Different E-box regulatory sequences are functionally distinct when placed within the context of the troponin I enhancer

Katherine E. Yutzey⁺ and Stephen F. Konieczny^{*}

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

Received June 29, 1992; Revised and Accepted September 9, 1992

ABSTRACT

Basic helix-loop-helix (bHLH) regulatory proteins are known to bind to a single DNA consensus sequence referred to as an E-box. The E-box is present in the regulatory elements of many developmentally controlled genes, including most muscle-specific genes such as troponin I (TnI). Although the E-box consensus is minimally defined as CANNTG, the adjacent nucleotides of functional E-boxes are variable for genes regulated by the bHLH proteins. In order to examine how E-box regulatory regions containing different internal and flanking nucleotides function when placed within the context of a single regulatory element, the E-box region (14 bp) present within the TnI enhancer was substituted with the corresponding E-box sequences derived from the muscle-specific M-creatine kinase (MCK) and cardiac α -actin regulatory elements as well as from the immunoglobulin κ (Ig κ) enhancer. Within the TnI enhancer, the E-box sequence derived from cardiac α -actin was inactive whereas the corresponding sequence from the MCK right E-box efficiently restored wild-type enhancer activity in muscle cells. Intermediate levels of gene activity were observed for TnI enhancers containing E-boxes derived from the MCK left E-box site or from the Ig κ E2 E-box. DNA binding studies of MyoD:E12 protein complexes with each substituted TnI enhancer confirmed that DNA binding activity *in vitro* mimics the relative strength of the enhancers *in vivo*. These studies demonstrate that the specific nucleotide composition of individual E-boxes, which are contained within the regulatory elements of most if not all muscle-specific genes, contributes to the complex regulatory mechanisms governing bHLH-mediated gene expression.

INTRODUCTION

Recently, a family of transcription factors has been defined based on a conserved motif known as the basic helix-loop-helix (bHLH) domain. The bHLH domain consists of ~60 amino acids and is responsible for the DNA binding and specific protein:protein

interactions which characterize this protein family. The most widely studied members of the bHLH family are the four mammalian muscle regulatory factors, MyoD (1), myogenin (2, 3), Myf-5 (4), and MRF4 (5, 6, 7). Each muscle regulatory factor, when over-expressed in fibroblasts, converts cells to a muscle lineage. Fibroblasts expressing the muscle regulatory factors differentiate into myofibers and transcriptionally activate the contractile protein gene set in response to low concentrations of peptide growth factors. Muscle-specific gene expression induced by MyoD, myogenin, Myf-5, or MRF4 is mediated via binding of the proteins to specific DNA consensus sequences referred to as E-boxes, which are contained within the regulatory elements of most contractile protein genes (8–11).

Many bHLH proteins, including the four muscle regulatory factors, bind with similar nucleotide contacts to E-box sequences *in vitro* (12–15). High affinity binding occurs, however, when the muscle regulatory factors form hetero-dimers with the more widely expressed E2A bHLH proteins E12 and E47 (16, 17). The E-box sequence, defined by the nucleotides CANNTG, is contained not only in the regulatory regions of most contractile protein genes but also in other genes regulated by bHLH proteins, including the B-cell specific immunoglobulin genes (18). The E-boxes within muscle-specific regulatory elements are crucial for the correct developmental timing of expression since specific mutations within these sites eliminate muscle-specific gene activation (14, 19–23).

Although the muscle regulatory factors bind to the CANNTG consensus sequence found within contractile protein genes, the nucleotides surrounding and internal to the consensus sequence are widely diverged. For example, of the approximately twenty characterized E-boxes from muscle-specific genes, including troponin I (TnI) (14), M-creatine kinase (MCK) (19), cardiac α -actin (23), myosin light chain (MLC) 1/3 (20, 24), desmin (25), and the acetylcholine receptor subunit genes (21, 22, 26, 27), no two sequences are identical within a ten nucleotide region encompassing the E-box consensus. In addition, aside from the invariant CA and TG nucleotides, neither the internal nucleotides nor the flanking nucleotides are conserved, although the internal nucleotides of most muscle-specific E-boxes consist of two G/C base pairs.

* To whom correspondence should be addressed

⁺ Present address: Department of Cell Biology and Anatomy, Cornell University Medical College, New York, NY 10021, USA

Techniques developed to assess *in vitro* binding of muscle regulatory factors to randomly generated oligonucleotides have been utilized recently in an effort to determine the preferred sequences to which bHLH proteins bind. Wright *et al.* (28) used the cyclic amplification and selection of targets (CASTing) technique to characterize a high affinity binding site consensus for myogenin-containing complexes immunoprecipitated from muscle cell nuclear extracts. Although it is clear that myogenin preferentially associates with the identified consensus sequence, TTGCACCTGTTNNTT, this sequence is not present in any of the E-box regions known to be involved in the regulation of muscle-specific gene expression. Similar techniques also were employed to determine the optimal binding sequence of the MyoD:E12 protein complex (29). Like the myogenin consensus site, this limited consensus, NXCACCTGTT/C (X = A, C, or G), is present in only one of eighteen known functional E-boxes (acetylcholine receptor δ -subunit) (26). Thus, the sequences with the highest binding affinities *in vitro* may not reflect the optimal sequences necessary for contractile protein gene activation.

Among the genes that are regulated by E-box containing elements is the muscle-specific gene troponin I (14). Developmental expression of the TnI gene is dependent on an internal regulatory element (IRE), a 148 bp muscle-specific enhancer located within the first intron (30). Three regions of the IRE are necessary for muscle-specific enhancer activity (14). These sequences include an E-box that interacts with heterodimers of E12 and the muscle regulatory factors plus two additional sites, Site I and Site II, that bind ubiquitously expressed proteins essential to TnI gene activity. Mutation of any one of

the three binding sites results in >85% reduction in TnI enhancer activity, suggesting that multiple protein interactions are required to achieve the proper expression of the TnI gene.

In an effort to determine the dependence of TnI enhancer activity on specific E-box nucleotides, we set out to substitute the E-box sequence of the TnI IRE with different E-box sequences known to be important for the regulation of a number of other genes. Thus, the fourteen nucleotides encompassing the E-box consensus of troponin I were substituted with the corresponding nucleotides from the MCK right and left MyoD binding sites, the cardiac α -actin E-box, and the immunoglobulin κ E2 E-box (19, 23, 31). By comparing the activation of the same regulatory element containing different E-boxes, variability due to interactions with other transcription factors was eliminated. Co-transfection of these mutants with a MyoD expression vector produced different levels of *trans*-activation in 10T1/2 cells, suggesting that the E-box sequence determines the extent to which a regulatory element is activated by muscle regulatory factors. Binding studies conducted *in vitro* also revealed that the affinity of the MyoD:E12 protein complex for each E-box element corresponds to the extent of *trans*-activation *in vivo*. Our studies suggest that the E-box sequence, excluding the binding consensus nucleotides, influences how the muscle regulatory factors interact with a regulatory element. Ultimately, specific interactions of the muscle regulatory factors with a targeted DNA sequence and with additional proteins bound to adjacent regulatory sites must function in concert to produce the appropriate expression pattern for each gene.

MATERIALS AND METHODS

Experimental gene constructions

Test plasmids were constructed using standard molecular techniques (32) and verified directly by dideoxy sequencing of double stranded plasmids (33). Each E-box containing regulatory element was inserted 5' to the basal promoter element of E1b TATA-CAT in a 3'–5' orientation (34). The TnI IRE and TnI mutant IRE were inserted into the Hind III site of E1b TATA-CAT as described previously (14). ActinE1b was generated by adding Hind III linkers to the (minus)CArG fragment (–100 to +68) isolated from the human cardiac α -actin reporter gene pHCA100CAT (35, 36) prior to cloning into E1b TATA-CAT. The MCK enhancer and mutant MCK enhancer were excised by Hind III digestion from p Δ E Δ (37) and from p(B3) Δ E Δ (38), respectively, and then inserted into E1b TATA-CAT to generate MCKE1b and mutMCKE1b. Sequences containing the mouse immunoglobulin κ enhancer (39) were amplified using the polymerase chain reaction (PCR) as described below from 1 μ g mouse genomic DNA in a reaction volume of 100 μ l. Oligonucleotide primers complementary to +185 to +199 and +422 to +435 of the mouse Ig κ enhancer (39) were used for the PCR reaction. After addition of Hind III linkers to the Ig κ enhancer (+185 to +435), the corresponding fragment was cloned into the Hind III site of E1b TATA-CAT to generate Ig κ E1b.

E-box substitutions were produced using the wild-type TnI IRE inserted into the Hind III site of pBluescript KS+ (14). Specific substitutions were introduced into the E-box region (+650 to +663) of the IRE using PCR mutagenesis. Briefly, oligonucleotides containing the desired E-box substitutions within the TnI sequence +642 to +671, as well as an oligonucleotide complementary to the multiple cloning site of pBluescript KS+

E-box Sequences

Troponin I	wildtype	7.5
	GAGACAGCTGCAGC	
Troponin I Mutant	substitutions	5.0
	GAGACAGCAAAAGC	
Cardiac α -actin	GCTCCAAGTGCACC	6.5
MCK Right Site	CCAAGACCTGCTGC	8.0
MCK Left Site	CAGACATGTGGCTG	7.0
Immunoglobulin κ E2	CAGGCAGGTGGCCC	7.5
<hr/>		
Consensus	C ^N GG ^{CA} CC ^{TG} CC ^N C G ^{AA} GG ^{TT} G	8.0

Figure 1. DNA sequences of individual E-boxes that are contained within the troponin I enhancer (14), the cardiac α -actin promoter (23), the M-creatine kinase (MCK) enhancer (19), and the immunoglobulin κ enhancer (44) are shown. Open letters represent identical nucleotides that are present within the troponin I E-box whereas solid letters indicate mismatches. The consensus sequence represents the MEF-1 consensus described by Buskin and Hauschka (38). Scores were determined by assigning the value of 1.0 to the consensus nucleotides (underlined) and a value of 0.5 to the degenerate nonconsensus nucleotides for a maximal possible score of 8.0.

(Stratagene) were hybridized to the pBS-IRE plasmid and subjected to 30 cycles of PCR. PCR reactions contained 1 ng template DNA, 1 μ M of each primer, 200 μ M dNTP, 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 1.5 mM MgCl₂, 0.001% gelatin, and 5 units Taq DNA polymerase (Promega). Each cycle consisted of 1 min at 94°C, 1 min at 40°C, and 1 min at 72°C. Hind III linkers were added to the amplified fragment, which then was digested with Hind III and inserted into the Hind III site of E1b TATA-CAT in a 5'-3' orientation. The oligonucleotides used in these studies consisted of the following sequences contained within the central region of a 30 nucleotide sequence representing TnI +642 to +671 (the core E-box sequence is underlined while lower case letters represent changes in the wild-type DNA sequence): wild type TnI, GAGACAGC TGCAGC; α -actin, GctcCAaCTGaccC; MCKR, ccaACAcC-TGcTGC; MCKL, cAGACAtGgctg; Ig κ cAGgCAGg-TGgccC.

Cell culture

C3H10T1/2 (10T1/2) fibroblasts were obtained from the American Type Culture Collection and maintained in growth medium containing BME (Basal Medium Eagle, Gibco) supplemented with 10% fetal bovine serum plus penicillin (100 units/ml) and streptomycin (100 μ g/ml). The myogenic cell line 23A2 was derived from 5-azacytidine treated 10T1/2 cells as described by Konieczny and Emerson (40). 23A2 myoblasts were maintained in growth medium (BME supplemented with 15% fetal bovine serum) or induced to differentiate by the addition of the defined medium ITS, which consists of low glucose DMEM, 5 μ g/ml insulin, 5 μ g/ml transferrin, 5 ng/ml selenium, and penicillin/streptomycin (30).

Cell transfections and CAT assays

Transient DNA transfections were performed as described previously (30, 35). For *trans*-activation studies, DNA precipitates containing 5 μ g of a CAT reporter gene and 5 μ g of the MyoD expression vector pEMC11s or 5 μ g of the control pEMSVscribe α 2 vector (1) were added to 1×10^6 10T1/2 cells per 100mm dish. Five hours after the addition of DNA, cultures were subjected to an osmotic shock for 2 minutes in serum-free medium containing 20% glycerol and then incubated in normal growth medium. Differentiation medium was added to the cultures 24 hr after transfection. Protein extracts derived from transfected cultures were isolated by standard procedures 48 hours later as described previously (30). Chloramphenicol acetyltransferase (CAT) assays were carried out as described by Gorman *et al.* (41). For *trans*-activation experiments, the amount of cell extract used in each CAT assay was normalized on the basis of protein concentration, which was determined by a modified Bradford assay (Bio-Rad) (42). In most cases, 25 μ g of protein was used for each assay.

23A2 myofibers were transfected by the addition of DNA precipitates containing 5 μ g of a test CAT reporter gene and 5 μ g of the reference plasmid, pRSVLacZ (37) to 1×10^6 myoblasts per 100 mm dish. An osmotic shock was performed as described above, after which the cells were maintained in differentiation medium (ITS) for 48 hr. Myofiber protein extracts were assayed for β -galactosidase activity as described previously (30, 43). The amount of protein extract used for each CAT assay was normalized to the specific activity of the co-transfected β -galactosidase gene product for each transfection group.

In vitro transcription and translation

MyoD and E12 RNA was synthesized *in vitro* using T3 RNA polymerase and a Maxiscript T7/T3 RNA polymerase *in vitro* transcription kit (Ambion) as per manufacturer's specifications. Transcription reactions included 1 μ g linearized pEMC11s (MyoD) (1) or 1 μ g linearized pE12R (E12) (18) DNA, and p¹-5¹-(7-methyl)-guanosine-P³-5'-guanosine triphosphate (m7GpppG) capping solution (Boehringer). After incubation at 37°C for 30 minutes, the reactions were treated with DNase I for an additional 10 minutes. RNA then was extracted with phenol/chloroform and precipitated in ethanol using 1 μ g yeast tRNA as a carrier. A rabbit reticulocyte lysate system (Promega) was used for *in vitro* co-translation of newly synthesized E12 and MyoD RNA. Translation reactions containing ~1 μ g of MyoD and ~1 μ g of E12 RNA were incubated at 30°C for one hour in a total volume of 50 μ l. Aliquots of the reactions were stored at -80°C.

Gel mobility shift assays

To generate ³²P-labelled DNA probes, each IRE mutant was excised from the E1b TATA-CAT vectors using Hind III, treated with alkaline phosphatase (Boehringer), and gel purified. Approximately 50 ng of each purified DNA fragment was labelled by filling-out Hind III digested ends using the Klenow fragment of DNA polymerase I (New England Biolabs) and high specific activity (6000 Ci/mMol) [α -³²P] ATP and [α -³²P] CTP (Amersham). Unincorporated nucleotides were removed after passage through a G-50 (Pharmacia) spin column.

For protein: DNA binding assays, end-labelled DNA probes (~10,000 cpm) were incubated with 0.5 μ g poly (dI-dC) and 5 μ l of *in vitro* translated protein products at room temperature for 20 min. Binding reactions also contained 25 mM HEPES (pH 7.9), 50 mM KCl, 0.5 mM DTT, 5 mM MgCl₂, and 10% glycerol. Samples were loaded onto 5% polyacrylamide gels (30:1 acrylamide:bisacrylamide) containing 2.5% glycerol and electrophoresed at 8 V/cm for 3.5 hr at room temperature with buffer recirculation in 12.5 mM Tris-Cl (pH 8.5), 95 mM glycine, and 0.5 mM EDTA. After electrophoresis, gels were fixed in 10% methanol, 10% acetic acid for 10 min, rinsed in water, vacuum dried, and exposed to Kodak XAR film.

RESULTS

Analysis of the E-box regions associated with the contractile protein gene enhancers TnI (14) and MCK (19), the cardiac α -actin promoter (23) and an immunoglobulin κ gene enhancer (44) revealed that although each contains a defined E-box consensus (CANNTG), the central and flanking nucleotides are not conserved (Figure 1). Based on a number of known functional E-boxes, Buskin and Hauschka (38) assigned a numerical value to each nucleotide within a fourteen base pair region surrounding the core E-box sequence and thereby proposed an E-box consensus that takes into account nucleotides flanking the four core positions. Each core nucleotide was assigned a value of 1.0 whereas matches with either of the degenerate flanking nucleotides were given a value of 0.5. The maximal match to the regional consensus is 8.0, which is the score attributed to the MCK right site (Figure 1). The TnI and Ig κ E2 E-boxes contain single mismatches from the regional consensus, giving each a score of 7.5. The MCK left site and the cardiac α -actin E-boxes are more widely diverged with two (score of 7.0) and

Figure 2. E-box containing regulatory elements are *trans*-activated by MyoD. The indicated regulatory sequences from the troponin I (14, 30), cardiac α -actin (35, 36), M-creatine kinase (MCK) (37, 38) and immunoglobulin κ (39) genes were ligated in a 3'–5' orientation relative to the E1b TATA-CAT gene. Each gene construct was co-transfected with the MyoD expression vector pEMC11s into 10T1/2 fibroblasts which subsequently were induced to differentiate in ITS medium. CAT activity is expressed as a percentage of the wild-type IRE co-transfected with MyoD in parallel transfections for each experiment. Each value represents the average of 2–4 independent transfections based on 25 μ g protein. Basal levels of activity in control pEMSVscribe α 2 transfections with each construct were less than 0.05% acetylated chloramphenicol/ μ g protein.

three (score of 6.5) mismatches, respectively. The relative strengths of these E-boxes within a specific regulatory element in promoting a transcriptional response have not been determined.

MyoD-induced expression of native E-box regulatory elements

Each E-box containing regulatory element is composed of a unique array of protein binding sites with which the bHLH proteins and other transcription factors interact. In some cases, such as for the MCK and Ig κ enhancers, multiple E-box sites exist whereas single E-boxes are present within the TnI and cardiac α -actin regulatory elements (Figure 2A). Additionally, the E-box sequences are distinct within these regulatory elements suggesting that different combinations of bHLH proteins, especially in the case of the Ig κ enhancer, may interact with these sites *in vivo*. Thus, it is likely that within a given regulatory

element, each E-box plays a slightly different role in the activation of gene expression. In order to compare directly how specific E-boxes function within their native regulatory elements, the enhancer regions associated with the TnI, MCK, and Ig κ genes, and the promoter region associated with the cardiac α -actin gene, were ligated in a 3'–5' orientation to the E1b TATA-CAT reporter gene (34) which normally is inactive in 10T1/2 cells (14). In addition to the wild-type regulatory regions, an MCK enhancer containing a mutated right E-box was examined for enhancer activity. The resulting mutant MCK enhancer contains the sequence CCAACACGGTAACC in place of the wild-type MCK right site E-box (38). A similar mutation of the TnI E-box consensus sequence also was generated for these studies (Figure 1). Co-transfection of 10T1/2 cells with each E1b TATA-CAT construct and with a MyoD expression vector allowed us

Table 1. Expression of E-box regulatory elements and TnI E-box substitutions in 23A2 myofibers

A. E-box Regulatory Elements		Relative Expression ^a
TnI		100.0
muTnI		14.5
Actin		317.9
MCK		392.3
muMCK		18.9
Ig κ		12.9
B. TnI E-box Substitutions		Relative Expression
TnI		100.0
muTnI		16.6
Actin		9.3
MCKR		137.9
MCKL		29.5
Ig κ E2		41.9

^aTranscriptional activity for each reporter gene construct described in Figures 2 and 3 was determined in transfected 23A2 myofibers. (A) Native E-box containing regulatory elements; (B) Individual E-box sequences substituted into the TnI IRE. CAT activity is expressed as a percentage of the activity of the wild-type IRE in parallel transfections which was set at 100%. Each value represents the average of 2–4 independent transfections. The amount of protein used in each assay was normalized to the activity of a co-transfected RSVlacZ gene. Basal levels of expression from the E1b TATA-CAT gene were approximately 8% of the activity of the TnI IRE TATA-CAT gene construct.

to determine whether the different reporter gene constructs were activated to similar levels by the bHLH transcription factor MyoD.

The levels of expression of the E1b TATA-CAT reporter genes containing each of the regulatory elements was at or below the minimum level of CAT enzyme detection when co-transfected with the pEMSVscribe α 2 control plasmid into 10T1/2 cells. 10T1/2 cells co-transfected with the TnI enhancer (IRE) reporter gene and the MyoD expression vector, however, exhibited a ~100-fold increase in CAT expression when compared to the pEMSVscribe α 2 control (Figure 2B). Mutation of the TG nucleotide pair of the TnI E-box consensus, along with the adjacent flanking nucleotide eliminated >95% of the observed *trans*-activation obtained with the wild-type TnI enhancer sequence. In contrast, the cardiac α -actin promoter element, when ligated in a 3'–5' orientation adjacent to the E1b TATA promoter, was *trans*-activated by MyoD to levels that were at least eight-fold higher than the *trans*-activation levels detected with the wild-type TnI enhancer (Figure 2B). Thus, while both TnI and cardiac α -actin regulatory elements contain single E-boxes, the cardiac α -actin promoter is *trans*-activated more efficiently by MyoD than the TnI enhancer element when placed adjacent to the E1b TATA-CAT reporter gene. Co-transfection of the wild-type MCK enhancer with MyoD also produced *trans*-activation levels that were approximately 5-fold higher than the wild-type TnI IRE. The strong levels of *trans*-activation attributed to the MCK enhancer were greatly reduced, however, when the MCK right E-box was mutated (Figure 2B). The right site has been described as the stronger of the two MCK E-boxes (19) and alteration of six nucleotides, including the TG consensus nucleotides, reduced the MyoD *trans*-activation response of the MCK enhancer to <5% of the original expression level, despite the presence of an intact MCK left E-box.

Although Ig κ gene expression *in vivo* is B-cell specific, Ig κ enhancer activity can be induced in certain cell types, such as

yeast, in response to over-expression of the E2A bHLH protein gene products (45). The Ig κ enhancer contains three binding sites for bHLH proteins, the strongest of which is the Ig κ E2 site (Figure 2A) since mutation of this site essentially eliminates enhancer function (44). In addition to the three E-boxes, the Ig κ enhancer contains a binding site for NF κ B which also is involved in B-cell specific expression (44). As expected, the Ig κ enhancer is silent in 10T1/2 cells co-transfected with the control pEMSVscribe α 2 vector, demonstrating that the endogenous levels of E2A proteins in 10T1/2 cells are not sufficient for the activation of the Ig κ reporter gene. However, E1b TATA-CAT reporter genes containing the Ig κ enhancer were activated approximately 100-fold when co-transfected with a MyoD expression plasmid (Figure 2B). The level of expression from the Ig κ enhancer was similar to that of the muscle-specific TnI enhancer, suggesting that in the presence of high levels of MyoD, the Ig κ enhancer is transcriptionally activated in converted muscle cells. These experiments demonstrate that the muscle-specific regulatory elements associated with the TnI, MCK, and cardiac α -actin genes, along with the enhancer element associated with the B-cell specific Ig κ gene, are activated by bHLH proteins but are unique in how they respond to high levels of MyoD.

In order to directly compare muscle-specific gene activation, each E-box containing regulatory element also was introduced into the myogenic cell line 23A2 (40). While 23A2 cells express MyoD, the levels of MyoD per cell are at least 20-fold lower in 23A2 myofibers than in MyoD-transfected 10T1/2 cells (35). Transfection of the TnI, MCK, and cardiac α -actin muscle-specific regulatory elements into 23A2 myofibers produced relative levels of expression that were similar to the levels obtained in the MyoD co-transfections of 10T1/2 cells (Table 1). For example, expression from the MCK enhancer and cardiac α -actin promoter (3'–5') was 3–4 fold higher than expression from the TnI enhancer, whereas mutations in the consensus nucleotides of the TnI E-box or mutations in the MCK right site essentially eliminated muscle-specific expression (Table 1). Interestingly, the B-cell specific Ig κ gene enhancer was relatively inactive in 23A2 myofibers, suggesting that high levels of MyoD expression in 10T1/2 cells are required for its activation, whereas in 23A2 myofibers B-cell specificity is preserved. These studies indicate that regulatory elements containing different E-box sequences display unique levels of CAT expression in both MyoD co-transfections and in 23A2 myofibers. Additionally, the Ig κ enhancer is differentially active in the two experimental systems.

Transcriptional activity of TnI enhancers containing specific E-box substitutions

In order to determine how individual E-box sequences affect the extent of gene activation, E-box substitutions were made in the TnI enhancer. The TnI enhancer was chosen for these studies because it contains a single E-box that must be linked to Sites I and II, which contain binding sites for ubiquitously expressed proteins (Figure 3A) (14). The ability of different E-box sequences to potentiate TnI enhancer activity was compared directly by replacing the E-box sequence of the TnI IRE with E-box sequences from the cardiac α -actin promoter, the MCK right and left sites as well as with the Ig κ E2 site. Five to seven point mutations were introduced within the fourteen base pair TnI E-box region in order to produce a sequence which was identical to the other E-box elements (Figure 1). As discussed above, within their native regulatory elements, the different E-boxes are responsible for different levels of gene expression, and

Figure 3. Troponin I (TnI) IREs containing specific E-box substitutions are *trans*-activated by MyoD. TnI IREs mutated to contain E-boxes from different origins as described in Figure 1 were ligated in a 5'–3' orientation relative to the E1b TATA-CAT reporter gene. Each gene construct was co-transfected into 10T1/2 cells with the MyoD expression vector as described in Figure 2. CAT activity is expressed as a percentage of the wild-type IRE co-transfected into 10T1/2 cells with MyoD in parallel transfections for each experiment. All constructs maintain the wild-type TnI Site I and Site II sequences. Note that the scale has been expanded when compared to Figure 2. Each value represents the average of 4 independent transfections based on 25 μ g protein. Basal levels of activity in control pEMSVscribe α 2 transfections for each construct were less than 0.02% acetylated chloramphenicol/ μ g protein.

the additional proteins that also bind to these elements likely play different roles in transcriptional activation. Thus, by changing the E-box sequence but preserving the precise spacing of E-box nucleotides with respect to Site I and Site II of the TnI enhancer, we were able to test whether the composition of nucleotides flanking the E-box consensus is important for gene activation. In addition, these studies facilitated a direct comparison of different E-boxes to determine whether specific sequences were more efficient than others in activating gene expression within a common genetic element.

To assess the relative abilities of E-boxes from different sources to function in TnI gene activation, 10T1/2 cells were co-

transfected with the MyoD expression vector and each mutant TnI IRE linked to E1b TATA-CAT. None of the substituted IREs was expressed in 10T1/2 cells in the absence of MyoD, confirming that the bHLH proteins endogenous to 10T1/2 cells are not sufficient by themselves to activate any of the troponin I E-box substituted IREs, including the immunoglobulin κ E2 E-box IRE. As in the previous experiments, co-transfection of the wild-type TnI IRE with MyoD resulted in *trans*-activation ~100 fold over the levels obtained in the pEMSVscribe α 2 control transfections. Surprisingly, the cardiac α -actin E-box, which is essential to the muscle-specific activation of the cardiac α -actin promoter (23), is not functional when placed within the context

Figure 4. MyoD:E12 protein complexes bind to the TnI E-box substitutions. TnI IREs with E-box substitutions (Figure 3) were used as DNA probes in gel mobility shift assays. MyoD:E12 proteins were co-translated in a rabbit reticulocyte lysate system from RNA transcribed *in vitro* and incubated with 32 P-labelled IRE probes as described in Materials and Methods. The specific MyoD:E12:DNA complexes are indicated.

of the TnI enhancer element (Figure 3B), even though the α -actin promoter linked to the E1b reporter gene is activated approximately 8-fold higher than the TnI enhancer (Figure 2B). The TnI IRE substituted with the cardiac α -actin E-box produced levels of expression comparable to the TnI mutant E-box consensus sequence which, as discussed previously, is not efficiently *trans*-activated by MyoD. These results suggest that substitution of the TnI E-box flanking nucleotides with the same nucleotides present within the α -actin E-box renders the IRE nonfunctional. Unlike the cardiac α -actin E-box, however, the relative strengths of the two MCK E-boxes were conserved when positioned within the TnI IRE. Substitution of the MCK right E-box site resulted in levels of *trans*-activation similar to the wild-type IRE (>90%), while substitution of the MCK left E-box site produced levels of *trans*-activation that were ~75% lower than the wild-type TnI or MCK right site IREs (Figure 3B). Thus the MCK right E-box remains the stronger of the two sites within the TnI IRE as well as within the MCK upstream enhancer.

Unlike the MCK right site, which can substitute efficiently for the TnI E-box, the Ig κ E2 E-box is not functionally equivalent to the TnI E-box. Although both the Ig κ E2 and TnI sites contain only single mismatches of the E-box consensus sequence, replacement of the TnI E-box with the Ig κ E2 sequence led to a ~60% decrease in gene activity (Figure 3B). Thus the nucleotide changes in the Ig κ E2-IRE, unlike the MCKR-IRE, are non-conservative substitutions. While the Ig κ E2 element is integral to the function of the Ig κ enhancer (44), it can not substitute efficiently for the E-box in the TnI IRE. However, the Ig κ E2 site is stronger than the MCK left site or the cardiac α -actin site when these are placed within the context of the TnI enhancer, suggesting that the Ig κ E2 site itself can function in the context of a muscle-specific element. Similar relative levels

of *trans*-activation also were observed when a myogenin expression vector was co-transfected into 10T1/2 cells with each reporter gene construct (data not shown). Thus each IRE E-box substitution produces a unique level of *trans*-activation by MyoD, suggesting that E-box sequences outside the CANNTG core consensus play an important role in determining the extent of gene activation *in vivo*.

The ability of the E-box substituted TnI IRE constructs to support enhancer activity also was tested in the myogenic cell line 23A2. As shown in Table 1, introduction of each mutant into 23A2 myofibers produced similar relative levels of gene activation as obtained by co-transfecting 10T1/2 cells with MyoD. For example, the MCK right site was functionally equivalent to the native TnI E-box sequence within the IRE. However, the presence of the cardiac α -actin E-box produced levels of expression comparable to the low levels obtained with the mutant TnI IRE (muTnI). Similarly, the MCK left site E-box also supported very low levels of gene activation when placed within the context of the TnI enhancer. Interestingly, substitution of the Ig κ E2 site into the TnI IRE produced moderate levels of muscle-specific gene activity as was observed for MyoD co-transfections, even though reporter genes containing the complete Ig κ enhancer are not efficiently expressed in differentiated myofibers (Table 1). Thus, within the context of the muscle-specific TnI enhancer, unlike within its native enhancer, the Ig κ E2 E-box is similarly effective in 23A2 myofibers and in 10T1/2 MyoD co-transfections. The ability of different E-boxes to function within the IRE enhancer element suggests that some flexibility in the sequences within and flanking the E-box core consensus exists. However, within this context, not all changes are equally recognized since some substitutions, especially those that produce the cardiac α -actin E-box, can greatly inhibit the activation of gene expression by the muscle regulatory factors.

TnI E-box substitutions exhibit variable binding of MyoD:E12 complexes

Although previous studies have shown that the muscle regulatory factors efficiently bind to various E-box sequences, we next examined whether the inability of some E-boxes to support transcription when positioned within the TnI IRE was due to an inability to generate stable protein:DNA complexes. To examine this possibility, binding of MyoD:E12 hetero-dimers to each E-box sequence within the context of the TnI IRE was assessed by gel mobility shift assays. MyoD:E12 proteins generated from mRNA transcribed and translated *in vitro* were incubated with radiolabelled mutant IREs and subjected to native gel polyacrylamide electrophoresis. As expected, no binding was detected when various IREs were incubated with MyoD or with E12 alone (data not shown, 14). However, MyoD:E12 protein complexes readily bound to the native TnI IRE while mutation of the E-box consensus nucleotides completely eliminated MyoD:E12 binding (Figure 4). Very strong binding of MyoD:E12 complexes to the IRE containing the MCK right site (MCKR) also was observed. Thus, the wild-type TnI and MCK right E-boxes, which support the highest levels of *trans*-activation when placed within the TnI IRE, also exhibit efficient binding of MyoD:E12 complexes *in vitro*. The ~60% lower levels of *trans*-activation obtained with the Ig κ substituted IRE also was reflected in a lower intensity of MyoD:E12 binding to the Ig κ E-box contained within the TnI IRE. In contrast, the MCK left site, which when substituted into the TnI IRE supported only very

low levels of gene expression in MyoD transfected 10T1/2 cells, inefficiently interacted with MyoD:E12 complexes since binding activity in these assays was apparent only with prolonged exposures of the autoradiograph. No MyoD:E12 binding was detected for the cardiac α -actin E-box IRE which contains an intact E-box consensus but was *trans*-activated to the same low background levels as observed with the TnI IRE mutant E-box consensus. Thus the binding activities of the mutant IREs *in vitro* correlate with their ability to support *trans*-activation by MyoD *in vivo*, suggesting that the nucleotide substitutions that were generated within the TnI IRE E-box significantly affect how the muscle regulatory factors interact with each regulatory element.

DISCUSSION

In these studies we have demonstrated that E-boxes consisting of different nucleotide sequences and derived from different genetic origins display distinct activities when placed within the context of a single regulatory element. Although each tested E-box is functional within its native gene, the response to over-expression of the bHLH protein MyoD was distinct for each sequence. Since each of the E-boxes contain the CANNTG core consensus, the differences in their *trans*-activation potentials when placed within the IRE demonstrate the importance of the flanking and internal nucleotides in determining E-box function. In addition, the ability of the E-box binding proteins to interact with other proteins that are essential to TnI gene activation also may be affected by specific E-box sequences. Thus the multiple protein:DNA interactions involved in activating E-box containing regulatory elements may require different binding properties and therefore different E-box target sites for the activation of specific genes.

The ability of some E-boxes to function more efficiently than others within the context of the TnI enhancer suggests that the nucleotide changes represent both conservative and non-conservative substitutions. For example, the five nucleotide changes that are represented in the MCK right E-box are comparable in activity to the native TnI E-box in supporting muscle gene activation. However, the six different nucleotide changes that are represented in the Ig κ E2 site produced a 60% reduction in enhancer activity. Interestingly, the TnI and Ig κ E-boxes are equally diverged from the consensus sequence based on E-box nucleotide similarities (see Figure 1), suggesting that the degree of similarity to the consensus E-box sequence may not be fully indicative of enhancer function. However the MCK left site, which diverges from the consensus more than the Ig κ E-box, produced a 70% reduction in TnI gene activity, while the cardiac α -actin E-box, which contains the least homology to the E-box consensus, displayed only background expression levels. Thus there may be a correlation between divergence from the consensus sequence at this level and the strength of the E-box sequence when placed within the context of the TnI enhancer.

Perhaps the most striking difference between E-box sequences that function within the TnI enhancer and those that do not is the presence of two G/C base pairs found between the consensus nucleotides of the TnI, MCKR, and Ig κ E2 sites and the presence of a single A/T base pair in the weaker MCKL and cardiac α -actin E-boxes. DNA binding studies have indicated that the base pairs internal to the consensus are important contact points for MyoD:E12 interactions (13–15). In addition, substitution of the two internal G/C base pairs of the IgH μ E5 E-box with two A/T base pairs strongly inhibits binding and activity of the μ E5 site

(45). Consensus sequences established by CASTing and SAABing techniques also have revealed a preference for internal G/C base pairs within the consensus nucleotides (28, 29). However, the presence of two G/C nucleotides is not sufficient to achieve complete activity since the Ig κ E2 site, when positioned within the TnI enhancer, resulted in a decrease in activity despite the presence of CAGGTG. A comparison of the E-boxes tested in this study also has revealed a second short palindromic sequence that shares the G nucleotide of the CANNTG palindrome (see Figure 1). In the TnI and MCKR sites the sequence is GCNGC and in the Ig κ E2 site the sequence is GGNCC. The MCKL and cardiac α -actin E-boxes, which display the lowest levels of activity within the TnI enhancer, do not contain a second palindromic sequence. These adjacent sequences could serve to stabilize protein binding to the E-box although the significance of these sequences has yet to be determined. In addition, we have been unable to detect a correlation between the predicted melting temperatures of the fourteen base pair E-box regions and their ability to function within the TnI enhancer.

Protein:DNA binding studies presented here and elsewhere suggest that the MyoD:E12 hetero-dimers bind with different affinities to different E-box sequences. For example, within the MCK enhancer it has been shown that MyoD and myogenin bind more readily to the right E-box site than to the left site (13, 19). Different binding affinities also have been observed for the different E-boxes within the myosin light chain 1/3 enhancer and the immunoglobulin μ enhancer (15, 18). In addition, we have shown that within the same regulatory element, different E-boxes have distinct binding affinities for MyoD. These studies demonstrate that despite the presence of the E-box consensus sequence in each of these binding sites, the adjacent nucleotides affect how the bHLH proteins interact with these sequences within the TnI enhancer and their ability to bind directly corresponds to the level of *trans*-activation obtained with each substituted IRE. However the binding studies *in vitro* do not take into account the interactions of additional proteins bound to individual regulatory elements that could also affect the stability of these interactions *in vivo*. For example, the apparent inability of MyoD to bind or *trans*-activate the cardiac α -actin E-box when placed within the TnI enhancer may not reflect how MyoD interacts with the native cardiac α -actin promoter *in vivo*, since within this context the E-box is necessary for the muscle-specific activation of cardiac α -actin gene expression (23). Thus the differential *trans*-activation of TnI enhancers containing specific E-boxes may reflect the ability of MyoD to activate genes containing different E-box sequences as well as the ability of MyoD to interact with other transcription factors bound to the same regulatory element in muscle cells (14).

The interactions of bHLH proteins with different E-box sequences have been shown to be important for the regulation of a wide variety of genes including the contractile protein (reviewed in 8), immunoglobulin (44), pancreatic (46), and SV40 (47) genes. The diversity of E-box proteins and of E-box sequences suggests that the different patterns of gene expression from these diverse genes could be established at least in part by specific E-box nucleotides. In addition, for many of these genes other proteins that bind to adjacent sequences also are required for their complete activation. Thus the specificity between bHLH proteins and their target sequences may have evolved to optimize these interactions. The studies presented here support this hypothesis since within the context of the same regulatory element, E-boxes from different sources are not equivalent.

ACKNOWLEDGEMENTS

We thank Haishan Lin, Tim Hinterberger, Yanfeng Kong, Lea Longcor, and Jennifer Mays for technical assistance, John Anderson and Ron Breaker for helpful discussions and Connie Philbrook for aiding in the preparation of this manuscript. K.E.Y. was supported by a pre-doctoral fellowship granted by the Indiana affiliate of the American Heart Association. This work was funded by an NIH grant (HD25518) to S.F.K. S.F.K. is an Established Investigator of the American Heart Association.

REFERENCES

- Davis, R.L., Weintraub, H., and Lassar, A. B. (1987) *Cell* 51:987–1000.
- Wright, W.E., Sassoon, D.A., and Lin, V.K. (1989) *Cell* 56:607–617.
- Edmondson, D.G., and Olson, E. N. (1989) *Genes Dev.* 3:628–640.
- Braun, T., Buschhausen-Denker G., Bober E., Tannich E., and Arnold, H.H. (1989) *EMBO J.* 8:701–709.
- Rhodes, S.J., and Konieczny, S.F. (1989) *Genes Dev.* 3:2050–2061.
- Miner, J. H., and Wold, B. (1990) *Proc. Natl. Acad. Sci. USA* 87:1089–1093.
- Braun, T., Bober, E., Winter, B., Rosenthal, N., and Arnold, H. H. (1990) *EMBO J.* 9:821–831.
- Olson, E.N. (1990) *Genes Dev.* 4:1454–1461.
- Emerson, C. P. (1990) *Cell Diff.* 2:1065–1075.
- Tapscott, S. J., and Weintraub, H. (1991) *J. Clin. Invest.* 87:1133–1138.
- Wright, W. E. (1992) *Curr. Opin. Genetics and Devel.* 2:243–248.
- Murre, C., McCaw, P. Schonleber, Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J. N., Hauschka, S.D., Lassar, A.B., Weintraub, H., and Baltimore, D. (1989) *Cell* 58:537–544.
- Brennan, T.J., and Olson, E. N. (1990) *Genes Dev.* 4:582–595.
- Lin, H., Yutzey, K.E., and Konieczny, S.F. (1991). *Mol. Cell. Biol.* 11:267–280.
- Braun, T., and Arnold, H. H. (1991) *Nucl. Acids Res.* 19:5645–5651.
- Lin, H., and Konieczny, S. F. (1992) *J. Biol. Chem.* 267:4773–4780.
- Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D., and Weintraub, H. (1991) *Cell* 66:305–315.
- Murre, C., McCaw, P. Schonleber, and Baltimore, D. (1989) *Cell* 56:777–783.
- Lassar, A.B., Buskin, J. N., Lockshon, D., Davis, R.L. Apone, S. Hauschka, S.D., and Weintraub, H. (1989) *Cell* 58:823–831.
- Wentworth, B.M., Donoghue, M., Engert, J.C., Berglund, E., and Rosenthal, N. (1991) *Proc. Natl. Acad. Sci. USA* 88:1242–1246.
- Piette, J., Bessereau, J.-L., Huchet, M., and Changeux, J.-P. (1990) *Nature* 345:353–355.
- Prody, C.A., and Merlie, J.P. (1991) *J. Biol. Chem.* 266:22588–22596.
- Sartorelli, V., Webster, K.A., and Kedes, L. (1990) *Genes Dev.* 4:1811–1822.
- Rosenthal, N., Berglund, E.B., Wentworth, B.M., Donoghue, M., Winter, B., Bober, E., Braun, T., and Arnold, H.-H. (1990). *Nucl. Acids Res.* 18:6239–6246.
- Li, Z., and Paulin, D. (1991) *J. Biol. Chem.*, 266:6562–6570.
- Baldwin, T.J., and Burden, S.J. (1989) *Nature*, 341:716–720.
- Numberger, M., Durr, I., Kues, W., Koenen, M., and Witzemann, V. (1991) *EMBO J.* 10:2957–2964.
- Wright, W.E., Binder, M., and Funk, W. (1991) *Mol. Cell. Biol.* 11:4104–4110.
- Blackwell, T.K., and Weintraub, H. (1991) *Science*, 250:1104–1110.
- Yutzey, K.E., Kline, R.L., and Konieczny, S.F. (1989) *Mol. Cell. Biol.* 9:1397–1405.
- Sen, R., and Baltimore D. (1986) *Cell* 46:705–716.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidmon, J. G., Smith, J. A., and Struhl, K. eds. (1987) *Curr. Protocols in Mol. Biol.* Vol 2. John Wiley and Sons, Inc., New York.
- Chen, E., and Seeburg, P. H. (1985) *DNA* 4:165–170.
- Lillie, J.W., and Green, M. R. (1989) *Nature* 338:39–44.
- Yutzey, K.E., Rhodes, S.J., and Konieczny, S.F. (1990) *Mol. Cell. Biol.* 10:3934–3944.
- Miwa, T., and Kedes, L. (1987) *Mol. Cell. Biol.* 7:2803–2813.
- Jaynes, J.B., Johnson, J. E., Buskin, J. N., Gartside, C. L., and Hauschka, S.D. (1988) *Mol. Cell. Biol.* 8:62–70.
- Buskin, J.N., and Hauschka, S. D. (1989) *Mol. Cell. Biol.* 9:2627–2640.
- Picard, D., and Schaffner, W. (1984) *Nature* 307:80–82.
- Konieczny, S.F., and Emerson, C.P. Jr. (1984) *Cell*, 38:791–800.
- Gorman, C.M., Moffat, L. F., and Howard, B. H. (1982) *Mol. Cell. Biol.* 2:1044–1051.
- Bradford, M.M. (1976) *Anal. Biochem.*, 72:248–254.
- Nielsen, D.A., Chou, J., MacKrell, A. J., Casadaban, M.J., and Steiner, D. F. (1983) *Proc. Natl. Acad. Sci. USA* 80:5198–5202.
- Lenardo, M., Pierce J.W., and Baltimore, D. (1987) *Science*, 236:1573–1577.
- Ruezinsky, D., Beckmann, H., and Kadesch, T. (1991) *Genes Dev.* 5:29–37.
- Nelson, C., Shen, L.-P., Meister, A., Fodor, E., Rutter, W. J. (1990) *Genes Dev.* 4:1035–1043.
- Hu, Y.-F., Luscher, B., Admon, A., Mermod, N., and Tjian, R. (1990) *Genes Dev.* 4:1741–1752.