Am J Physiol Cell Physiol 291: C300–C307, 2006.
First published March 22, 2006; doi:10.1152/ajpcell.00345.2005.

An E box in the exon 1 promoter regulates insulin-like growth factor-I expression in differentiating muscle cells

A. S. McLellan,[1] T. Kealey,[1,2] and K. Langlands[1,2]
[1]*University of Cambridge, Department of Clinical Biochemistry, Addenbrooke's Hospital, Cambridge; and* [2]*Clore Laboratory, University of Buckingham, Buckingham, United Kingdom*

Submitted 12 July 2005; accepted in final form 17 February 2006

McLellan, A. S., T. Kealey, and K. Langlands. An E box in the exon 1 promoter regulates insulin-like growth factor-I expression in differentiating muscle cells. *Am J Physiol Cell Physiol* 291: C300–C307, 2006. First published March 22, 2006; doi:10.1152/ajpcell.00345.2005.—Insulin-like growth factor (IGF)-I expression is subject to complex temporal and spatial regulation. Endocrine synthesis occurs in the liver, where transcription is initiated from promoters located in either exon 1 (P1) or in exon 2 (P2), whereas local transcription is mainly initiated from P1. IGF-I is expressed in a range of tissues and, in particular, is an important regulator of skeletal muscle mass, although the mechanisms of tissue-specific regulation remain to be fully characterized. Gene regulation in skeletal muscle is associated with the E box DNA element (5′-CANNTG-3′) recognized by myogenic regulatory factors (MRFs), such as MyoD1. Transcription element profiling identified a hypothetical myogenic E box (sequence 5′-CAGCTG-3′) within P1, immediately upstream of the major muscle transcriptional start site, and we sought to test its activity in differentiating C2C12 myoblasts. We found P1-driven IGF-I mRNA expression to be associated with myogenic differentiation and, moreover, that a single base-pair mutation in the E box specifically reduced expression in myofibers. A synthetic enhancer construct containing a triplet repeat of the E box was active in muscle cells and strongly induced in myofibers. The capacity of a double-stranded *IGF-I* E box probe (but not one bearing a single-base pair alteration) to bind C2C12 nuclear lysates increased with myogenesis, and a transactivation assay demonstrated that the E box was recognized by E protein-MRF heterodimers. Mechanisms of tissue-specific gene activation are of increasing biological interest, and we have identified a *cis*-element able to direct muscle-specific *IGF-I* gene expression.

muscle development; gene expression regulation; transcription factor; myogenic regulatory factor

INSULIN-LIKE GROWTH FACTOR I (IGF-I or somatomedin C) is a small, basic, single-chain polypeptide of 70 amino acids with relative molecular mass of 7.5 kDa with both autocrine and paracrine actions in numerous cell types (5, 17). IGF-I is implicated in a variety of physiological processes, including, notably, the regulation of both proliferation and differentiation, and its expression is subject to exquisite control at both the transcriptional and translational levels (37). Endocrine IGF-I is synthesized in the liver, with its message being transcribed from promoters located in either exon 1 (P1) or exon 2 (P2) (36). IGF-I synthesis also occurs in peripheral tissues where its transcription is generally initiated from the exon 1 promoter, in which case exon 2 is spliced from the pre-mRNA (15, 42). While exons 3 and 4 encode the mature peptide, the alternative usage of exons 5 or 6 (which encode the E peptide) is associated with the synthesis of the propeptides IGF-Ib or IGF-Ia, respectively (34). The functionally inactive propeptides are cleaved at a unique pentabasic prohormone cleavage motif in the NH_2-terminus of each E peptide to yield the functional, secreted forms of IGF-I (7).

Deletional analysis has identified a number of sequences upstream of exon 1 involved in tissue-specific *IGF-I* expression; indeed, this 5′-untranslated region (UTR) of the *IGF-I* gene maintains the highest degree of conservation through evolutionary time (25). Wang et al. (42) identified two minimal promoters within a 495-bp region responsible for maximal basal exon 1 activity in a range of cell lines. This region contains four potential transcriptional start sites (TSSs) and a number of transcription factor-binding sites, including those recognizing hepatocyte nuclear factor 1 (Hnf1), CCAAT/enhancer-binding proteins (C/EBP), and GATA elements (26, 27, 41). In particular, core promoter elements responsible for gene expression in a variety of cell types appear to cluster around TSS3 (42). Interestingly, a potential MyoD1-like E box sequence lies ~76 nt downstream of the first TSS (TSS1, numbered according to the human sequence) and ~62 nt upstream of TSS3, which is also the major TSS used in skeletal muscle (11, 18, 36). Furthermore, this element and its flanking nucleotides are highly conserved through evolutionary time; thus we hypothesized that this element may have a role in the regulation of *IGF-I* expression in skeletal muscle.

IGF-I is a key regulator of skeletal muscle mass, and *IGF-I* knockout mice display skeletal muscle hypoplasia, whereas ectopic IGF-I expression is associated with hypertrophy (2, 37). A transcriptional variant of IGF-I known as mechano-growth factor (MGF) is transiently expressed in skeletal muscle immediately after mechanical damage and is thought to be involved in satellite cell activation (12). Myogenic differentiation or muscle loading is also accompanied by an increase in systemic IGF-I mRNA (IGF-Ia), although the *cis*-elements regulating this process remain to be elucidated (8, 24). It has been previously shown that ectopic MyoD1 increased activity from exon 1 promoter constructs in vitro, although the elements mediating this effect are also unknown (24).

The differentiation of myoblasts into contractile myofibers is specified by a sequential pattern of basic helix-loop-helix (bHLH) transcription factor activation (for a review, see Ref. 29). The myogenic regulatory factors (MRFs; MyoD1, Myf5, myogenin, and Myf6) form heterodimers with the widely expressed E proteins (E12, E47, ITF2, and HEB), and these complexes bind to distinct E box DNA elements (with the

Address for reprint requests and other correspondence: K. Langlands, Clore Laboratory, Univ. of Buckingham, Hunter St., Buckingham, MK18 1EG, UK (e-mail: kenneth.langlands@buckingham.ac.uk).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

 0363-6143/06 $8.00 Copyright © 2006 the American Physiological Society

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

consensus sequence 5′-CANNTG-3′) located in the promoters of muscle-associated genes to provide temporal and spatial control of gene expression (23). A number of Web-based computational tools can assist in the identification of potential regulatory elements in noncoding regions of DNA, notably, MatInspector (32), TRANSFAC (19), and the transcription element search system (TESS; http://agave.humgen.upenn.edu/tess/index.html). Informed by previously characterized *cis*-elements, such applications use precompiled weight matrixes to assess the likelihood that any DNA sequence is a potential target for DNA binding proteins. Scanning any promoter sequence inevitably returns multiple hits, and thus it is incumbent upon the investigator to refine the search strategy to minimize this background noise and so reduce the number of targets for subsequent analysis. Rather than using a preexisting weight matrix populated by a broad range of myogenic E box elements, we derived a bespoke matrix from functional skeletal muscle elements sharing a specific core, thereby improving both the specificity and relevance of our search. Moreover, as genetic drift is strongly selected against in regulatory noncoding regions over evolutionary time, conservation provides an additional clue to function. We sought to combine these approaches to improve our discrimination of any uncharacterised E box located in the *IGF-I* exon 1 promoter.

IGF-I expression in peripheral tissues is specified by a great many factors and so provides an excellent model for the study of gene activation. The characterization of *cis*-elements in the *IGF-I* promoter will, therefore, contribute to our understanding of IGF-I biology in particular and gene regulation in general. We tested the ability of differentiating mouse muscle cell line (C2C12) preparations to activate a range of reporter constructs and bind E box DNA elements to determine if a hypothetical E box in the 5′-UTR of *IGF-I* exon 1 was involved in the temporal and spatial regulation of *IGF-I* expression.

MATERIALS AND METHODS

All reagents were obtained from Sigma-Aldrich (Poole, UK) unless otherwise indicated. Wherever possible, HUGO standard gene nomenclature is used (www.gene.ucl.ac.uk/nomenclature).

Tissue culture. C2C12 mouse myoblasts, OVCAR-3 human ovarian carcinoma, and HeLaB human cervical adenocarcinoma cancer cell lines were obtained from the American Type Culture Collection. Cells were maintained as standard in either RPMI (OVCAR-3) or DMEM (C2C12 and HelaB cells) supplemented with 10% FBS. C2C12 differentiation was induced by substituting 2% horse serum for 10% bovine serum, and cultures were maintained for 2 days (enriching for myotubes) or 4 days (myofibers).

Real-time RT-PCR. RNA was extracted from C2C12 myoblasts, myotubes, and myofibers in triplicate using Qiagen RNAeasy columns (Qiagen; Crawley, UK). Adult female CD1 mouse livers and skeletal muscle (hindlimb) were homogenized in 1 ml TRI-reagent, and total RNA was isolated as a standard. First-strand cDNA was synthesised using 1 μg total RNA as a template in a 20-μl reaction using random hexamer priming and Moloney murine leukemia virus reverse transcriptase (Invitrogen; Paisley, UK). PCR primers for *IGF-I* promoters located in either exon 1 (P1) or in exon 2 (P2) were each paired with primers for *IGF-I* exon 4 (Table 1) to measure *IGF-I* P1- and P2-derived expression, respectively. *IGF-I* expression data were normalized to the expression of the housekeeping gene hypoxanthine-guanine phosphorybosyl transferase-1 (*HPRT1*; Table 1). Real-time

Table 1. *Oligonucleotide sequences*

Identifier	Use	Sequence
IGF-I P1		
Forward	PCR	5′-TCTCCCATCTCTCTGGATTTC-3′
IGF-I P2		
Forward	PCR	5′-TACCCACYCTGACCTGCTGT-3′
IGF-I P4		
Reverse	PCR	5′-AGATCACAGCTCCGGAAGCA-3′
HPRT1		
Forward	PCR	5′-TGCTGACCTGCTGGATTACAT-3′
Reverse	PCR	5′-GGAATTTCAAATCCAACAAAG-3′
CKM − 1256		
Forward	PCR	5′-GCGCTCGAGCCACTACGGGTCTAGGCTGC-3′
CKM − 1050		
Reverse	PCR	5′-GCGAGATCTGATCCACCAGGGACAGGGTT-3′
QuickChange		
Forward	Mutagenesis	5′-AGGCAAATGTTCCCCAAGCTGTTTCCTGTCTACAGT-3′
Reverse	Mutagenesis	5′-ACTGTAGACAGGAAACAGCTTGGGGAACATTTGCCT-3′
Multimerized E box		
Sense	Synthetic enhancer	5′-GATCCAATGTTCCCCCAGCTGTTTCCTGTCTAATGTTCCCCCAGCTGTTTCCTGTCTAATGTTCCCCCAGCTGTTTCCTGTCTG-3′
Antisense	Synthetic enhancer	5′-TCGACAGACAGGAAACAGCTGGGGGAACATTAGACAGGAAACAGCTGGGGGAACATTAGACAGGAAACAGCTGGGGGAACATTG-3′
CKM-R		
Sense	EMSA	5′-TCCCCCCAACACCTGCTGCCTGA-3′
Antisense	EMSA	5′-TCAGGCAGCAGGTGTTGGGGGGA-3′
WT		
Sense	EMSA	5′-ATGTTCCCCCAGCTGTTTCCTGTC-3′
Antisense	EMSA	5′-GACAGGGGGCAGCTGGGGGAACAT-3′
Mutant		
Sense	EMSA	5′-ATGTTCCCCAAGCTGTTTCCTGTC-3′
Antisense	EMSA	5′-GACAGGGGGCAGCTTGGGGAACAT-3′

IGF-I P1, P2, and P4, insulin-like growth factor-I promoters in exon 1, 2, and 4, respectively; HPRT1, hypoxanthine-guanine phosphorybosyl transferase-1; CKM, muscle creatine kinase; WT, wild type. Underlined sequences indicate restriction enzyme recognition sites.

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

PCR was performed using an ABI PRISM 7900 sequence detection system (Applied Biosystems; Warrington, UK) and a QuantiTect SYBR Green PCR Kit (Qiagen) in a total volume of 12 μl in triplicate wells, with each containing 1 μl cDNA and 600 nM of each primer. A no-template control was also prepared in triplicate for each primer set. Furthermore, a duplicate 12-step 1:2 serial dilution series was prepared for each primer set, starting with a mixture of 1 μl liver and 1 μl skeletal muscle cDNA per 12-μl reaction. This allowed the ABI SDS software to prepare a standard curve from which cycle threshold values were transformed into arbitrary quantity data. The following PCR protocol was used: a denaturation program (95°C for 15 min) with 40 cycles of an amplification and quantification program (95°C for 15 s, 55°C for 30 s, and 72°C for 45 s with a single fluorescence measurement taken during the extension stage) and a melting curve program (60–95°C with a heating rate of 1.0°C/30 s and continuous fluorescence measurement). Thereafter, PCR product quality was assessed by generating a melting curve, which was also used to verify the absence of PCR artifacts (primer-dimers) or nonspecific PCR products. Results are described as mean *IGF-I* expression relative to mean *HPRT1* expression.

Promoter analysis. Human, mouse, rat, cow, goat, sheep, salmon, and zebrafish exon 1 sequences were aligned using CLUSTALW, and a consensus sequence was created from the resulting multiple alignment using the European Molecular Biology Open Software Suite program CONS (33, 38). A bespoke weight matrix was constructed from 11 functional skeletal muscle-specific E box elements sharing the 5′-CAGCTG-3′ core, along with four flanking nucleotides each side (Table 2). TESS software readily accepts the use of user-defined weight matrixes, and this package was used to scan the consensus promoter sequence for high scoring matches (scores > 10).

Plasmids. The IGF-I630/Luc construct, containing nucleotides −1630 to +322 of the human *IGF-I* exon 1 sequence cloned upstream of a luciferase (Luc) reporter gene, was a kind gift from Peter Rotwein (Oregon Health Sciences University, Portland, OR) (18). The *IGF-I* promoter was subcloned as a *Pst*I fragment into the pGL3-Basic reporter vector (Promega; Southampton, UK), which lacks promoter and enhancer elements, to create pGL1630. The pGL3-basic polylinker was previously modified by the insertion of a cassette containing *Bgl*II, *Nde*I, *Eco*RV, *Pst*I, and *Hin*dIII sites into the *Bgl*II/*Hin*dIII-digested vector, creating pGL3+. The wild-type (WT) 5′-CAGCTG-3′ E box sequence in the pGL1630 construct was altered to 5′-AAGCTG-3′ with QuickChange site-directed mutagenesis (Stratagene; Cambridge, UK) according to the manufacturer's recommendations to generate pGL1630E−. This substitution removed a *Pvu*II site, facilitating ready identification of mutants, although the integrity of the altered sequence was also verified by sequencing. The synthetic E box enhancer vector containing three copies of the putative *IGF-I* E box (pGLEbox) was created by inserting a double-stranded oligonucleotide containing the multimerized E box sequence (5′-CAGCTG-3′ core flanked by 10 nucleotides) into a pGL3-promoter vector (Promega). The muscle creatine kinase (*CKM*) enhancer region [which contains three E boxes as well as myocyte-specific enhancer factor (MEF)-2 binding sites (13)] was PCR amplified from rat genomic DNA and cloned as a *Sal*1/*Bam*H1 fragment into the pGL3-promoter vector. All other plasmids (full-length MyoD1, Myf6, and E47 cDNAs in a pRcCMV background) were as previously described (20). Oligonucleotides (MWG Biotech; Munich, Germany) are detailed in Table 1.

Transient transfections and reporter assays. Cells were cotransfected with 1 μg of each relevant plasmid and 1 μg of a β-galactosidase reporter (BD Clontech; Oxford, UK) using FuGeneII reagent (Roche; Lewes, UK) at a 3:1 ratio of FuGeneII-DNA in six-well tissue culture plates. Unless a time course experiment was being performed, cells were harvested 48 h posttransfection. Luc activity in cell lysates was determined using a Promega Luciferase Assay Kit as recommended. Transfection efficiencies were normalized relative to β-galactosidase activity, and all experiments were performed in duplicate.

EMSAs. Nuclear proteins were prepared using the method of Schreiber et al. (35), and shifts were performed using the Pearce LightShift Chemiluminescent EMSA kit (Pearce). Binding was performed in 1× EMSA buffer [25 mM HEPES (pH 7.9), 40 mM KCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, and 10% glycerol] containing 50 ng/μl polydIdC, 4 μg nuclear extract, and 1 fmol/μl double-stranded biotinylated probe (Table 1; each strand was synthesized with a single 5′-biotin moiety). After the samples were incubated for 20 min at room temperature, 5× loading buffer (Pearce) was added, and reactions were resolved by 6% polyacrylamide gel electrophoresis (Novex, Invitrogen) before a semidry transfer to Hybond N+ membranes (Amersham Pharmacia; Little Chalfont, UK). Shifts were visualized with LightShift reagents, according to the manufacturer's recommendations, and subsequent exposure to X-ray film (BioMax, Amersham Pharmacia). Densitometric analysis of images was performed using AlphaEase imaging software (Alpha Innotech; San Leandro, CA). Quantitative binding values were extracted from triplicate independent experiments.

Statistical analysis. Unless otherwise stated in the text, Luc levels obtained from promoter constructs and promoterless parental vectors in the ectopic transfection assays were compared using paired *t*-tests. For real-time RT-PCR and quantitative EMSA analysis, values obtained from either myotubes or myofibers were compared with those obtained from myoblasts. All graphs represent mean values ± SE, and statistical analysis was performed using Excel (Microsoft).

RESULTS

IGF-I expression increases with myogenesis. Real-time RT-PCR analysis indicated a 4.7-fold increase in P1 transcripts with myogenic differentiation ($P = 0.0008$; Fig. 1) and, moreover, that P1-derived transcripts were the most abundant species in adult mouse skeletal muscle. Whereas higher levels of P1 transcripts were evident in liver RNA compared with muscle, the predominant RNA species in the former tissue were transcribed from the P2 promoter. Low levels of P2

Table 2. *Construction of a myogenic E box weight matrix*

	Position													
Nucleotide	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
A	2	2	4	1	0	11	0	0	0	0	1	0	2	1
C	2	4	2	4	11	0	0	11	0	0	5	6	2	6
G	4	3	4	2	0	0	11	0	0	11	1	0	4	1
T	3	2	1	4	0	0	0	0	11	0	4	5	3	3

Frequencies of nucleotides at each position in a consensus E box were calculated from E box sequences located in the following skeletal muscle-expressed genes: cholinergic receptor, nicotonic, α_1, mouse (30); cholinergic receptor, nicotonic, β, mouse (31); cholinergic receptor, nicotonic, α, mouse (10); cholinergic receptor, nicotonic, ε; mouse (28); *CKM,* mouse (6); cytochrome *c* oxidase subunit VIa mouse (40); cytochrome *c* oxidase subunit VIII, rat (21); desmin, human (22); myosin, light polypeptide 1, chicken (9); myosin, light polypeptide 4, mouse (1); and slow troponin T, chicken (44).

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

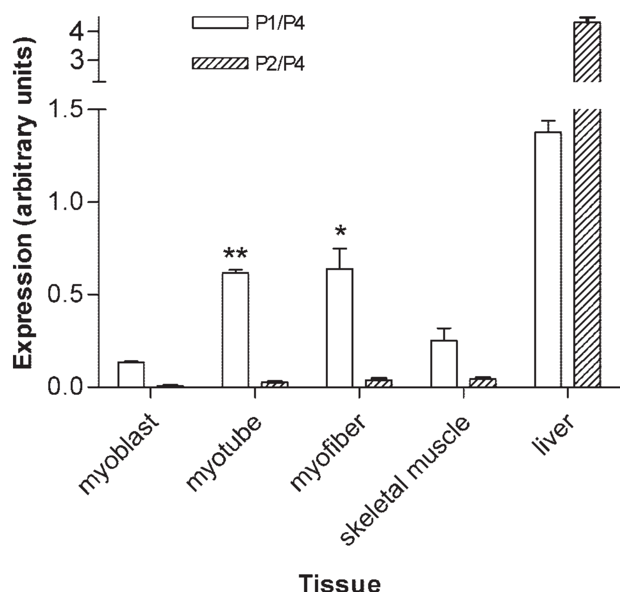

Fig. 1. *IGF-I* expression in C2C12 cells and primary mouse tissues. Real-time RT-PCR was performed using either exon 1 transcript-specific (P1) or exon 2 (P2) transcript-specific primers in conjunction with an exon 4 reverse primer (P4). A 4.7-fold increase in P1 activation was detected in both myotubes and myofibers relative to myoblasts. Significant increases in gene expression are indicated as follows: *$P < 0.05$ and **$P < 0.001$.

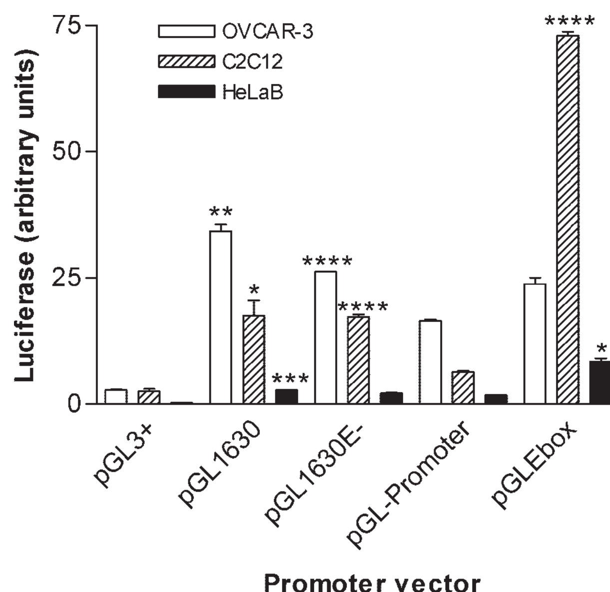

Fig. 2. *IGF-I* promoter activity in cell lines. Activation of *IGF-I* promoter constructs and parental vector controls (pGL3+ for the pGL1630 vectors and pGL-promoter for the pGLEbox vector) in OVCAR-3, C2C12, and HeLaB cells is shown. The significance of luciferase induction relative to parental vector is indicated as follows: *$P < 0.05$, **$P < 0.02$, ***$P < 0.01$, and ****$P < 0.005$.

transcripts were detected in C2C12 RNAs, both before and after differentiation, and in skeletal muscle samples.

Analysis of a candidate E box in the IGF-I exon 1 promoter. Running a promoter scan under TESS using standard TRANSFAC matrixes identified 12 potential E boxes with the consensus 5′-CANNTG-3′ in a 2.5-kb region of the human exon 1 promoter. A subsequent TESS search using the TRANSFAC myogenic E box weight matrixes identified four high-scoring elements (each with a likelihood score > 10). However, before any experimental evaluation was performed, we sought to increase the likelihood that any elements identified were functional. Multiple alignments of exon 1 promoter sequences from diverse species identified a region of ~2-kb that is highly conserved through evolutionary time. A subsequent TESS scan of a consensus promoter derived from this multiple alignment identified two highly conserved 5′-CAG-CTG-3′ elements (at positions −837 and +76 relative to the first TSS of the human *IGF-I* gene). A final TESS search using our bespoke myogenic matrix (Table 2) identified one high-scoring element at position +76 as significant (generating a likelihood score 14.7 of a maximum of 18). The sequence of this element, as well as its conservation in a range of other species, is shown in Table 3. For clarity, the results of the initial searches are not shown.

IGF-I P1 is active in OVCAR-3, HeLaB, and C2C12 myoblasts. The 1952-bp exon 1 fragment significantly increased Luc activity in OVCAR-3 (12-fold, $P = 0.014$), C2C12 (7-fold, $P = 0.05$), and HeLaB cells (9-fold, $P = 0.007$) over background (Fig. 2). A single-base pair mutation in the putative E box region had no significant effect on Luc expression in any cell line relative to the WT promoter; however, the synthetic enhancer construct containing three tandem repeats of this E box was highly induced in C2C12 myoblasts (11-fold, $P = 0.003$). Little induction of the multimerized E box enhancer construct over background was detected in OVCAR-3 cells, although an increase in activity was also detected in HeLaB cells (6.5-fold, $P = 0.033$).

IGF-I promoter activity during C2C12 differentiation. Luc expression induced from the exon 1 promoter fragment was detected in proliferating myoblasts and maintained during myogenesis (Fig. 3). However, the destruction of the putative *IGF-I* E box in the exon 1 promoter resulted in a significant

Table 3. *Phylogenetic relationships of the IGF-I E box and flanking sequences*

Organism	Accession No.	Sequence
Human	M12659	5′-AGGCAAATGTTCCCCCAGCTGTTTCCTGTCTACAGTGTCT-3′
Rat	M15647	5′-AGGCGAATGTTCCCCCAGCTGTTTCCTGTCTACAGTGTCT-3′
Mouse	Y18062	5′-AGGCGAATGTTCCCCCAGCTGTTTCCTGTCTACAGTGTCT-3′
Goat	D26116	5′-AGGCAAGCGTTCCCCCAGCTGTTTCCTGTCTACAGTGTCT-3′
Cow	AF210383	5′-AGGCAAGCGTTCCCCCAGCTGTTTCCTGTCTACAGTGTCT-3′
Sheep	X69472	5′-AGGCAAGCGTTCCCCCAGCTGTTTCCTGTCTACAGTGTCT-3′
Zebrafish	BX510924	5′-AGGCAAATGCTGCCCCAGCTGTTTCCTGTTGAAAATGTCT-3′
Consensus		5′-AGGCaAatGtTcCCCCAGCTGTTTCCTGTctAcAgTGTCT-3′

The E box core is underlined, and mismatches are shown in bold. Lowercase letters indicate consensus nucleotides.

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

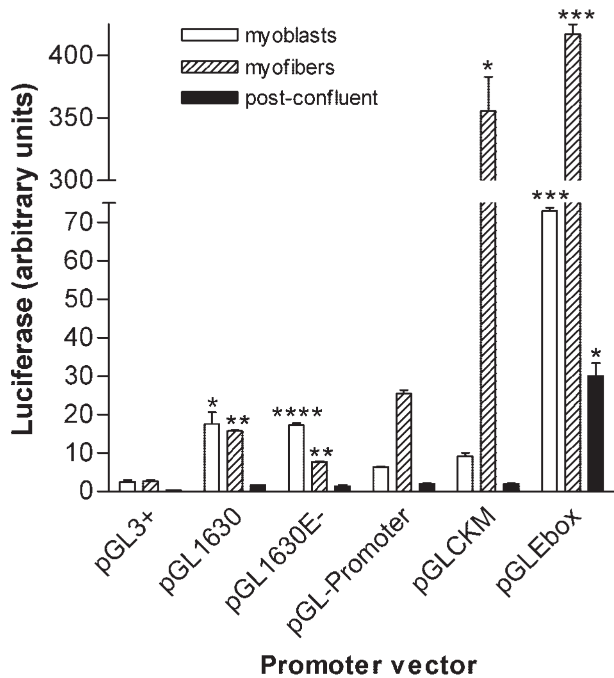

Fig. 3. *IGF-I* promoter activity during myogenic differentiation. Luciferase activation resulting from ectopic *IGF-I* exon 1 and muscle creatine kinase (*CKM*) enhancer construct activity in C2C12 myoblasts and myofibers is shown. The significance of increased luciferase levels relative to parental vector is indicated as follows: *$P < 0.05$, **$P < 0.02$, ***$P < 0.01$, and ****$P < 0.001$. Results from C2C12 cells maintained in nondifferentiating medium (DMEM supplemented with 10% bovine serum) for 4 days after reaching confluence (postconfluent) are also shown.

decrease in Luc activity in myofibers (2.2-fold, $P = 0.015$). The increase in ectopic Luc expression stimulated by the multimerized *IGF-I* E box enhancer element during C2C12 differentiation was more dramatic (6-fold, $P = 0.007$), although the magnitude of the change was not as great as *CKM* enhancer-driven expression (134-fold, $P = 0.03$) due to a higher level of E box enhancer vector transcription in myoblasts.

EMSAs. The ability of C2C12 nuclear lysates prepared at various differential stages to retard the migration of biotinylated probes derived from the *CKM* right E box, the WT *IGF-I* E box, or the mutated *IGF-I* E box bearing a single-base pair alteration was evaluated using a nonradioactive EMSA method (Fig. 4). As expected, an increase in *CKM* E box binding characterized myogenic differentiation, whereas only low levels of binding to the mutated *IGF-I* probe were seen. Densitometric analysis of WT *IGF-I* probe EMSAs indicated a 1.5-fold increase in binding of C2C12 myofiber nuclear proteins relative to myoblasts ($P = 0.045$).

bHLH protein activation of IGF-I promoters. Reporter gene activation by any ectopic bHLH factor in isolation led to only small increases in Luc activation over background (relative to parental promoter vectors) with the exception of MyoD1 homodimers, which were able to significantly activate the *CKM* (5-fold, $P = 0.006$), *IGF-I* exon 1 (8-fold, $P = 0.017$), and multimerized E box constructs (1.6-fold, $P = 0.015$) (Fig. 5). Myf6 was also able to activate the *CKM* promoter over background in the absence of an ectopic binding partner (2.2-fold, $P = 0.017$). However, higher levels of transactivation were obtained by the expression of heterodimers with E47-MyoD1 complexes inducing the *CKM* enhancer (4-fold, $P = 0.002$), multimerized *IGF-I* E box (2.6-fold, $P = 0.011$), and WT promoter 1 construct (21-fold, $P = 0.0004$). E47-Myf6 complexes also activated the exon 1 promoter (6-fold, $P = 0.007$) and *CKM* enhancer (4-fold, $P = 0.02$), although the effect of this complex on the synthetic *IGF-I* E box enhancer was not found to be significant. The single-base pair change in the exon 1 promoter E box element consistently resulted in a significant decrease in Luc activity relative to the WT sequence in response to MyoD1 homodimer and heterodimer and Myf6 heterodimer action (2-fold, $P = 0.04$; 3-fold, $P = 0.03$; and 1.5-fold, $P = 0.007$, respectively).

DISCUSSION

An increasing number of *cis*-regulatory elements are known to initiate both local and endocrine *IGF-I* expression, particularly those associated with the well-conserved exon 1 5′-UTR (18, 26, 39, 41, 42). However, a clear understanding of the mechanisms specifying compartmental *IGF-I* expression is yet to be achieved. Particularly, IGF-I is an important regulator of skeletal muscle physiology, and its study is important toward the treatment of muscle wasting or atrophy associated with age, with wider implications for the elucidation of tissue-specific gene regulatory networks.

The exon 1 5′-UTR is the most highly conserved region of the *IGF-I* gene, suggesting that the most fundamental regulatory elements map to this area (25). While promoter-scanning programs provide valuable tools in the identification of those regions capable of specifying gene expression, it is imperative that a clear understanding of the underlying biology informs any search strategy. The use of such a rational approach facilitated our characterization of a potential myogenic E box element immediately upstream of the major exon 1 TSS. The conservation of this E box and its flanking nucleotides from zebrafish to humans increased our confidence that this element was functional and so merited functional investigation. Real-time PCR analysis showed that the induction of IGF-I mRNA

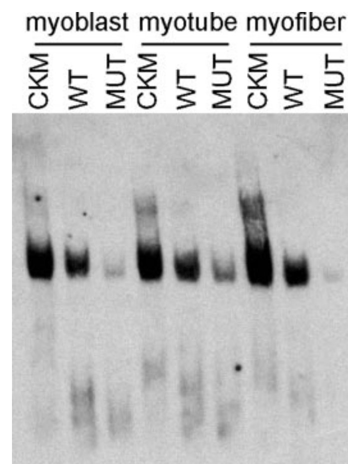

Fig. 4. Binding of C2C12 nuclear lysates to E box probes. EMSAs were performed using biotinylated *CKM*-R E box (CKM), *IGF-I* E box [wild type (WT)] and mutated (MUT) *IGF-I* E box probes incubated with C2C12 nuclear lysates prepared from cells at preconfluence (myoblast) and at 2 and 4 days of differentiation (myotube and myofiber, respectively). A typical experiment is shown.

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

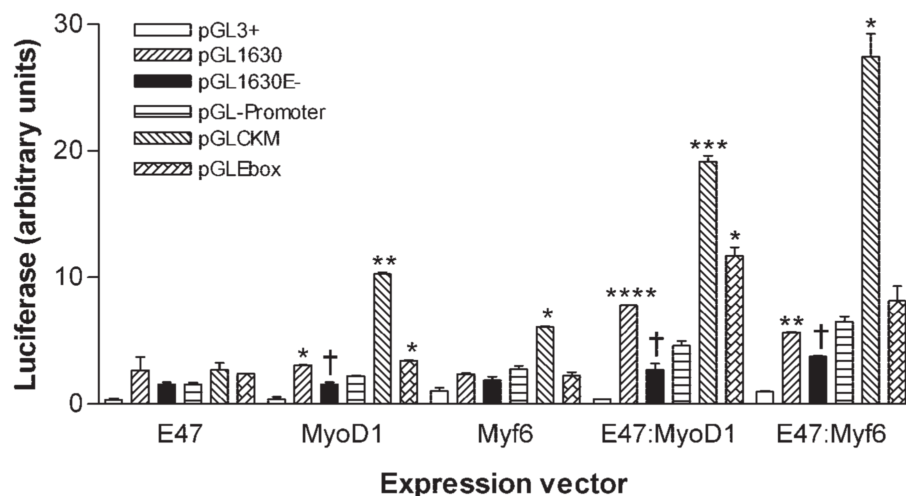

Fig. 5. Transactivation of *IGF-I* and *CKM* promoter vectors by basic helix-loop-helix (bHLH) factors. HeLaB cells were cotransfected with various promoter constructs and full-length myogenic regulatory factors (MyoD1 or Myf6) or E protein (E47) expression vectors in either homo- or heterodimeric combinations. The significance of increased luciferase expression relative to the appropriate parental vector controls is indicated as follows: *$P < 0.02$, **$P < 0.01$, ***$P < 0.002$, and ****$P < 0.0005$. †Significant reduction in reporter activity resulting from a single-base pair mutation in the *IGF-I* exon 1 promoter ($P < 0.05$).

species characterized C2C12 differentiation and, moreover, that this transcription was initiated from P1, the majority peripheral promoter. An ~1.9-kb fragment of the human *IGF-I* exon 1 5′-UTR highly active in the neuroblastoma cell line SK-N-MC, which express mainly *IGF-I* exon 1 transcripts, was also shown to be active (albeit to a lesser extent) in OVCAR-3 cells, which express mainly *IGF-I* exon 2 transcripts (15). We used a related 1952-bp fragment (which includes additional 3′-exon 1 DNA), which was also shown to be highly active in SK-N-MC cells (18). However, we found reporter activity to be low when we introduced this construct into OVCAR-3 cells. Assuming this to be a limitation of the vector backbone, we subcloned the entire fragment into a pGL3-basic Luc reporter vector. This produced detectable Luc activity in OVCAR-3 cells with minimal activition in HeLaB cells, which express little IGF-I (data not shown), whereas proliferating C2C12 cells activated the reporter to an intermediate level. The E box knockout had little influence in any of these cell lines; however, the synthetic E box enhancer construct strongly induced Luc expression in C2C12 cells. This suggests that the E box is not active in myoblasts but has the potential to be activated by myogenic factors.

It appears contradictory that the upregulation in IGF-I message was not accompanied by P1 activation during myogenic differentiation. Indeed, McCall et al. (24) reported that a series of P1 fragments were also unresponsive during C2C12 differentiation. This is consistent with the 1952-kb *IGF-I* fragment spanning a minimal basal promoter region, and its activation in myoblasts is, at least in part, as a consequence of the lack of distal repressor elements and not due to activation of the E box per se (further confirmed by the minimal effect of destroying this element). Differentiation is, however, accompanied by myogenic activation of the E box because its mutation did significantly reduce P1 activity in myofibers but not in any other cell line studied. More compelling evidence for E box-mediated expression was provided by the muscle-specific activation of the multimerized *IGF-I* E box construct, which generated Luc levels comparable with those obtained by the *CKM* enhancer. Strong E box-driven activation of skeletal muscle-associated genes is usually associated with occupancy of one or more E box elements as well as the cooperative binding of proximal accessory factors such MEF2, serum response factor, or Sp1 (for a review, see Ref. 43). Intuitively, one would predict that searching for clusters of closely mapping promoter elements would increase predictive power. However, such clustering analysis was shown to identify <50% of known skeletal muscle promoters; thus such a strategy is of only limited value in the evaluation of previously uncharacterised *cis*-elements (43). Our observations are consistent with a relatively weak (in its natural context) yet highly specific regulatory sequence that would ensure the fine control of IGF-I protein levels in skeletal muscle.

C2C12 myoblast nuclear extracts were able to bind synthetic olignucleotides derived from both *CKM* and, to a lesser extent, *IGF-I* E box (WT) sequences. Because myoblasts express high levels of ID proteins, which sequester free E proteins, these shifted bands may indicate MyoD1 homodimer binding (16). Consistent with the differentiation-associated downregulation in ID expression, increased binding was detected in myofiber protein preparations, presumably as a result of E protein-MRF heterodimer formation. However, it is difficult to predict the precise composition of these DNA-binding complexes as any of the four E proteins can form heterodimers with the four MRFs (20). Furthermore, DNA binding does not necessarily indicate efficient transactivation, so we used an ectopic expression assay to assess the influence of E proteins and MRFs, either in isolation or as pairs. MyoD1 expression is associated with commitment to myogenic differentiation, and Myf6 is a later-acting protein whose expression is maintained in terminally differentiated muscle fibers, whereas E47 is more widely expressed (23, 29). These experiments were performed in a HeLaB background because these cells have a relatively low ID and E protein background [Refs. 4 and 14, and data not shown], but we cannot exclude the possibility that apparent MyoD1 homodimer transactivation was facilitated by endogenous E protein dimerization.

Ectopic MyoD1 was able to transactivate each promoter above background levels, with the greatest induction seen in the activity of *IGF-I* P1, followed by the *CKM* enhancer, with a small but significant increase in the activity of the pGLEbox construct. As expected, dimerization with E47 greatly facilitated activation, with both native P1 and multimerized E box construct activity increasing in particular. Myf6 was able to transactivate only weakly as a homodimer, but E47 dimeriza-

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

tion facilitated a dramatic rise in *CKM* and P1 activation. The predominant MRF after 4 days of C2C12 differentiation is myogenin; however, we were unable to generate a myogenin vector with comparable expression levels to those of MyoD1 and Myf6 (data not shown). Myf6 expression is sustained in mature myofibers [indeed, it is the only MRF detected at this stage (29)], and thus it has the potential to maintain low levels of IGF-I expression in terminally differentiated tissue.

There has been a postgenomic perspective shift from gene function to gene regulation with the increasing availability of distal regulatory elements that may map many kilobases from the transcriptional start site of a gene. Myogenic regulation of *IGF-I* activity is complex, but we have identified a single element conferring muscle-specific gene expression, although the full constellation of promoter-binding factors remains to be elucidated. The study of *IGF-I* expression provides an excellent model for tissue-specific gene activation as its transcription is regulated by multiple elements located in multiple promoters active in multiple cell types. Furthermore, phylogenetic analysis facilitates the identification of those sequences subject to the strongest selection pressure and so central to physiological control. A greater understanding of those *cis*-acting *IGF-I* elements active in the myogenic program will expand the promoter map and so enrich our understanding of regulatory networks. Moreover, the growth in gene regulatory network analysis must be accompanied by progressively sophisticated methods to confidently identify *cis*-elements if predictive developmental maps are to be realized (3). It is hoped that the rational approach described herein to the identification of a myogenic control element will complement such analysis.

ACKNOWLEDGMENTS

We are grateful to Dr. P. Rotwein for the kind gift of the IGF-I630/Luc construct and to Dr. Tom Moore (University of Cork) for providing real-time RT-PCR facilities and support.

Present address of A. S. McLellan: Dept. of Biochemistry, Univ. College Cork, Cork, Ireland.

GRANTS

This work was funded by a grant from Unilever Research.

REFERENCES

1. **Catala F, Wanner R, Barton P, Cohen A, Wright W, and Buckingham M.** A skeletal muscle-specific enhancer regulated by factors binding to E and CArG boxes is present in the promoter of the mouse myosin light-chain 1A gene. *Mol Cell Biol* 15: 4585–4596, 1995.
2. **Coleman ME, DeMayo F, Yin KC, Lee HM, Geske R, Montgomery C, and Schwartz RJ.** Myogenic vector expression of insulin-like growth factor I stimulates muscle cell differentiation and myofiber hypertrophy in transgenic mice. *J Biol Chem* 270: 12109–12116, 1995.
3. **Davidson E and Levin M.** Gene regulatory networks. *Proc Natl Acad Sci USA* 102: 4935, 2005.
4. **Deed RW, Bianchi SM, Atherton GT, Johnston D, Santibanez-Koref M, Murphy JJ, and Norton JD.** An immediate early human gene encodes an Id-like helix-loop-helix protein and is regulated by protein kinase C activation in diverse cell types. *Oncogene* 8: 599–607, 1993.
5. **D'Ercole AJ, Stiles AD, and Underwood LE.** Tissue concentrations of somatomedin C: further evidence for multiple sites of synthesis and paracrine or autocrine mechanisms of action. *Proc Natl Acad Sci USA* 81: 935–939, 1984.
6. **Donoviel DB, Shield MA, Buskin JN, Haugen HS, Clegg CH, and Hauschka SD.** Analysis of muscle creatine kinase gene regulatory elements in skeletal and cardiac muscles of transgenic mice. *Mol Cell Biol* 16: 1649–1658, 1996.
7. **Duguay SJ, Lai-Zhang J, and Steiner DF.** Mutational analysis of the insulin-like growth factor I prohormone processing site. *J Biol Chem* 270: 17566–17574, 1995.
8. **Frost RA, Nystrom GJ, and Lang CH.** Regulation of IGF-I mRNA and signal transducers and activators of transcription-3 and -5 (Stat-3 and -5) by GH in C2C12 myoblasts. *Endocrinology* 143: 492–503, 2002.
9. **Fujisawa-Sehara A, Hanaoka K, Hayasaka M, Hiromasa-Yagami T, and Nabeshima Y.** Upstream region of the myogenin gene confers transcriptional activation in muscle cell lineages during mouse embryogenesis. *Biochem Biophys Res Commun* 191: 351–356, 1993.
10. **Gilmour BP, Fanger GR, Newton C, Evans SM, and Gardner PD.** Multiple binding sites for myogenic regulatory factors are required for expression of the acetylcholine receptor gamma-subunit gene. *J Biol Chem* 266: 19871–19874, 1991.
11. **Hall LJ, Kajimoto Y, Bichell D, Kim SW, James PL, Counts D, Nixon LJ, Tobin G, and Rotwein P.** Functional analysis of the rat insulin-like growth factor I gene and identification of an IGF-I gene promoter. *DNA Cell Biol* 11: 301–313, 1992.
12. **Hill M and Goldspink G.** Expression and splicing of the insulin-like growth factor gene in rodent muscle is associated with muscle satellite (stem) cell activation following local tissue damage. *J Physiol* 549: 409–418, 2003.
13. **Horlick RA and Benfield PA.** The upstream muscle-specific enhancer of the rat muscle creatine kinase gene is composed of multiple elements. *Mol Cell Biol* 9: 2396–2413, 1989.
14. **Hu JS, Olson EN, and Kingston RE.** HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. *Mol Cell Biol* 12: 1031–1042, 1992.
15. **Jansen E, Steenbergh PH, van Schaik FM, and Sussenbach JS.** The human IGF-I gene contains two cell type-specifically regulated promoters. *Biochem Biophys Res Commun* 187: 1219–1226, 1992.
16. **Jen Y, Weintraub H, and Benezra R.** Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Genes Dev* 6: 1466–1479, 1992.
17. **Jones JI and Clemmons DR.** Insulin-like growth factors and their binding proteins: biological actions. *Endocr Rev* 16: 3–34, 1995.
18. **Kim SW, Lajara R, and Rotwein P.** Structure and function of a human insulin-like growth factor-I gene promoter. *Mol Endocrinol* 5: 1964–1972, 1991.
19. **Knuppel R, Dietze P, Lehnberg W, Frech K, and Wingender E.** TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol* 1: 191–198, 1994.
20. **Langlands K, Yin X, Anand G, and Prochownik EV.** Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J Biol Chem* 272: 19785–19793, 1997.
21. **Lenka N, Basu A, Mullick J, and Avadhani NG.** The role of an E-box binding basic helix loop helix protein in the cardiac muscle-specific expression of the rat cytochrome oxidase subunit VIII gene. *J Biol Chem* 271: 30281–30289, 1996.
22. **Li Z and Paulin D.** Different factors interact with myoblast-specific and myotube-specific enhancer regions of the human desmin gene. *J Biol Chem* 268: 10403–10415, 1993.
23. **Massari ME and Murre C.** Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* 20: 429–440, 2000.
24. **McCall GE, Allen DL, Haddad F, and Baldwin KM.** Transcriptional regulation of IGF-I expression in skeletal muscle. *Am J Physiol Cell Physiol* 285: C831–C839, 2003.
25. **Mittanck DW, Kim SW, and Rotwein P.** Essential promoter elements are located within the 5′ untranslated region of human insulin-like growth factor-I exon I. *Mol Cell Endocrinol* 126: 153–163, 1997.
26. **Nolten LA, Steenbergh PH, and Sussenbach JS.** Hepatocyte nuclear factor 1 alpha activates promoter 1 of the human insulin-like growth factor I gene via two distinct binding sites. *Mol Endocrinol* 9: 1488–1499, 1995.
27. **Nolten LA, van Schaik FM, Steenbergh PH, and Sussenbach JS.** Expression of the insulin-like growth factor I gene is stimulated by the liver-enriched transcription factors C/EBP alpha and LAP. *Mol Endocrinol* 8: 1636–1645, 1994.
28. **Numberger M, Durr I, Kues W, Koenen M, and Witzemann V.** Different mechanisms regulate muscle-specific AChR gamma- and epsilon-subunit gene expression. *EMBO J* 10: 2957–2964, 1991.
29. **Olson EN and Klein WH.** bHLH factors in muscle development: dead lines and commitments, what to leave in and what to leave out. *Genes Dev* 8: 1–8, 1994.

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017

30. **Piette J, Bessereau JL, Huchet M, and Changeux JP.** Two adjacent MyoD1-binding sites regulate expression of the acetylcholine receptor alpha-subunit gene. *Nature* 345: 353–355, 1990.
31. **Prody CA and Merlie JP.** A developmental and tissue-specific enhancer in the mouse skeletal muscle acetylcholine receptor alpha-subunit gene regulated by myogenic factors. *J Biol Chem* 266: 22588–22596, 1991.
32. **Quandt K, Frech K, Karas H, Wingender E, and Werner T.** MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* 23: 4878–4884, 1995.
33. **Rice P, Longden I, and Bleasby A.** EMBOSS: the European molecular biology open software suite. *Trends Genet* 16: 276–277, 2000.
34. **Rotwein P.** Two insulin-like growth factor I messenger RNAs are expressed in human liver. *Proc Natl Acad Sci USA* 83: 77–81, 1986.
35. **Schreiber E, Matthias P, Muller MM, and Schaffner W.** Rapid detection of octamer binding proteins with 'mini-extracts,' prepared from a small number of cells. *Nucleic Acids Res* 17: 6419, 1989.
36. **Shemer J, Adamo ML, Roberts CT Jr, and LeRoith D.** Tissue-specific transcription start site usage in the leader exons of the rat insulin-like growth factor-I gene: evidence for differential regulation in the developing kidney. *Endocrinology* 131: 2793–2799, 1992.
37. **Stewart CE and Rotwein P.** Growth, differentiation, and survival: multiple physiological functions for insulin-like growth factors. *Physiol Rev* 76: 1005–1026, 1996.
38. **Thompson JD, Higgins DG, and Gibson TJ.** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680, 1994.
39. **Umayahara Y, Ji C, Centrella M, Rotwein P, and McCarthy TL.** CCAAT/enhancer-binding protein delta activates insulin-like growth factor-I gene transcription in osteoblasts. Identification of a novel cyclic AMP signaling pathway in bone. *J Biol Chem* 272: 31793–31800, 1997.
40. **Wan B and Moreadith RW.** Structural characterization and regulatory element analysis of the heart isoform of cytochrome *c* oxidase VIa. *J Biol Chem* 270: 26433–26440, 1995.
41. **Wang L, Wang X, and Adamo ML.** Two putative GATA motifs in the proximal exon 1 promoter of the rat insulin-like growth factor I gene regulate basal promoter activity. *Endocrinology* 141: 1118–1126, 2000.
42. **Wang X, Yang Y, and Adamo ML.** Characterization of the rat insulin-like growth factor I gene promoters and identification of a minimal exon 2 promoter. *Endocrinology* 138: 1528–1536, 1997.
43. **Wasserman WW and Fickett JW.** Identification of regulatory regions which confer muscle-specific gene expression. *J Mol Biol* 278: 167–181, 1998.
44. **Watanabe T, Takemasa T, Yonemura I, and Hirabayashi T.** Regulation of troponin T gene expression in chicken fast skeletal muscle: involvement of an M-CAT-like element distinct from the standard M-CAT. *J Biochem (Tokyo)* 121: 212–218, 1997.

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.5 on April 4, 2017