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


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 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).
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 matrices 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

<table>
<thead>
<tr>
<th>Identifier</th>
<th>Use</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF-I P1</td>
<td>PCR</td>
<td>5’-TCTCCCATCTCTGCTGATTTTC-3’</td>
</tr>
<tr>
<td>IGF-I P2</td>
<td>PCR</td>
<td>5’-TACCCACGCTGCTGTTCT-3’</td>
</tr>
<tr>
<td>IGF-I P4</td>
<td>PCR</td>
<td>5’-AGATCGACGCTC GGAAAGCA-3’</td>
</tr>
<tr>
<td>HPRT1</td>
<td>PCR</td>
<td>5’-TCGTCAGCTGCTGATTCAAT-3’</td>
</tr>
<tr>
<td>CKM — 1256</td>
<td>PCR</td>
<td>5’-GGATTTCTAAATCCACAAAG-3’</td>
</tr>
<tr>
<td>CKM — 1050</td>
<td>PCR</td>
<td>5’-GGGAGTTCTGATCCCAAGGACAGGTT-3’</td>
</tr>
<tr>
<td>QuickChange</td>
<td>PCR</td>
<td>5’-AGGGAAATTGTGCCAAGGACAGGACAGGTT-3’</td>
</tr>
<tr>
<td>Multimerized E box</td>
<td>Synth enhancer</td>
<td>5’-TCGACAGGCCAGGGCAAGGGGGA-3’</td>
</tr>
<tr>
<td>Mutant</td>
<td>EMSA</td>
<td>5’-ATGTTCCCAAGCCTGTTGCTCAG-3’</td>
</tr>
</tbody>
</table>

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.
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). Therefore, 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 at each side (Table 2). TESS software readily accepts the use of user-defined weight matrices, and this package was used to scan the consensus promoter sequence for high scoring matches (scores ≥ 10). 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 at each side (Table 2). TESS software readily accepts the use of user-defined weight matrices, and this package was used to scan the consensus promoter sequence for high scoring matches (scores ≥ 10).

Table 2. Construction of a myogenic E box weight matrix

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

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, nicotinic, α1, mouse (30); cholinergic receptor, nicotinic, β, mouse (31); cholinergic receptor, nicotinic, α, mouse (10); cholinergic receptor, nicotinic, 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).

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
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 scan 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

<table>
<thead>
<tr>
<th>Organism</th>
<th>Accession No.</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>M12659</td>
<td>5’-AGGCAAACTTTCCGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
<tr>
<td>Rat</td>
<td>M15647</td>
<td>5’-AGGGAATTCTCCGGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
<tr>
<td>Mouse</td>
<td>Y18062</td>
<td>5’-AGGGAATTCTCCGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
<tr>
<td>Goat</td>
<td>D26116</td>
<td>5’-AGGCAAGGTCTCCGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
<tr>
<td>Cow</td>
<td>AF210383</td>
<td>5’-AGGCAAGGTCTCCGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
<tr>
<td>Sheep</td>
<td>X69472</td>
<td>5’-AGGCAAGGTCTCCGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
<tr>
<td>Zebrafish</td>
<td>BX510924</td>
<td>5’-AGGCAAGGTCTCCGCGAGCTTTCTGCCTTTGCTCTACAGTCTCCT-3’</td>
</tr>
</tbody>
</table>

The E box core is underlined, and mismatches are shown in bold. Lowercase letters indicate consensus nucleotides.
activation resulting from ectopic IGF-I

**Regulation of**

**Promoter vector**

**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.

**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 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).

**E-box proteins activation of IGF-I promoters.** Reporter gene activation by any ectopic bHLH factor in isolation led to only small increases in Luc activity 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
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 activation 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 myoblasts 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 oligonucleotides 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 (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-
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. Myogenin mutant mice display a deficit in skeletal muscle mass during development and have reduced expression of IGF-I. In myogenic cells, IGF-I 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-I-630/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


