Published online 16 October 2007 *Nucleic Acids Research, 2007, Vol. 35, No. 21* ***7087–7095***
doi:10.1093/nar/gkm746

MyoD uses overlapping but distinct elements to bind E-box and tetraplex structures of regulatory sequences of muscle-specific genes

Jeny Shklover, Shulamit Etzioni, Pnina Weisman-Shomer, Anat Yafe, Eyal Bengal and Michael Fry*

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

Received August 1, 2007; Revised September 6, 2007; Accepted September 9, 2007

ABSTRACT

Muscle differentiation and expression of muscle-specific proteins are initiated by the binding of heterodimers of the transcription factor MyoD with E2A proteins to E-box motif d(CANNTG) in promoters or enhancers of muscle-specific genes. MyoD homodimers, however, form tighter complexes with tetraplex structures of guanine-rich regulatory sequences of some muscle genes. In this work, we identified elements in MyoD that bind E-box or tetraplex structures of promoter sequences of the muscle-specific genes α7 integrin and sarcomeric Mitochondrial Creatine Kinase (*sMtCK*). Deletions of large domains of the 315 amino acids long recombinant MyoD indicated that the binding site for both E-box and tetraplex DNA is its basic region KRKTTNADRRKAATMRERRR that encompasses the three underlined clusters of basic residues designated R_1, R_2 and R_3. Deletion of a single or pairs of R triads or R111C substitution completely abolished the E-box-binding capacity of MyoD. By contrast, the MyoD deletion mutants Δ102–114, ΔR_3, ΔR_1R_3 or ΔR_2R_3 maintained comparable tetraplex DNA-binding capacity as reflected by the similar dissociation constants of their protein–DNA complexes. Only deletion of all three basic clusters abolished the binding of tetraplex DNA. Implications of the binding of E-box and tetraplex DNA by non-identical MyoD elements are considered.

INTRODUCTION

Gene transcription is tightly regulated at multiple levels. Two expansively investigated mechanisms are the epigenetic modification by methylation of gene regulatory sequences and the remodeling of chromatin by enzymatic modifications of histones and disruption of histone–DNA interactions. Structural transitions from B-DNA to non-B-DNA that are generated by positive or negative superhelical stress in DNA constitute a third level of transcription regulation (1,2). Of the non-B-DNA structures, tetraplex or G-quadruplex configurations of guanine-rich sequences are of special interest. Evidence showed that the expression of multiple genes such as chicken β-globin, mouse *MCK* and α7 integrin and human insulin, *c-myc*, *sMtCK* and *FMR1* was affected by tetraplex structures that were formed in their promoter or enhancer regions or that the tetrahelical DNA served as target for transcription factors (3). We reported recently that segments of promoter and enhancer regions of several muscle-specific genes had a disproportional high prevalence of clusters of contiguous guanine residues and that these sequences readily folded *in vitro* into hairpin and parallel-stranded G′4 unimolecular and G′2 bimolecular tetraplex structures (4). We also found that homodimers of the myogenic master transcription factor MyoD bound preferentially to these tetrahelical structures (5). Based on these observations, we proposed that tetraplex domains in regulatory regions of muscle-specific genes may contribute to their expression during embryonic differentiation.

Skeletal muscle tissue differentiates from embryonic omnipotent mesodermal stem cells in a series of successive steps. Cells that commit to myogenic precursors initially divide as myoblasts that in turn cease to proliferate and initiate the expression of muscle-specific genes. In a final step, the cells fuse to form fully differentiated syncitial myotubes (6–8). Coordinated activation of the various muscle-specific genes during myogenesis is regulated by four myogenic MRF transcription regulatory factors; MyoD, Myf-5, MRF4 (Myf-6) and myogenin that comprise a subgroup within the superfamily of basic

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

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

helix-loop-helix (bHLH) proteins (9,10). Targeted inactivation of the various MRFs in mouse germ line showed initially that MyoD and Myf-5 act as determination factors that control the commitment of proliferating somitic cells to the myogenic lineage (11–13), whereas MRF4 and myogenin direct the subsequent differentiation of committed myoblasts into myocytes and myotubes (14–17). More recent data suggested, however, that MRF4-like Myf5 also operates as a determination factors upstream of MyoD by directing omnipotent embryonic cells into the myogenic lineage (18). Being tissue-specific (class II) bHLH proteins, the MRFs either self-associate through their HLH segment to form homodimers or link with class I bHLH proteins that include HEB/HTF4, E2-2/ITF-2 and E2A proteins (E12 and E47) to form heterodimers (10). Structure–function analysis of MRFs revealed that their basic region serves as the DNA-binding site (19). MyoD forms heterodimers with the bHLH proteins E12, E47 and ITF1 at greater efficiency than its self-association into homodimers (19–21). Studies of myogenesis in cell cultures showed that transcription of muscle-specific genes is initiated by the binding of MyoD-E12 or MyoD-E47 heterodimers to a conserved E-box motif d(CANNTG) in promoters or enhancers of the activated genes. Although homodimers of the 60 amino acids long bHLH domain of MyoD were also reported to bind specifically to E-box DNA *in vitro* (20) and to induce myogenesis in stably transfected mouse fibroblasts (22), homodimers of full-length MyoD displayed significantly lower affinity for E-box than the MyoD-E12 heterodimers (20,23).

In an earlier work it was reported that recombinant MyoD bound tetrahelical structures of a guanine-rich mouse creatine kinase enhancer sequence and of *Tetrahymena* telomeric DNA (24). Measurement of the dissociation constants of MyoD–DNA complexes revealed that the association of MyoD with tetraplex DNA was 4- to 5-fold tighter than with E-box DNA. More recently we demonstrated that MyoD homodimers bound tightly to bimolecular DNA tetraplexes of the muscle gene DNA sequences but did not associate with their single-stranded, hairpin, double-stranded or intramolecular tetraplex forms (5). Moreover, measurements of dissociation constants, K_d, of protein–DNA complexes revealed that MyoD homodimers formed significantly tighter complexes with the G′2 DNA tetraplexes than with E-box DNA. Conversely, MyoD-E47 heterodimers bound E-box more tightly than G′2 tetraplex DNA structures. We proposed that the preferential binding of the relatively inactive MyoD homodimers to tetraplex domains in regulatory regions of muscle-specific genes may prevent unproductive occupation of the E-box by MyoD homodimers (5).

The differential binding of MyoD homo- and heterodimers to E-box and to tetraplex DNA invited structure–function analysis of the interaction of this protein with the two DNA elements. We thus identified in this study MyoD elements that participate in the binding of E-box and tetraplex structures of promoter sequences of two muscle-specific genes. We report that the basic region of MyoD serves as the binding site for both DNA types. However, whereas a point mutation or minimal deletions in this region inactivate the capacity of MyoD to bind E-box, tetraplex DNA can be bound by MyoD variant proteins that possess just a single cluster of three basic amino acids within their mutated basic region. The contrasting stringent structural requirements of MyoD for the binding of E-box as opposed to the minimal demands for its association with tetraplex DNA may serve in the binding of MyoD to alternate genomic targets prior to the activation of muscle-specific genes during muscle differentiation.

MATERIALS AND METHODS

Preparation of double-stranded E-box and bimolecular tetraplex DNA structures

The synthetic DNA oligomers Integrin and sMtCK (Table 1) whose nucleotide sequences were derived from guanine-rich promoter regions of the genes *sarcomeric Mitochondrial Creatine Kinase* and α7 *integrin* (4), respectively, were purified by denaturing gel electrophoresis in 8.0 M urea, 12% polyacrylamide (acryl/bisacrylamide, 19:1) (25), and were subsequently 5′-^{32}P labeled in bacteriophage T4 polynucleotide kinase-catalyzed reaction.

Table 1. DNA oligomers used in this work

Oligomer	Bases	Nucleotide sequence
5′-E-box	26	5′-d[TCGATCCCCCAA**CACCTG**CTGCCTGA]-3′
3′-E-box	26	5′-d[TCAGGCAG**CAGGTG**TTGGGGGATCGA]-3′
Integrin	26	5′-d[CAT**GGGGGC**G**GGG**AA**GGGGC**G**GGGG**TCT]-3′
sMtCK	24	5′-d[CTGA**GGA**G**GGGG**CT**GG**A**GGG**ACCAC]-3′
5′-Δ119–121	31	5′-d[GCCACCATGCGCGAGCTGAGCAAAGTGAATG]-3′
3′-Δ119–121	31	5′-d[CATTCACTTTGCTCAGCTCGCGCATGGTGGC]-3′
5′-Δ102–104C	29	5′-d[CTGCAAGGCGTGCACCACCAACGCTGATC]-3′
3′-Δ102–104G	29	5′-d[GATCAGCGTTGGTGGTGCACGCCTTGCAG]-3′
5′-Δ110–112	25	5′-d[GACCACCAACGCTGATGCCGCCACC]-3′
3′-Δ110–112	25	5′-d[GGTGGCGGCATCAGCGTTGGTGGTC]-3′
5′-Δ102–104G	29	5′-d[CTGCAAGGCGTGCACCACCAACGCTGATG]-3′
3′-Δ102–104C	29	5′-d[CATCAGCGTTGGTGGTGCACGCCTTGCAG]-3′

Complementary E-box core sequences are underlined in the 5′ and 3′ E-box oligomers. Tetrad forming clusters of guanine residues are underlined in the integrin and sMtCK oligomers that were, respectively, derived from guanine-rich promoter regions of the genes α7 integrin or sarcomeric mitochondrial creatine kinase.

Bimolecular quadruplex structures of the two oligomers were formed as we described (4). A DNA double strand that contained the E-box CACCTG–CAGGTG motif was prepared by annealing equimolar amounts of the 5′- and 3′-E-box oligomers, (Table 1), as previously detailed (26).

Preparation, purification and expression of full-length and mutant recombinant MyoD

GST-fused full-length *Mus musculus* MyoD cDNA was ligated into a pRK171α vector and cloned in *Escherichia coli XL-1*. Plasmids harboring MyoD Δ102–114 mutant DNA or its bHLH domain (residues 102–162) (see Figure 1 for a map), were generously contributed by Dr S. J. Tapscott (FHCRC, Seattle). Large regions of MyoD DNA were deleted by PCR amplification of a desired fragment of the full-length cDNA using primers that consisted of 5′ or 3′ sequences of the MyoD fragment and pGEX-6P sequences, which had EcoRI and XhoI restriction sites, respectively. An R111C point mutation was generated in MyoD cDNA by PCR amplification using primers that contained an R to C substitution in codon 111. The R_3 cluster of the three amino acids RRR was deleted from the MyoD basic region by PCR using full-length MyoD cDNA template and 5′ and 3′ Δ119–121 primers (Table 1). Doubly deleted ΔR_1R_3 MyoD cDNA was generated by PCR using ΔR_3 MyoD template DNA and 5′-Δ102–104C and 3′-Δ102–104G primers (Table 1). The ΔR_2R_3 MyoD mutant was similarly prepared except that the primers 5′-Δ110–112 and 3′-Δ110–112 were used. A triple $\Delta R_1R_2R_3$ MyoD mutant was generated by PCR employing a ΔR_2R_3 MyoD cDNA template and the 5′-Δ102–104G and 3′-Δ102–104C primers (Table 1). Because of the high guanine–cytosine content of sequences in the vicinity of the R_1, R_2 and R_3 clusters a specialized PCR protocol devised by Ralser *et al.* (27) was employed

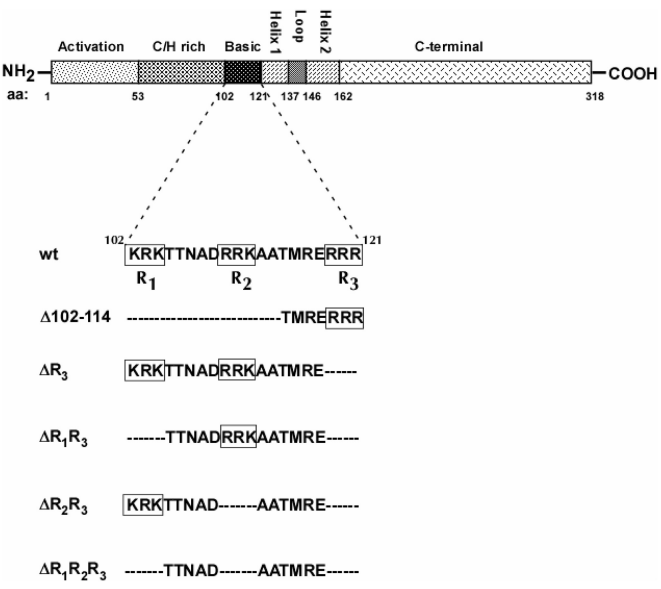

Figure 1. Scheme of MyoD domains and deletion mutations in its basic region. Deletion mutations were generated within the basic region, (residues 102–121), as described under Materials and Methods section. The triads of basic amino acids, R_1, R_2 and R_3 are boxed.

to produce the various deletion mutations. Briefly, reaction mixtures contained in a final volume of 50 µl: 10 ng pGEX-6P full-length or mutant MyoD DNA template; 2.5 units Pfu-Ultra DNA polymerase; 5 µl 10× polymerase buffer; 20 pmol each of 3′ and 5′ primers; 1 mM dNTPs and 6.6 µl of enhancer solution consisting of 83 µg/ml BSA, 10 mM DTT, 10% DMSO and 4 M Betaine. The amplification program included 2 min at 95°C, followed by 30 cycles of DNA melting at 95°C for 30 s and elongation and annealing at 72°C for 6 min and concluded with a single step of additional elongation at 72°C for 10 min. Following selection and isolation of mutant clones and verification of the desired mutation by DNA sequencing, full-length and mutant MyoD proteins were expressed in *E. coli BL21(DE3)pLysS* cells as we described (5). The recombinant proteins were purified to >95% homogeneity from the bacterial cell extracts by glutathione-agarose (Sigma) affinity column chromatography. The GST residue was cleaved by incubating 100 µg of fusion protein for 4 h and at 4°C with 2.0 U preScission protease (Amersham Biosciences).

Electrophoretic mobility shift assay of protein binding to DNA and determination of dissociation constants of the protein–DNA complexes

Homodimers of full-length or mutant MyoD were formed prior to their binding to DNA probes by incubating specified amounts of purified recombinant protein 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 buffer, pH 8.0, and 0.5 µg HeLa whole cell extract. Reaction mixtures for protein–DNA binding contained in a final volume of 10 µl: specified amounts of full-length or mutant MyoD homodimers and 5′-^{32}P labeled DNA probe, 14.5 mM KCl, 0.45 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT, 20% glycerol and 0.05 µg HeLa whole cell extract in 20 mM Tris–HCl buffer, pH 8.0. Reaction mixtures for the binding of 5′-^{32}P labeled double-stranded E-box DNA also contained 100-fold (w/w) excess of unlabeled poly d(I-C) (Sigma). Mixtures for the binding of end-labeled G′2 bimolecular tetraplex DNA structures of the integrin or sMtCK 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 (1.2 mM EDTA in 0.54 mM Tris–borate buffer, pH 8.3). 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 were quantified by phosphor imaging analysis.

To determine dissociation constants, K_d, of complexes of normal or mutant MyoD with E-box DNA or with G′2 tetraplex structures of integrin, increasing amounts of ^{32}P-labeled DNA were incubated with a constant amount of protein under the above described conditions.

Table 2. The MyoD basic region is required for the binding of both E-box and G′2 tetraplex integrin DNA

Protein	DNA binding	
	G′2 integrin DNA	E-box DNA
Full-length MyoD	+	+
Deleted activation region (Δ3–56)	+	+
Deleted activation and C/H regions (Δ1–93)	+	+
Deleted activation, C/H and basic regions (Δ1–121)	−	−
Isolated bHLH domain (residues 102–162)	+	+

Increasing amounts of full-length or mutant MyoD proteins (0–30 pmol) were incubated with 65 fmol of 5′-^{32}P labeled E-box or G′2 integrin DNA under binding conditions and protein–DNA complexes were resolved from free DNA by non-denaturing 4% polyacrylamide electrophoresis as described in the Materials and Methods section. Marking of (−) signifies absence of detectable protein–DNA complex at the highest amount of added protein.

Following electrophoretic mobility shift resolution of the protein–DNA complexes from free DNA, their relative amounts were determined by phosphor imaging quantification of the dried gel. K_d values were derived from the negative reciprocal of the slope of a Scatchard plot of the results as we detailed elsewhere (28).

RESULTS

The MyoD basic region is the binding site for both E-box and tetraplex DNA

Homodimeric MyoD associates more tightly with tetraplex forms of regulatory sequences of muscle-specific genes than with E-box DNA motif, which is the preferred binding target for MyoD-E47 heterodimers (5,24). MyoD domains include the N-terminal transcription activation region, a cysteine–histidine C/H-rich domain, a basic region which was shown to be the E-box-binding site (19), a helix-loop-helix (HLH) domain that mediate oligomerization, and a C-terminal stretch (Figure 1, top). To identify the region in MyoD to which tetraplex DNA binds, we assessed the capacity of mutant MyoD proteins that lacked defined domains to associate with E-box and G′2 tetraplex integrin DNA. Data summarized in Table 2 indicated that the activation domain and the C/H region were not required for the binding of E-box and G′2 tetraplex DNA. However, extending the deletion to the end of basic region abolished the binding of both types of DNA. Conversely, isolated bHLH domain (residues 102–162) formed complexes with both E-box and G′2 integrin DNA (Table 2). These results suggested that similar to E-box, the binding of tetraplex DNA was also mediated by the basic domain of MyoD.

Mutated MyoD basic region binds tetraplex DNA but not E-box

To inquire whether or not the E-box and tetraplex-binding sites completely overlap, we compared the DNA-binding capacity of a MyoD mutant that contained a Δ102–114 partial deletion within the 20 amino acids long basic region that extends from residue 102–121 (Figure 1). As shown in Figure 2A, the mutant protein failed to detectably associate with E-box DNA whereas it bound G′2 integrin DNA to almost the same extent as full-length MyoD (Figure 2B). The different MyoD structure requirements for the binding of E-box and tetraplex DNA were further underscored by a comparison of their ability to associate with a MyoD R111C mutant protein. As shown in the left panel of Figure 3, substituting the 111 residue in the center of the basic region (Figure 1) from ariginine to cysteine completely abolished the capacity of MyoD to bind E-box DNA. By clear contrast, the R111C mutant protein associated with G′2 tetraplex integrin DNA to practically the same extent as did native MyoD (Figure 3, right panel). Put together, results shown in Table 2 and in Figures 2 and 3 indicated that although the MyoD basic region served as the common binding site for both E-box and tetraplex DNA, binding of E-box required an intact basic region whereas tetraplex DNA could associate with a partially deleted or mutated basic region.

A single cluster of three basic amino acids suffices for the binding of tetraplex DNA

Since mutated basic region of MyoD maintained its capacity to bind tetraplex DNA, we undertook to define the minimum requirements for the binding of tetrahelical structures of integrin and sMtCK regulatory sequences. A prominent feature of the basic region is that it includes three clusters of three basic amino acids each. These clusters, KRK at positions 102–104, RRK at 110–112 and RRR at 119–121, were designated R_1, R_2 and R_3, respectively (Figure 1). The capacity of Δ102–114 MyoD protein to bind G′2 integrin DNA, (Figure 2B) indicated that a largely deleted basic region with only a short stretch of 7 amino acids remaining at its C-terminus was capable of binding the tetraplex structure. Since this remainder of the basic region included the R_3 cluster, we speculated that any single cluster of three basic amino acids may be necessary and sufficient for the binding of tetraplex DNA. To test this hypothesis, we assessed the capacities of a series of mutant MyoD proteins that lacked one, two or three basic amino acids clusters to bind E-box and G′2 tetraplex structures of integrin or sMtCK DNA. Representative results of electrophoretic mobility shift analysis shown in the first panel of Figure 4A indicated that whereas MyoD with an intact basic region formed a complex with E-box DNA, deletion of the R_3 cluster alone or in combination with R_1, R_2 or both resulted in a complete loss of the E-box-binding capacity. By contrast, full-length MyoD as well as its mutants ΔR_3, ΔR_1R_3 and ΔR_2R_3 formed complexes with G′2 tetraplex structures of integrin or sMtCK DNA and only the triply deleted mutant protein $\Delta R_1R_2R_3$ lost the capacity to bind the two tetrahelices (Figure 4A, second and third panels). Notably, these data also showed that the deletion mutation partially compromised the tetraplex DNA-binding capacity of MyoD. This was confirmed by following the binding of a constant amount of G′2 tetraplex forms

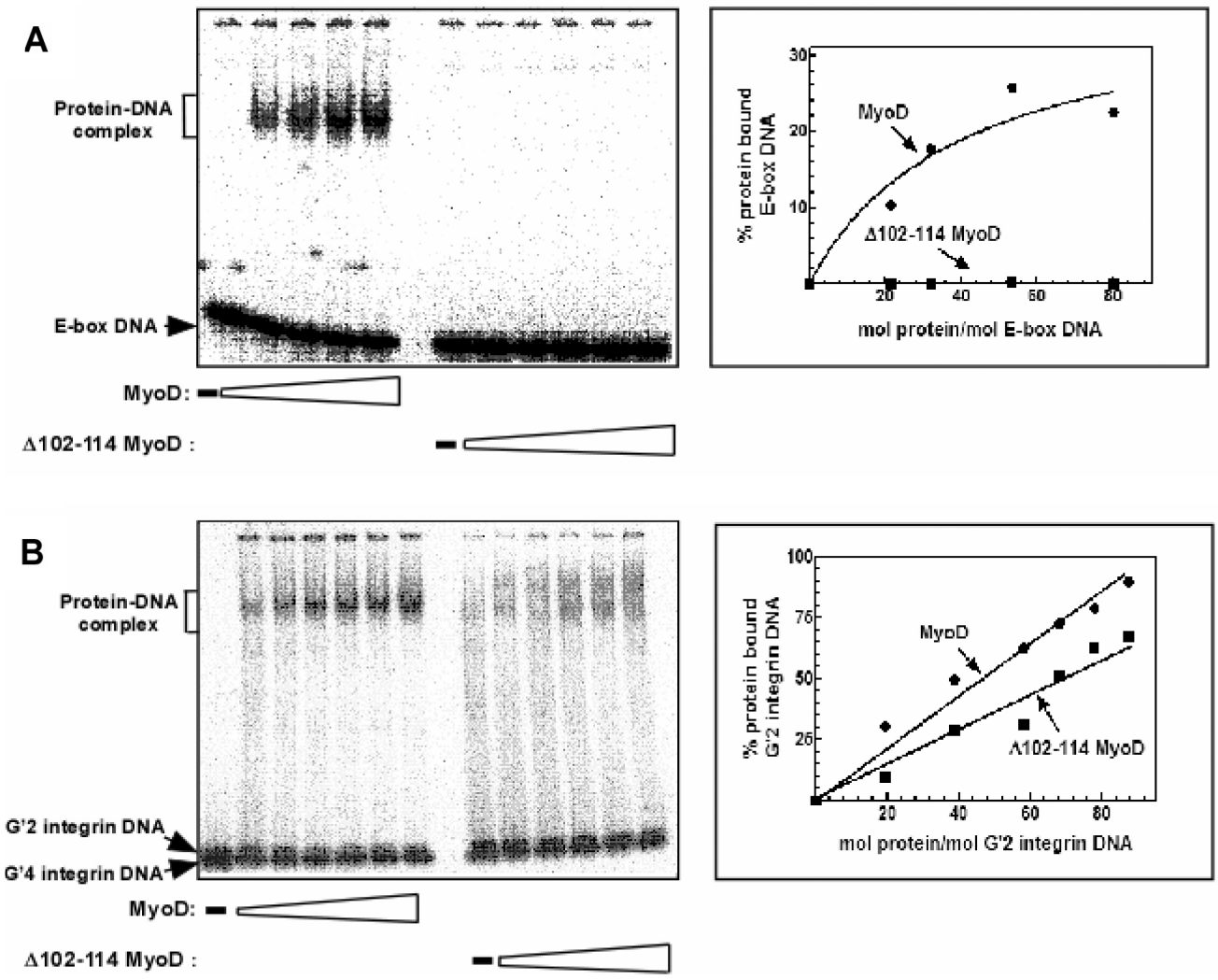

Figure 2. Deletion of MyoD residues 102–114 abolishes the binding of E-box but not of G′2 tetraplex integrin DNA. Increasing amounts of full-length MyoD or its Δ102–114 mutant were incubated under binding conditions with 5′-^{32}P labeled double-stranded E-box or G′2 integrin DNA. Protein–DNA complexes were resolved from unbound DNA by non-denaturing polyacrylamide gel electrophoresis and the relative proportions of protein-bound DNA were quantified by phosphor imaging analysis (see Materials and Methods section). Quantification indicated that the bimolecular G′2 integrin DNA was in equilibrium with monomolecular G′4 structure and that MyoD formed complexes solely with the G′2 tetraplex (5). (**A**) Binding of E-box DNA by full-length and Δ102–114 MyoD. Left—autoradiogram of electrophoretically resolved protein–DNA complexes. Right—plot of the quantified results. (**B**) Binding of G′2 integrin DNA by full-length and Δ102–114 MyoD. Left—autoradiogram of electrophoretically resolved protein–DNA complexes. Whereas full-length MyoD generated a single protein–DNA complex, the Δ102–114 mutant protein formed two complexes. Right—plot of the quantified results. Percent G′2 integrin DNA bound to Δ102–114 MyoD was the sum of the two types of formed complexes.

of either integrin or sMtCK DNA to increasing amounts of full-length or mutant MyoD proteins. As shown in Figure 4B, in this experiment, the binding of G′2 integrin DNA was only minimally diminished by deletion of the R_3 cluster and removal of this triad of basic residues even increased complex formation with G′2 sMtCK. Combined deletion of two clusters, R_1 and R_3 or R_2 and R_3, was more detrimental, significantly diminishing the protein-binding capacity for G′2 integrin binding and even more so for G′2 sMtCK DNA. Similar titration showed that any added amount of the triple deletion mutant $\Delta R_1R_2R_3$ failed to detectably bind either G′2 integrin or sMtCK DNA (data not shown).

The affinity of MyoD for tetraplex DNA is moderately reduced by removing a single or pairs of basic amino acids triads

To assess more accurately the contribution of specific basic amino acids clusters to the affinity of MyoD for tetraplex DNA, we determined the dissociation constants, K_d, of complexes of the various MyoD deletion mutants with G′2 tetraplex integrin DNA. Typical Scatchard plots of the association of constant amounts of full-length or ΔR_3 MyoD with increasing amounts of 5′ end-labeled G′2 integrin DNA are presented in Figure 5. These analyses indicated that in this particular experiment deletion of the

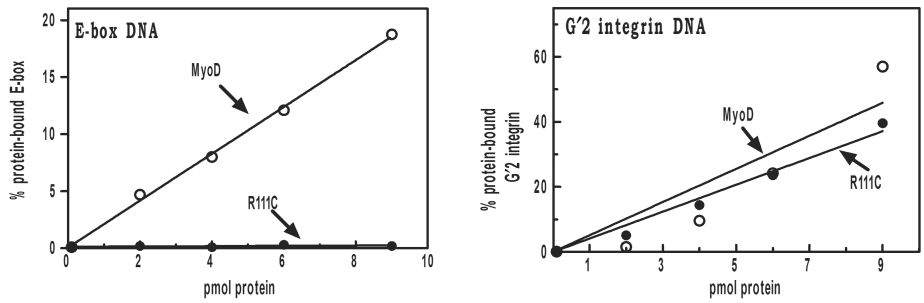

Figure 3. An R111C mutation in MyoD abolishes its E-box-binding activity without affecting the G′2 tetraplex integrin DNA-binding capacity. Full-length or R111C MyoD proteins were bound to 5′-^{32}P labeled E-box or G′2 integrin DNA and protein–DNA complexes were resolved by non-denaturing gel electrophoresis and quantified as detailed in the legend to Figure 2. Presented are plots of percent E-box or G′2 integrin DNA bound as a function of the amounts of added full-length or mutant MyoD.

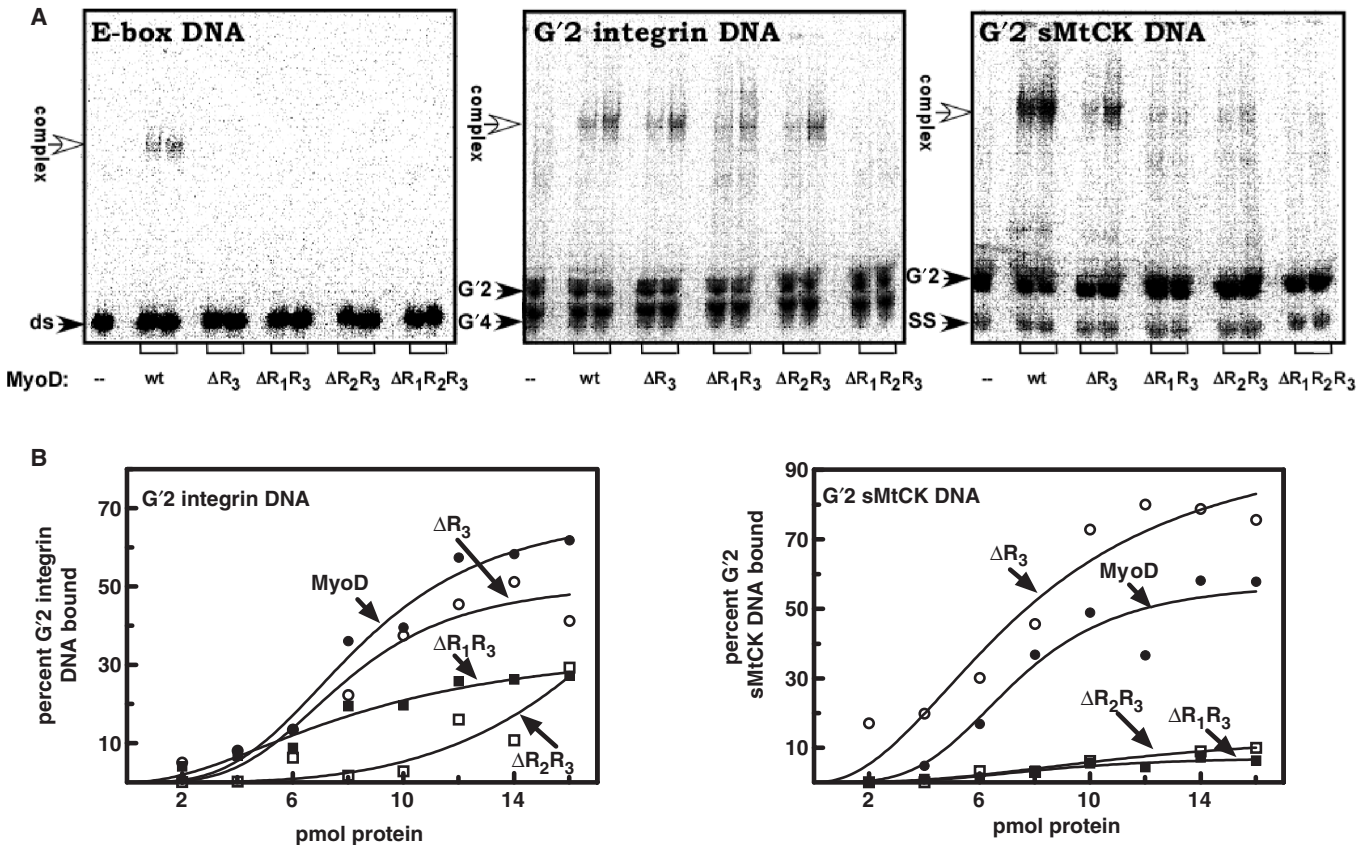

Figure 4. Deletion of basic amino acid triads from the MyoD basic region abolishes binding of E-box but not of G′2 tetraplex DNA. The 5′-^{32}P labeled double-stranded E-box or G′2 tetraplex structures of integrin or sMtCK DNA were bound to different amounts of full-length or the indicated mutant MyoD proteins. Protein–DNA complexes were resolved from free DNA by non-denaturing gel electrophoresis as detailed in the legend to Figure 2. (**A**) Autoradiograms of electrophoretically resolved protein–DNA complexes. Shown are results of DNA binding to 6 and 13 pmol of each examined MyoD protein. (**B**) Quantified results of the binding of increasing amounts of full-length and mutant MyoD proteins to G′2 tetraplex structures of integrin and sMtCK DNA.

R_3 triad of basic amino acids slightly elevated the affinity of MyoD for G′2 integrin DNA. To obtain more complete data, we conducted replicate similar determinations of the K_d values of complexes of full-length and of MyoD deletion mutant proteins with G′2 integrin DNA. Results of these measurements are compiled in Table 3. The measured K_d of 5.8 ± 1.8 nM for complexes of full-length MyoD with G′2 integrin DNA was in the same range as our previously published value of 2.3 ± 1.6 nM for these complexes (5). The measured K_d value of 3.3 ± 1.2 nM of complexes formed by the ΔR3 MyoD mutant (Table 3) indicated that presence of the R_1 and R_2 clusters in the basic region without R_3 was sufficient to maintain an uncompromised affinity of the protein for the tetraplex DNA. Measurements of K_d values of complexes of mutant proteins with deleted pairs of triads revealed that each remaining single basic amino acids cluster sufficed for a relatively tight binding of G′2 integrin DNA.

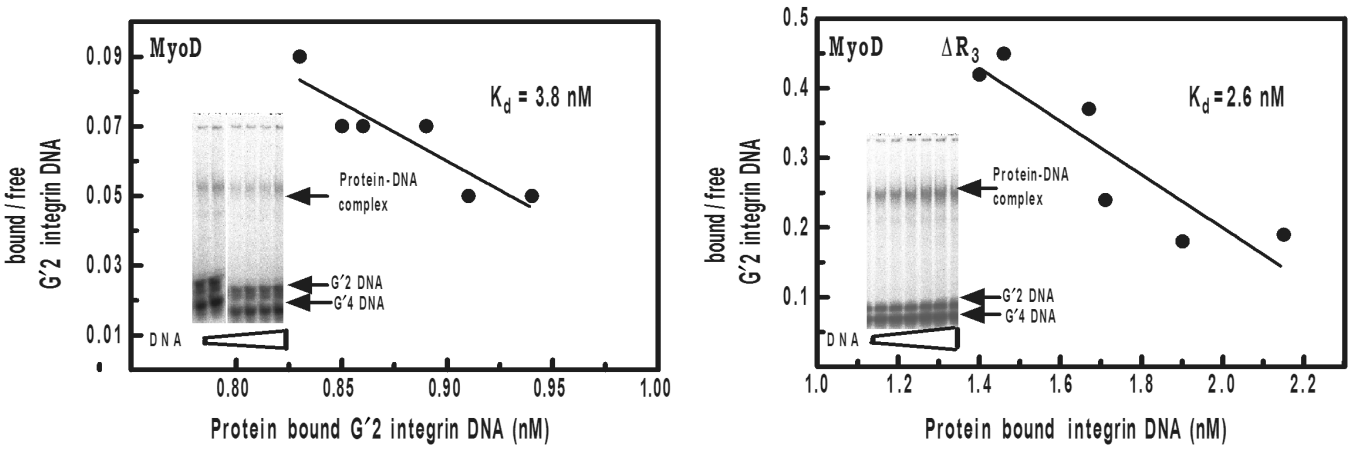

Figure 5. Representative Scatchard plots of the binding of G′2 integrin DNA to full-length and to ΔR_3 mutant MyoD. DNA binding, electrophoretic separation of protein–DNA complexes and their quantification by phosphor imaging were performed as described in the legend to Figure 2. Shown are autoradiograms (insets) and Scatchard plots of the quantified results. The dissociation constants, K_d, were calculated as detailed under Materials and Methods section.

Table 3. MyoD mutants with a remaining single basic amino acid triad maintain moderately reduced affinity for G′2 integrin DNA

Protein	Remaining motifs	K_d of protein-G′2 integrin DNA complex (nM) [N]
Full-length MyoD	R_1; R_2; R_3	5.8 ± 1.8 [6]
ΔR_3 MyoD	R_1; R_2	3.3 ± 1.2 [4]
ΔR_2R_3 MyoD	R_1	26.4 ± 6.9 [4]
ΔR_1R_3 MyoD	R_2	11.8 ± 4.1 [4]
Δ102–114 MyoD	R_3	9.1 ± 1.0 [3]

Dissociation constants, K_d, of complexes of 5′-^{32}P labeled G′2 integrin DNA with the listed MyoD variant proteins were determined as described under Materials and Methods section and in the legend to Figure 5. Listed are average K_d values and SD for the indicated numbers [*N*] of independent determinations.

However, not every cluster contributed equally to MyoD and tetrahelical DNA complex formation. Relative to the full-length protein, MyoD with R_3 as its only existing triad displayed only a 1.6-fold reduction in its affinity for the tetraplex DNA. Proteins that had R_2 or R_1 as their single remaining cluster displayed modest relative diminution of affinity having, respectively, 2.0- and 4.6-fold higher K_d values than full-length MyoD (Table 3).

DISCUSSION

The principal finding of this article is that overlapping but distinct structural elements of MyoD homodimers are employed in the binding of double-stranded E-box or tetraplex structures of promoter sequences of muscle-specific genes. Our results indicated that an intact MyoD basic region is essential for the binding of E-box DNA. The ability of MyoD to form a complex with E-box DNA was completely lost by deleting the N-terminal two-thirds of this region (Figure 2), in the absence of a single or pairs of basic amino triads within this domain (Figure 4A) or by introducing an R to C point mutation at residue 111 (Figure 3). By clear contrast, as illustrated in Figures 2, 3 and 4, MyoD molecules that carried any of these mutations maintained a capacity to associate with G′2 tetraplex structures of the integrin and sMtCK DNA sequences. Notably, the double mutations ΔR_1R_3 and ΔR_2R_3 decreased the binding of G′2 sMtCK DNA to a greater extent than the binding of G′2 integrin DNA (Figure 4B). This difference may be due to the different geometry of the two tetraplexes and their different accommodation within the basic region of MyoD (*vide infra*). Only deletion of all the three basic amino acids triads in the MyoD basic region inactivated its tetraplex DNA-binding capability. Data pointed to any one of the three clusters R_1, R_2 and R_3 of basic amino acids in the basic region of MyoD as an essential element in the binding of tetraplex DNA structures. Thus, the presence of a single cluster of three basic amino acids in a mutated basic region was a necessary and sufficient condition for the binding of the tetrahelical DNA structures (Figure 4 and Table 3).

To evaluate the significance of the R_1, R_2 and R_3 basic clusters, we surveyed the MyoD basic region by applying the Web-based ConSurf 3.0 program which identifies evolutionarily conserved residues in functional domains of proteins (29). Results of the analysis of a database consisting of all MRF proteins as plotted in Figure 6A indicated that except for residues 104 and 112 whose conservation scores could not be significantly determined, other residues that comprised the R_1, R_2 and R_3 triads had scores that ranged between 6 and 9, with 9 being the highest achievable rank. Hence, it appeared that the triads that were necessary for the binding of tetraplex DNA were under strong evolutionary constraints. Figure 6B depicts the crystal structure of the complex of the MyoD bHLH domain with E-box DNA (30) with a color-coded conservation score overlay. These data indicate that the highly conserved R_3 arginine residues 119 and 121 and R_2 arginine 111 maintain direct contact with the DNA. By contrast, none of the residues that comprise the R_1 cluster are in contact with the E-box DNA (Figure 6B). As no crystal structure is available yet of a complex

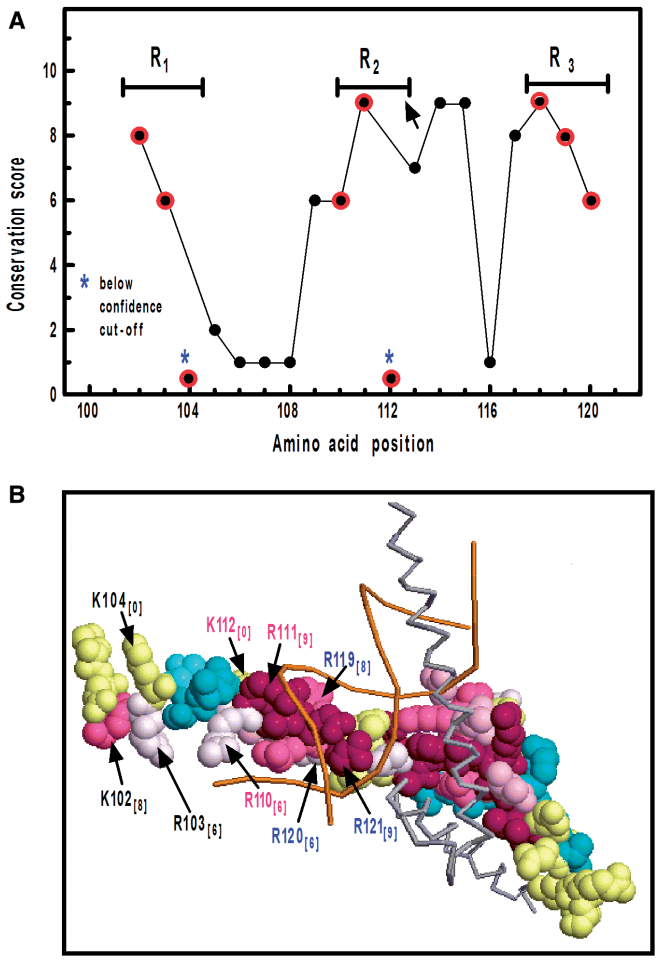

Figure 6. Residues of the three basic triads are highly conserved. (**A**) Plot of the conservation score of amino acids that comprise the MyoD basic region. The conservation score of each residue was obtained by the ConSurf 3.0 Web-based program (see Discussion section). Residues of the R_1, R_2 and R_3 basic clusters are marked in red. Scores of residues 104 and 112 were below the confidence cutoff. (**B**) Conserved residues in the crystal structure of the complex of E-box with the MyoD basic region. Amino acids are color coded in one MyoD monomer according to their conservation score. Residues comprising the basic triads R_1 (black font), R_2 (red font) and R_3 (blue font) are indicated with the conservation score of each logged in parentheses. Also shown are the E-box double helix (orange double ribbon) and the paired MyoD monomer (gray ribbon).

of the MyoD basic region with tetraplex DNA, we used the molecular visualization applications PyMol (Delano Scientific LLC) and DeepView Swiss-Pdb Viewer (Glaxo-Smith and Swiss Institute of Bioinformatics) to superimpose an image of G′2 bimolecular tetraplex structure of the telomeric sequence $(TTAGGG)_2$ on the crystal structure of the MyoD basic region. This modeling suggested that to accommodate the tetraplex DNA, which has wider dimensions than E-box, the dimeric basic region should possess greater flexibility. Thus, for instance, it was observed that the loss of E-box-binding capacity by the R111C mutant (Figure 2) was likely to be due to interference by the substituting cysteine with the positioning of the adjacent R110 residue relative to the E-box. By contrast, the smaller dimensions of cysteine relative to arginine made accommodation of the tetraplex DNA possible. Interestingly, however, although each of the basic triads was sufficient for the binding of tetraplex DNA, G′2 integrin DNA was most tightly bound by mutant MyoD that had as its sole cluster the R_3 triad which is most intimately associated with E-box (Figure 6B). Accordingly, the MyoD mutant whose only cluster was the more remote R_1 triad displayed the weakest association with the tetraplex DNA and the midway positioned R_2 cluster had intermediate affinity for the DNA (Table 3). These results raised the possibility that despite their different geometry, both E-box and tetraplex DNA are similarly positioned most closely to the R_3 triad and most distantly to the R_1 cluster.

Activation by MyoD of the transcription of muscle-specific genes depends on two highly conserved amino acids, alanine at position 114 and threonine at 115, termed the myogenic code (31–33). These two residues together with a lysine in the junction of the first helix of MyoD are sufficient to induce myogenesis (32,34). An A114N mutation was reported to decrease by 3-fold the binding of MyoD homodimers to E-box and to completely abolish transcription activation by MyoD/E47 heterodimers (35). Our results showed that homodimers of the Δ102–114 mutant MyoD protein failed to bind E-box whereas their ability to bind G′2 integrin DNA was minimally affected (Figure 2). This minor effect of the absence of A114 on complex formation with tetraplex DNA contrasted the contribution of this residue to the binding of MyoD homodimers to E-box and its essential role in transcription activation and underscored the different interaction of the two DNA types with MyoD.

We recently proposed that tetrahelical structures in regulatory sequences of muscle-specific genes may trap MyoD homodimers to limit their competition with MyoD/E2A heterodimers on E-box occupancy (5). This idea gains support both by the preferential binding of MyoD homodimers to tetraplex DNA over E-box (5) as well as by the presently reported permissive versus stringent protein structure requirements for their association with tetrahelical DNA and E-box, respectively.

ACKNOWLEDGEMENTS

We thank Dr Noam Adir (Technion, Faculty of Chemistry) for his help with the PyMol and DeepView Swiss-Pdb Viewer modeling programs. This study was supported by grants to M.F. from the Israel Science Foundation and the United States-Israel Binational Science Foundation. Funding to pay the Open Access publication charges for this article was provided by the United States-Israel Binational Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Holstege,F.C., Tantin,D., Carey,M., van der Vliet,P.C. and Timmers,H.T. (1995) The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.*, **14**, 810–819.

2. Liang,C.P. and Garrard,W.T. (1997) Template topology and transcription: chromatin templates relaxed by localized linearization are transcriptionally active in yeast. *Mol. Cell. Biol.*, **17**, 2825–2834.
3. Dexheimer,T.S., Fry,M. and Hurley,L.H. (2006) In Neidle,S. and Balasubramanian,S. (eds), *Quadruplex Nucleic Acids*, RSC Publishing, London, pp. 180–207.
4. Yafe,A., Etzioni,S., Weissman-Shomer,P. and Fry,M. (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res.*, **33**, 2887–2900.
5. Etzioni,S., Yafe,A., Khateb,S., Weisman-Shomer,P., Bengal,E. and Fry,M. (2005) Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes. *J. Biol. Chem.*, **280**, 26805–26812.
6. Tajbakhsh,S. (2003) Stem cells to tissue: molecular, cellular and anatomical heterogeneity in skeletal muscle. *Curr. Opin. Genet. Dev.*, **13**, 413–422.
7. Schnorrer,F. and Dickson,B.J. (2004) Muscle building; mechanisms of myotube guidance and attachment site selection. *Dev. Cell*, **7**, 9–20.
8. Shi,X. and Garry,D.J. (2006) Muscle stem cells in development, regeneration, and disease. *Genes Dev.*, **20**, 1692–1708.
9. Sabourin,L.A. and Rudnicki,M.A. (2000) The molecular regulation of myogenesis. *Clin. Genet.*, **57**, 16–25.
10. Berkes,C.A. and Tapscott,S.J. (2005) MyoD and the transcriptional control of myogenesis. *Semin. Cell Dev. Biol.*, **16**, 585–595.
11. Braun,T., Rudnicki,M.A., Arnold,H.H. and Jaenisch,R. (1992) Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. *Cell*, **71**, 369–382.
12. Rudnicki,M.A., Braun,T., Hinuma,S. and Jaenisch,R. (1992) Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development. *Cell*, **71**, 383–390.
13. Rudnicki,M.A., Schnegelsberg,P.N., Stead,R.H., Braun,T., Arnold,H.H. and Jaenisch,R. (1993) MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell*, **75**, 1351–1359.
14. Nabeshima,Y., Hanaoka,K., Hayasaka,M., Esumi,E., Li,S. and Nonaka,I. (1993) Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature*, **364**, 532–535.
15. Hasty,P., Bradley,A., Morris,J.H., Edmondson,D.G., Venuti,J.M., Olson,E.N. and Klein,W.H. (1993) Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene. *Nature*, **364**, 501–506.
16. Rawls,A., Morris,J.H., Rudnicki,M., Braun,T., Arnold,H.H., Klein,W.H. and Olson,E.N. (1995) Myogenin's functions do not overlap with those of MyoD or Myf-5 during mouse embryogenesis. *Dev. Biol.*, **172**, 37–50.
17. Zhang,W., Behringer,R.R. and Olson,E.N. (1995) Inactivation of the myogenic bHLH gene MRF4 results in up-regulation of myogenin and rib anomalies. *Genes Dev.*, **9**, 1388–1399.
18. Kassar-Duchossoy,L., Gayraud-Morel,B., Gomes,D., Rocancourt,D., Buckingham,M., Shinin,V. and Tajbakhsh,S. (2004) Mrf4 determines skeletal muscle identity in Myf5:Myod double-mutant mice. *Nature*, **431**, 466–471.
19. Davis,R.L., Cheng,P.F., Lassar,A.B. and Weintraub,H. (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell*, **60**, 733–746.
20. Murre,C., McCaw,P.S., Vaessin,H., Caudy,M., Jan,L.Y., Jan,Y.N., Cabrera,C.V., Buskin,J.N., Hauschka,S.D. *et al.* (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell*, **58**, 537–544.
21. Rudnicki,M.A. and Jaenisch,R. (1995) The MyoD family of transcription factors and skeletal myogenesis. *Bioessays*, **17**, 203–209.
22. Tapscott,S.J., Davis,R.L., Thayer,M.J., Cheng,P.F., Weintraub,H. and Lassar,A.B. (1988) MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science*, **242**, 405–411.
23. Lassar,A.B., Davis,R.L., Wright,W.E., Kadesch,T., Murre,C., Voronova,A., Baltimore,D. and Weintraub,H. (1991) Functional activity of myogenic HLH proteins requires hetero-oligomerization with E12/E47-like proteins in vivo. *Cell*, **66**, 305–315.
24. Walsh,K. and Gualberto,A. (1992) MyoD binds to the guanine tetrad nucleic acid structure. *J. Biol. Chem.*, **267**, 13714–13718.
25. 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.
26. 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.
27. Ralser,M., Querfurth,R., Warnatz,H.J., Lehrach,H., Yaspo,M.L. and Krobitsch,S. (2006) An efficient and economic enhancer mix for PCR. *Biochem. Biophys. Res. Commun.*, **347**, 747–751.
28. 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.
29. Landau,M., Mayrose,I., Rosenberg,Y., Glaser,F., Martz,E., Pupko,T. and Ben-Tal,N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.*, **33**, W299–W302.
30. Ma,P.C., Rould,M.A., Weintraub,H. and Pabo,C.O. (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell*, **77**, 451–459.
31. Brennan,T.J., Chakraborty,T. and Olson,E.N. (1991) Mutagenesis of the myogenin basic region identifies an ancient protein motif critical for activation of myogenesis. *Proc. Natl Acad. Sci. USA*, **88**, 5675–5679.
32. Davis,R.L. and Weintraub,H. (1992) Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science*, **256**, 1027–1030.
33. 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.
34. Weintraub,H., Dwarki,V.J., Verma,I., Davis,R., Hollenberg,S., Snider,L., Lassar,A. and Tapscott,S.J. (1991) Muscle-specific transcriptional activation by MyoD. *Genes Dev.*, **5**, 1377–1386.
35. Huang,J., Weintraub,H. and Kedes,L. (1998) Intramolecular regulation of MyoD activation domain conformation and function. *Mol. Cell. Biol.*, **18**, 5478–5484.