DNA (or RNA) and with nonspecific sites in DNA and RNA. Equation 1 describes the partitioning of factor between these three classes of binding sites.

$$P \cdot \theta = \frac{P_t(\theta_t - P \cdot \theta)}{K_\theta\left(1 + \dfrac{RNA_t - P \cdot RNA}{K_{RNA}} + \dfrac{DNA_t - P \cdot DNA}{K_{DNA}}\right) + (\theta_t - P \cdot \theta)} \quad \text{(Eq. 1)}$$

In Equation 1, P_t, θ_t, DNA_t, and RNA_t are the total concentrations of the nucleic acid-binding protein, the specific binding site, DNA, and RNA, respectively. $P \cdot \theta$, $P \cdot DNA$, and $P \cdot RNA$ are the concentrations of the different protein-nucleic acid complexes, and K_θ, K_{DNA}, and K_{RNA} are the equilibrium dissociation constants for specific binding, nonspecific binding to DNA, and nonspecific binding to RNA, respectively. Equation 1 is similar to the competition equation derived by Lin and Riggs (22), but additional terms appear to account for nonspecific binding to RNA, a situation that is likely to occur with single-stranded nucleic acid-binding proteins. Under conditions where the total concentration of nonspecific sites in cellular nucleic acids greatly exceeds the concentration of binding protein and $DNA_t \gg P \cdot DNA$ and $RNA_t \gg P \cdot RNA$, Equation 1 can be simplified, and substituting the term for fractional specific site occupancy ($\gamma = P \cdot \theta/\theta_t$) gives Equation 2.

$$\gamma = \frac{P_t(1 - \gamma)}{K_\theta\left(1 + \dfrac{RNA_t}{K_{RNA}} + \dfrac{DNA_t}{K_{DNA}}\right) + \theta_t(1 - \gamma)} \quad \text{(Eq. 2)}$$

Equation 2 was used to evaluate how the occupancy of a specific site by MF3 or any other protein can be influenced by the abundance of factor and sites and by nonspecific binding to DNA and RNA (Fig. 8).

The parameters used for calculations of site occupancy were derived from previous measurements of cellular DNA and RNA content and volume, estimates of telomere-binding protein levels, and a value for the volume of the nucleus that has been used by others. Similar assumptions have been used to estimate the extent of the adenovirus major late promoter by the MLTF DNA-binding protein (23). The content of DNA and RNA in chicken embryo fibroblasts is 1.3 and 1.55 pg/cell, respectively (24). Thus, the calculated concentration of total intracellular RNA is 2.2 mg ml^{-1} (RNA_t = 6.7 mM), and the calculated concentration of total intranuclear DNA is 19 mg ml^{-1} (DNA_t = 58 mM) assuming that the nucleus is one-tenth the total volume of the cell. Estimates for telomeric probe-binding protein in higher eucaryotes range from <200 to ~10^7 copies/cell (15–17), and the calculated concentration of telomere-binding protein (P_t) will range from 5 nM to ~10^{-4} M assuming that the nucleus is a sphere with a 5-μm diameter (25). Likewise, the concentration of specific binding sites (θ_t) may vary from ~100 per nucleus (2.5 nM), assuming a single binding site/telomere end, to 100,000 per nucleus (2.5 μM) or more if the protein binds specifically to abundant non-telomeric sites. The binding constants for MF3 were used for many of these calculations, where the dissociation constant for the specific site (K_θ) is 1 nM, and the dissociation constants for nonspecific binding to DNA and RNA are 100 and 4 μM, respectively (Table I) (3).

Fig. 8A shows that the concentration of the telomere-binding protein and the relative affinity for a specific site are critical parameters that determine telomere site occupancy. Under the conditions assumed for these calculations, 50% telomere site occupancy occurs with ~10^4 binding protein molecules/nucleus when the telomere nucleoprotein dissociation constant is 10^{-10} M, ~10^5 binding protein molecules/nucleus when the dissociation constant is 10^{-9} M, and ~10^6 binding protein molecules/ nucleus when the dissociation constant is 10^{-8} M (with K_{RNA} and K_{DNA} values of 4 and 100 μM, respectively). For the telo-

meric probe-binding activity described here, we estimate that the factor level in avian erythrocytes is ~10^6 molecules/cell based upon purification data and calculations of nucleoprotein complex levels at saturating concentrations of telomeric probe assuming a 1:1 binding stoichiometry (3). If this factor is confined to the nucleus, the calculated level of telomere site occupancy is >92% with 10^6 MF3 molecules/cell and 53% with 10^5 MF3 molecules/cell. However, if MF3 was present at 10^4 copies/ cell, which is the estimated level of the MLTF transcription factor (26), the calculated level of telomere site occupancy would only be 10%. These calculations indicate that telomere end-binding proteins may have to be present at high levels to achieve a significant level of site occupancy despite the relatively low level of telomere ends. We note, however, that these estimates of telomere site occupancy may represent minimal values because much of the total DNA or RNA may not be free in solution, or large portions of the nucleic acid may be masked by bound proteins. Despite uncertainties about the levels of nucleic acid that are available for nonspecific competition, these calculations indicate that the thermodynamic properties of MF3 are reasonably consistent with the hypothesis that it can occupy sites at telomere ends or other sequences of altered DNA structure within the cell.

These calculations also indicate that the extent of telomere site occupancy is relatively unaffected by the total number of specific binding sites within the nucleus (θ_t) as long as θ_t does not exceed the concentration of binding proteins (P_t). For example, in Fig. 8B, the extent of site occupancy is plotted as a function of the level of specific binding sites for three concentrations of telomere-binding protein. At the higher levels of telomere-binding protein (10^6 and 10^7 molecules/nucleus), the calculated site occupancies are similar when the total number of sites is varied from 10 to 10^6 per nucleus. The effects of nonspecific binding to RNA and DNA on telomere site occupancy are shown in Fig. 8 (C and D, respectively). As in the case of gene regulatory proteins (25, 27), the interaction of telomere-binding proteins with nonspecific sites in nucleic acids is predicted to have a significant influence on the extent of specific site occupancy. For the telomere-binding activity described here, the affinity for nonspecific sites in single-stranded RNA is 25-fold higher than the affinity for nonspecific sites in double-stranded DNA, and a greater fraction of this protein is predicted to be bound nonspecifically to RNA than to DNA under *in vivo* conditions.

DISCUSSION

Sequence-specific, Single-stranded Nucleic Acid-binding Activities—The telomere ends of ciliates contain short G-rich extensions of single-stranded DNA, and the nucleoprotein structures formed by these sequences have been the subject of numerous investigations. It is not known whether telomere ends in animal cells possess single-stranded extensions, but extracts made from these cells contain activities that specifically bind to single-stranded probes of telomeric repeats *in vitro* (3, 14–17). Comparing these reports, it appears that at least two general classes of nucleic acid-binding proteins have been identified. One class of telomeric probe-binding activity represents loosely associated proteins of the heterogeneous nuclear ribonucleoprotein complex (14, 17). These proteins bind with higher affinity to RNA oligonucleotides with the repeat sequence r(UUAGGG) than to their DNA counterpart, and the first four positions in the repeat, rUUAG, are most critical for specific factor recognition. The second class of telomeric probe-binding activity is specific for DNA oligonucleotides (3, 16). The activity reported here, MF3, specifically binds to telomeric and non-telomeric DNA probes that possess two or more tracts of consecutive guanosine residues. A telomeric probe-binding ac-
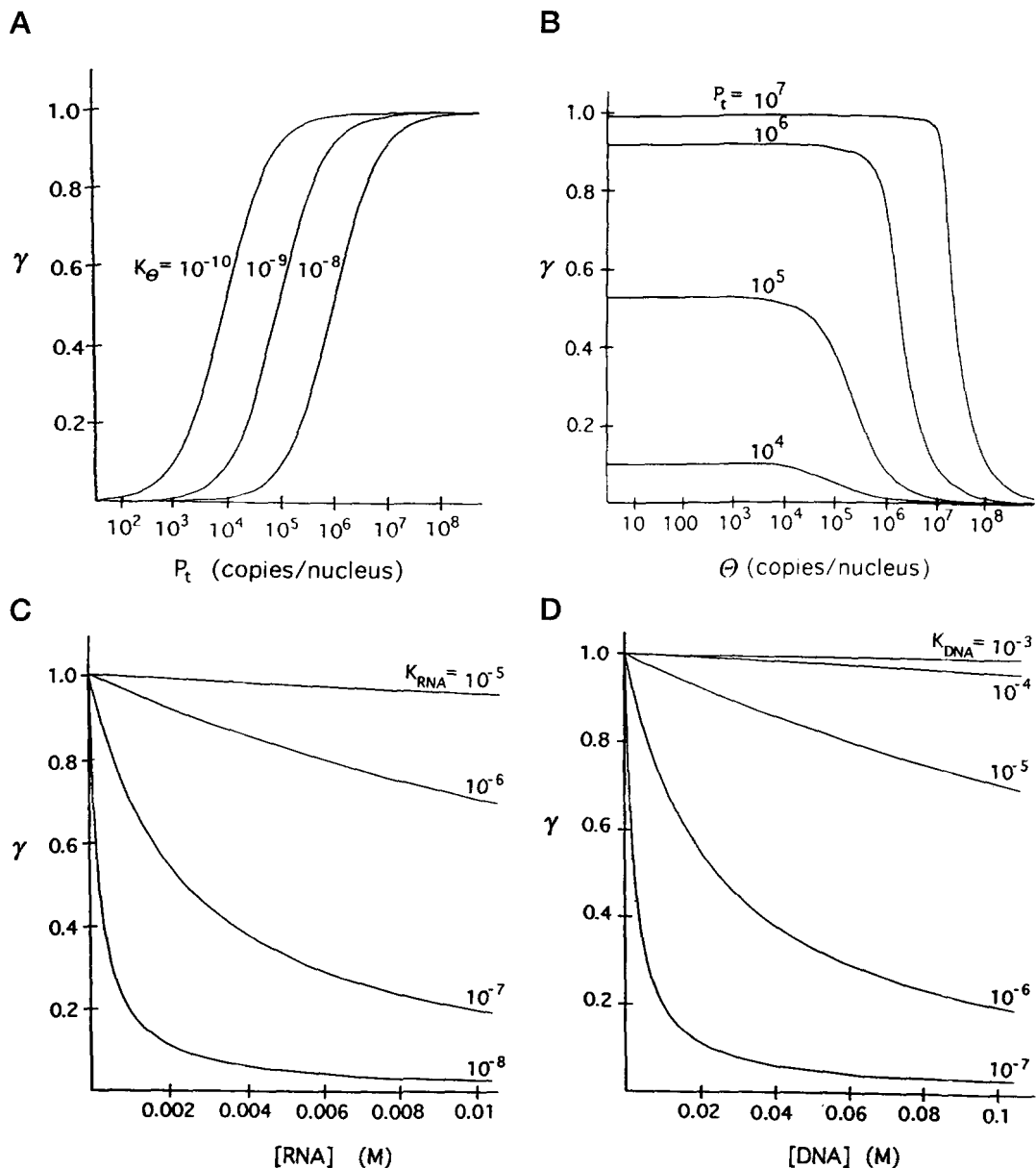
A

B

C

D

FIG. 8. **Parameters that influence estimated specific site occupancy by factor *in vivo*.** Plots were constructed from the relationships expressed in Equation 2. Estimated site occupancy (γ) was calculated using the parameters determined for the factor described in this study except when that parameter was varied for the calculations. The dissociation constant for specific sites (K_θ) was 1×10^{-9} M, and the dissociation constants for nonspecific binding to RNA and DNA (K_{RNA} and K_{DNA}) were 1×10^{-4} and 4×10^{-6} M, respectively. The intracellular RNA concentration is 2.2 mg ml^{-1}, and the intranuclear DNA concentration is 19 mg ml^{-1} for a chicken embryo fibroblast; and thus, the calculated concentrations of nonspecific sites in RNA (RNA$_t$) and DNA (DNA$_t$) are 6.7×10^{-3} and 5.8×10^{-2} M, respectively. For many of these calculations, it was assumed that the intranuclear concentration of factor (P$_t$) was 2.5×10^{-5} M (1×10^6 copies/nucleus) and that the concentration of specific sites (θ_t) was 2.5×10^{-9} M (100 per nucleus). *A*, estimated specific site occupancy as a function of the concentration of binding protein calculated with different values for the dissociation constant for the specific protein-nucleic acid complex; *B*, estimated specific site occupancy as a function of the concentration of specific binding sites in the nucleus calculated for different nuclear concentrations of binding protein; *C*, estimated specific site occupancy as a function of total RNA concentration in the cell calculated with different values for the dissociation constant for the nonspecific interaction with RNA (for these calculations, the contribution of nonspecific binding to DNA was assumed to be negligible (DNA$_t$ = 0)); *D*, estimated specific site occupancy as a function of total DNA concentration in the nucleus calculated with different values for the dissociation constant for the nonspecific interaction with DNA (for these calculations, the contribution of nonspecific binding to RNA was assumed to be negligible (RNA$_t$ = 0)).

tivity detected in *Xenopus* egg and somatic cell extracts is also specific for single-stranded DNA sequences (16). Similar to the telomeric probe-binding activity reported here, the *Xenopus* factor displays a relatively high tolerance for sequence alterations outside of the tracts of guanosine residues in that it binds with high affinity to single-stranded overhang repeats of 5'-TTAGGG-3' and 5'-GGGTTA-3' and displays a slightly reduced affinity for single-stranded repeats of 5'-AAAGGG-3' and 5'TTTGGG-3'.

Specificity for DNA—Guanine-guanine base pairing appears

to be a critical feature in specific telomeric probe recognition by MF3. Here we demonstrate that only a single guanine N-2 group is required for specific MF3 binding and that this critical group can occur at any of the three guanine positions within the internal repeat. These data suggest that a single G·G base pair utilizing this hydrogen bond donor is required for specific factor recognition at the core-binding site (also see Ref. 3). Chemical protection assays of the dI-substituted probes revealed that the residues of the terminal telomeric repeat were undermethylated when the exocyclic N-2 group was present in

the internal telomeric repeat, and the extents of chemical protection correlated with the number of dG residues within the internal repeat (Fig. 3). These data provide further correlative evidence that the dG·dG base pairing, such as may occur in a fold-back structure, can have a role in specific MF3 binding. Additionally, it is also possible that the binding of MF3 can promote the formation of the dG·dG base-paired structure similar to what has been described for the β-subunit of the *Oxytricha* telomere-binding protein (5).

Because G·G base pairs are likely to occur frequently in cellular RNAs, the specificity of MF3 for DNA may be a feature that is critical for its function. To determine the structural features that enable MF3 to distinguish between DNA and RNA, dU was substituted for dT in one of the high affinity DNA probes. The dU substitution had little or no effect on binding (Fig. 7). These data rule out the possibility that the methyl group at position 5 of the pyrimidine ring is a key determinant in specific recognition, and they indicate that the factor distinguishes between DNA and RNA on the basis of differences in ribose ring structure. One possibility is that MF3 makes specific contacts with the 2'-positions of the sugar molecules. An alternative hypothesis is that the 2'-hydroxyl group of RNA prevents the nucleic acid from attaining a conformation that is specifically recognized by MF3. The 2'-hydroxyl group will influence overall nucleic acid conformation by altering the preferred pucker conformation of the ribose ring (21). In complementary RNA double helices, the 2'-hydroxyl group predominantly confines the ribose to the $C^{3'}$-*endo* pucker, and the helix acquires the A conformation. However, in DNA helices, the $C^{2'}$-*endo* and $C^{3'}$-*endo* ribose puckers occur, allowing numerous double-stranded conformations. Thus, the extra flexibility of the deoxyribose ring, permitted by the absence of the 2'-hydroxyl group, may be an important feature in the specific recognition of DNA by MF3. Based on these data, we propose that MF3 recognizes a G·G base-paired structure that occurs in a DNA-specific conformation and that RNA is not bound with high affinity because the 2'-hydroxyl group prevents the nucleic acid from attaining this conformation. Through this type of mechanism, MF3 can selectively bind to G·G base pairs in DNA and minimize interactions with cellular RNAs that can adopt a multitude of tertiary structures through non-canonical base pairing.

Parameters Influencing Specific Site Occupancy—Estimates of dissociation constants for specific and nonspecific binding sites permit calculations regarding the extent of MF3 binding to telomere ends or other G-rich regions of altered chromatin structure *in vivo*. Due to the uncertainty of conditions within the cell, we used a wide range of parameters in these calculations to identify thermodynamic properties that are critical for telomere site occupancy by protein. For the telomeric probe-binding activity described here, the dissociation constant for the specific DNA site (K_9) is 1 nM, and the dissociation constants for nonspecific DNA (K_{DNA}) and RNA (K_{RNA}) are 100 and 4 μM, respectively (Table I); these parameters were utilized for many of the calculations.

The concentration of the binding protein is a critical parameter in determining the extent of telomere site occupancy (Fig. 8A). MF3 is a relatively abundant factor, and in avian erythrocytes and embryonic skeletal muscle, we estimate that the level of this factor may be as high as 10^6 copies/cell. Given these parameters, the calculated level of telomere site occupancy is >90% assuming that MF3 is present at 10^6 molecules/cell and >50% assuming that MF3 is present at 10^5 molecules/cell. These calculations also indicate that a significant fraction of MF3 (>99%) is likely to be bound to nonspecific nucleic acid sequences *in vivo*. Similar behavior has also been predicted for

sequence-specific B-DNA-binding proteins, such as the *lac* repressor, where >98% of the total factor is bound to nonspecific sites in the genome (25, 27). However, for MF3, RNA is likely to be the predominant nonspecific competitor (Fig. 8, compare C and D), and nonspecific sites in RNA are predicted to complex with 75% of this protein given our assumptions about physiological conditions. The *Xenopus* telomere end-binding factor (XTEF) is reported to be present at <100 binding units/nucleus in somatic tissues (16). If this estimation of factor abundance is correct, it can be calculated that an extraordinarily low dissociation constant ($\sim 10^{-13}$ M) would be required for significant (~90%) telomere site occupancy (where K_{DNA} and K_{RNA} are 100 and 4 μM, respectively). On the other hand, if the dissociation constant of XTEF for telomeric sites was in the nanomolar range, the extent of telomere site occupancy would be <1% due to the overwhelming nonspecific interactions with nucleic acids. Thus, measurements of the XTEF affinity for specific and nonspecific sites in nucleic acids will be required to assess the physiological significance of the XTEF-telomere end interaction.

The extent of specific site occupancy is relatively unaffected by the total concentration of specific sites in that significant competition for binding only occurs when the concentration of specific sites exceeds the concentration of the factor (Fig. 8B). In view of these considerations, it is interesting to compare the levels of telomere end-binding proteins that are required for significant site occupancy in ciliates that have up to $\sim 2 \times 10^7$ telomere ends and in animals that have $\sim 1 \times 10^2$ telomere ends. The calculated level of telomere-binding protein required for 90% telomere end occupancy is 1.9×10^7 in ciliates, where most of the protein is bound to telomeric sites, and 8×10^5 in animals, where most of the protein is bound nonspecifically to nucleic acid. This suggests that despite enormous differences in the abundance of telomere ends in these two species (>10^5-fold), the binding proteins must be present at relatively high levels in both species to occupy a significant fraction of these telomere end sites *in vivo*.

Collectively, these calculations permit an initial assessment of a protein's ability to bind to telomere ends *in vivo* based on measurements of factor interactions with nucleic acids performed *in vitro*. These estimates can be revised if it is determined that the telomeric sites are blocked by the binding of a protein with more favorable thermodynamic characteristics or if other preassembled telomeric proteins facilitate the binding of the factor to the telomere end. In this regard, we note that the *Oxytricha* telomeric protein is a heterodimer of subunits that bind cooperatively to telomeric DNA probes to form a highly stable ternary complex (28). Whether similar subunit interactions occur at chromosome ends in multicellular organisms remains to be determined, but we note a number of similarities between MF3 and the macronuclear telomere end-binding protein of ciliates. For example, both proteins recognize oligonucleotide probes to telomeric repeats with similar apparent affinities, both nucleoprotein complexes are relatively resistant to high concentrations of salt in the binding reaction, and the substitution of dU residues for dT does not significantly effect complex formation (Fig. 7) (29, 30).[2]

The data presented here are consistent with the hypothesis that MF3 specifically binds to unusual non-Watson-Crick base-paired structures of DNA. A key feature of the specific binding site is the occurrence of two tracts of guanosine residues where the exocyclic N-2 groups of the 3'-tract of guanosines are critical in nucleoprotein complex formation. Here we show that a

---

single exocyclic N-2 group at any position within the internal repeat of a telomeric probe is essential for factor binding, while all other exocyclic N-2 groups are expendable for this interaction. These binding data can be correlated with chemical protection assays showing that probes containing exocyclic N-2 groups in the internal telomeric repeat were undermethylated relative to a probe that did not. Furthermore, the extent of methylation protection correlated with the total number of exocyclic N-2 groups within the internal repeat. These data indicate that MF3 has a relatively low specificity for nucleotide sequence, but a high specificity for a tertiary structure that may be stabilized by a specific dG·dG base pair configuration. This tertiary structure may be rare in chromatin and only occur at chromosome ends or at other G-rich sequences that are sensitive to S1 nuclease digestion. However, G·G base pairs may occur more frequently in cellular RNAs, and the specificity of MF3 for deoxyribose may be a feature that enables this factor to occupy a significant fraction of telomere ends or other sites of altered DNA structure *in vivo* despite an overwhelming abundance of non-Watson-Crick base-paired RNA. Finally, the thermodynamic properties and binding specificity measurements reported here may prove useful in evaluating the physiological roles of putative telomere end-binding activities encoded by molecular clones.

REFERENCES

1. Henderson, E. R., and Blackburn, E. H. (1989) *Mol. Cell. Biol.* **9**, 345–348
2. Klobutcher, L. A., Swanton, M. T., Donini, P., and Prescott, D. M. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 3015–3019
3. Gualberto, A., Patrick, R. M., and Walsh, K. (1992) *Genes & Dev.* **6**, 815–824
4. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) *Cell* **59**, 871–880
5. Fang, G., and Cech, T. R. (1993) *Cell* **74**, 875–885
6. Veselkov, A. G., Malkov, V. A., Frank-Kamenetskii, M. D., and Dobrynin, V. N. (1993) *Nature* **364**, 496
7. Walsh, K., and Gualberto, A. (1992) *J. Biol. Chem.* **267**, 13714–13718
8. Liu, Z., Frantz, J. D., Gilbert, W., and Tye, B. K. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3157–3161
9. Henderson, E. R., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., and Blackburn, E. H. (1987) *Cell* **51**, 899–908
10. Gottschling, D. E., and Zakian, V. A. (1986) *Cell* **47**, 195–205
11. Gray, J. T., Celander, D. W., Price, C. M., and Cech, T. R. (1991) *Cell* **67**, 807–814
12. Fang, G. W., and Cech, T. R. (1991) *Nucleic Acids Res.* **19**, 5515–5518
13. Hicke, B. J., Celander, D. W., MacDonald, G. H., Price, C. M., and Cech, T. R. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1481–1485
14. McKay, S. J., and Cooke, H. (1992) *Nucleic Acids Res.* **20**, 6461–6464
15. McKay, S. J., and Cooke, H. (1992) *Nucleic Acids Res.* **20**, 1387–1391
16. Cardenas, M. E., Bianchi, A., and de Lange, T. (1993) *Genes & Dev.* **7**, 883–894
17. Ishikawa, F., Matunis, M. J., Dreyfuss, G., and Cech, T. R. (1993) *Mol. Cell. Biol.* **13**, 4301–4310
18. Santoro, I. M., Yi, T., and Walsh, K. (1991) *Mol. Cell. Biol.* **11**, 1944–1953
19. Evans, T., Reitman, M., and Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 5976–5980
20. Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) *Nucleic Acids Res.* **15**, 8783–8798
21. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, pp. 51–104, Springer-Verlag New York Inc., New York
22. Lin, S., and Riggs, A. D. (1972) *J. Mol. Biol.* **72**, 671–690
23. Chodosh, L. A., Carthew, R. W., and Sharp, P. A. (1986) *Mol. Cell. Biol.* **6**, 4723–4733
24. Golde, A. (1962) *Virology* **16**, 9–20
25. Lin, S., and Riggs, A. D. (1975) *Cell* **4**, 107–111
26. Carthew, R. W., Chodosh, L. A., and Sharp, P. A. (1985) *Cell* **43**, 439–448
27. Kao-Huang, Y., Revzin, A., Butler, A. P., O'Conner, P., Noble, D. W., and von Hippel, P. H. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 4228–4232
28. Fang, G., Gray, J. T., and Cech, T. R. (1993) *Genes & Dev.* **7**, 870–882
29. Raghuraman, M. K., Dunn, C. J., Hicke, B. J., and Cech, T. R. (1989) *Nucleic Acids Res.* **17**, 4235–4253
30. Raghuraman, M. K., and Cech, T. R. (1990) *Nucleic Acids Res.* **18**, 4543–4552