Nucleic Acids Research, 2005, Vol. 33, No. 9 ***2887–2900***
doi:10.1093/nar/gki606

Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes

Anat Yafe, Shulamit Etzioni, Pnina Weisman-Shomer and Michael Fry*

Department of Biochemistry, Rappaport Faculty of Medicine, Technion–Israel Institute of Technology, PO Box 9649 Bat Galim, Haifa 31096, Israel

Received March 1, 2005; Revised April 4, 2005; Accepted May 2, 2005

ABSTRACT

Clustered guanine residues in DNA readily generate hairpin or a variety of tetrahelical structures. The myogenic determination protein MyoD was reported to bind to a tetrahelical structure of guanine-rich enhancer sequence of muscle creatine kinase (*MCK*) more tightly than to its target E-box motif [K. Walsh and A. Gualberto (1992) *J. Biol. Chem.*, 267, 13714–13718], suggesting that tetraplex structures of regulatory sequences of muscle-specific genes could contribute to transcriptional regulation. In the current study we show that promoter or enhancer sequences of various muscle-specific genes display a disproportionately high incidence of guanine clusters. The sequences derived from the guanine-rich promoter or enhancer regions of three muscle-specific genes, human sarcomeric mitochondrial creatine kinase (*sMtCK*), mouse *MCK* and α*7 integrin* formed diverse secondary structures. The *sMtCK* sequence folded into a hairpin structure; the α*7 integrin* oligonucleotide generated a unimolecular tetraplex; and sequences from all three genes associated to generate bimolecular tetraplexes. Furthermore, two neighboring non-contiguous guanine-rich tracts in the α*7 integrin* promoter region also paired to form a tetraplex structure. We also show that homodimeric MyoD bound bimolecular tetraplex structures of muscle-specific regulatory sequences more efficiently than its target E-box motif. These results are consistent with a role of tetrahelical structures of DNA in the regulation of muscle-specific gene expression.

INTRODUCTION

Clusters of contiguous guanine residues in DNA can associate *in vitro* under physiological-like conditions to form four-stranded structures designated DNA tetraplexes or quadruplexes. At the core of these structures are Hoogsteen hydrogen-bonded, cation coordinated stacked guanine quartets [for reviews see (1,2)]. Tetrahelical DNA species are grouped into three major classes according to the stoichiometry of the DNA strands: monomolecular tetraplexes, bimolecular tetraplexes and G4 four-molecular tetraplexes. In addition, different types of tetraplex DNA are also distinguished by the orientation of their strands. The DNA strands of monomolecular and bimolecular tetraplexes may be positioned in antiparallel (1,2) or parallel (3) orientation, whereas the four strands of G4 DNA are oriented parallel to one another. The various types of quadruplex DNA are in addition set apart by parameters, such as the molecular geometry of the tetrahelix, their glycosidic torsion angles, participation of non-guanine nucleotides in tetrad formation, type of the coordinating cation and the base composition of spacer stretches that separate guanine clusters (1,2).

Although the formation of DNA tetraplexes *in vitro* is well established, direct indications for their existence *in vivo* are just beginning to appear (4–6). At the same time, several indirect lines of evidence point to the existence of tetrahelical structures in genomic DNA and to their potential participation in various physiological and pathological processes (2,7). Tetraplex structures that might be formed by biologically important guanine-rich DNA sequences were implicated in the execution of specific roles *in vivo*. For example, tetrahelices formed by the telomeric G-strand were proposed to contribute to the regulation of telomere extension and to the protection of chromosome ends (8,9). Furthermore, the pairing of homologous chromosomes was purported to be mediated by the transient formation of interchromosomal tetraplex structures (10,11). In another case, folding of the d(CGG) trinucleotide

*To whom correspondence should be addressed. Tel: +972 4 829 5328; Fax: +972 4 851 0735; Email: mickey@tx.technion.ac.il

repeat sequence in the *FMR-1* gene into tetraplex structures was suggested to cause polymerase pausing and slippage that result in the expansion of the repeat sequence and in the silencing of the *FMR-1* gene, setting off the fragile X syndrome (12,13). Arguments for the existence of tetraplex DNA and RNA structures *in vivo* are supported by the identification of multiple cellular proteins that interact with tetrahelical nucleic acids. Proteins from diverse organisms bind tetraplex DNA preferentially and at high affinity (14–20). Other proteins process or modulate the structure of tetraplex DNA. Such are the nucleases in fission yeast (21,22), mouse (23) and human cells (24) that specifically hydrolyze DNA (19,20,22) and RNA (21) next to quadruplex structures. Other proteins modify the equilibrium between single-stranded DNA and tetraplex guanine-rich DNA. For instance, the β-subunit of an *Oxytricha* telomere binding protein (25,26) and the yeast RAP1 protein (27,28) enhance the formation of tetraplex structures by the guanine-rich strand of telomeric DNA. Conversely, yeast and human helicases of the RecQ family (29–32) as well as members of the hnRNP family (33,34) and other proteins (35) were shown to unwind or destabilize tetraplex structures of guanine-rich sequences in DNA or RNA.

The myogenic determination protein MyoD was reported to bind a tetrahelical structure of a guanine-rich tract from muscle creatine kinase (*MCK*) enhancer sequence more tightly than its target E-box motif (36). Tetraplex structures of guanine-rich stretches in regions upstream to genes, such as *c-MYC* (37) and insulin (38), were implicated in the regulation of their transcription. In analogy, it is possible that the preferential binding of MyoD to tetraplex structures of regulatory DNA sequences modulates the expression of muscle-specific genes. We thus examined in this work, the formation *in vitro* of secondary structures of guanine-rich DNA sequences derived from enhancer or promoter regions of muscle-specific genes and studied their properties. We show that guanine clusters in regulatory sequences of several muscle-specific genes readily form hairpin and monomolecular or bimolecular tetraplex structures. We also report that MyoD homodimers bind to bimolecular tetraplex structures of muscle-specific gene regulatory sequences more efficiently than to their E-box target motif.

MATERIALS AND METHODS

Preparation of hairpin, and monomolecular and bimolecular tetraplex DNA structures

Synthetic DNA oligonucleotides listed in Table 1 were supplied by Genosys and purified by denaturing gel electrophoresis in 8.0 M urea, 12% polyacrylamide (acryl/bisacrylamide, 19:1) (29). The single-stranded oligomers were labeled with ^{32}P at their 5′ ends in bacteriophage T4 polynucleotide kinase-catalyzed reaction. Hairpin structures of the DNA oligomers were formed by the boiling and rapid cooling on ice of solutions of 2.0 ng/μl end-labeled single-stranded oligonucleotides in TE buffer (1.0 mM EDTA in 10 mM Tris–HCl, pH 8.0). Monomolecular tetraplex structure of guanine-rich DNA oligomer was formed as described for hairpin DNA except that the oligomer was boiled and rapidly cooled in TE buffer that contained 50 mM KCl. Bimolecular tetraplex DNA structures were generated by the incubation of 0.2 or 0.6 μg/μl 5′-^{32}P-labeled oligomer in 300 mM KCl in TE buffer at 37°C for 16–20 h.

Table 1. DNA oligomers used in this work

Oligomer	Sequence
sMtCK	5′-d(CTGA**GGA**G**GGG**CT**GGA**G**GG**ACCAC)-3′
5′-Tail *sMtCK*	5′-d(CGTGACGCTGA**GGA**G**GGG**CT**GG**A-**GGG**ACCAC)-3′
Integrin 26	5′-d(CAT**GGGGG**C**GGG**AA**GGGG**C**GGGG**TCT)-3′
Integrin 26-I	5′-d(CAT**GGGI**GC**GI**GAA**GG**IIC**GI**G**TCT)-3′
3′-Tail Integrin 26	5′-(CAT**GGGGG**C**GGG**AA**GGGG**C**GGGG**TCTCGT-**GG**ACTC)-3′
Integrin 29	5′-d(AAA**GGT**G**GGGG**C**GG**CA**GGG**CGCAA**GG**-CAAT)-3′
MCK	5′-(TCC**GGA**G**GGGG**CA**GG**CTGA**GGG**C**GG**C)-3′
3′-Tail MCK	5′-d(TCC**GGA**G**GGGG**CA**GG**CTGA**GGG**C**GG**-CCACTTCCCTCAGC)-3′
18mer	5′-d(CCCTAAAAAAAAGAATTC)-3′
24mer 1	5′-d(CCCTAACCCTAACCCTAACCCTAA)-3′
24mer 2	5′-d(CCCTTACCCTTACCCTTACCCTAA)-3′
26mer	5′-d(CGGGATCCATGTTGGGCGACCTCACC)-3′
30mer	5′-d(GGGGTACCAGACATAAAAATTTTTTGCTTG)-3′

Clusters of contiguous guanine residues are marked in boldface and underlined.

Non-denaturing gel electrophoresis of DNA

The preparation of polyacrylamide gels, loading of the DNA samples and electrophoresis (39) were conducted at 4°C. Single-stranded DNA and hairpin DNA were resolved by non-denaturing gel electrophoresis in the absence of salt. End-labeled DNA oligomer in H_2O was mixed with an equal volume of buffer D (0.5 mM DTT, 1.0 mM EDTA, 20% glycerol in 25 mM Tris–HCl, pH 8.0) and aliquots of 10 μl were loaded onto 10 or 15% polyacrylamide gel (acryl/bisacrylamide, 19:1) in 0.5× TBE buffer (1.2 mM EDTA in 0.54 mM Tris–borate, pH 8.3). Electrophoresis of the DNA was conducted at 200–250 V until a bromophenol blue marker dye migrated 7.5 cm into the gel. Monomolecular and bimolecular tetraplex DNA structures in 50 or 300 mM KCl in TE buffer, were resolved by non-denaturing gel electrophoresis at 150–180 V in 0.5× TBE buffer, 20 mM KCl. The gels were dried and the relative proportions of bands of the resolved DNA structures were quantified by phosphor imaging analysis.

Methylation-protection analysis of tetraplex DNA structures

Single-stranded DNA oligomers or their monomolecular or bimolecular tetraplex forms were analyzed for the accessibility of their guanine N7 position to methylation by dimethylsulfate (DMS). In a final volume of 10 μl, the reaction mixture contained 0.6 μg of 5′-^{32}P-labeled single-stranded or tetraplex DNA, 50 mM of KCl and 1.0 or 0.75% DMS in TE buffer (pH 8.0). Control mixtures were devoid of DMS. The mixtures were incubated at 20°C for periods of time as specified in Results and the methylation reactions were terminated by the addition of 2 vol of chilled buffer D and rapid cooling of the mixtures on ice. The DNA samples were resolved by non-denaturing electrophoresis through 10% polyacrylamide gel in 0.5× TBE buffer and 20 mM KCl. Gel bands that contained radioactive single-stranded or tetraplex DNA were visualized by autoradiography and excised, and the DNA was

extracted from the gel slices into 100 mM KCl in TE buffer by gentle shaking overnight at 4°C. The DNA was precipitated by ethanol, suspended in 20 µl TE buffer, 50 mM KCl and the mixtures were heated at 90°C for 10 min following the addition of an equal volume of 2.0 M pyrrolidine. The DNA samples were cooled to 4°C, dried by SpeedVac centrifugation and residual pyrrolidine was removed by five washes with 50 µl aliquots of H_2O. Protection of guanine residues against methylation was assessed by the resistance of their phosphodiester bonds to breakage by pyrrolidine as identified by denaturing gel electrophoresis of the DNA in 12–17% polyacrylamide gel and 8.0 M urea.

Circular dichroism spectroscopy

Tetraplex DNA structures of 5′-^{32}P-labeled DNA oligomers were generated as described above. DNA samples were resolved through non-denaturing polyacrylamide gel to ascertain the formation of rapidly migrating or retarded unimolecular or bimolecular tetraplexes. The CD spectra of these tetraplex DNA structures in 10 mM Tris–HCl, pH 8.0 containing 50 mM KCl, were determined at room temperature in 1 cm quartz cuvette using Jasco J-810 spectropolarimeter. The spectra were measured at a band width of 1 nm over a range of 350–220 nm.

Measurement of the effect of cationic porphyrins on the thermal stability of tetraplex DNA

Cationic porphyrins, 5,10,15,20-tetra(*N*-methyl-2-pyridyl)-porphin (TMPyP2), 5,10,15,20-tetra(*N*-methyl-3-pyridyl)-porphin (TMPyP3) and 5,10,15,20-tetra(*N*-methyl-4-pyridyl)-porphin (TMPyP4) were the gift of Dr Laurence H. Hurley (University of Arizona, AZ). The mixture used for the measurement of the effect of each porphyrin on the thermal stability of tetraplex structures of guanine-rich DNA sequences contained in a final volume of 10 µl, 180 fmol of 5′-^{32}P-labeled bimolecular tetraplex DNA, 320 fmol of a specified cationic porphyrin and 20 mM KCl in buffer D. Control mixtures were devoid of porphyrin. The mixtures were incubated for 10 min at specified temperatures, rapidly cooled on ice and the DNA was resolved by non-denaturing gel electrophoresis at 4°C in 10% polyacrylamide, 20 mM KCl, 0.5× TBE buffer. Proportions of the remaining bimolecular tetraplex DNA relative to DNA samples that were kept at 4°C were quantified on the dried gel by phosphor imaging analysis.

Expression and purification of recombinant MyoD

Full-length MyoD protein encoding cDNA in pRK171α vector was PCR amplified using primers that contained EcoR1 or Xho1 restriction sequences at their respective 5′ or 3′ ends. The product molecules of cDNA were purified (QIAquick; Qiagen) and ligated to Xho1 and EcoR1 digested pGEX-6P vector, recombinant plasmids were electroporated into *Escherichia coli XL-1* and the presence of an intact cDNA insert was verified by determination of its nucleotide sequence. To isolate recombinant MyoD protein, pGEX-6P plasmids harboring MyoD cDNA was electroporated into competent *E.coli BL21(DE3)pLysS* cells, the cells were grown to A_{600} of ~0.6 in Luria–Bertani medium that contained ampicillin and chloramphenicol and synthesis of the glutathione *S*-transferase (GST) fused proteins was induced by the addition of 100 µM isopropyl-β-D-thiogalactopyranoside for 3 h. The recombinant protein was purified to >95% homogeneity from the bacterial cell extracts by glutathione–agarose (Sigma) affinity column chromatography. The glutathione *S*-transferase residue was cleaved by incubation of 100 µg fusion protein with 2.0 U preScission protease (Amersham Biosciences) for 4 h at 4°C.

Electrophoretic mobility shift assay of protein binding to DNA

MyoD homodimers were formed by incubating purified recombinant MyoD for 10 min at 37°C in reaction mixtures that contained in a final volume of 10 µl, 45 mM KCl, 4.5 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 20% glycerol, 20 mM Tris–HCl (pH 8.0) and 0.5 µg of HeLa whole cell extract protein prior to their binding to the different DNA probes. Reaction mixtures for the binding of protein to DNA contained in a final volume of 10 µl, specified amounts of MyoD homodimer, 0.2 pmol 5′-^{32}P-labeled DNA probe, 14.5 mM KCl, 0.45 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 20% glycerol in 20 mM Tris–HCl (pH 8.0) and 0.05 µg of HeLa whole cell extract protein. Reaction mixtures for the binding of end-labeled double-stranded E-box DNA also contained 100-fold (w/w) excess of unlabeled poly(dI–dC) (Sigma). Mixtures for the binding of end-labeled bimolecular tetraplex DNA structures of muscle-specific regulatory sequences contained 100-fold (w/w) excess of unlabeled single-stranded oligomer of the same sequence. The mixtures were incubated for 20 min at 30°C and protein–DNA complexes were resolved from free DNA by electrophoresis at 4°C and 200–250 V in non-denaturing 4% polyacrylamide gel (acryl/bisacrylamide, 19:1) in 10 mM KCl, 0.25× TBE buffer. Results of control experiments indicated that bimolecular tetraplex integrin 26 DNA remained fully stable in the presence of 10 mM KCl under the described DNA-binding reaction conditions. Electrophoresis of the DNA was conducted until a bromophenol blue marker dye migrated 7.5 cm into the gel. The gels were dried on DE81 filter paper and the relative proportions of bands of free and protein-bound DNA structures were quantified by phosphor imaging analysis.

RESULTS

Single-stranded DNA sequences that include runs of contiguous guanine residues readily fold into hairpin and tetraplex structures. The muscle-specific master transcription factor MyoD was reported to bind a tetrahelical structure of guanine-rich tract from *MCK* enhancer more tightly than its E-box target motif (36). We undertook, therefore, to examine in detail the formation *in vitro* of secondary structures of DNA sequences derived from enhancer or promoter regions of several muscle-specific genes.

Regulatory sequences of muscle-specific genes contain regions with disproportionate high content of guanine clusters

Initial inspection of nucleotide sequences of regulatory elements of several muscle-specific genes identified guanine–cytosine rich sections in every surveyed gene that contained disproportionate numbers of guanine clusters. Examples of

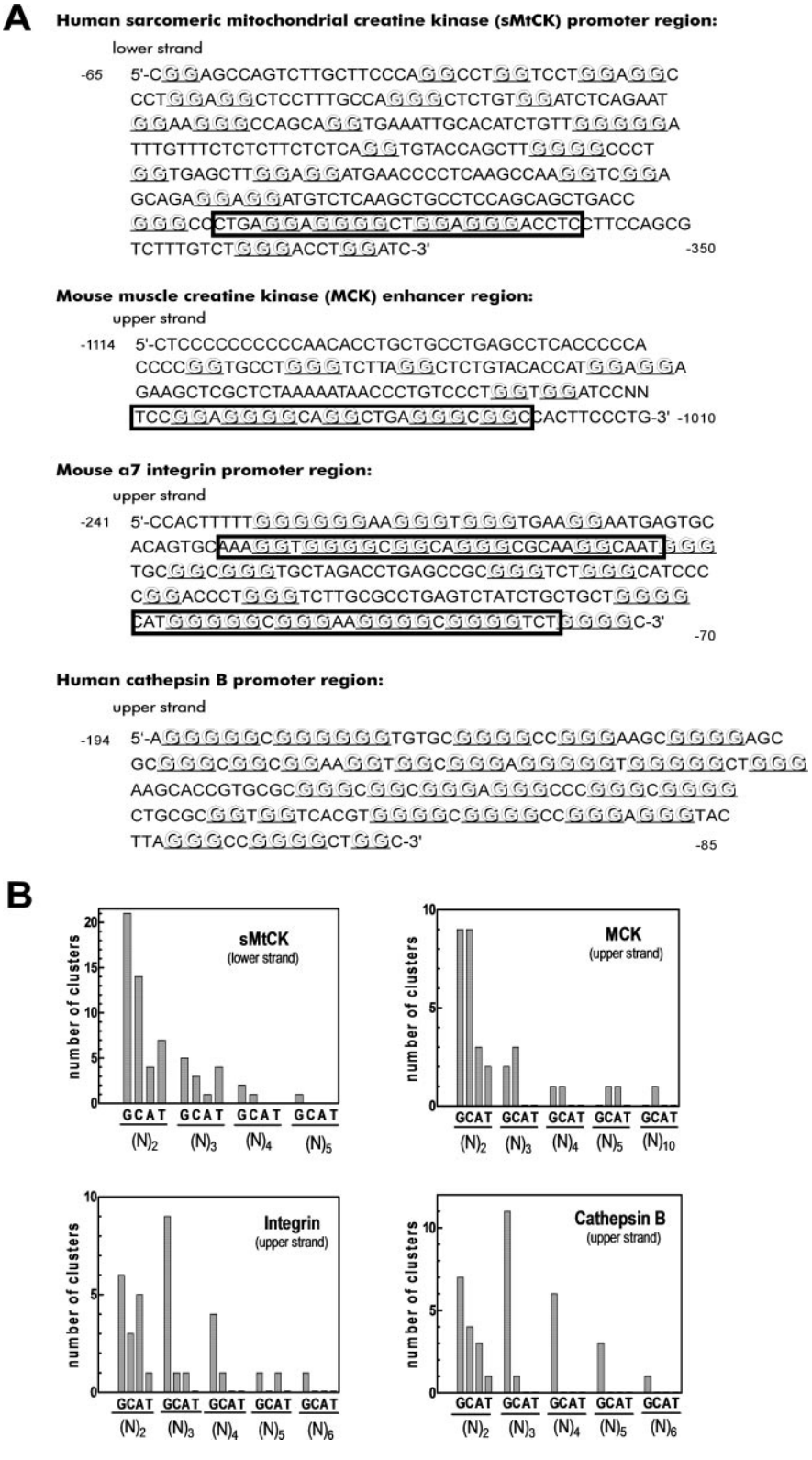

Figure 1. High relative incidence of guanine clusters in regulatory regions of four muscle-specific genes. (**A**) Promoter or enhancer regions of human *sMtCK* (accession no. M72981), mouse *MCK* (accession no. M21390), mouse α*7 integrin* (accession no. U60419) and human *cathepsin B* (accession no. AF086639) have a high proportion of contiguous guanine residues. Clusters of two or more contiguous guanine residues are marked in boldface and underlined. Sequences of the DNA oligomers used in this work (Table 1) are enclosed in boxes. (**B**) Occurrence of clusters of two or more contiguous residues of each nucleotide in the muscle-specific regulatory sequences shown in (A). The number of guanine residues is exclusive for each cluster group such that $N = 3$, for instance, does not include clusters of four or more guanines.

such segments in promoter or enhancer regions of four muscle-specific genes are presented in Figure 1A. The clusters of contiguous guanine residues were prominent in ~100–300 nt long runs along the upper or lower strand of regulatory regions of the muscle-specific mouse α*7 integrin*, *MCK*, human *sMtCK* and *cathepsin B*. Quantification established a clear predominance of contiguous guanine residues over clusters of any one of the other three nucleotides (Figure 1B). In the lower strand of the *sMtCK* promoter region that had a relative GC-content of 61.6%, the ratios of adjacent identical residues $(G)_n$:$(C)_n$:$(A)_n$:$(T)_n$ (n = 2, 3, 4 or 5) were 20:14:4:7, 5:3:1:4, 2:1:0:0 or 1:0:0:0. Similarly, cytosine and guanine clusters were predominant in the upper and lower strands, respectively, of the *MCK* enhancer region which had a GC ratio of 63.0%. Thus, the respective ratios of $(G)_n$:$(C)_n$:$(A)_n$:$(T)_n$ (n = 2, 3, 4, 5 or 10) in the upper strand of *MCK* were 9:9:3:2, 2:3:0:0, 1:1:0:0, 0:1:1:0 or 0:1:0:0. Similarly, the region examined in the upper strand of the *α7 integrin* promoter sequence (relative GC content 63.4%) had respective ratios of $(G)_n$:$(C)_n$:$(A)_n$:$(T)_n$ (n = 2, 3, 4, 5 or 6) of 6:3:5:1, 9:1:1:0, 4:1:0:0, 1:0:1:0 or 1:0:0:0. Finally, the ratios of $(G)_n$:$(C)_n$:$(A)_n$:$(T)_n$ (n = 2, 3, 4, 5 or 6) in an upper strand tract of the *cathepsin B* promoter region (GC ratio 82.2%) were 7:4:3:1, 11:1:0:0, 6:0:0:0, 3:0:0:0 or 1:0:0:0. Notably, enrichment in guanine clusters was restricted to select segments of the muscle-specific promoter regions. Other randomly picked upper strand promoter sequences of *α7 integrin* (positions −1351 to −1151), *MCK* (−600 to −400) or *cathepsin B* (−746 to −546) did not display preference for clusters of guanine or cytosine residues over those of other nucleotides and the number of contiguous guanines per cluster did not exceed three. Furthermore, computer search did not identify regulatory regions of non-muscle genes that displayed substantial homology with the guanine-rich sequences of the examined myogenic genes. Finally, examination of the nucleotide sequences of promoters of select non-muscle genes, such as serum albumin or γ-globin, did not reveal regions with disproportionately high frequency of guanine clusters.

Sequences that have high content of guanine clusters are prone to fold into hairpin or tetrahelical structures. We, thus, examined the formation of secondary structures by selected 24–29 nt long guanine-rich tracts in regulatory regions of the *sMtCK*, *MCK* and α*7 integrin* genes.

sMtCK promoter sequence forms hairpin and bimolecular tetraplex structures

sMtCK is expressed only *in vivo* in striated muscle and is induced during differentiation of muscle cells *in vitro* (40–42). Encoding a muscle-specific energy generating mitochondrial protein, this gene belongs to the global class of muscle-specific proteins. Similar to other muscle-specific genes, the expression of *sMtCK* is closely regulated. Four likely MEF1/MyoD binding motifs were identified at positions −146, −161, −192 and −285 of the mouse *sMtCK* promoter region (43). As demonstrated in Figure 1, the region between positions −65 and −350 of the lower strand of this sequence also possesses a disproportionately high number of clusters of guanine residues.

In order to investigate whether guanine clusters in a promoter-derived tract form secondary structures, we chose the 24 nt long sequence 5′-d(CTGAGGAGGGGCTGGAGGGACCAC)-3′ representing positions −296 to −320 of the *sMtCK* regulatory region (Table 1 and Figure 1A). To assess the capacity of this sequence to fold into a unimolecular secondary structure, a solution of 2.0 ng/μl DNA in water was boiled, rapidly cooled on ice and resolved by salt-free non-denaturing PAGE. The ^{32}P-labeled 24mer *sMtCK* oligonucleotide migrated in the gel almost as rapidly as an 18mer marker oligomer, displaying higher mobility than that of the two 24mer marker oligonucleotides (Figure 2A). However, substitution of its guanine residues by inosines caused this sequence to migrate in the non-denaturing gel true to size. Since the C2 position of inosine lacks an NH_2 group that participates in both canonical Watson–Crick and Hoogsteen hydrogen bonds, these results suggested that the *sMtCK* oligomer folded into hydrogen-bonded compact structure. This structure was formed instantaneously and its rate of accumulation was independent of the concentration of DNA (data not shown). The two likely types of compact structures that could possibly be formed with such apparent zero order kinetics are either unimolecular tetraplex or hairpin. Since a coordinating cation is essential for the formation and stability of tetraplex DNA and since the compact DNA structure was formed in the absence of the cation, it was most likely a hairpin.

In order to examine the capacity of the *sMtCK* sequence to fold into cation-dependent multimolecular secondary structure, 600 ng/μl end-labeled oligomer was incubated at 37°C for 16–20 h in the presence of 300 mM KCl. Resolution of the reaction products by non-denaturing PAGE revealed that a major portion of the single-stranded DNA was converted into a slowly migrating species which was presumed to represent a multimolecular complex (Figure 2B). This structure which was converted back to single-strand by heat denaturation (Figure 2B) was not formed in the absence of KCl (data not shown). Incubation of 3′-tail *sMtCK* oligonucleotide under similar conditions also resulted in the formation of an electrophoretically retarded band. An equimolar mixture of the *sMtCK* and 3′-tail *sMtCK* DNA oligomers that was incubated under the same conditions yielded in addition to the respective slowly migrating forms of each sequence, a third species with an intermediate mobility (Figure 2B). This result was interpreted as reflecting the formation of bimolecular complexes of two identical strands of either the *sMtCK* or 3′-tail *sMtCK* DNA as well as a third complex consisting of one molecule of each oligomer. The salt-dependence of the formation of the complexes and their stoichiometry suggested that they are bimolecular tetraplexes. Such structures are stabilized by non-canonical guanine–guanine Hoogsteen hydrogen bonds. To identify residues that may engage in such hydrogen bonding, we modified isolated bimolecular complexes of *sMtCK* DNA with DMS; the treated DNA was hydrolyzed with pyrrolidine and the reaction products were resolved by denaturing PAGE. The results presented in Figure 2C demonstrated that all the guanine residues in the single-stranded oligomer were modified by DMS and hydrolyzed by pyrrolidine. In contrast, except for the three 3′-terminal guanines, all the residues in the bimolecular complex resisted modification and hydrolysis, indicating their engagement in guanine quartet stabilizing Hoogsteen hydrogen bonds. Thus, two *sMtCK* hairpins appeared to pair in the presence of KCl into Hoogsteen hydrogen-bonded bimolecular tetraplex DNA structures. The

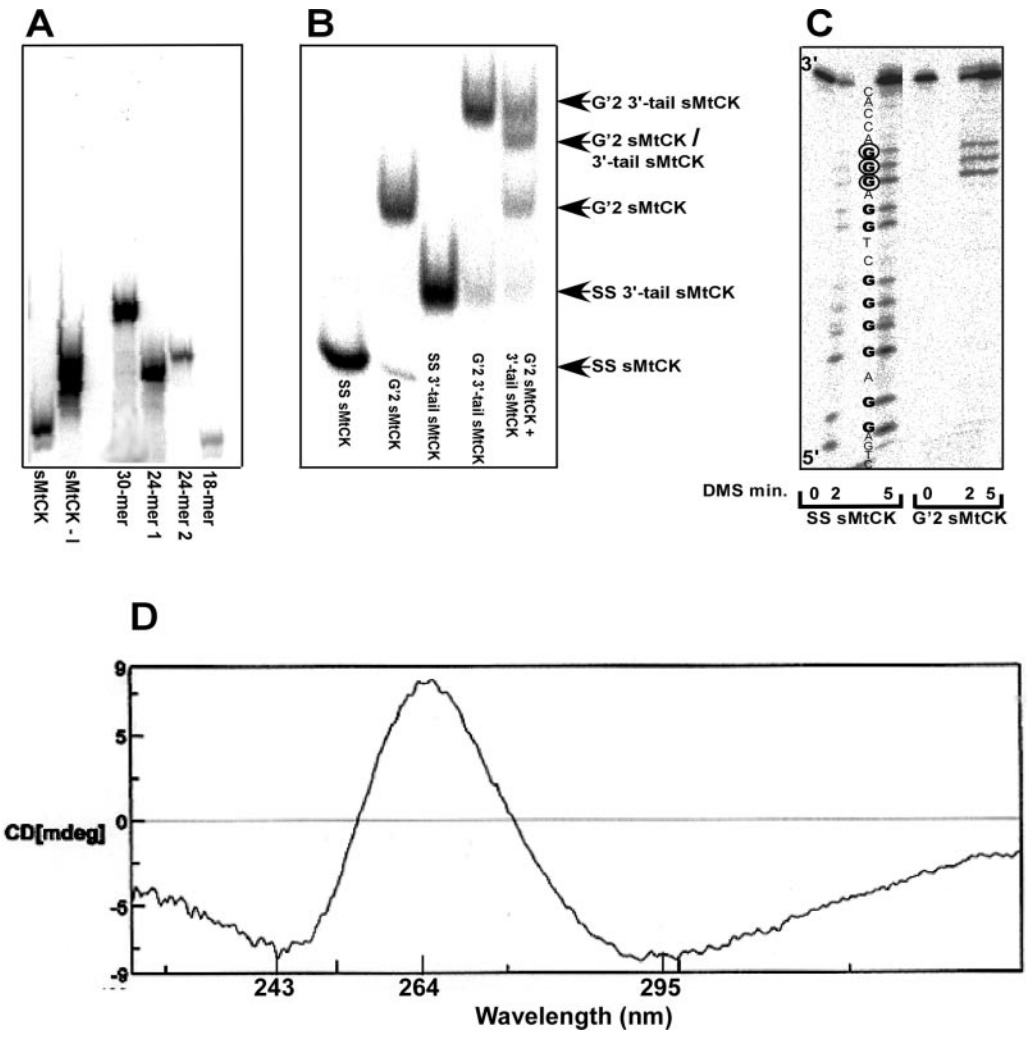

Figure 2. The *sMtCK* oligonucleotide forms hairpin and G′2 bimolecular tetraplex structures. (**A**) *sMtCK* DNA oligomer but not *sMtCK*-I DNA behaves as a compact structure in the absence of KCl. 5′-^{32}P-labeled *sMtCK* DNA or *sMtCK*-I DNA oligomers, each at 0.2 ng/µl in H_2O, were boiled for 5 min, rapidly cooled on ice and resolved by non-denaturing electrophoresis through 15% polyacrylamide gel in 0.5× TBE buffer side-by-side with labeled molecular size marker DNA oligomers (see Materials and Methods). (**B**) *sMtCK* DNA forms bimolecular complexes. An aliquot of 0.6 µg/µl of 5′-^{32}P-labeled *sMtCK* or 3′-tail *sMtCK* DNA oligomer or an equimolar mixture thereof were incubated at 37°C for 16 h in TE buffer, 300 mM KCl and resolved by electrophoresis through a non-denaturing 10% polyacrylamide gel in 0.5× TBE buffer containing 20 mM KCl. Positions of boiled, single-stranded *sMtCK* or 3′-tail *sMtCK* DNA oligomers and of their respective slowly migrating G′2 complexes are indicated. Note the presence of a hybrid band of G′2 *sMtCK*/3′-tail *sMtCK* DNA in the lane containing an equimolar mixture of the two oligomers. (**C**) Patterns of methylation protection of single-stranded and G′2 tetraplex sMtCK DNA. 5′-^{32}P-labeled single-stranded *sMtCK* DNA oligomer or its G′2 bimolecular tetraplex DNA structure were exposed to 1% DMS at 20°C for the specified periods of time. The DMS treated single-stranded and tetraplex DNA structures were hydrolyzed by 2.0 M pyrrolidine as described in Materials and Methods. The position of each nucleotide in the sequence is marked on the phosphor image of the gel and DMS modified guanines in the tetraplex DNA structure are circled. (**D**) CD spectrum of the *sMtCK* bimolecular tetraplex DNA.

strand orientation of this tetraplex structure was assessed by determining its CD spectrum. The results shown in Figure 2D revealed a positive peak at 264 nm and a negative one at 243 nm. This spectrum is characteristic of parallel-stranded tetraplex (44,45). In such DNA formation each guanine residue is *anti* and interacts with three other guanine bases to form cyclic G-quartets via Hoogsteen hydrogen bonds as indicated by the methylation protection data shown in Figure 2C.

A guanine-rich sequence in the *MCK* enhancer forms bimolecular tetraplex structures

The *MCK* gene encodes MCK which serves as a major energy generating protein in the muscle tissue. The expression of this gene is increased during muscle differentiation by more than 1000-fold through the binding of MyoD to its enhancer (46). The enhancer of *MCK* includes a region rich in guanine clusters that may form secondary structures (Figure 1). To examine the capacity of an *MCK* enhancer tract to form DNA secondary structure, we chose the 5′-d(TCCGGAGGGGCAGGCTGA-GGGCGGC)-3′ sequence which represents positions −1045 to −1021 of the upper strand of the *MCK* enhancer region (Table 1 and Figure 1A). Using end-labeled *MCK* and 3′-tail *MCK* oligonucleotides (Table 1) we found that under reaction conditions that allowed the generation of bimolecular *sMtCK* tetraplex structure (Figure 2), two *MCK* or 3′-tail *MCK* DNA molecules or one of each oligomer also paired to form bimolecular tetraplex complexes (data not shown).

An α*7 integrin* promoter sequence folds into both monomolecular and bimolecular tetraplex structures

The superfamily of integrin proteins consists of multiple transmembrane heterodimeric receptors that mediate cell–cell and cell–extracellular matrix interactions (47). Each integrin is composed of non-covalently paired α-subunits and β-subunits. The α7 subunit, which forms a distinct integrin with β1 (48), is expressed mainly in skeletal and cardiac muscles in which it serves as the major laminin-binding integrin. The expression of α*7 integrin* mRNA is upregulated by MyoD during the differentiation of myoblasts to myotubes (49,50). The upper strand of the mouse α*7 integrin* gene includes multiple guanine clusters which can potentially fold into secondary structures (Figure 1). To examine the ability of an α*7 integrin* promoter sequence to form monomolecular secondary structure, we first chose a 26mer sequence termed integrin 26; 5′-d(CATGGGGGCGGGAAGGGGCGGGGTCT)-3′ that represented positions −101 to −75 of the promoter region (Table 1 and Figure 1A). A ^{32}P terminally labeled oligomer at low concentration (2.0 ng/μl) was boiled and rapidly cooled to 4°C with or without 50 mM KCl. DNA oligomers that were treated in the absence or presence of salt were resolved by non-denaturing gel electrophoresis in a buffer devoid of or containing 20 mM KCl. In comparison with the mobility of marker DNA oligomers, the integrin 26 oligomer migrated in the salt-free gel true to its size (Figure 3A). In contrast, integrin 26 DNA that was incubated and resolved in the presence of

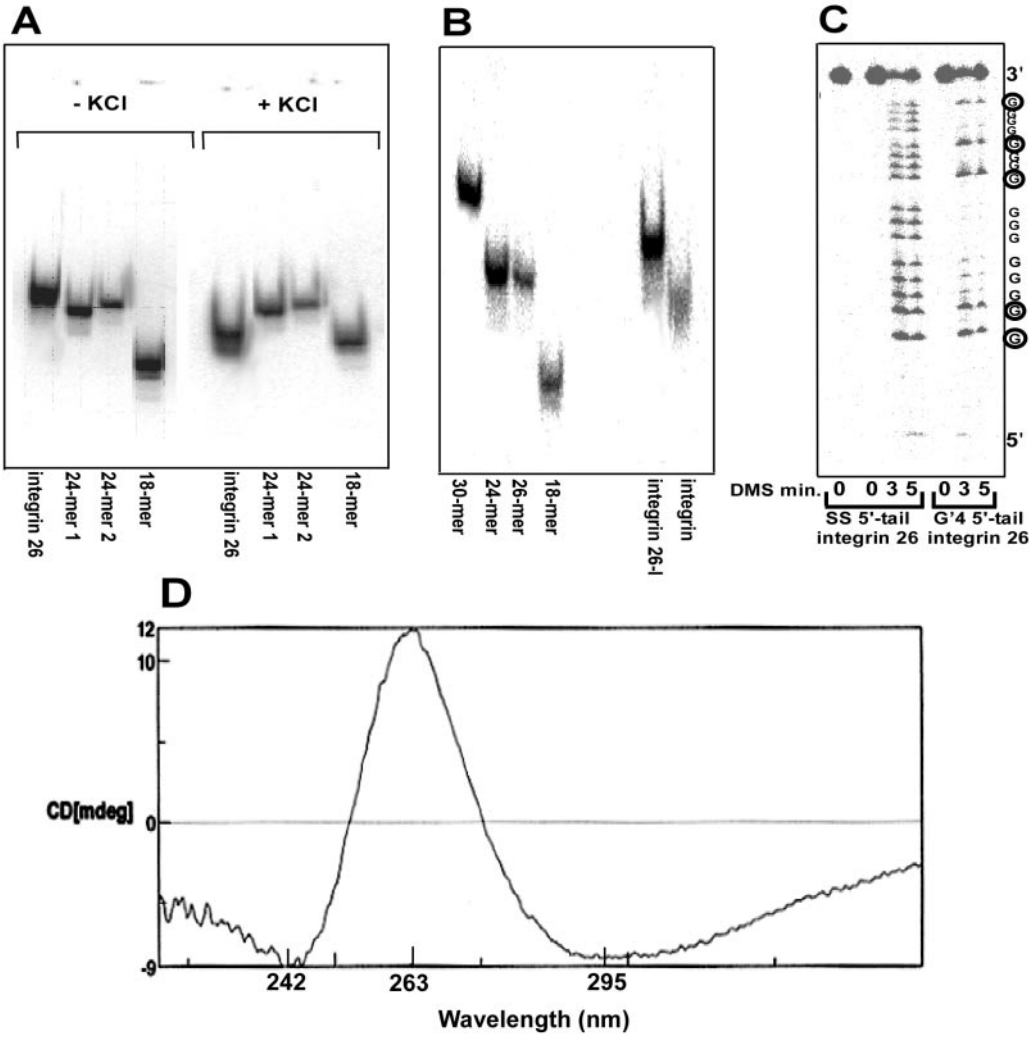

Figure 3. The integrin 26 DNA oligonucleotide folds into a monomolecular tetraplex structure. (**A**) Integrin 26 DNA oligomer behaves as a compact structure in the presence of KCl. 5′-^{32}P-labeled integrin 26 DNA oligomer at 2.0 ng/μl in H_2O was boiled for 5 min in the absence or presence of 50 mM KCl, rapidly cooled on ice and resolved by a non-denaturing 15% PAGE in 0.5× TBE buffer devoid of or containing 20 mM KCl, as the case be (see Materials and Methods). Shown is a composite of phosphor images of two gels one without and the other containing KCl. (**B**) Substitution of the guanine residues in integrin 26 DNA by inosines prevents its folding into a compact form. 5′-^{32}P-labeled integrin 26 DNA or integrin 26-I DNA oligomers at 2.0 ng/μl each were boiled in the presence of 50 mM KCl, rapidly cooled and resolved by non-denaturing 15% PAGE in 0.5× TBE buffer, 20 mM KCl, side-by-side with 5′-^{32}P-labeled molecular size marker DNA oligomers. (**C**) Patterns of methylation protection of single-stranded and monomolecular tetraplex integrin DNA. 5′-^{32}P-labeled single-stranded integrin 26 DNA oligomer or its monomolecular tetraplex form were prepared without or with 50 mM KCl, respectively, as described above. The DNA samples were exposed to 1% DMS and hydrolyzed by 2.0 M pyrrolidine (see Materials and Methods). Shown is a phosphor image of DNA oligomers resolved by electrophoresis in denaturing 12% polyacrylamide gel and 8.0 M urea. The position of each guanine in the sequence is marked in the phosphor image of the gel and DMS modified, pyrrolidine hydrolyzed residues are circled. (**D**) CD spectrum of integrin 26 monomolecular tetraplex.

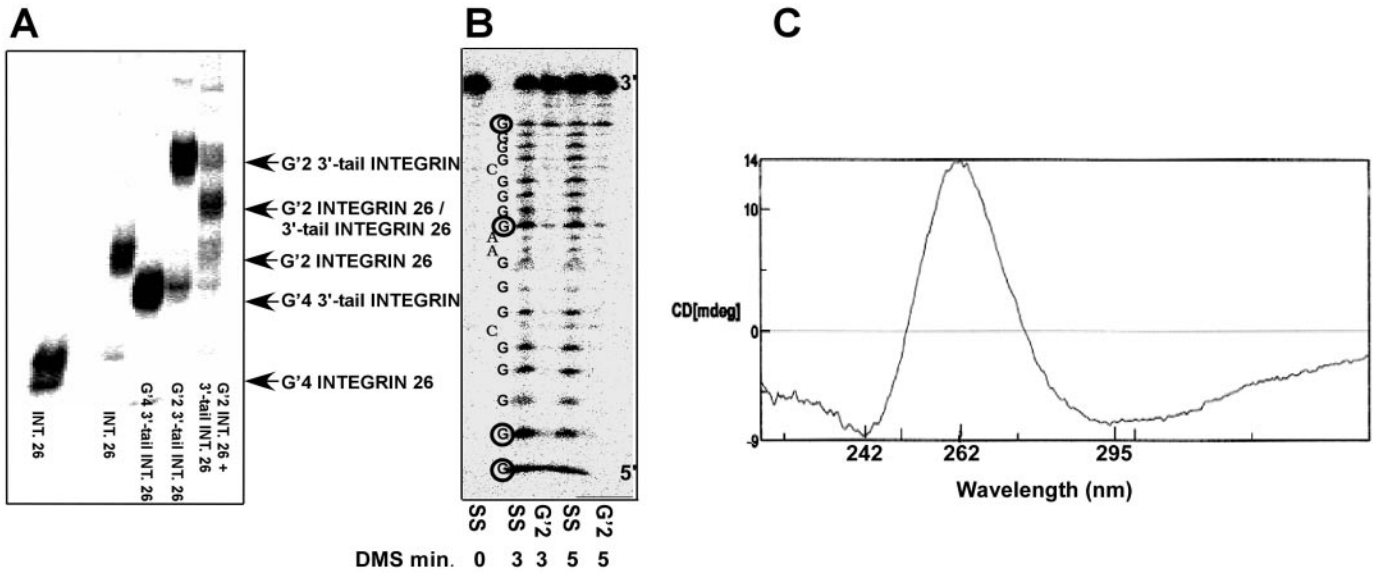

Figure 4. The integrin 26 DNA oligonucleotide forms a G′2 bimolecular tetraplex structure. (**A**) Formation and stoichiometry of bimolecular complexes of integrin 26 DNA. 5′-^{32}P-labeled integrin 26 or 3′-tail integrin 26 DNA oligomers or an equimolar mixture thereof at a final concentration of 0.6 μg/μl were incubated at 37°C for 18 h in TE buffer containing 300 mM KCl and resolved by a non-denaturing 10% PAGE in 0.5× TBE buffer containing 20 mM KCl. Positions of boiled integrin 26 DNA or 3′-tail integrin 26 DNA and their respective slowly migrating G′2 complexes are indicated. Notice the presence of a hybrid band of G′2 integrin 26 DNA/3′-tail integrin 26 DNA in the lane containing an equimolar mixture of the two oligomers. (**B**) Patterns of methylation protection of single-stranded and G′2 tetraplex integrin 26 DNA. 5′-^{32}P-labeled single-stranded integrin 26 DNA oligomer and its G′2 bimolecular tetraplex form that were prepared as described above were exposed to 1% DMS at 20°C for the specified periods of time, and hydrolyzed by 2.0 M pyrrolidine as described in Materials and Methods and in the legend to Figure 3). DMS modified, pyrrolidine hydrolyzed guanine residues in the tetraplex DNA structure are circled. (**C**) CD spectrum of G′2 tetraplex integrin 26.

KCl migrated ahead of 24mer marker sequences, suggesting that it assumed a compact structure (Figure 3A). However, the gel mobility of integrin 26 DNA in which six guanine residues were substituted by inosines (integrin 26-I oligomer, Table 1) was true to size (Figure 3B). The failure of the integrin 26-I oligomer to form a rapidly migrating structure suggested that the folding of the integrin 26 oligonucleotide required hydrogen bonding. In addition, the apparent zero order kinetics of the generation of the compact structure of the integrin 26 DNA sequence, as reflected by its time and DNA concentration independent kinetics (data not shown), indicated that it was a unimolecular entity and the dependence of its formation on K^+ ions suggested that it represented tetraplex rather than hairpin structure (51,52). To verify the tetrahelical nature of the folded structure, the pattern of its resistance to DMS methylation was compared with that of the single-stranded oligomer. Specific residues in the folded structure were protected against DMS modification, whereas every guanine in the single-stranded oligomer was modified by DMS and was consequently hydrolyzed by pyrrolidine (Figure 3C). The unmodified residues were most probably involved in guanine–guanine Hoogsteen hydrogen bonding and in the formation of guanine quartets. The CD spectrum of this tetraplex showed a strong positive maximum at 263 nm and a negative peak at 242 nm (Figure 3D), indicating that it was a parallel-stranded tetrahelix (44,45). All in all, these results indicated that the examined guanine-rich α7 *integrin* promoter sequence readily folded into K^+ dependent, Hoogsteen hydrogen-bonded, parallel-stranded, intramolecular tetraplex structure.

We next inquired whether the integrin 26 DNA oligomer was also capable of forming an intermolecular tetraplex complex. Solutions containing ^{32}P-labeled integrin 26 DNA or 3′-tail integrin 26 oligomer (Table 1) or an equimolar mixture thereof, at a final concentration of 600 ng/μl were incubated at 37°C for 16–20 h in the presence of 300 mM KCl. Non-denaturing PAGE of the various DNA mixtures revealed that each monomolecular tetraplex structure was largely converted under these reaction conditions into species that exhibited retarded mobility (Figure 4A). Both integrin 26 and 3′-tail integrin 26 oligonucleotides displayed a single slowly migrating band whose mobility was inversely related to the length of the oligomer. In addition to the retarded band of each oligomer, a third hybrid band with an intermediate mobility was generated in their 1:1 mixture (Figure 4A), indicative of bimolecular stoichiometry of the slowly migrating complexes. To corroborate its tetrahelical nature the retarded band of the integrin 26 DNA was isolated and subjected to DMS modification. Four clusters of three guanines each in the bimolecular complex resisted DMS modification indicating their involvement in Hoogsteen hydrogen bonds, whereas all the guanine residues in the single-stranded oligonucleotide were modified by DMS and were consequently hydrolyzed by pyrrolidine (Figure 4B). These results were consistent with the complex being a guanine quartet stabilized bimolecular DNA tetraplex. Similar to the monomolecular tetraplex structure of the integrin 26 DNA, the bimolecular quadruplex of this sequence also displayed CD spectrum typical to a parallel-stranded tetrahelix (Figure 4C). Hence, under conditions that allowed intermolecular interaction, the integrin 26 DNA sequence formed Hoogsteen hydrogen-bonded, parallel-stranded, bimolecular tetraplex structure.

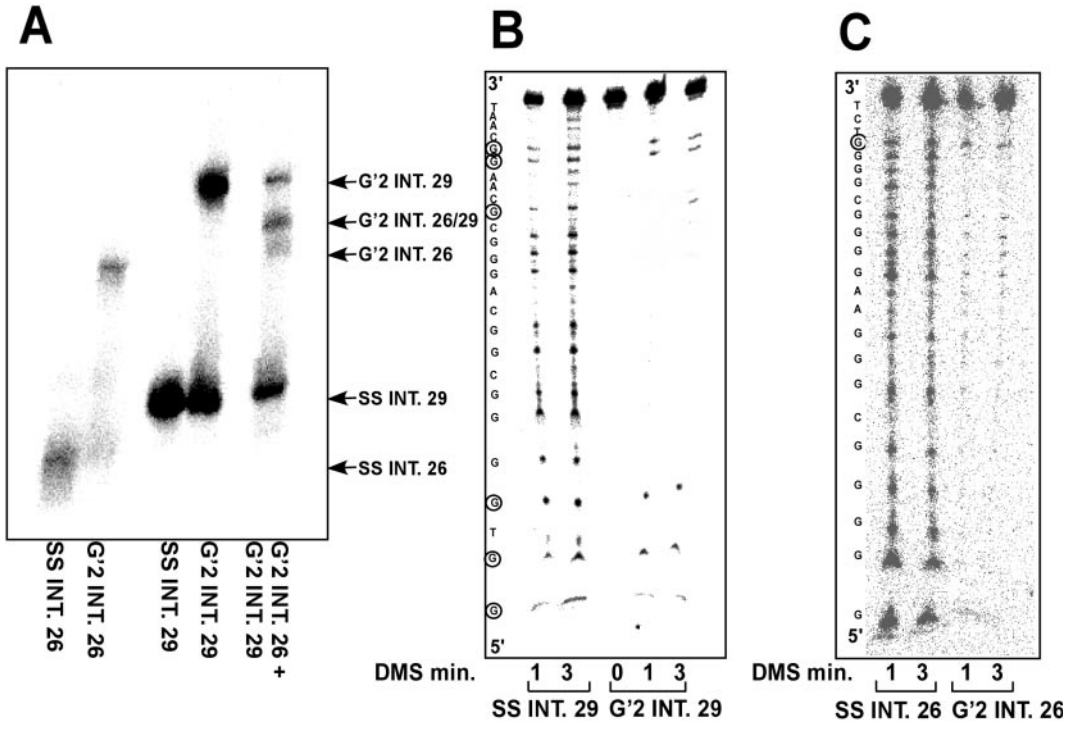

Figure 5. The neighboring integrin 26 and integrin 29 DNA sequences associate to form a G′2 bimolecular tetraplex structure. (**A**) The integrin 26 and integrin 29 DNA oligomers pair to form bimolecular complexes. 5′-^{32}P-labeled integrin 26 DNA, integrin 29 DNA or their equimolar mixture at a final concentration of 0.2 μg/μl were incubated at 37°C for 18 h in TE buffer that contained 300 mM KCl and resolved by a non-denaturing 10% PAGE in 0.5× TBE buffer containing 20 mM KCl. Positions of boiled single-stranded integrin 26 or integrin 29 DNA and their respective slowly migrating G′2 complexes are indicated. Notice the presence of a hybrid band of G′2 integrin 26 DNA/integrin 29 DNA in the lane containing an equimolar mixture of the two oligomers. (**B**) Patterns of methylation protection of single-stranded integrin 29 DNA and of the integrin 29 component the G′2 integrin 29 DNA/integrin 26 DNA tetraplex. 5′-^{32}P-labeled single-stranded integrin 29 DNA oligomer and its G′2 bimolecular tetraplex complex with unlabeled integrin 26 DNA were exposed to 0.75% DMS at 20°C for the specified periods of time and hydrolyzed by 2.0 M pyrrolidine as described in Materials and Methods and in the legend to Figure 3. DMS modified and pyrrolidine hydrolyzed guanine residues in the tetraplex DNA structure are circled. (**C**) Patterns of methylation protection of single-stranded integrin 26 DNA and of the integrin 26 DNA component the G′2 integrin 29 DNA/integrin 26 DNA tetraplex complex. DMS modification, isolation of the treated DNA, pyrrolidine hydrolysis and denaturing gel electrophoresis were as in (B) except that the integrin 26 DNA oligomer was 5′-^{32}P-labeled, whereas the integrin 29 DNA was unlabeled. The circled 3′ guanine residue of integrin 26 DNA in its tetraplex complex with integrin 29 DNA was sensitive to DMS modification.

Two neighboring guanine-rich α7 *integrin* promoter sequences pair to form a bimolecular tetraplex structure

We next asked whether two neighboring guanine-rich sequences were capable of folding and pairing into a bimolecular tetraplex structure. For this we used the integrin 26 DNA sequence located at positions −101 to −75 of the mouse α7 *integrin* promoter region and a 29mer oligonucleotide designated integrin 29 [5′-d(AAAGGTGGGGCGGCAGGGCGCAAGGCAAT)-3′] that represented positions −215 to −186 of the α7 *integrin* promoter region. These two sequences are separated from one another in the mouse genome by 85 nt (Figure 1A). Folding of integrin 26 and integrin 29 tracts into hairpin structures and looping out of the 85 bases long spacer stretch in single-stranded DNA could potentially allow pairing of the two hairpins to generate a tetraplex structure. To investigate whether the formation of such a structure is feasible, 200 ng/μl of each of the ^{32}P-labeled DNA oligomers or their equimolar mixtures were incubated at 37°C for 16–20 h in the presence of 300 mM KCl. Resolution of the DNA by non-denaturing gel electrophoresis revealed that the oligomers were largely converted into their respective slowly migrating species. In addition to the retarded structures of each sequence the 1:1 mixture of integrin 26 and integrin 29 DNA yielded a third species of intermediate mobility that was assumed to represent a heterodimer of the two oligomers (Figure 5A). To inquire whether this complex was a Hoogsteen hydrogen-bonded tetraplex structure, it was examined for the resistance of its guanine N7 position to methylation. Bimolecular complexes of 5′-^{32}P integrin 26 DNA with unlabeled integrin 29 oligomer or of unlabeled integrin 26 with 5′-^{32}P integrin 29 DNA were modified by DMS; resolved by non-denaturing gel electrophoresis; the hybrid radioactive bands were excised; the DNA was extracted and hydrolyzed with pyrrolidine (see Materials and Methods). Data presented in Figure 5B and C showed that defined residues in both integrin 29 and integrin 26 resisted DMS modification and hydrolysis, whereas all the guanine residues in the single-stranded forms of both oligomers were hydrolyzed by pyrrolidine. These results indicated that the two sequences associated to form a guanine quartet stabilized bimolecular tetraplex structure. The different patterns of DMS-resistant guanine residues in the integrin 26/integrin 29 DNA quadruplex (Figure 5C) and in the bimolecular tetraplex of integrin 26 (Figure 4B) indicated that the hetero-tetraplex differed from tetraplexes formed by each oligomer.

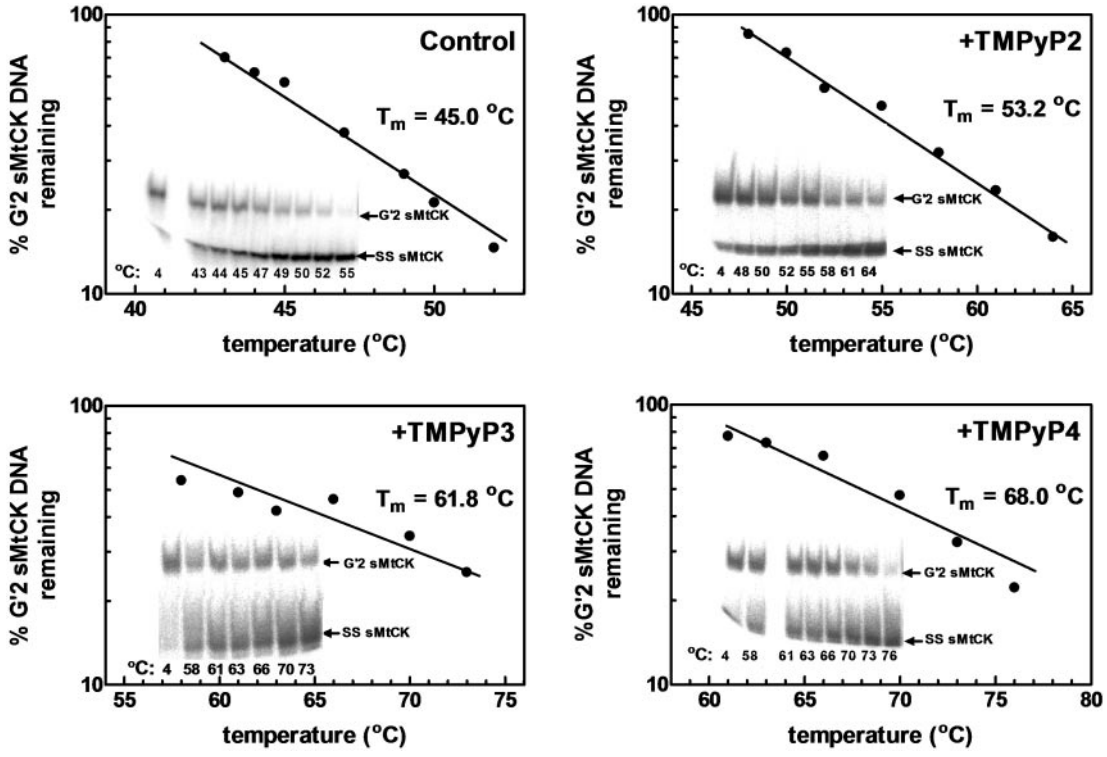

Figure 6. G′2 *sMtCK* DNA is stabilized to different extents by three cationic porphyrin isomers. Aliquots of 5′-^{32}P-labeled G′2 tetraplex structure of the *sMtCK* DNA oligomer were incubated at the specified temperatures for 10 min without or with the indicated cationic porphyrin isomer as detailed in Materials and Methods. Single-stranded and the G′2 tetraplex form of *sMtCK* DNA were resolved by a non-denaturing 10% PAGE in 0.5× TBE buffer containing 20 mM KCl and their proportions were quantified by phosphor imaging analysis. Shown are semi-logarithmic plots of the fraction of remaining G′2 tetraplex DNA as a function of increasing temperature relative to an initial 100% value determined for G′2 DNA that was maintained at 4°C. Phosphor images of the electrophoretically resolved *sMtCK* DNA structures are presented in insets within each panel.

Thermal stabilities of tetraplex structures of the different muscle-specific regulatory sequences are differentially affected by cationic porphyrins

Small molecules of diverse composition and structure were found to interact with tetraplex DNA structures and to affect their stability. Of special interest are the three positional cationic porphyrin isomers TMPyP2, TMPyP3 and TMPyP4 whose geometry and size approximate those of guanine quartet (37,53–56). These porphyrins vary in their binding affinity for different tetraplex DNA molecules (55,56). In addition, by binding to telomeric DNA tetraplexes, they inhibit to different degrees the activities of telomerase (57) or DNA helicase (56). A probable source for this variance is the variable arrangement of folded strands, groove sizes, base composition and dimensions of loops of unpaired nucleotides among tetraplex structures of different guanine-rich sequences (55,58).

Our results yielded structures of bimolecular tetraplexes that were significantly dissimilar for the different muscle gene sequences. To examine whether the structural heterogeneity of these tetraplexes resulted in their divergent interaction with cationic porphyrins, we compared the effect of TMPyP2, TMPyP3 and TMPyP4 on the thermal stabilities of bimolecular tetraplex structures of integrin 26, integrin 26/integrin 29 and *sMtCK* DNA. Terminally ^{32}P-labeled bimolecular tetraplex forms of these sequences were incubated at increasing temperatures for 10 min without or in the presence of each of the three cationic porphyrins. The fraction of remaining intact bimolecular tetraplex DNA in each sample was quantified by phosphor imaging analysis of electrophoretically resolved DNA in non-denaturing polyacrylamide gel. The results of typical experiments presented in Figure 6 show phosphor images of non-denaturing electrophoregrams of bimolecular tetraplex *sMtCK* DNA incubated at increasing temperatures for 10 min without or with each of the three cationic porphyrins. Quantification of these results yielded the shown plots of the fraction of remaining native bimolecular tetraplex DNA as a function of the incubation temperature. Melting temperatures, T_m, were defined as the temperatures at which 50% of the initial amount of bimolecular tetraplex *sMtCK* DNA was denatured. The shown T_m values as inferred from the presented results indicated that every tested porphyrin increased the heat resistance of bimolecular tetraplex *sMtCK* DNA with an order of effectiveness of TMPyP4 > TMPyP3 > TMPyP2. Average results of multiple determinations of the T_m values of the three examined G′2 tetraplex structures of muscle-specific regulatory sequences are summarized in Table 2. Clearly, the three cationic porphyrins differed in their capacity to protect the different tetraplex DNA structures against thermal denaturation. The T_m of G′2 integrin 26 remained virtually unaffected by TMPyP2 and TMPyP3 but was somewhat decreased by TMPyP4. In this context, G′2 integrin 26 was reminiscent of G′2 d(CGG)$_n$, which was destabilized

Table 2. Melting temperatures of G′2 tetraplex DNA structures of muscle-specific regulatory sequences are differentially affected by three cationic porphyrin isomers

	T_m (°C) [*N*] —	+TMPyP2	+TMPyP3	+TMPyP4
G′2 integrin 26 DNA	52.9 ± 5.7 [7]	52.8 ± 5.8 [3] Δ = −0.1	51.0 ± 4.3 [3] Δ = −1.9	48.0 ± 4.2 [6] Δ = −4.9
G′2 integrin 26/29 DNA	42.7 ± 1.7 [3]	46.7 ± 1.1 [2] Δ = +4.0	42.7 ± 2.4 [2] Δ = +0.0	48.5 ± 3.5 [2] Δ = +5.8
G′2 *sMtCK* DNA	48.0 ± 1.8 [6]	54.8 ± 1.3 [4] Δ = +6.8	64.8 ± 2.8 [4] Δ = +16.7	71.0 ± 2.3 [3] Δ = +23.0

The G′2 tetraplex DNA structures were heat denatured in the absence or presence of the indicated cationic porphyrin isomers as described in Materials and Methods. The remaining fraction of G′2 DNA was quantified by phosphor imaging analysis and the results were plotted semi-logarithmically to determine average T_m values (see Materials and Methods and the legend to Figure 6). Listed are the number of independent determinations of each T_m value, [*N*] and the Δ elevation (+) or reduction (−) in °C relative to T_m values of the respective G′2 DNA control samples that were maintained at 4°C.

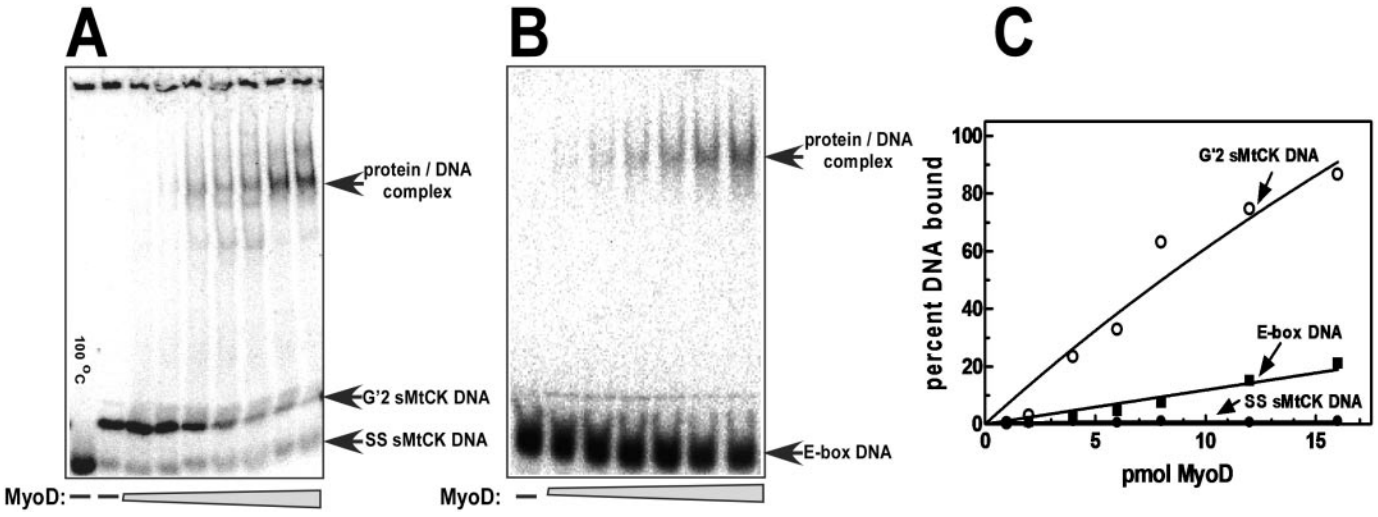

Figure 7. MyoD associates preferentially with bimolecular G′2 tetraplex structure of integrin 26 DNA. Recombinant MyoD protein was expressed in *E.coli*, purified and incubated to form homodimers as described in Materials and Methods. Increasing amounts of the protein were incubated under binding conditions with 0.2 pmol 5′-^{32}P-labeled E-box DNA or G′2 *sMtCK* DNA and protein–DNA complexes were resolved from free DNA by non-denaturing 4% PAGE in 0.25× TBE buffer containing 10 mM KCl. (**A**) Binding of MyoD to G′2 *sMtCK* DNA. Rising amounts of MyoD bound progressively increasing proportions of the G′2 *sMtCK* DNA whereas the residual hairpin shaped single-stranded DNA did not associate with the protein. (**B**) Binding of MyoD to E-box DNA. (**C**) Plots of phosphor imaging quantified results shown in (A and B).

by TMPyP4 (59). The thermal stability of G′2 integrin 26/integrin 29 DNA was not altered by TMPyP3 but was to an extent augmented by TMPyP2 and TMPyP4. In contrast, every tested porphyrin increased the thermal stability of G′2 *sMtCK* DNA to a significant extent with a rising level of stabilization by the porphyrins in the order TMPyP4 > TMPyP3 > TMPyP2 (Table 2).

MyoD binds G′2 tetraplex structures of muscle gene regulatory sequences more efficiently than E-box DNA motif

In view of the reported preferential binding of the myogenic determination protein MyoD to a tetrahelical structure of guanine-rich enhancer sequence of *MCK* (36), we next inquired whether this transcription factor associated with any of the secondary structures of muscle gene regulatory sequences that were identified in this work. The results shown in Figure 7 indicated that homodimeric recombinant MyoD bound a G′2 bimolecular tetraplex structure of the sMtCK sequence more efficiently than its target E-box DNA. In contrast, these data also pointed out that the protein did not form detectable complexes with single-stranded *sMtCK* DNA which, spontaneously folded into a hairpin structure (Figure 2A). Similar experiments revealed that MyoD bound bimolecular tetraplex forms of integrin 26 or integrin 26/integrin 29 more efficiently than E-box DNA, but that it did not detectably bind a monomolecular tetraplex structure of integrin 26 DNA (data not shown). The observed preference of MyoD for multi-molecular tetraplex structures of muscle-specific DNA sequences over E-box DNA was in line with a previous report (36), implying a biological role for such DNA secondary structures.

DISCUSSION

The key observation of this work is that promoter and enhancer sequences of several muscle-specific genes are enriched with clusters of guanine residues which readily form hairpin, unimolecular and bimolecular tetraplex structures.

The disproportionately high content of guanine clusters in select regions of regulatory sequences of several muscle-specific genes (Figure 1 and related text) raised the prospect that these sequences may fold into secondary structures.

This was tested and confirmed in our analysis of sequences derived from promoter or enhancer regions of three muscle-specific genes. A guanine-rich oligomer representing positions −296 to −320 of the lower strand of the *sMtCK* promoter region folded in an apparent zero order reaction and in the absence of K^+ cation into compact hydrogen-bonded structure that was consistent with a hairpin (Figure 2A). Incubation of this DNA sequence in the presence of 300 mM KCl resulted in a second-order pairing reaction of two hairpins to generate a Hoogsteen-bonded bimolecular tetraplex (Figure 2B–D). The guanine-rich sequence that represented positions −1045 to −1021 of the upper strand of the *MCK* enhancer region, similarly, formed a bimolecular tetrahelical complex (see Results). The DNA sequence dubbed integrin 26 that represented positions −101 to −75 of the *α7 integrin* promoter region folded in the presence of K^+ cation in an apparent zero order reaction into a Hoogsteen-bonded unimolecular tetraplex structure (Figure 3). At the same time, extended incubation of high concentrations of this oligonucleotide yielded in a second-order reaction a bimolecular tetraplex complex (Figure 4). CD spectra revealed parallel-stranded orientation for the monomolecular tetraplex form of integrin 26 DNA and for the G′2 bimolecular tetraplex structures of both *sMtCK* DNA and integrin 26 DNA (Figures 2D, 3D and 4D). Interestingly, X-ray crystallography indicated that monomolecular and bimolecular tetraplex structures of the telomeric DNA repeat sequence that were similarly formed in the presence of KCl were also parallel stranded (3).

It is obvious that the formation of bimolecular quadruplex structures of the *sMtCK*, *MCK* or integrin 26 DNA cannot be feasible *in vivo* since each sequence is represented as a single copy in its respective region of the gene. At the same time, when DNA is unwound during transcription, the folding of two proximate guanine-rich tracts along the single-stranded DNA stretch into hairpins might potentially lead to their pairing and to the formation of a unimolecular tetraplex structure. The feasibility of such a scenario was partially demonstrated by showing that the oligomeric DNA sequences integrin 26 and integrin 29 that represented tracts situated 85 bases away from each other in the *α7 integrin* promoter region formed a heterodimeric tetraplex complex (Figure 5). In summary, our results indicated that single-stranded guanine-rich runs derived from regulatory sequences of three muscle-specific genes were capable of folding into one or more secondary structures: hairpins, and unimolecular and bimolecular tetraplexes. Furthermore, as was amply demonstrated for tetraplex structures of telomeric DNA (55,57,60), results summarized in Figure 6 and Table 2 showed that the stability of some of the tetraplex DNA structures of muscle-specific genes could be modulated by cationic porphyrins. The different effect that the cationic porphyrins exerted on the thermal stability of the tetrahelical structures of different muscle gene sequences underlined the distinctive structural characteristics of each tetraplex. Thus, if these tetraplexes affect transcription, their activity may potentially be accentuated or depressed by porphyrins that increase or decrease their stability.

Tetraplex structures in DNA have been implicated in the regulation of the expression of at least two genes. In one case, a tetraplex structure formed upstream to the P1 promoter of *c-MYC* functioned to suppress its expression. It was found that the guanine-rich strand of the nuclease hypersensitive element III (1) upstream to the P1 *c-MYC* promoter which controls 85–90% of the transcriptional activation of the gene, folded into a chair-form unimolecular tetraplex (37,61). Destabilization of this DNA secondary structure by a single G→A transition mutation led to a 3-fold increase in the basal transcriptional activity of the P1 *c-MYC* promoter (37). Conversely, stabilization of the tetraplex structure by the cationic porphyrin TMPyP4 (37,62) or by a 2,6-pyridin-dicarboxamide derivative (63) further inhibited the transcriptional activation of *c-MYC*. In a converse case, tetraplex structures in the human insulin-linked polymorphic region (ILPR) acted to enhance transcription. Guanine-rich sequences in the promoter region of the chicken β *globin* gene (38,51) and in the ILPR located in the promoter of the human insulin gene (64), formed tetraplex structures *in vitro*. Single nucleotide differences in the ILPR that affected insulin expression were found to correlate with the capacity of the guanine-rich sequence to fold into intermolecular and intramolecular tetraplex structures. Hence, high transcription of a linked reporter gene was positively associated with the propensity of the ILPR repeats to fold into tetraplex formations (65).

The high incidence of guanine clusters in promoter and enhancer sequences of muscle-specific genes and their proclivity to fold into tetraplex structures, raise the possibility that such structures, in analogy with their counterparts in *c-MYC* or ILPR, take part in the regulation of myogenic gene expression. This speculation gained support by the reported preferential binding of the myogenic transcription factor MyoD to a tetraplex DNA structure of an *MCK* enhancer sequence (36). Indeed, evidence presented in Figure 7 indicated that homodimeric MyoD associated more efficiently with bimolecular tetraplex structures of the muscle-specific regulatory sequences than with E-box. This observation suggests that tetraplex structures formed in muscle-specific promoter or enhancer sequences may serve to trap MyoD homodimers, thus decreasing the probability of their association with E-box elements and an untimely myogenic gene expression. Regulated unwinding of the tetrahelical DNA structures may release MyoD and allow it to associate with E proteins. By forming tight complexes with E-box motifs, MyoD/E-protein heterodimers may in turn initiate timely transcriptional activation of the myogenic genes.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Eyal Bengal for useful discussions. This study was supported by grants to M.F. from the Israel Science Foundation, the Conquer Fragile X Foundation and the Fund for Promotion of Research in the Technion. Funding to pay the Open Access publication charges for this article was provided by a grant from the Israel Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Simonsson,T. (2001) G-quadruplex DNA structures—variations on a theme. *Biol. Chem.*, **382**, 621–628.
2. Shafer,R.H. and Smirnov,I. (2000) Biological aspects of DNA/RNA quadruplexes. *Biopolymers*, **56**, 209–227.

3. Parkinson,G.N., Lee,M.P. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
4. Schaffitzel,C., Berger,I., Postberg,J., Hanes,J., Lipps,H.J. and Pluckthun,A. (2001) *In vitro* generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc. Natl Acad. Sci. USA*, **98**, 8572–8577.
5. Chang,C.C., Kuo,I.C., Ling,I.F., Chen,C.T., Chen,H.C., Lou,P.J., Lin,J.J. and Chang,T.C. (2004) Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative. *Anal. Chem.*, **76**, 4490–4494.
6. Duquette,M.L., Handa,P., Vincent,J.A., Taylor,A.F. and Maizels,N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.*, **18**, 1618–1629.
7. Arthanari,H. and Bolton,P.H. (2001) Functional and dysfunctional roles of quadruplex DNA in cells. *Chem. Biol.*, **8**, 221–230.
8. Arthanari,H. and Bolton,P.H. (2003) Did quadruplex DNA play a role in the evolution of the eukaryotic linear chromosome? *Mini. Rev. Med. Chem.*, **3**, 1–9.
9. Zahler,A.M., Williamson,J.R., Cech,T.R. and Prescott,D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718–720.
10. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
11. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
12. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)$_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA*, **91**, 4950–4954.
13. Usdin,K. and Woodford,K.J. (1995) CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.*, **23**, 4202–4209.
14. Weisman-Shomer,P. and Fry,M. (1993) QUAD, a protein from hepatocyte chromatin that binds selectively to guanine-rich quadruplex DNA. *J. Biol. Chem.*, **268**, 3306–3312.
15. Uliel,L., Weisman-Shomer,P., Oren-Jazan,H., Newcomb,T., Loeb,L.A. and Fry,M. (2000) Human Ku antigen tightly binds and stabilizes a tetrahelical form of the Fragile X syndrome d(CGG)$_n$ expanded sequence. *J. Biol. Chem.*, **275**, 33134–33141.
16. Sarig,G., Weisman-Shomer,P., Erlitzki,R. and Fry,M. (1997) Purification and characterization of qTBP42, a new single-stranded and quadruplex telomeric DNA-binding protein from rat hepatocytes. *J. Biol. Chem.*, **272**, 4474–4482.
17. Muniyappa,K., Anuradha,S. and Byers,B. (2000) Yeast meiosis-specific protein Hop1 binds to G4 DNA and promotes its formation. *Mol. Cell. Biol.*, **20**, 1361–1369.
18. Lu,Q., Schierer,T., Kang,S.G. and Henderson,E. (1998) Purification, characterization and molecular cloning of TGP1, a novel G-DNA binding protein from *Tetrahymena thermophila*. *Nucleic Acids Res.*, **26**, 1613–1620.
19. Frantz,J.D. and Gilbert,W. (1995) A yeast gene product, G4p2, with a specific affinity for quadruplex nucleic acids. *J. Biol. Chem.*, **270**, 9413–9419.
20. Schierer,T. and Henderson,E. (1994) A protein from *Tetrahymena thermophila* that specifically binds parallel-stranded G4-DNA. *Biochemistry*, **33**, 2240–2246.
21. Liu,Z., Frantz,J.D., Gilbert,W. and Tye,B.K. (1993) Identification and characterization of a nuclease activity specific for G4 tetrastranded DNA. *Proc. Natl Acad. Sci. USA*, **90**, 3157–3161.
22. Liu,Z. and Gilbert,W. (1994) The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of *in vivo* functions for this novel DNA structure. *Cell*, **77**, 1083–1092.
23. Bashkirov,V.I., Scherthan,H., Solinger,J.A., Buerstedde,J.M. and Heyer,W.D. (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. *J. Cell Biol.*, **136**, 761–773.
24. Sun,H., Yabuki,A. and Maizels,N. (2001) A human nuclease specific for G4 DNA. *Proc. Natl Acad. Sci. USA*, **98**, 12444–12449.
25. Fang,G. and Cech,T.R. (1993) The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell*, **74**, 875–885.
26. Fang,G. and Cech,T.R. (1993) Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein. *Biochemistry*, **32**, 11646–11657.
27. Giraldo,R. and Rhodes,D. (1994) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.*, **13**, 2411–2420.
28. Giraldo,R., Suzuki,M., Chapman,L. and Rhodes,D. (1994) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc. Natl Acad. Sci. USA*, **91**, 7658–7662.
29. Fry,M. and Loeb,L.A. (1999) Human werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)$_n$. *J. Biol. Chem.*, **274**, 12797–12802.
30. Huber,M.D., Lee,D.C. and Maizels,N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.*, **30**, 3954–3961.
31. Sun,H., Bennett,R.J. and Maizels,N. (1999) The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G–G paired DNAs. *Nucleic Acids Res.*, **27**, 1978–1984.
32. Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **273**, 27587–27592.
33. Fukuda,H., Katahira,M., Tsuchiya,N., Enokizono,Y., Sugimura,T., Nagao,M. and Nakagama,H. (2002) Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc. Natl Acad. Sci. USA*, **99**, 12685–12690.
34. Khateb,S., Weisman-Shomer,P., Hershco,I., Loeb,L.A. and Fry,M. (2004) Destabilization of tetraplex structures of the fragile X repeat sequence (CGG)$_n$ is mediated by homolog-conserved domains in three members of the hnRNP family. *Nucleic Acids Res.*, **32**, 4145–4154.
35. Lin,Y.C., Shih,J.W., Hsu,C.L. and Lin,J.J. (2001) Binding and partial denaturing of G-quartet DNA by Cdc13p of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **276**, 47671–47674.
36. Walsh,K. and Gualberto,A. (1992) MyoD binds to the guanine tetrad nucleic acid structure. *J. Biol. Chem.*, **267**, 13714–13718.
37. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
38. Howell,R.M., Woodford,K.J., Weitzmann,M.N. and Usdin,K. (1996) The chicken beta-globin gene promoter forms a novel 'cinched' tetrahelical structure. *J. Biol. Chem.*, **271**, 5208–5214.
39. Weisman-Shomer,P., Naot,Y. and Fry,M. (2000) Tetrahelical forms of the fragile X syndrome expanded sequence d(CGG)$_n$ are destabilized by two heterogeneous nuclear ribonucleoprotein-related telomeric DNA-binding proteins. *J. Biol. Chem.*, **275**, 2231–2238.
40. Payne,R.M., Haas,R.C. and Strauss,A.W. (1991) Structural characterization and tissue-specific expression of the mRNAs encoding isoenzymes from two rat mitochondrial creatine kinase genes. *Biochim. Biophys. Acta*, **1089**, 352–361.
41. Payne,R.M. and Strauss,A.W. (1994) Expression of the mitochondrial creatine kinase genes. *Mol. Cell. Biochem.*, **133-134**, 235–243.
42. Payne,R.M. and Strauss,A.W. (1994) Developmental expression of sarcomeric and ubiquitous mitochondrial creatine kinase is tissue-specific. *Biochim. Biophys. Acta*, **1219**, 33–38.
43. Klein,S.C., Haas,R.C., Perryman,M.B., Billadello,J.J. and Strauss,A.W. (1991) Regulatory element analysis and structural characterization of the human sarcomeric mitochondrial creatine kinase gene. *J. Biol. Chem.*, **266**, 18058–18065.
44. Chen,F.M. (1992) Sr^{2+} facilitates intermolecular G-quadruplex formation of telomeric sequences. *Biochemistry*, **31**, 3769–3776.
45. Lu,M., Guo,Q. and Kallenbach,N.R. (1993) Thermodynamics of G-tetraplex formation by telomeric DNAs. *Biochemistry*, **32**, 598–601.
46. Weintraub,H., Genetta,T. and Kadesch,T. (1994) Tissue-specific gene activation by MyoD: determination of specificity by *cis*-acting repression elements. *Genes Dev.*, **8**, 2203–2211.
47. Hynes,R.O. (1992) Integrins: versatility, modulation, and signaling in cell adhesion. *Cell*, **69**, 11–25.
48. Kramer,R.H., Vu,M.P., Cheng,Y.F., Ramos,D.M., Timpl,R. and Waleh,N. (1991) Laminin-binding integrin alpha 7 beta 1: functional characterization and expression in normal and malignant melanocytes. *Cell Regul.*, **2**, 805–817.
49. Ziober,B.L. and Kramer,R.H. (1996) Identification and characterization of the cell type-specific and developmentally regulated alpha7 integrin gene promoter. *J. Biol. Chem.*, **271**, 22915–22922.

50. Bergstrom,D.A., Penn,B.H., Strand,A., Perry,R.L., Rudnicki,M.A. and Tapscott,S.J. (2002) Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. *Mol. Cell*, **9**, 587–600.
51. Woodford,K.J., Howell,R.M. and Usdin,K. (1994) A novel K^+-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.*, **269**, 27029–27035.
52. Nadel,Y., Weisman-Shomer,P. and Fry,M. (1995) The fragile X syndrome single strand d(CGG)$_n$ nucleotide repeats readily fold back to form unimolecular hairpin structures. *J. Biol. Chem.*, **270**, 28970–28977.
53. Anantha,N.V., Azam,M. and Sheardy,R.D. (1998) Porphyrin binding to quadrupled T4G4. *Biochemistry*, **37**, 2709–2714.
54. Wheelhouse,R.T., Sun,D., Han,H., Han,F.X. and Hurley,L.H. (1998) Cationic porphyrins as telomerase inhibitors: the interaction of tetra-(*N*-methyl-4-pyridyl)porphine with quadruplex DNA. *J. Am. Chem. Soc.*, **120**, 3261–3262.
55. Haq,I., Trent,J.O., Chowdhry,B.Z. and Jenkins,T.C. (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin1. *J. Am. Chem. Soc.*, **121**, 1768–1779.
56. Han,H., Langley,D.R., Rangan,A. and Hurley,L.H. (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. *J. Am. Chem. Soc.*, **123**, 8902–8913.
57. Shi,D.F., Wheelhouse,R.T., Sun,D. and Hurley,L.H. (2001) Quadruplex-interactive agents as telomerase inhibitors: synthesis of porphyrins and structure–activity relationship for the inhibition of telomerase. *J. Med. Chem.*, **44**, 4509–4523.
58. Arthanari,H., Basu,S., Kawano,T.L. and Bolton,P.H. (1998) Fluorescent dyes specific for quadruplex DNA. *Nucleic Acids Res.*, **26**, 3724–3728.
59. Weisman-Shomer,P., Cohen,E., Hershco,I., Khateb,S., Wolfovitz-Barchad,O., Hurley,L.H. and Fry,M. (2003) The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)$_n$. *Nucleic Acids Res.*, **31**, 3963–3970.
60. Fedoroff,O.Y., Rangan,A., Chemeris,V.V. and Hurley,L.H. (2000) Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism. *Biochemistry*, **39**, 15083–15090.
61. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
62. Grand,C.L., Han,H., Munoz,R.M., Weitman,S., Von Hoff,D.D., Hurley,L.H. and Bearss,D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. *Mol. Cancer Ther.*, **1**, 565–573.
63. Lemarteleur,T., Gomez,D., Paterski,R., Mandine,E., Mailliet,P. and Riou,J.F. (2004) Stabilization of the c-myc gene promoter quadruplex by specific ligands' inhibitors of telomerase. *Biochem. Biophys. Res. Commun.*, **323**, 802–808.
64. Hammond-Kosack,M.C., Dobrinski,B., Lurz,R., Docherty,K. and Kilpatrick,M.W. (1992) The human insulin gene linked polymorphic region exhibits an altered DNA structure. *Nucleic Acids Res.*, **20**, 231–236.
65. Lew,A., Rutter,W.J. and Kennedy,G.C. (2000) Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc. Natl Acad. Sci. USA*, **97**, 12508–12512.