Nucleic acid specificity of a vertebrate telomere-binding protein: evidence for G-G base pair recognition at the core-binding site

Antonio Gualberto, Robert M. Patrick, and Kenneth Walsh

Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio 44106 USA

A factor from avian cells formed complexes with telomeric sequences and other single-stranded probes that contained tracts of guanine residues. Nucleoprotein complexes with telomere probes required two or more of the telomeric repeats that were incapable of Watson–Crick base-pairing. Methylation interference and protection experiments identified guanine N7 residues that were critical for the formation of the nucleoprotein complex and for the formation of a higher-order structure that occurred in the absence of the protein. Substitutions of deoxyinosine (dI) for deoxyguanosine (dG) demonstrated that the exocyclic N2 amino groups in the internal telomeric repeat, but not the terminal repeat, were required for the formation of the chemically protected structure and for protein binding. On the basis of these data we propose that the factor specifically recognizes a hairpin DNA structure that is stabilized by intramolecular G-G base-pairing between the telomere repeats. The positions of the critical guanine N2 and N7 groups indicate a G-G base-pairing configuration, where guanines function as hydrogen bond donors at the internal telomeric repeat and hydrogen bond acceptors at the terminal telomeric repeat.

[Key Words: Avian cells; nucleoprotein complexes; telomere-binding protein; G-G base-pairing]

Received January 7, 1992; revised version accepted March 5, 1992.

Telomeres at the natural chromosome ends are sites of altered DNA structure (Zakian 1989; Greider 1990; Blackburn 1991). Telomeres consist of short tandem repeats of 5'-TTAGGG-3' in vertebrates, slime molds, and trypanosomes, and 5'-TTGGGG-3' in the widely studied ciliates. Short 3' extensions of single-stranded repeats have been found at extreme chromosome termini in some organisms. These single-stranded and duplex repeats are required for the complete replication of the chromosome and may contribute to chromosome stability and nuclear architecture. One approach to understanding higher-order telomere structure is to study the unique recognition properties of the proteins that bind to these sequences.

Most sequence-specific DNA-binding proteins make primary contacts with the major groove of Watson–Crick base-paired DNA in the B conformation (Ollis and White 1987). Several nonspecific single-stranded DNA-binding activities have been characterized (Chase and Williams 1986), but relatively little is known about the proteins that bind to specific sequence motifs within single-stranded DNA probes. These site-specific factors are usually referred to as single-stranded DNA-binding proteins because they preferentially bind to probes of single-stranded nucleic acid. However, it is generally not known whether these proteins recognize sites within random coils of single-stranded DNA or higher-order structures stabilized by base-stacking and intramolecular base-pairing. Included in this group of proteins are the factors that bind to the 3' overhang of telomeres. In ciliates, which have $>10^7$ telomere ends, the sequence-specific single-stranded DNA-binding proteins can be released from 2 M salt-extracted macronuclear DNA by treatment with a nuclease (Gottschling and Zakian 1986). Other proteins of this class can specifically bind single-stranded probes of replication origins (Hofmann and Gasser 1991), recombination hot spots (Edelmann et al. 1989), or gene regulatory regions (Jost et al. 1987; Herget et al. 1989; Lannigan and Notides 1989; Rajavashisth et al. 1989; Gaillard and Strauss 1990, Wilkinson et al. 1990). Presumably these proteins interact with chromatin structures that are sensitive to conformation-dependent nucleases, such as S1, and to bromoacetaldehyde (Larsen and Weintraub 1982, Kohwi-Shigematsu et al. 1983).

We identified a protein, referred to previously as MF3, that has a potent single-stranded DNA-binding activity
and is specific for a subset of probes of gene regulatory sequences (Santoro et al. 1991). Here, we report that this factor also recognizes ciliate and vertebrate telomeric sequences. High-affinity binding occurred with probes containing two or more single-stranded telomeric repeats, but not with a probe containing a single telomere unit. A subset of the guanine N7 and exocyclic N2 positions within the telomeric repeats was found to be essential for factor binding. The locations of these critical guanine groups suggest that the factor recognizes a specific nucleic acid structure that is stabilized by Hoogsteen base-pairing between guanines of the two telomeric repeats. The nucleic acid recognition properties of this factor indicate that it is capable of occupying sites at chromosome termini or at other guanine-rich regions of altered DNA structure.

Results

A factor specific for telomere probes and other single-stranded, guanine-rich sequences

A fraction containing avian MF3 was tested for its ability to bind to telomeric sequences because these short tandem repeats resemble the G-rich binding sites identified previously (Santoro et al. 1991). For these studies, partially purified fractions from adult erythrocytes and embryonic skeletal muscle were utilized because the crude extracts contained several nonspecific single-stranded DNA-binding proteins as well as an inhibitor of the binding activity described here. The single-stranded telomere probes VET0, VET1, VET2, VET3, and VET4 contain zero, one, two, three, and four copies, respectively, of the vertebrate telomere repeat at the 3' end of an unrelated sequence (Fig. 1). VET0, the control probe, did not bind the factor in the electrophoretic mobility-shift assay (EMSA). Also, a single telomere repeat (VET1) was not sufficient for binding, but probes containing two or more telomere repeats formed nucleoprotein complexes. These complexes typically appeared as doublets, with the highest mobility band having the greatest intensity in the autoradiograph. Binding also occurred with the single-stranded TET2 probe, which contains two repeats of the telomere sequence from Tetrahymena.

MF3 was identified initially as a sequence-specific single-stranded nucleic acid-binding activity that bound to probes of gene regulatory sequences. To demonstrate that this protein, or group of proteins, also bound to the telomeric sequences, gene regulatory sequences were tested for their ability to compete for the nucleoprotein complex formation with the VET2 telomere probe (Fig. 2). The single-stranded probes used for the competition experiment were from the noncoding strand of the chicken skeletal actin muscle regulatory element (MREnc), the noncoding strand of the mouse c-fos serum response element (SREnc), the coding strand of the chicken skeletal actin conserved muscle gene DNA element (MCATcod), and the noncoding strand of the two E-box motifs (2E-BOXnc) from the mouse creatine kinase M enhancer. MREnc, MCATcod, and 2E-BOXnc were able to compete for the formation of the nucleoprotein complex. However, SREnc was not an effective competitor even at 10- to 100-fold higher concentrations than were used in Figure 2 (data not shown). These data demonstrate that this factor can bind specifically to at least two classes of single-stranded DNA sites, telomeric and gene regulatory sequences. We note that the probes binding to this protein are similar in that they contain two or more stretches of consecutive dG residues and little or no other regions of sequence similarity.

To characterize this protein–telomere interaction in greater detail, synthetic chromosome termini were tested for binding in the EMSA (Fig. 3). Probes with a "natural terminus" have a single duplex region of T2AG3/A2TC3 followed by a tail of one, two, or three telomere repeats that cannot form Watson–Crick base pairs. On the basis of competition experiments with these probes, binding occurred if the termini had overhangs of two or more telomere repeats. The terminus...
Specific G-G base pair recognition

Competitor: None MREnc SREnc MCATcod 2E-BOXnc

Gene Regulatory Sequence Probes
MREnc GGCCGTGGCCATATTTGGGTGGCTGGGC SREnc GCCGGCAGCGGTATAAACAAATTC
MCATcod GCCGGGACGGGTATAAACAAATTC 2E-BOXnc CGGCCGTCGCCATATTTGGGTGGCTGGGC

Figure 2. Competition of the telomeric nucleoprotein complex by gene regulatory sequences. Complexes between factor and radiolabeled VET2 probe were formed in the presence or absence of a 50-fold molar excess of the indicated single-stranded probes of gene regulatory sequences. Gene regulatory sequences are from MREnc, SREnc, MCATcod, and 2E-BOXnc, as described in the text. The tracts of dG are indicated in boldface type. Approximately 0.08 pmole (10,000 cpm) of the VET2 probe was used in each assay with or without 4 pmols of competitor and 1.0 μg of poly[d(I-C)] ∙ poly[d(I-C)], poly[d(A-T)] ∙ [A-T], poly[d(G-T)] ∙ [d(C-A)], and poly[d(A-C)] ∙ [d(C-T)] with five, five, three, and three orders of magnitude lower than for the specific site, respectively (data not shown). At this time no specific RNA-binding site has been identified for this factor. The factor recognized the alternating copolymers, poly[d(I-C)] ∙ [d(I-C)], poly[d(A-T)] ∙ [A-T], poly[d(G-T)] ∙ [d(C-A)], and poly[d(A-C)] ∙ [d(C-T)] with five, five, three, and three orders of magnitude lower affinities than for the specific site, respectively. Collectively, these experiments dem-

with one T2AG3 single-stranded tail did not compete for complex formation. Synthetic termini with random duplex sequences, and tails of two, three, and four telomere repeats were all capable of competition. Collectively, these experiments indicate that telomerases are potential binding sites for this factor and that the interaction requires two or more G-rich repeats that cannot form Watson–Crick base pairs. In subsequent experiments we focus primarily on the VET2 probe because it was the least complex telomere probe that could bind to the factor.

Initial estimates of binding affinities were determined for the nucleoprotein complexes with the VET2 and the MREnc probes. Binding affinities were determined by measuring the quantity of the specific protein–DNA complex formed as a function of DNA concentration in the EMSA (Fig. 4). This method was advantageous because the concentration of input DNA was known and the concentration of active nuclear factor did not have to be estimated. Binding constants were calculated by Scatchard-type analysis. Assuming a 1 : 1 stoichiometry between the specific protein and the single-stranded DNA sites, dissociation constants were 0.4 × 10⁻⁹ M for MREnc and 1.5 × 10⁻⁹ M for the telomere probe VET2. The same binding affinities were obtained over a 15-fold range of protein concentrations and with partially purified factor from avian skeletal muscle or erythrocyte extracts. The specificity of this binding activity was analyzed further by comparing the relative amounts of specific and nonspecific nucleic acid required to compete for complex formation in quantitative competition experiments (data not shown). Relative differences in the dissociation constants for the specific and nonspecific complexes could be calculated on the basis of molar concentration of nucleotides for the nonspecific sites. The affinity of this factor for nonrelated, single-stranded DNA oligonucleotides and RNA was four and three orders of magnitude lower than for the specific site, respectively (data not shown). At this time no specific RNA-binding site has been identified for this factor. The factor recognized the alternating copolymers, poly[d(I-C)] ∙ [d(I-C)], poly[d(A-T)] ∙ [A-T], poly[d(G-T)] ∙ [d(C-A)], and poly[d(A-C)] ∙ [d(C-T)] with five, five, three, and three orders of magnitude lower affinities than for the specific site, respectively. Collectively, these experiments dem-

Figure 3. Competition for complex formation using synthetic chromosome termini. Complexes were formed between factor and 0.05 pmole of the radiolabeled MREnc probe in the presence or absence of an 80-fold molar excess of the indicated double-stranded competitors. Binding reactions were performed with 1.0 μg of poly[d(I-C)] ∙ poly[d(I-C)], partially purified protein from embryonic skeletal muscle, and in the presence or absence of 4 pmoles of competitor. The nucleoprotein complex (C) and the free DNA (F) are indicated by arrows. Natural terminus refers to the presence of the T2AG3/A2TC3 duplex, not meant to imply in vivo telomere structure.
Figure 4. Scatchard analysis of factor binding to a single-stranded telomere and gene regulatory sequence probes. Binding constants were determined for the single-stranded VET2 telomere and the MREnc probes. The slope of the plots is the negative reciprocal of the dissociation constant. EMSAs were performed with a fixed amount of protein and variable amounts of DNA probes. For the MREnc determination, the factor was from an embryonic chicken skeletal muscle fraction; for the VET2 determination, the factor was from an adult turkey erythrocyte fraction; for VET2, the amounts of complexed and free DNA were determined for different concentrations of the VET2 probe from the EMSA shown below. The levels of bound (B) and free (F) probe are indicated. For MREnc, the concentrations of complexed and free DNA represent mean values from three separate EMSAs. All assays had 1.0 μg of poly[d(I-C)]·poly[d(I-C)].

onstrate that the factor has a high degree of specificity for single-stranded DNA probes of specific sequence and that its affinity for a telomeric sequence is comparable to that of transcription factors that bind to specific sites in B DNA.

Structure of the specific DNA-binding site

Base modification and substitution experiments were performed to identify additional structural requirements for factor binding to the single-stranded telomere probe. Deoxyinosine (dI) was substituted for deoxyguanosine (dG) to analyze the role of the exocyclic N2 amino groups in the VET2 probe [Fig. 5]. Substitutions were made in the terminal telomere repeat (VET 3'T), the internal telomere repeat (VET 5'T), or in both telomere repeats (VET 5'/3'T). The avian factor bound to the VET 3'T and the parental VET2 probe, but no specific complex could be detected with the other dI-substituted VET2 probes in an EMSA [Fig. 5]. This experiment revealed that the guanine residues of the telomeric repeats are not equivalent with regard to factor binding and demonstrated that only the exocyclic N2 groups of guanine in the internal telomeric repeat are essential for nucleoprotein complex formation.

A possible interpretation of the asymmetric effects of inosine substitution on factor recognition is that the protein may recognize a specific G-G base-pairing configuration at the core binding site [see Discussion]. To test whether these telomere probes can form a folded structure with a specific G-G geometry, chemical protection experiments were performed on these VET2 probe derivatives in the absence of protein. The N7 positions were reproducibly undermethylated by dimethylsulfate (DMS) in the terminal repeats of the VET2 and VET 3'T probes, but this methylation protection pattern was not observed in the VET2 5'I or the VET 5'/3'I probes [Fig. 6A]. The methylation protection patterns between these probes could be enhanced on the autoradiograph under conditions of more extensive methylation [Fig. 6B]. Because the extents of methylation between dG and dI are very similar in probes VET 5'I and VET 5'/3'I, the differential effects of the chemical protection for VET2 and VET 3'I can be attributed to differences in nucleotide position and not to differences in purine structure. These methylation protection data indicate that the telomere probe requires exocyclic N2 amino groups at the internal repeat to form the higher-order structure. Furthermore, the probes that formed the methylation-resistant structure were also capable of forming specific nucleoprotein complexes [Fig. 5], suggesting that the protected structure may be a key feature in factor recognition.

Methylation interference footprinting of the VET2 nucleoprotein complex was performed to examine directly
the involvement of guanine N7 positions in protein binding. The N7 methylation of any one of five guanine residues in the telomere repeats interfered with factor binding (Fig. 7). Only the methylation of the guanine at the most 3' position of the telomere had no effect. These data provide further evidence that the factor specifically interacts with telomeric sequences within the probe. Interference from the methylation of the 3' telomeric repeat could result from the disruption of the structure identified by the methylation protection experiments described above. Interference from the methylation of the internal telomeric repeat may indicate that the protein makes specific contacts with these positions.

Discussion

Telomeric sequence—protein interaction

Here, we describe a vertebrate factor that binds to telomeric sequences and features of the binding site that are required for specific recognition. This factor formed nucleoprotein complexes with vertebrate (TTAGGG) and
residues that were incapable of Watson-Crick base-pairing at the N7 position.

Figure 7. Methylation interference footprint of the nucleoprotein complex with a telomere probe. The single-stranded probe with two telomeric repeats, VET2, was partially methylated with DMS and used in an EMSA with factor from turkey erythrocytes. The bound (B) and free (F) DNA was excised from the gel, purified, and cleaved with piperidine. The portion of the VET2 nucleotide sequence with the telomeric repeats is indicated, and guanine residues whose methylation interferes with factor binding are shown [●]. The five indicated residues were the only guanines in the VET2 probe that interfered with factor binding following methylation at the N7 position.

Tetrahymena (TTGGG) telomere probes and with other single-stranded probes of specific sequence. High-affinity binding required two or more stretches of dG residues that were incapable of Watson–Crick base-pairing. The importance of these G-rich repeats was demonstrated by binding experiments with the VET series of telomere probes (Fig. 1). A single telomeric repeat of TTAGGG was not sufficient for binding, but nucleoprotein complexes could form with probes containing two or more of the telomeric repeat units. Critical guanine N7 and exocyclic N2 positions within these telomeric repeats were identified by chemical modification and nucleotide substitution experiments (Figs. 5–7).

The telomeric single-stranded DNA-binding proteins have been studied primarily in Oxytricha and other ciliates (Gottschling and Zakian 1986; Price and Cech 1987; Hicke et al. 1990; Price 1990; Gray et al. 1991). In Oxytricha, the single-stranded overhangs of the telomere interact with a heterodimer comprised of 41- and 56-kD subunits. The ciliate heterodimer has properties in common with the vertebrate-binding activity described here, indicating that we have identified a functional telomere-binding protein. Both the vertebrate and the Oxytricha nucleoprotein complexes can be reconstituted with single-stranded DNAs synthesized to the telomeric repeats and to synthetic termini that mimic natural chromosome ends. The dissociation constant of 1.5 × 10^-9 M for the vertebrate nucleoprotein complex is comparable to that of the Oxytricha protein interaction with macronuclear telomerases (~1 × 10^-9 M) (Raghuraman et al. 1989, Raghuraman and Cech 1990), and both activities produce somewhat similar N7 methylation footprints on telomeric sequences. The detection of a protein with these features in vertebrate cell extracts suggests the telomere overhang may be conserved between lower and higher eukaryotes. Prior to this study the vertebrate vimentin protein was reported to bind to telomere probes (Shoeman et al. 1988). Vimentin bound to telomeric sequences containing either a random duplex sequence or a random 3’ overhang with a dissociation constant of 2.5 × 10^-8 M. These properties distinguish vimentin from the telomere-binding activity described here.

Comparing factor-binding sites indicates a core single-stranded consensus sequence of (G)_{n}N_{n}(G)_{n}, where n >2 and N is any nucleotide (Figs. 1 and 2; Santoro et al. 1991). Because this motif occurs frequently, it appears that the specificity for binding relies on the ability of the protein to recognize a unique conformation of nucleic acid. It is possible that this protein is capable of occupying G-rich sites of altered chromatin structure that occur outside of the telomere repeats. Numerous studies indicate that the genome contains small regions of non-B DNA that are hypersensitive to digestion by S1 nuclease. Fine-structure mapping has revealed that many S1 nuclease-sensitive sites occur at tracts of consecutive dG residues, which are common in gene promoters and enhancers (e.g., see Nickol and Felsenfeld 1983; Schon et al. 1983; Yu and Manley 1986; Boles and Hogan 1987; Usdin and Furano 1988). The precise structure of the S1-sensitive conformation is not known, but the prevalence of dG at these sites suggests that the unique base-pairing properties of this nucleotide may be involved. Furthermore, titration experiments (Fig. 4) indicate that this protein is present at a vast molar excess relative to telomere ends and this suggests that other nucleic acid sites may be occupied. We note that clear connections between telomerases and other cellular processes are found in the lower eukaryotes. For example, yeast telomerases reversibly repress the expression of adjacent polymerase II genes (Gottschling et al. 1990), and the RAP1 protein from yeast functions in the control of telomere length, gene transcription, and meiotic recombination (Conrad et al. 1990; Lustig et al. 1990; Kurtz and Shore 1991; White et al. 1991). The multifunctional RAP1 protein binds directly to gene regulatory sequences as well as to the double-stranded yeast telomeric repeat in vitro (Berneman et al. 1986; Buchman et al. 1988). Similarly, proteins specific for the single-stranded termini of the telomere may have other roles at nontelomeric-binding sites.

Features of the specific binding site

G-rich oligonucleotides can form unusual folded structures that are stabilized by non-Watson–Crick base-pairing (Henderson et al. 1987; Williamson et al. 1989; Henderson et al. 1990; Jin et al. 1990; Panyutin et al. 1990; Sen and Gilbert 1990). The data reported here indicate that G-G base pairs between the telomeric repeats
are an important feature for specific factor binding. Four G-G base-pairing schemes are possible (Hobza and Sandorfy 1987) (Fig. 8). In the hairpin duplex with antiparallel strands, configurations I, III, and IV occur between syn and anti guanines, but in configuration II both guanines have the same orientation about the glycosyl bond. Because either guanosine conformation can occur at the terminal or the internal telomere repeat, at least eight structures are possible for the simple hairpin duplex. Information about potential G-G base-pairing at the binding site comes from DMS methylation assays, which test the involvement of the guanosine N7 position, and from dl substitution experiments, which test the involvement of the guanosine exocyclic N2 group. These modifications can block binding either by interfering directly with the protein–nucleic acid contacts or by disrupting a specific nucleic acid structure that is recognized by the protein.

Higher-order DNA structure in the absence of protein was demonstrated by chemical protection assays performed on probes containing only two telomeric repeats. The guanine residues in the terminal telomere repeat were less sensitive to DMS modification, but this pattern of protection was not observed when inosine was substituted for guanosine at the internal repeat (Fig. 6). This pattern of methylation protection provided clues about the hydrogen bonds within the telomere probe by revealing the importance of the N7 groups of the 3'-terminal Gs and the N2 amino groups of the internal Gs in structure formation. Collectively, these data are consistent with the G-G base-pairing configurations I and II because both structures utilize a single N7 and N2 group per base pair. Configurations III and IV can be ruled out because none and two exocyclic N2 groups are required for hydrogen bonding, respectively, and because N7 groups are not involved in either configuration. These data demonstrate further that the guanines at the internal telomeric repeat function as hydrogen bond donors through the N2 group, and the guanines in the terminal telomeric repeat function as hydrogen bond acceptors through the N7 group (Fig. 8). Thus, six of the eight possible hydrogen bond configurations can be ruled out for this structure. These experiments provide a framework for further analyses of the telomeric nucleoprotein complex.

Analyses of the nucleoprotein complex revealed that

Figure 8. Possible schemes for guanine–guanine base-pairing. In the hairpin duplex model, the intramolecular G-G base pairs occur between bases in the syn and anti conformations in configurations I, III, and IV, but configuration II occurs between bases in the same conformation. Chemical modification and nucleotide substitution experiments on telomeric sequences revealed the critical roles of the guanine N7 and N2 positions in the absence of protein and in the nucleoprotein complex. The data presented here are only consistent with G-G base pairs in configurations I or II. The nonequivalent roles of the N7 and N2 groups demonstrate further that the guanines function as hydrogen bond donors at the internal telomeric repeat and hydrogen bond acceptors at the terminal telomeric repeat.
dl substitutions at the terminal repeat of VET2 had no effect on binding, but dl substitutions at the internal repeat completely blocked binding (Fig. 5). The finding that the protein only binds to telomere probes that are protected from DMS methylation in the absence of protein (Fig. 6) provides correlative evidence that the chemically protected structure is recognized by the protein. Additional structural information about the nucleoprotein complex was provided by the methylation interference footprint (Fig. 7). The methylation of any one of five telomeric guanine N7 positions in the VET2 probe interfered with protein binding. On the basis of the G-G pairing hypothesis outlined above, methylation of the 3′- terminal Gs will disrupt binding because the specific nucleic acid structure is prevented from forming. However, the methylation of Gs at the internal repeat is not predicted to disrupt nucleic acid structure (Fig. 8). Interference from these modifications may result from the direct steric inhibition of binding, indicating that the protein occupies a G-G base pair groove.

The nucleic acid structure and protein-binding data presented here are consistent with the model that the telomere probes form a simple hairpin duplex structure that is stabilized, in part, by a specific G-G base pair configuration. However, additional structures have been proposed for telomeres. G-G base pairs have been detected in the Tetrahymena telomeric repeat by 31P and two-dimensional 1H nuclear magnetic resonance (NMR) spectroscopy (Henderson et al. 1987). The probe used for this NMR study contained four telomeric repeat units and was more complex than the VET2 probe used for most of the experiments described here. Probes with four telomeric repeats can form G-strand structures, which display aberrant electrophoretic mobilities in nondenaturing gels containing monovalent cations. It has been proposed that these G-strand structures are quadruple helices with four guanine residues in a square-planar array around a central cation (Williamson et al. 1989; Panyutin et al. 1990; Sen and Gilbert 1990). These four-stranded structures contain G-G base pairs stabilized by N1-donor/O6-acceptor and N2-donor/N7-acceptor hydrogen bond combinations [as in configuration I of Fig. 8]. Under the conditions of our assays none of the probes exhibited altered mobility in a nondenaturing gel that would be indicative of the quadruple helix structure (A. Gualberto and K. Walsh, unpubl.). We note, however, that hairpin duplexes are probably folding intermediates for the G-quartet structure (Panyutin et al. 1990; Sen and Gilbert 1990), and the binding activity described here may inhibit or facilitate the assembly of G quartets at telomeres by interacting with a folding intermediate.

The base modification and substitution experiments described here for telomeric DNA sequences provide evidence for specific G-G base pair configurations that occur in the absence of protein and within the nucleoprotein complex. Recently, a G-G base pair within an RNA bulge structure was found to be critical for REV protein binding to a viral site (Bartel et al. 1991). These findings provide evidence for a novel group of nucleic acid-binding proteins that function by interacting with noncanonical base-paired structures. Elucidating the recognition properties of these proteins may provide structural information about biologically relevant nucleic acid conformations within these complexes.

Materials and methods

Preparation of factor

The telomere-binding activity was partially purified from breast and thigh muscle of day 12–14 embryonic chickens or from the nuclear extracts of adult turkey erythrocytes. The disruption of 30–50 grams (wet weight) of embryonic tissue and all purification steps were performed in chromatography buffer [10% glycerol, 25 mM Tris (pH 7.5), 1 mM EDTA, 2 mM DTT]. One part dissected tissue was added to one part chromatography buffer with 100 mM NaCl and 1 μg/ml of the protease inhibitors leupeptin, aprotinin, chymostatin, and pepstatin (Boehringer Mannheim). The tissue was typically disrupted by ultrasonic vibration at 4°C using a Branson Sonifier at setting 3–5 with a ½-inch tip. The homogenate was clarified by centrifugation at 25,000g for 30 minutes, and the supernatant was decanted and filtered. The cell extract was applied to a 120-μl heparin-Sepharose column (Pharmacia). The telomere-binding activity was eluted with a step gradient from 550 to 800 mM NaCl. This column was run with a linear flow rate of 0.18 cm/min. The active fractions were pooled and frozen at −80°C. Alternatively, the factor was partially purified from a 250-ml nuclear extract of turkey erythrocyte that was prepared as in Evans et al. (1988). The activity was purified 31-fold from this nuclear extract by sequential chromatography through columns of DEAE-Sepharose (Sigma) and heparin-Sepharose. The nuclear extract was applied to a 180-ml DEAE-Sepharose column at 150 mM NaCl, and the telomere-binding activity was eluted by a step to 400 mM NaCl. These active fractions were applied to a 30-ml heparin-Sepharose column, and the telomere-binding activity was eluted with a step gradient from 500 to 850 mM NaCl.

DNA-binding assay

EMSAs for the nucleic acid competition experiments were typically performed with 0.02–0.6 ng of 32P-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). Binding reactions also 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[d(I-C)] · poly[d[I-C]] in 10 μl. Binding reactions were initiated by the addition of protein. Following a 5 to 10 min incubation at room temperature, mixing was loaded on a 5% polyacrylamide gel and electrophoresis was performed at 20 V/cm in 22 mM Tris-borate buffer with 0.5 mM EDTA. No changes in the extent of complex formation were detected by longer incubations prior to the electrophoresis step. Gels were dried and exposed to film for 6–24 hr at −70°C with an intensifying screen. Probes used in the binding assays were labeled with polynucleotide kinase and [γ-32P]ATP (>4500 Ci/mmole, Amersham). Probes used for the binding constant determinations were purified with Etüti-p columns [Schleicher & Schuell] according to the directions of the manufacturer. DNA oligonucleotides were prepared with an Applied Biosystems DNA synthesizer using the phosphoramidite method and purified by denaturing polyacrylamide gel electrophoresis and Sep-Pak C18 cartridges (Waters Associates). Gene regulatory sequence probes have been described previously [Santoro et al. 1991].
Methylation interference footprinting was performed on the nucleoprotein complex formed with the VET2 probe. The 5’ phosphorylated VET2 oligonucleotide was partially methylated in 0.2 ml of 50 mM sodium cacodylate (pH 6.5) and 1 mM EDTA, with 1 µl of fresh DMS for 5 min at room temperature. The reaction was stopped by the addition of 50 µl of 1.5 M sodium acetate (pH 6.5) and 9 µl 2-mercaptoethanol. The modified DNA was used as a probe in an EMS with a partially purified factor. The nucleoprotein complex and free DNA bands were excised from the 5% nondenaturing gel and eluted into 20 mM Tris, 1 mM EDTA, and 200 mM NaCl. This material was purified with Elutip-d columns and ethanol precipitated. Specific cleavage at the modified guanine residues was performed in 100 µl of 1 M piperidine at 90°C for 30 min. The samples were dried under vacuum, dissolved in deionized formamide, and loaded onto a denaturing 20% polyacrylamide gel. Following electrophoresis the gel was exposed to film overnight at 4°C.

Chemical protection assays
Telomere DNA probes were methylated by incubation with DMS in the absence of protein. Reactions were usually performed in the standard DNA-binding buffer (see above), and DMS levels were varied from 1 µl in a 20-µl reaction mixture, formed in the standard DNA-binding buffer (see above), and Telomere DNA probes were methylated by incubation with from the 5% nondenaturing gel and eluted into 20 mM Tris, 1 mM EDTA, and 200 mM NaCl. This material was purified with Elutip-d columns and ethanol precipitated. Specific cleavage at the modified guanine residues was performed in 100 µl of 1 M piperidine at 90°C for 30 min. The samples were dried under vacuum, dissolved in deionized formamide, and loaded onto a denaturing 20% polyacrylamide gel. Following electrophoresis the gel was exposed to film overnight at 4°C.

Estimations of binding affinities
Dissociation constants for the specific nucleoprotein complexes were determined by the EMSA method. Constant amounts of partially purified fraction from embryonic chicken skeletal muscle or adult turkey erythrocytes were incubated with increasing amounts of specific DNA sequence for 5–10 min at 25°C prior to gel electrophoresis. After electrophoresis and autoradiography, the regions corresponding to the protein–NA complex and the free DNA were excised from the gel and quantified by liquid scintillation counting. In these measurements the signal was linearly proportional to the input radioactivity for the range of probe concentrations studied (700–35,000 cpm). The developed autoradiographs were also scanned with an LKB densitometer to determine peak areas. The results of the densitometry were generally in agreement with those obtained by liquid scintillation counting, but the signal deviated from linearity at low probe concentrations. Competition experiments were carried out with a fixed concentration of MREnc noncoding probe and varying levels of nonlabeled competitor. Different competitors were tested for their ability to reduce the intensity of the specific protein–32P-labeled DNA complex by 50%. Relative affinities were determined by comparing the test competitor with the nonlabeled, noncoding strand of the MREnc in side-by-side experiments. The single-stranded control DNA was synthesized to a nonrelated sequence.

Acknowledgments
This work was supported by grants from the National Institutes of Health [AR40197], the American Heart Association, and the Mathers Foundation to K.W. A.G. is a fellow of the Ministerio de Educacion y Ciencia, Spain. K.W. is an established investigator of the American Heart Association. The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References
Bartel, D.P., M.L. Zapp, M.R. Green, and J.W. Szostak. 1991. HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell 67: 529–536.
Berman, J., C.Y. Tachibana, and B.-K. Tye. 1986. Identification of a telomere-binding activity from yeast. Proc. Natl. Acad. Sci. 86: 3713–3717.
Blackburn, E.H. 1991. Structure and function of telomerases. Nature 350: 569–573.
Boles, C. and M. Hogan. 1987. DNA structure equilibria in the human c-myc gene. Biochemistry 26: 367–376.
Buchman, A.R., N.F. Lee, and R.D. Kornberg. 1988. Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. Mol. Cell. Biol. 8: 5086–5099.
Champoux, J.J. and R.K. Williams. 1986. Single-stranded DNA binding proteins required for DNA replication. Annu. Rev. Biochem. 55: 103–136.
Conrad, M.N., J.H. Wright, A.J. Wolf, and V.A. Zakian. 1990. RAP1 protein interacts with yeast telomeres in vivo: Overproduction alters telomere structure and decreases chromosome stability. Cell 63: 739–750.
Edelmann, W., B. Kroger, M. Goller, and I. Horak. 1989. A recruitment of DNA to the LTR of a mouse retrotransposon identified in an in vitro system. Cell 57: 937–946.
Evans, T., M. Reitman, and G. Felsenfeld. 1988. An erythrocyte-specific DNA binding factor recognizes a regulatory sequence common to all chicken globin genes. Proc. Natl. Acad. Sci. 85: 5976–5980.
Gaillard, C. and F. Strauss. 1990. Sequence specific single-strand-binding protein for the simian virus 40 early promoter stimulates transcription in vitro. J. Mol.Biol. 215: 245–255.
Gottschling, D.E. and V.A. Zakian. 1986. Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA. Cell 47: 195–205.
Gottschling, D.E., O.A. Aparicio, B.L. Billington, and V.A. Zakian. 1990. Position effect at S. cerevisiae telomeres: Reversible repression of Pol II transcription. Cell 63: 751–762.
Gray, J.T., D.W. Clander, C.M. Price, and T.R. Cech. 1991. Cloning and expression of genes for the Oxytricha telomer-binding protein: Specific subunit interactions in the telomeric complex. Cell 67: 807–814.
Greider, C.W. 1990. Telomeres, telomerase and senescence. BioEssays 12: 363–369.
Henderson, E., C.C. Hardin, S.K. Walk, I. Tinoco Jr., and E.H. Blackburn. 1987. Telomeric DNA oligonucleotides from novel intramolecular structures containing guanine-guanine base pairs. Cell 51: 899–908.
Henderson, E., M. Moore, and B.A. Malcolm. 1990. Telomere G-strand structure and function analyzed by chemical protection, base analogue substitution, and utilization by telomerase in vitro. Biochemistry 29: 732–737.
Herget, T., M. Bubra, M. Schmoll, K. Zimmermann, and A. Starzinski-Powitz. 1989. Regulated expression of nuclear protein[s] in myogenic cells that binds to a conserved 3’ untranslated region in Pro[II] collagen cDNA. Mol. Cell. Biol. 9: 2828–2836.
Hicke, B.J., D.W. Clander, G.H. MacDonald, C.M. Price, and T.R. Cech. 1990. Two versions of the gene encoding the
41-kilodalton subunit of the telomere binding protein of *Oxytricha nova*. *Proc. Natl. Acad. Sci.* 87: 1481–1485.

Hobza, P. and C. Sandorfy. 1987. Nonempirical calculations on all the 29 possible DNA base pairs. *J. Am. Chem. Soc.* 109: 1302–1307.

Hofmann, F.-X. and S.M. Gasser. 1991. Identification and purification of a protein that binds the yeast ARS consensus sequence. *Cell* 64: 951–960.

Jin, B., K.I. Breslauer, B.A. Jones, and B.L. Graffney. 1990. Tetraplex formation of a guanine-containing nonamer DNA fragment. *Science* 250: 543–546.

Jost, J.-P., H. Saluz, J. Irticny, and B. Moncharmont. 1987. Estradiol-dependent trans-acting factor binds preferentially to a dyad-symmetry structure within the third intron of the avian vitellogenin gene. *I. Cell. Biochem.* 35: 69–82.

Kohwi-Shigematsu, T., R. Gelinans, and H. Weintraub. 1983. Detection of altered DNA conformations at specific sites in chromatin and supercoiled DNA. *Proc. Natl. Acad. Sci.* 80: 4389–4393.

Kurtz, S. and D. Shore. 1991. RAP1 protein activates and silences transcription of mating-type genes in yeast. *Genes & Dev.* 5: 616–628.

Lamming, D.A. and A.C. Notides. 1989. Estrogen receptor selectively binds to the "coding strand" of an estrogen-responsive element. *Proc. Natl. Acad. Sci.* 86: 863–867.

Larsen, A. and H. Weintraub. 1982. An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. *Cell* 29: 609–622.

Lustig, A.I., S. Kurtz, and D. Shore. 1990. Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science* 250: 549–553.

Nickol, J.M. and G. Felsenfeld. 1983. DNA conformation at the 5' end of the chicken adult β-globin gene. *Cell* 35: 467–477.

Ollis, D.L. and S.W. White. 1987. Structural basis of protein-nucleic acid interactions. *Chem. Rev.* 87: 981–995.

Panyutin, I.G., O.I. Kovalsky, E.I. Budowsky, R.E. Dickerson, M.E. Rikhirev, and A.A. Lipanov. 1990. G-DNA: A twice folded DNA structure adopted by single-stranded oligo(dG) and its implications for telomeres. *Proc. Natl. Acad. Sci.* 87: 867–870.

Price, C.M. 1990. Telomere structure in *Euplotes crassus*: Characterization of DNA-protein interactions and isolation of a telomere-binding protein. *Mol. Cell. Biol.* 10: 3421–3431.

Price, C.M. and T.R. Cech. 1987. Telomeric DNA-protein interactions of *Oxytricha* macronuclear DNA. *Genes & Dev.* 1: 783–793.

Raghuraman, M.K. and T.R. Cech. 1990. Effect of monovalent cation-induced telomeric DNA structure on the binding of *Oxytricha* telomeric protein. *Nucleic Acids Res.* 18: 4543–4552.

Raghuraman, M.K., C.I. Dunn, B.I. Hicke, and T.R. Cech. 1989. *Oxytricha* telomeric nucleoprotein complexes reconstituted with synthetic DNA. *Nucleic Acids Res.* 17: 4235–4253.

Rajavashisth, T.B., A.K. Taylor, A. Andalibi, K.L. Svenson, and A.J. Lusis. 1989. Identification of a zinc-finger protein that binds to the sterol regulatory element. *Science* 245: 640–643.

Santoro, I.M., T.-M. Yi, and K. Walsh. 1991. Identification of single-stranded-DNA-binding proteins that interact with muscle gene elements. *Mol. Cell. Biol.* 11: 1944–1953.

Schon, E., T. Evans, J. Welsh, and A. Efstratiadis. 1983. Conformation of promoter DNA: Fine mapping of S1-hypersensitive sites. *Cell* 35: 837–848.

Sen, D. and W. Gilbert. 1990. A sodium-potassium switch in the formation of four-stranded G-DNA. *Nature* 344: 410–414.

Shoeman, R.L., S. Wadle, A. Scherbarth, and P. Traub. 1988. The binding in vitro of the intermediate filament protein vimentin to synthetic oligonucleotides containing telomere sequences. *J. Biol. Chem.* 263: 18744–18749.

Usdin, K. and A.V. Furano. 1988. Rat L ([long interspersed repeated DNA] elements contain guanine-rich homopurine sequences that induce unpairing of contiguous duplex DNA. *Proc. Natl. Acad. Sci.* 85: 4416–4420.

White, M.A., M. Wierdl, P. Detloff, and T.D. Petes. 1991. DNA-binding protein RAP1 stimulates meiotic recombination at the HIS4 locus in yeast. *Proc. Natl. Acad. Sci.* 88: 9755–9759.

Wilkinson, W.O., H.Y. Min, K.P. Claffey, B.L. Satterberg, and B.M. Spiegelman. 1990. Control of the adipin gene in adipocyte differentiation: Identification of distinct nuclear factors for binding to single- and double-stranded DNA. *J. Biol. Chem.* 265: 477–482.

Williamson, J.R., M.K. Raghuraman, and T.R. Cech. 1989. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell* 59: 871–880.

Yu, Y.-T. and J.L. Manley. 1986. Structure and function of the S1 nuclease-sensitive site in the adenovirus late promoter. *Cell* 45: 743–751.

Zakian, V.A. 1989. Structure and function of telomeres. *Annu. Rev. Genet.* 23: 579–604.
Nucleic acid specificity of a vertebrate telomere-binding protein: evidence for G-G base pair recognition at the core-binding site.

A Gualberto, R M Patrick and K Walsh

*Genes Dev.* 1992, *6*:
Access the most recent version at doi:10.1101/gad.6.5.815