Development 130, 3415-3426

doi:10.1242/dev.00552

Analysis of a key regulatory region upstream of the *Myf5* gene reveals multiple phases of myogenesis, orchestrated at each site by a combination of elements dispersed throughout the locus

Juliette Hadchouel[1,*,†], Jaime J. Carvajal[2,*], Philippe Daubas[1], Lola Bajard[1], Ted Chang[1], Didier Rocancourt[1], David Cox[2], Dennis Summerbell[2], Shahragim Tajbakhsh[1], Peter W. J. Rigby[2] and Margaret Buckingham[1,‡]

[1]Département de Biologie du Développement, CNRS URA 2578, Institut Pasteur, 25 rue du Dr Roux, 75015 Paris, France
[2]Section of Gene Function and Regulation, The Institute of Cancer Research, Chester Beatty Laboratories, 237 Fulham Road, London SW3 6JB, UK

*These authors contributed equally to this work
†Present address: INSERM U36, Médecine Expérimentale, Collège de France, 11 place Marcelin Berthelot, 75005 Paris Cedex 231, France
‡Author for correspondence (e-mail: margab@pasteur.fr)

Accepted 16 April 2003

SUMMARY

Myf5 is the first myogenic regulatory factor to be expressed in the mouse embryo and it determines the entry of cells into the skeletal muscle programme. A region situated between –58 kb and –48 kb from the gene directs *Myf5* transcription at sites where muscles will form. We now show that this region consists of a number of distinct regulatory elements that specifically target sites of myogenesis in the somite, limbs and hypoglossal cord, and also sites of *Myf5* transcription in the central nervous system. Deletion of these sequences in the context of the locus shows that elements within the region are essential, and also reveals the combinatorial complexity of the transcriptional regulation of *Myf5*. Both within the –58 kb to –48 kb region and elsewhere in the locus, multiple sequences are present that direct transcription in subdomains of a single site during development, thus revealing distinct phases of myogenesis when subpopulations of progenitor cells enter the programme of skeletal muscle differentiation.

Key words: Myf5, Mouse embryo, Myogenesis, Transcriptional regulation, BACs, Myotome, Limb, CNS

INTRODUCTION

The emergence of different cell types during embryogenesis depends on signals, usually from adjacent tissues, leading to the expression of genes that determine the differentiation programme the cell will enter. Activation of a regulatory cascade culminates in the acquisition of a tissue-specific phenotype. Skeletal myogenesis conforms to this model. In the mouse embryo, Myf5 is the myogenic determination factor that is responsible for directing cells into the skeletal muscle programme (Braun et al., 1992; Tajbakhsh et al., 1996b). In keeping with this role, Myf5 is capable of remodelling the chromatin of muscle genes, rendering them accessible to transcriptional activation (Bergstrom and Tapscott, 2001; Gerber et al., 1997). At later stages of development, the expression of MyoD can also activate the myogenic cascade. The double null mutation of *Myf5* and *MyoD* prevents the formation of skeletal muscle precursor cells (Rudnicki et al., 1993). However *Myf5*, together with *Pax3*, lies upstream of *MyoD* (*Myod1* – Mouse Genome Informatics) in the genetic hierarchy that regulates myogenesis (Tajbakhsh et al., 1997). Mrf4 (Myf6 – Mouse Genome Informatics) and myogenin, the other two members of the family of basic-helix-loop-helix transcription factors to which MyoD and Myf5 belong, are implicated in muscle cell differentiation in the embryo. Myogenin is directly involved in the transcriptional activation of muscle genes and in its absence many skeletal muscles are compromised (Venuti et al., 1995), whereas Mrf4 appears to play this role for the differentiation of the myotome, the earliest muscle mass to form in the embryo (Buckingham, 1994; Nabeshima et al., 1993). MyoD can also act as a differentiation factor, as evidenced by the phenotype of the *MyoD/Mrf4* double mutant where myoblasts are present but fail to differentiate (Rawls et al., 1998).

During embryogenesis, the spatiotemporal expression of *Myf5* marks skeletal muscle precursor cells and sites of myogenesis (see Tajbakhsh and Buckingham, 2000). In newly formed somites, it is expressed before the other myogenic factors, in the epaxial part of the dermomyotome, adjacent to the neural tube. This is the first source of muscle precursor cells, which involute and migrate from this epithelium to form the first differentiated skeletal muscle, the myotome. The hypaxial extremity of the dermomyotome and later somitic bud is also a source of muscle precursors, which express *Myf5* and

form the hypaxial myotome, the origin of intercostal and body wall muscles (Christ et al., 1983). This part of the dermomyotome, in somites at the appropriate axial level, also gives rise to cells which migrate to other sites of myogenesis, to the limb buds or, via the hypoglossal chord (Mackenzie et al., 1998; Noden, 1983), to the pharynx, tongue and probably also diaphragm (Tremblay et al., 1998). Cells that migrate to the limb do not express *Myf5* until they reach their destination (Tajbakhsh and Buckingham, 1994). Myf5-positive cells are already present in the hypoglossal cord where, in contrast to the limb progenitors, they do not migrate as separate mesenchymal cells, but rather as a coherent cell mass (Noden, 1983). *Myf5* is also expressed in the branchial arches in cells derived from anterior paraxial mesoderm which will contribute to the formation of facial muscles. An unexpected site of *Myf5* expression is in neurones, in prosomeres p1 and p4 of the brain (Tajbakhsh and Buckingham, 1995) and in a ventral domain of the neural tube (Tajbakhsh et al., 1994). The Myf5 protein does not accumulate in the central nervous system and there is no detectable neuronal perturbation in the *Myf5* mutant embryos (Daubas et al., 2000).

The molecular mechanisms that lead to *Myf5* activation at these multiple sites in the embryo are largely unknown. Signals from the axial structures, neural tube and notochord, and from surface ectoderm are required for the initiation of myogenesis in the somite in the mouse (Summerbell and Rigby, 2000; Tajbakhsh and Buckingham, 2000), as in the avian embryo (Christ and Ordahl, 1995). Sonic hedgehog from the notochord and floor plate of the neural tube, as well as Wnt proteins produced by the dorsal neural tube and surface ectoderm, have been implicated in *Myf5* activation, which is particularly responsive to β-catenin-dependent Wnt1 signalling, as shown in mouse embryo explants of presomitic mesoderm or immature somites (Borycki et al., 1999; Tajbakhsh et al., 1998). Expression of *Myf5* in explants from the region of the brain where the gene is transcribed also responds in this way to Wnt1, leading to the suggestion that this may reflect misfiring of a regulatory element that normally directs *Myf5* transcription in the somite, in response to the Wnt1 signal (Daubas et al., 2000).

The multiple regulatory elements that direct the complete spatiotemporal expression pattern of *Myf5* in the embryo have been mapped within 145 kb of genomic DNA by YAC (Hadchouel et al., 2000) and BAC (Carvajal et al., 2001a) transgenic analysis. Another gene, which encodes the myogenic regulatory factor Mrf4, lies 7 kb 5′ of the *Myf5* gene (Braun et al., 1990; Miner and Wold, 1990). Within this intergenic region, just 3′ of the *Mrf4* gene, an enhancer element has been identified that directs early expression of *Myf5* in the epaxial dermomyotome (Summerbell et al., 2000; Teboul et al., 2002). Gustafsson et al. have reported that this element responds to sonic hedgehog signalling, which is required for the expression of *Myf5* specifically in this epaxial domain (Borycki et al., 1999; Gustafsson et al., 2002). Also in the intergenic region between *Mrf4* and *Myf5*, further regulatory regions have been described, including one that directs expression to the branchial arches (Patapoutian et al., 1993; Summerbell et al., 2000) and another that leads to expression in the neural tube (Summerbell et al., 2000). Within the *Myf5* gene itself, an enhancer that directs expression to the hypaxial domain of the somite has been identified (Summerbell et al., 2000). In the genomic DNA lying upstream of the *Mrf4-Myf5* genes, there are a number of regulatory regions, including one responsible for maintenance of *Myf5* expression (–88 kb to –81 kb) in some muscles of the trunk and head and others responsible for aspects of *Myf5* expression in the hypaxial somite, arches and hypoglossal cord (Carvajal et al., 2001a; Carvajal et al., 2001b; Hadchouel et al., 2000). In particular, a sequence lying between –58 kb and –48 kb from the *Myf5* gene (Hadchouel et al., 2000) directs expression of a transgene to sites of myogenesis in the somites, hypoglossal cord and limb buds, as suggested by deletions in this region (Carvajal et al., 2001a; Zweigerdt et al., 1997), and also to the brain. In this paper, we present the analysis of this region and show that it is necessary for transcription of *Myf5* at these sites in the embryo. Different sites are targeted by distinct sequences and, furthermore, within a single site, regulatory subdomains emerge that are identified by dissection of the region or revealed after its deletion. Unexpectedly, a distinct sequence targets transcription of *Myf5* to the central nervous system. At least three sequences are active at different times and to varying extents in fore- versus hindlimbs. In the somite, *Myf5* expression is regulated by a minimum of six different sequences, with more than one regulatory module required even within an ostensibly uniform structure such as the myotome. The fine analysis of the region located between –58 kb and –48 kb from *Myf5*, and the consequences of its deletion add a further dimension to our appreciation of the complexity of the information that has to be integrated to ensure the spatiotemporal regulation of this myogenic determination gene. Furthermore, this analysis reveals subpopulations of cells that contribute to myogenesis as the embryo develops.

MATERIALS AND METHODS

Plasmid transgene constructions

All plasmid constructs used in this study, except p-58/-48*tk-nlacZ* and p8.8 (Summerbell et al., 2000), are derived from p*baMyf5-nlacZ*, (Hadchouel et al., 2000), containing the *nlacZ* reporter gene with 2.6 kb of *Myf5* 5′ flanking sequence that includes the *Myf5* branchial arch element located between –1.7 kb (*Nhe*I site) and –561 bp (*Bsa*BI site), the neural tube element located between –561 bp and –141 bp (*Xba*I site) (Summerbell et al., 2000), and the *Myf5* minimal promoter (–141 bp from the Cap site). We refer to this construct as *pMyf5-nlacZ*. p-58/-48*Myf5-nlacZ* (Hadchouel et al., 2000) was totally digested by *Xho*I then partially digested by *Sau*IIIA and religated to itself to create a random 5′ deletion series of the –58/–48 fragment. Among the resulting constructs, p-56.4/-48*Myf5-nlacZ* and p-53.3/-48*Myf5-nlacZ* were selected for micro-injection. p-58/-48*Myf5-nlacZ* was digested by *NcoI* and *Hin*dIII, and religated on itself to obtain p-58/-54*Myf5-nlacZ*. A 1.4 kb *Ava*I or a 2.9 kb *Ava*I-*Acc*I fragment was subcloned in p*Myf5-nlacZ* to create p-58/-56.6*Myf5-nlacZ* or p-56.6/-53.7*Myf5-nlacZ*, respectively. A 5 kb *Cla*I-*Xho*I fragment located between –63 and –58 was subcloned in p-58/-54*Myf5-nlacZ* to obtain p-63/-54*Myf5-nlacZ*. The fragments –58/–57, –57.5/–56.6 and –57.5/–57 were synthesized by PCR using the Advantage 2 Taq polymerase (Clontech), with *Xho*I and *Hin*dIII sites added to 5′ and 3′ primers, respectively. The forward and reverse primers used to generate the –58/–57, –57.5/–56.6 and –57.5/–57 fragments were, respectively: fwd (5′-GGG CTC GAG CAA TAT AAT GTC-3′), rev (5′-AAA AAG CTT TTT CTC TTA AGA GAG AGC TTG GGC ACC-3′); fwd (5′-AAA CTC GAG GTA TGT TTG TTG GAA AGG CC-3′), rev (5′-AAA AAG CTT GGG ATT TGT ACC TCC ATC AGA TGG G-3′); and fwd (5′-AAA CTC

GAG GTA TGT TTG TTG GAA AGG CC-3′), rev (5′-AAA AAG CTT TTT CTC TTA AGA GAG AGC TTG GGC ACC-3′). The PCR products were checked by sequencing (Genome Express) and cloned into p*Myf5-nlacZ*, in front of the transgene.

p-58/-48*tk-nlacZ* was created by subcloning the 10 kb *Xho*I-*Hin*dIII fragment located between –58 and –48 kb into p*tk-nlacZ*, which contains the *nlacZ* reporter gene under the control of the thymidine kinase (*tk*) promoter isolated from pBL*CAT2* (Luckow and Schütz, 1987).

BAC transgene constructions

All BAC deletion constructs are based on BAC195*APZ* (Carvajal et al., 2001a). For BAC195Δ54-49, homology arms were synthesized by PCR amplification of BAC195*APZ* with the following primer pairs: 5′ homology arm (1220 bp), Δ54-49.5F (5′-TTA GAT CTA TTG TCA GAA GAA TAG AGA AAA GGA-3′) and Δ54-49.5R (5′-AAG GAT CCG ATC TTG AAG AAA TTT TGG TAA TTC C-3′); 3′ homology arm (1209 bp), Δ54-49.3F (5′-GTT TTG ATA GAG GAT GAA TAC TCA A-3′) and Δ54-49.3R (5′-TCA TTT GAA TAG AGA CCT AAA GAT C-3′). The 5′ homology arm fragment was cloned into pCRII (Invitrogen) to generate pΔ54-49.5. The 3′ homology arm fragment was digested with *Xho*II and cloned into a *Bam*HI-digested pΔ54-49.5 to give pCasΔ54-49.

For BAC195Δ59-54, homology arms were synthesized by PCR amplification of BAC195*APZ* with the following primer pairs: 5′ homology arm (140 bp), Δ59-54.5F (5′-CTG ATG CAT GCT TGT CAT GGT-3′) and Δ59-54.5R (5′-TGG ATC CTG AAA ACG TGA GGC ACC GGA GG-3′); 3′ homology arm (140 bp), Δ59-54.3F (5′-CCA TAG GAA TTA CCA AAA TTT CTT C-3′) and Δ59-54.3R (5′-CGT AAA CCA TTA AGA TGG TGG-3′). To generate the deletion cassette, 5′ and 3′ homology arms were digested with *Bam*HI and *Xho*II, respectively, ligated and re-amplified using Δ59-54.5F and Δ59-54.3R.

For BAC195Δ63-48, homology arms were synthesized by PCR amplification of BAC195*APZ* with the following primer pairs: 5′ homology arm (145 bp), Δ63-48.5F (5′-AAA TGT GCT AAT GTG GAG AGG-3′) and Δ63-48.5R (5′-CAC ATA CAC AAC TTC ACA AAA GCT ATG CCA GGT TGC TAT CCC TCC-3′, including a 24mer tail homologous to the 5′-end of the 3′ homology arm); 3′ homology arm (144 bp), Δ63-48.3F (5′-AGC TTT TGT GAA GTT GTG TAT GTG-3′) and Δ63-48.3R (5′-GTC TGC ATG GAA CTA GTG TAA-3′). To generate the deletion cassette, 5′ and 3′ homology arm fragments were gel purified, mixed in a 1:1 ratio, denatured for 5 minutes at 95°C and left to re-anneal at 37°C for 30 minutes. Standard PCR-mix, not including primers, was added and the reaction incubated at 72°C for 30 minutes. The extended products were then re-amplified by PCR using Δ63-48.5F and Δ63-48.3R.

For BAC195Δ54-49, pCasΔ54-49 was digested with *Not*I to excise the deletion cassette. Fragments were isolated by gel electrophoresis and subcloned into a *pSV-RecA* vector (Yang et al., 1997), which had been modified by introducing a single *Not*I site replacing the *Sal*I cloning site (D.C., J.J.C. and P.W.J.R., unpublished). Generation of the deleted BAC with this plasmid construct was carried out as previously described (Cox et al., 2002). For the generation of other BAC deletion constructs, we used our own modification of the linear recombination method (Lee et al., 2001; Swaminathan et al., 2001). BAC195Δ59-54 was generated by introducing the 5 kb deletion in BAC195*APZ*, while BAC195Δ63-48 was generated by deleting the entire region from BAC195Δ59-54.

For BAC195Δ59-54 and BACΔ63-48, 25 ng of DNA from each BAC was used to transform electrocompetent *E. coli* DY380 cells. Single colonies growing under chloramphenicol (CAM) selection were analysed to check that BACs transferred to the new host were not rearranged. Single colonies were isolated and grown on LB-CAM media overnight at 32°C with constant agitation. This culture (220 μl) was used to seed 11 ml of LB-CAM. Cells were incubated at 32°C with shaking (>250 rpm) until the OD_{600} was between 0.5 and 0.7. Cultures were then incubated at 42°C for 15 minutes, transferred to wet ice and left to cool down for at least 20 minutes. Cells were washed three times in ice-cold H_2O and electroporated immediately after the last wash. Deletion cassettes were denatured by 10 minutes incubation in 300 mM NaOH, ethanol precipitated, resuspended in 20 μl of cold H_2O and mixed with the electrocompetent DY380 cells carrying the BAC of interest. Electroporation conditions were as follows: 1.75 kV, 200 Ohms, 25 μF. After electroporation, cells were diluted and plated into a single 96-well plate at a density of 10-30 cells/well. After overnight growth at 32°C, colony pools were screened by PCR using primers outside the deletion cassette. Positive pools were diluted and plated on LB-agar CAM to obtain ~250 colonies/plate and incubated overnight at 32°C. Single colonies were picked onto 96 well plates, grown overnight at 32°C with constant agitation and PCR-screened with the same primer pair to identify positive clones. All clones were sequenced to confirm the deletion, and the integrity of the sequences corresponding to the homology arms. Furthermore, positive clones were digested with a panel of restriction enzymes to ensure no additional deletions, insertions or rearrangements had occurred.

Generation of transgenic mice

BAC DNA purification was carried out as described previously (Carvajal et al., 2001a) and plasmid fragments as described elsewhere (Kelly et al., 1995). Transgenic mice were generated by microinjection of purified BAC or plasmid DNA into fertilized (C57BL/6J×SJL) or (CBA×C57BL/6J) F_2 eggs at a concentration of

Table 1. Transient transgenic embryos

Transgene	E9.5/E9.75	E10.5/E.11	E11.5/E12	E12.5	E13.5/E14.5
–58/–48 *tk-nlacZ*			8 (8)		
–56.4/–48 *Myf5-nlacZ*			8 (12)		
–53.3/–48 *Myf5-nlacZ*		2 (4)	7 (10)		3 (4)
–56.6/–53.7 *Myf5-nlacZ*	4 (6)		4 (6)		
–58/–54 *Myf5-nlacZ*	8 (8)	2 (2)	6 (8)		
–58/–56.6 *Myf5-nlacZ*			8 (9)	3 (3)	
–58/–57 *Myf5-nlacZ*			6 (6)		
–57.5/–56.6 *Myf5-nlacZ*			6 (8)		
–57.5/–57 *Myf5-nlacZ*	4 (9)		3 (3)		3 (4)
–63/–54 *Myf5-nlacZ*	8 (13)	3 (3)	5 (6)		
BAC 195*APZ*			3 (3)		
BAC 195*APZ*Δ63-48			3 (3)		
BAC 195*APZ*Δ59-54	8 (8)	1 (2)	4 (4)		3 (3)
BAC 195*APZ*Δ54-49			6 (6)		

Numbers are given at each stage for embryos that showed *Myf5*-specific X-gal staining. The numbers in brackets indicate the total number of transgenic embryos examined.

~1-2 ng/μl using standard techniques (Hogan et al., 1994). Injected eggs were reimplanted the same day or the day after the injection into pseudopregnant (C57BL/6J × CBA) F_1 foster mothers.

Identification of transgenic animals

DNA was prepared from mouse tails or, for transient transgenics, a region of the embryo, and analysed by PCR, using standard techniques.

Analysis of transgene expression

Heterozygous transgenic males were crossed with non-transgenic females ([C57BL/6J×SJL or C57BL/6J×CBA] F_1). Embryos were dated, taking E0.5 as the day of the appearance of the vaginal plug. Transient transgenic embryos were dated taking the day of reimplantation into the pseudo-pregnant foster mothers as E0.5. Numbers of positive transient transgenics analysed at each time point are indicated in Table 1. The following transgenic lines, with numbers given in brackets, were also analysed: –58/–48*Myf5-nlacZ* (2), BAC*195Z* (6), BAC*195APZ* (5), BAC*195Δ63-48* (3), BAC195Δ59-54 (3) and BAC195Δ54-49 (4).

X-gal staining: embryos were dissected in PBS, fixed in 4% paraformaldehyde for 5 to 60 minutes depending on the age of the embryo, or in Mirsky's Fixative (National Diagnostics) for 1 hour to overnight, rinsed twice in PBS and stained in X-gal solution (Summerbell et al., 2000; Tajbakhsh et al., 1996a) at 37°C from 2 hours to overnight. Transgenic embryos were examined by whole-mount microscopy or cryostat sectioning, as described previously (Kelly et al., 1995).

Immunofluorescence: cryostat sectioning and immunofluorescence experiments were performed as described elsewhere (Teboul et al., 2002), with use of a monoclonal anti-myosin heavy chain antibody (MF20, dilution 1:50) from the Developmental Studies Hybridoma Bank.

RESULTS

In order to investigate the spatiotemporal regulation of *Myf5* expression by the sequence that we had identified between –58 kb and –48 kb (Hadchouel et al., 2000), we constructed transgenes containing fragments of this region as indicated in Fig. 1A. They were placed in front of a proximal *Myf5* sequence (2.6 kb), which includes the promoter, and ventral neural tube and branchial arch elements (Summerbell et al., 2000), which serve as positive controls for expression of the *nlacZ* transgene. In Fig. 1B, we show the *nlacZ* expression at embryonic day (E)11.5 of a transgene where the reporter is regulated by the –58/–48 sequence with a heterologous thymidine kinase promoter. The expression pattern at this stage was similar to that previously described with the *Myf5* promoter (Hadchouel et al., 2000). The fact that the sequence acts with a heterologous promoter, independently of position relative to the promoter suggests that it has the properties of an enhancer.

The –53.3/–48 region directs expression mainly to the most anterior somites and to the hindlimb

The region between –53.3 kb and –48 kb placed in front of the proximal *Myf5* sequence directed reporter gene expression to musculature in the cervical region and in the limb buds at E12 as shown in Fig. 2A. There was also labelling at this stage of some facial muscles derived from the branchial arches because of the branchial arch element. Hindlimb muscle masses were more strongly labelled than those in the forelimb, despite the anteroposterior developmental gradient (see Fig. 1B). With a –56.4/–48 *Myf5-nlacZ* transgene, the hindlimb bud was also more strongly labelled than the forelimb (Fig. 2B) at E11. β-galactosidase-positive cells were present in anterior somites, and the section shown in Fig. 2C,D indicates that these represent a subpopulation of differentiated muscle cells in the myotome, which is marked by myosin heavy chain expression. The myotomal expression is also seen with the –53.3/–48 *Myf5-nlacZ* transgene (results not shown). In some embryos where transgene expression was particularly strong, labelled cells were also detected in more posterior myotomes. At E10.5, when *Myf5* is normally expressed in the forelimb, no β-galactosidase positive cells were detectable with these transgenes in the limb buds. At this earlier stage, occasional embryos showed weak myotomal labelling, varying in its extent on the anteroposterior axis. At later stages (e.g. E13.5) shoulder and limb muscles continued to be positively labelled with this transgene (results not shown).

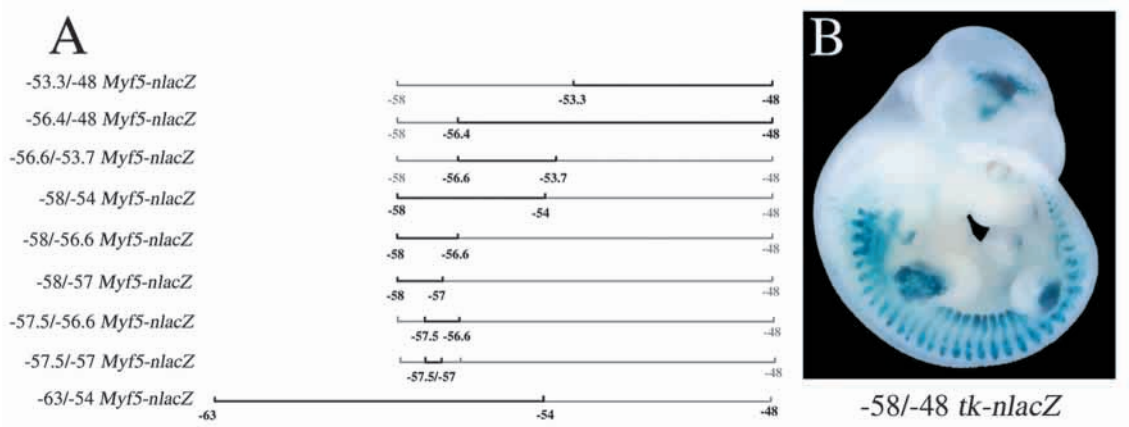

Fig. 1. The regulatory region at –58/–48 kb upstream of *Myf5*. (A) Schematic representation of the different fragments of the –58/–48 kb region and its upstream extension to –63 kb, which are subcloned in front of the *Myf5-nlacZ* cassette (2.6 kb, including a branchial arch element, neural tube element and the *Myf5* promoter) (Hadchouel et al., 2000; Summerbell et al., 2000). For each construct, the cloned region is indicated by a black line. Distances in kb are from the *Myf5* Cap site. (B) The *nlacZ* expression pattern, revealed by X-gal staining, is shown for an embryo at E11.5 (40-45 somites) with a –58/–48 *tk-nlacZ* transgene, where the region between –58 kb and –48 kb has been placed in front of a heterologous thymidine kinase promoter.

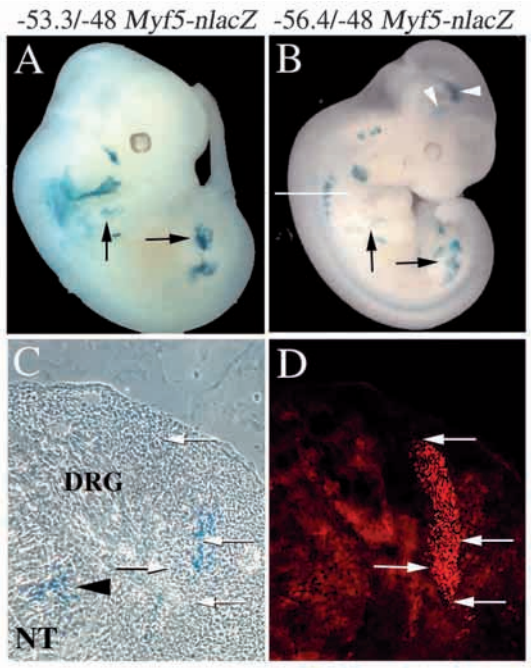

Fig. 2. The –53.3/–48 region drives *Myf5-nlacZ* expression mainly in cervical somites and the hindlimb bud. X-gal staining at E12, with a –53.3/–48 *Myf5-nlacZ* transgene (A), and at E11.0, with a –56.4/–48 *Myf5-nlacZ* transgene (B). Black arrows indicate X-gal staining in fore- and hind-limbs. β-galactosidase-positive cells are detected in the brain in prosomeres p1 and p4 (white arrowheads). The embryo shown in B was sectioned at the level of the cervical somites (white line) and the corresponding transverse cryostat section is shown under phase contrast (C) or fluorescence (D), after immunoreaction with an antibody against muscle myosin heavy chain (red labelling) to show differentiated muscle cells in the myotome, the extent of which is indicated by white arrows. Black arrowhead in C indicates β-galactosidase-positive neurones within the ventral neural tube (NT). DRG, dorsal root ganglia.

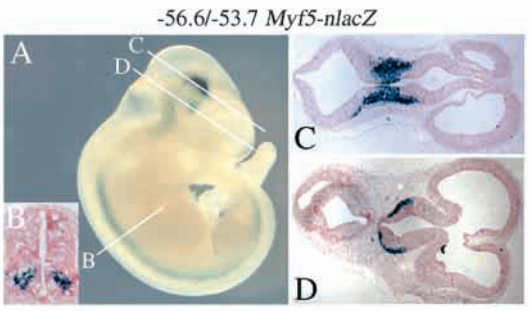

Fig. 3. The –56.6/–53.7 region drives *Myf5-nlacZ* expression only in the central nervous system. (A) X-gal staining at E11.5 of an embryo, with a –56.6/–53.7 *Myf5-nlacZ* transgene, showing *nlacZ* expression in the embryonic brain (prosomeres p1 and p4) and in the neural tube. The embryo in A was sectioned at levels indicated by the white lines, and sections at these levels are shown in B-D. β-galactosidase-positive cells are detected in the ventral neural tube (B) and in the neuroepithelium at the levels of prosomeres p1 (C) and p4 (D). In these brain sections, telencephalic vesicles are on the right.

The –56.6/–53.7 region directs expression in the central nervous system

In Fig. 2B, expression in the brain and ventral neural tube (Fig. 2C) was detected, suggesting that an element responsible for the transcription of *Myf5* in the central nervous system (Tajbakhsh and Buckingham, 1995) is located between –56.4 kb and –53.3 kb (compare Fig. 2A with 2B). Fig. 3A shows that this is indeed the case. With a –56.6/–53.7 *Myf5-nlacZ* transgene, prosomeres p1 and p4 were labelled. A section through prosomere p1 (Fig. 3C) shows β-galactosidase-positive cells in this region of the diencephalon, where the endogenous gene is expressed, extending from the mantle layer into the subventricular zone where neuronal precursors are present. In contrast, the labelled cells in prosomere p4 in a section through the mesencephalon (Fig. 3D) were confined to the outer layer of the neuroepithelium where differentiated neurones are present, as seen with the endogenous gene (Daubas et al., 2000; Tajbakhsh and Buckingham, 1995). Expression of the transgene was also seen in cells in the ventral neural tube (Fig. 3B). As in the case of prosomere p1, expression was in the same region, but not identical to that seen with the endogenous gene, possibly because further regulatory elements which refine and restrict transcription are absent from the 2.9 kb fragment of the enhancer present in this transgene. As the 2.6 kb proximal region of *Myf5*, which is present in this and other transgenes, contains a neural tube element (Summerbell et al., 2000) and the complete –58/–48 region, when placed in front of a minimal *tk* promoter (Fig. 1B), also directs expression to the ventral neural tube, we conclude that two distinct elements are responsible for this site of expression, and probably cooperate to target the subset of neurones marked by transcription of the endogenous gene (Tajbakhsh et al., 1994).

The –58/–56.6 region contains elements which target limb buds, hypoglossal cord, and distinct regions of the somite

A transgene with the region between –58 and –54 kb directed *Myf5-nlacZ* expression in the central nervous system, as expected, and at additional sites in the embryo (Fig. 4). At E9.75 (Fig. 4A), there was no labelling in the most caudal immature somites, where myogenesis is initiated by the early epaxial enhancer that activates *Myf5* transcription in the dermomyotome (Teboul et al., 2002). In more mature somites, β-galactosidase positive cells were present in the myotome, underlying the epithelial dermomyotome in a central position (Fig. 4B) where muscle differentiation first takes place (Spörle, 2001). This corresponds to the region of the myotome that is revealed when the earlier epaxial enhancer is deleted (Teboul et al., 2002). At E10.5 (Fig. 4C), labelled cells continued to be located in the central part of the myotome. By E11.5 the expression pattern in the somite has changed (Fig. 4D). Cells in the epaxial myotome were still β-galactosidase-positive in a region that is intercalated between epaxial most and hypaxial domains of the myotome (Fig. 4E). Labelling was now also detectable in more mature anterior somites in the epaxial lip of the dermomyotome and, in the interlimb region, the hypaxial somitic bud also contained β-galactosidase-positive cells (Fig. 4E). At this stage, the epithelial structure of the dermomyotome is no longer present except at the extremities of the somite. Some labelled cells could also be seen in the hypaxial myotome. In previous studies, the –58 kb end point was chosen arbitrarily. Given that sequences up to this point conferred muscle expression, we tested more upstream sequences for activity in transgenic assays. In Fig. 4F, an E11.5

embryo with a –63/–54 *Myf5-nlacZ* transgene is shown. The X-gal labelling is essentially similar to that with the –58/–54 sequence. This additional 5 kb of 5′ sequence therefore did not extend the expression pattern.

Fig. 5A shows that a –58/–56.6 *Myf5-nlacZ* transgene, which no longer contains the distal central nervous system element, directed expression at E11.5 to the sites of myogenesis shown in Fig. 4D. The –58/–57 part of this fragment (Fig. 5B) continued to direct expression in the hypoglossal cord and limb buds, whereas labelling in the somite was confined to the epaxial and hypaxial extremities (Fig. 5C), where β-galactosidase-positive cells were clearly seen in the epithelial structures of the epaxial lip (Fig. 5D) and hypaxial somitic bud (Fig. 5E). In contrast, a –57.5/–56.6 fragment directed expression to the intercalated domain of the epaxial myotome (Fig. 5F,G), showing that this site of *Myf5* transcription in the somite is under distinct regulation. This transgene showed variable expression in the epaxial lip and somitic bud. Both the –58/–57 and –57.5/–56.6 fragments directed expression in the limb buds and hypoglossal cord (Fig. 5B,F). Consistent with this, the overlapping 500 bp fragment (–57.5/–57) also directed expression to these sites (Fig. 5H). This fragment also showed strong expression in the hypaxial somitic bud and in the epaxial lip of the dermomyotome.

Deletion of the –58/–48 region reveals its essential role in directing *Myf5* expression in the embryo

The preceding observations demonstrate that the –58/–48 region directs transcription of *Myf5* to the brain, limbs, hypoglossal cord and to specific domains of myogenesis in the somite. In order to establish whether this region is necessary for *Myf5* expression at these sites in the context of the locus, it was deleted

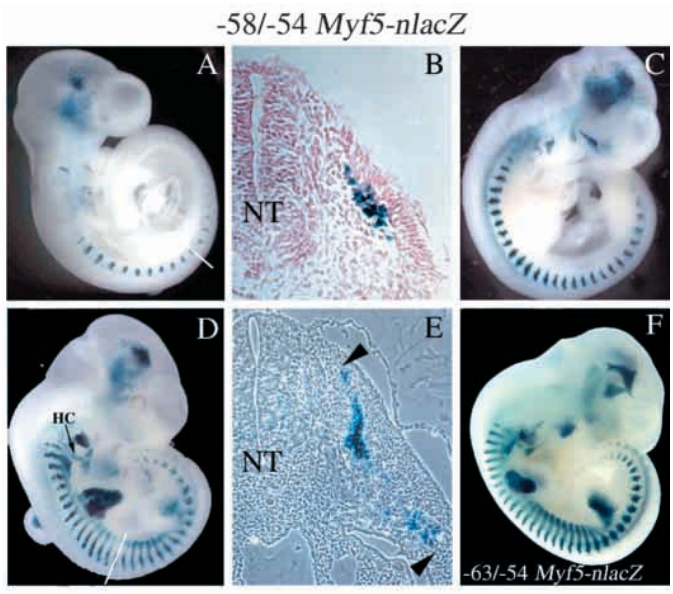

Fig. 4. The –58/–54 region drives *Myf5-nlacZ* expression to several sites of myogenesis in the somites, limb buds and hypoglossal cord. (A,C,D) X-gal staining of an embryo with a –58/–54 *Myf5-nlacZ* transgene at E9.75 (A, 26 somites), E10.5 (C, 30 somites) and E11.5 (D, 36 somites). At E9.75, β-galactosidase-positive cells are detected in the developing myotome in the central epaxial region, as shown in a transverse cryostat section (stained with Eosin) (B) of the same embryo shown in A (white line indicates level of sectioning). At E10.5 (C), *Myf5-nlacZ* expression is still restricted to the same area of the myotome. It is detected in the extremities of the more mature somites at E11.5 (D). (E) Phase contrast microscope view of a section in the interlimb region of the embryo shown in D (white line indicates level of sectioning) where both the epaxial lip of the epithelial dermomyotome and the hypaxial somitic bud contain cells with blue nuclei (black arrowheads). At this stage, there is strong X-gal staining in the forelimb buds and labelling is also detectable in the less mature hindlimb buds. Cells in the hypoglossal cord are also positive for β-galactosidase. (F) X-gal staining of an embryo at E11.5 (40 somites), with a –63/–54 *Myf5-nlacZ* transgene showing a pattern which is essentially similar to that shown in D. With these two constructs, *Myf5-nlacZ* activity is also detected in the brain and neural tube, owing to the central nervous system element (–56.6/–53.7), and in the branchial arches, owing to the control branchial arch sequence. There is also ectopic labelling of head mesenchyme, which is particularly notable in C. NT, neural tube; HC, hypoglossal cord.

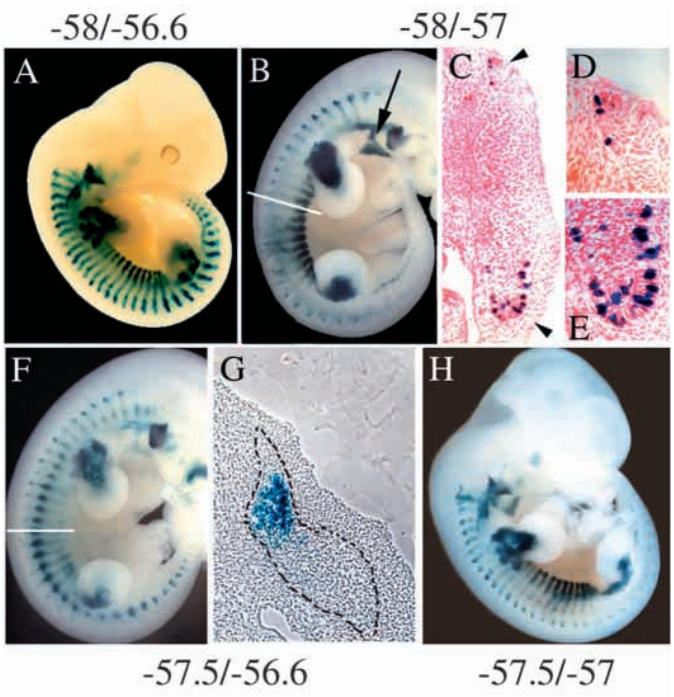

Fig. 5. The –58/–56.6 region contains elements that direct transgene expression to distinct regions of the somite, as well as to the limb buds and hypoglossal cord. X-gal staining of embryos at E11.5 (38-40 somites) in which the *Myf5-nlacZ* transgene contains a –58/–56.6 kb fragment (A) or subfragments of it: –58/–57 (B), –57.5/–56.6 (F) or –57.5/–57 (H). In A, expression is no longer seen in the brain (compare with Fig. 4D), but in somites, limb buds and hypoglossal cord. In B, the –58/–57 subfragment directs expression to the limb buds and hypoglossal cord (black arrow) and to the epaxial and hypaxial extremities of the somite. The cryostat section in C (stained with Eosin), corresponds to a transverse section (white line in B) at the thoracic level of embryo (B). Higher magnifications (D,E) show β-galactosidase-positive cells in the epithelial structures of the epaxial lip (D) and hypaxial somitic bud (E). In F, somitic expression is mainly seen in the epaxial intercalated region of the myotome, shown on a transverse cryostat section at the thoracic level (white line) in G. The outline of the myotome, which is marked by immune staining with a myosin heavy chain antibody (not shown) is presented as a broken line. In H, the 500 bp (–57.5/–57) fragment directs expression to the limb buds, hypoglossal cord, and epaxial and hypaxial extremities of the somite.

from BAC195*APZ*, which contains 195 kb of genomic DNA 5′ to the *Myf5* gene and 39.6 kb 3′ to it. This BAC recapitulates the complete pattern of *Myf5* expression, as monitored with an *nlacZ* reporter sequence, introduced at the ATG of the gene (Carvajal et al., 2001a). The deletion extends from –63 kb to –48 kb, in order to include any potential sequences that reinforce activity at the 5′ end. As expected, comparison of the *nlacZ* expression profile in the deleted BAC195*APZ*Δ63-48 and wild-type BAC195*APZ* constructs at different developmental stages showed no difference at E8.5 when β-galactosidase-positive cells, which are located in the epaxial dermomyotome, were present in both cases (results not shown). At this stage, *Myf5* expression is directed by the early epaxial enhancer situated at –6.5 kb from the *Myf5* gene (Teboul et al., 2002). At E9.0 (Fig. 6A,B), *Myf5* expression in immature caudal somites was unaffected by the deletion. In more mature somites, the pattern of dermomyotomal expression was skewed towards the caudal edge of the somite (Fig. 6B), compared with that seen in the undeleted BAC (Fig. 6A). At E11.5, myotomal expression was perturbed in BAC195*APZ*Δ63-48 with foreshortening of the epaxial and hypaxial extent (Fig. 6C-F), as shown on interlimb level sections (Fig. 6G,H), which also suggest loss of β-galactosidase-positive cells in the central epaxial domain. At this stage, it was clear that all brain and limb expression of *Myf5* was abolished by the deletion (Fig. 6D). At E10.5, when expression of the transgene begins to be detectable in the forelimb bud, this is not seen with the deleted BAC (results not shown). Expression in the hypoglossal cord is also affected at this stage. These results demonstrate that the region deleted in the BAC is essential for *Myf5* expression in the limb buds, in the hypoglossal cord and in the brain, and for aspects of (dermo)myotomal expression, notably those in the epaxial and hypaxial extremities of more mature somites. The epaxial intercalated expression in the myotome is more difficult to follow in the deleted BAC, probably because it tends to be obscured by more proximal elements that also target β-galactosidase in the myotome, and in particular by the perdurance of β-galactosidase as a result of the activity of the early epaxial enhancer (Teboul et al., 2002). This is certainly also the case for the element at –53.3/–48 that targets anterior somites.

Further deletions of the 5′ (BAC195*APZ*Δ59-54) and 3′ (BAC195*APZ*Δ54-49) parts of the region are shown in Fig. 7 and compared with p8.8, a plasmid that contains the intragenic hypaxial element (Summerbell et al., 2000) and the *Mrf4*/*Myf5* intergenic region, which includes the early epaxial enhancer (Summerbell et al., 2000; Teboul et al., 2002). At E9.5, BAC195*APZ*Δ63-48 (Fig. 7A) and BAC195*APZ*Δ59-54 (Fig. 7B) showed the skewed dermomyotomal expression with a pronounced caudal somitic expression already noted in Fig. 6, which is due to this intragenic element (Fig. 7D). By contrast, epaxial dermomyotome expression in newly formed somites, directed by the early epaxial enhancer (Teboul et al., 2002), is maintained (Fig. 7A-C). Deletion of only the 3′ half of the region in BAC195*APZ*Δ54-49 (Fig. 7C) corrects this pattern, probably owing to the dermomyotomal suppressor element

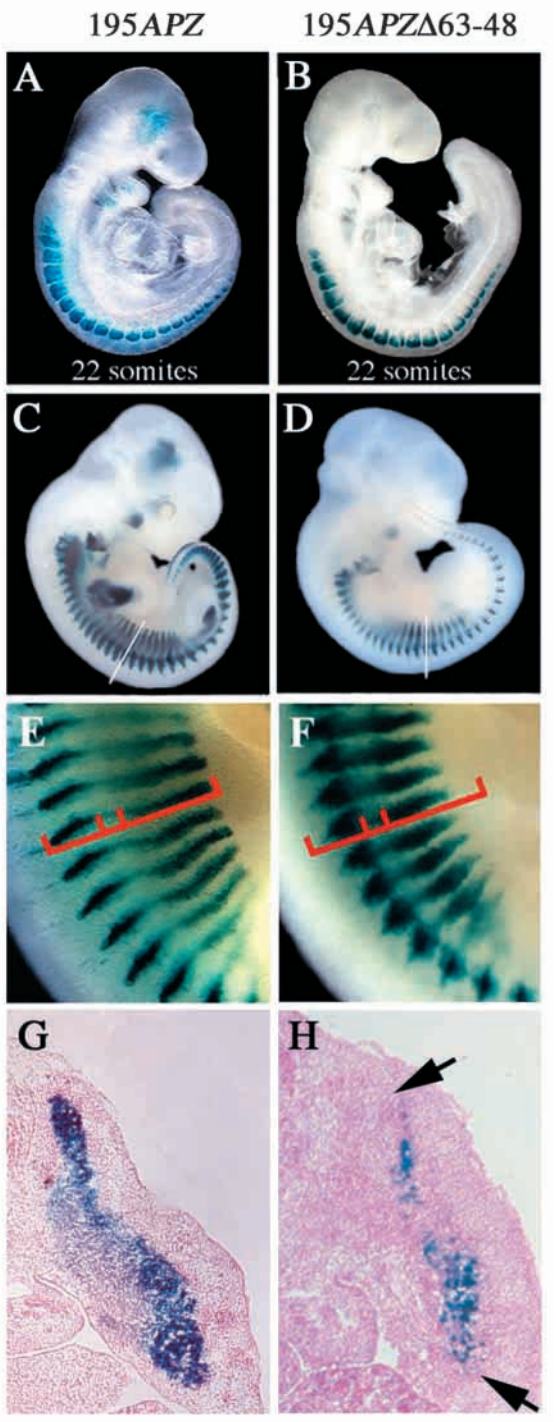

Fig. 6. The –63kb to –48kb region is necessary for *Myf5* expression in limb buds, central nervous system and domains of the somite. Comparison of X-gal staining profiles between embryos of BAC195*APZ* (A,C,E) and BAC195*APZ*Δ63-48 (B,D,F) transgenic lines at different stages of embryonic development. (A,B) At E9.0, the transgene is activated in the epaxial somitic lip as directed by the early epaxial enhancer (Teboul et al., 2002). (C,D) At E11.5, limb and brain expression are missing from the deleted BAC line. Expression in the thoracic somites at this stage does not comprise the full epaxial-hypaxial myotome extension, whereas staining shows a clear caudal bias as shown on higher magnification of these regions of E11.5 embryos (E,F). Expression in the thoracic somites at this stage does not comprise the full epaxial-hypaxial extent of the myotome (indicated by the red lines in E,F). (G,H) Transverse cryostat sections, stained with Eosin and X-gal from the thoracic level (white lines in C,D) of the embryos shown in C,D. Arrows in H indicate the full extent of the myotome (see G).

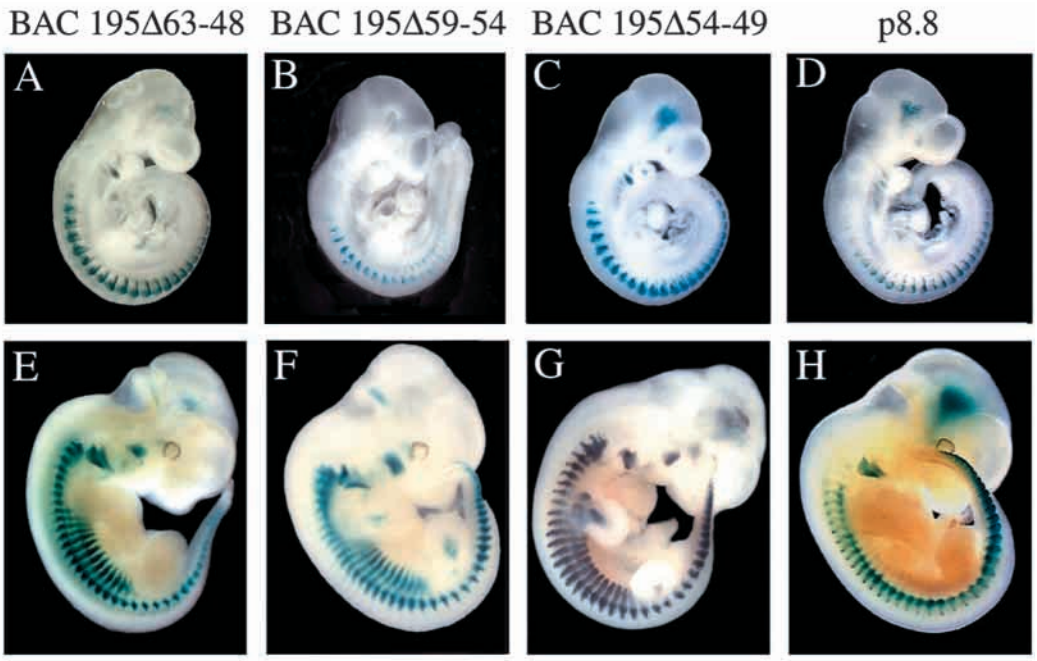

Fig. 7. Multiple deletions within the –63 kb to –48 kb interval reveal interactions between different regulatory regions. Comparison of *Myf5-nlacZ* expression profiles between BAC195*APZ*Δ63–48 (A,E), Δ59-54 (B,F), Δ54-49 (C,G) and p8.8 (D,H) constructs. At E9.5 (A-D), deletion of the entire –63 kb to –48 kb region (A) or the –59 kb to –54 kb region (B) results in a marked caudal somitic expression in the dermomyotome, reminiscent of the expression pattern obtained with plasmid p8.8 (D), and driven by the intragenic enhancer element (Summerbell et al., 2000). Deletion of the –54 kb to –49 kb region does not seem to affect the expression pattern at this stage, as it results in a similar pattern to the wild type. This suggests that the –59 kb to –54 kb interval contains the predicted suppressor of the dermomyotome element. At E11.5 (E-H), limb expression is missing in the –63 kb to –48 kb deletion (E), whereas expression in both limbs is seen with both the half-deletion constructs (F,G). In the –59 kb to –54 kb deletion (F), expression in the hindlimb is stronger, whereas in the –54 kb to –48 kb deletion, the relative intensity of expression between fore- and hindlimbs is reversed. This suggests that both regions are required for the full recapitulation of limb expression at this stage. (H) The intragenic region in p8.8 directs a pronounced band of labeling in the myotome of more mature somites.

previously mapped between –59 kb and –8.8 kb (Carvajal et al., 2001a), and maintains the epaxial myotome expression directed by the –58/–54 region (Fig. 4A,B). At E11.5, BAC195*APZ*Δ54-49 (Fig. 7G) showed the full epaxial/hypaxial extent of *Myf5* expression with labelling of the epaxial lip and notably the hypaxial somitic bud, which is missing in BAC195*APZ*Δ63-48 and BAC195*APZ*Δ59-54 (Fig. 7E,F). Notably in these BACs, expression in prosomeres p1 and p4 was missing (Fig. 7E,F, compare with 7G). At this stage the intragenic region in p8.8 still directed caudal somitic expression and a pronounced band of labelling in the myotome (Fig. 7H). At E11.5, the BAC deletions clearly confirmed observations on limb bud expression. In the absence of the 5′ half of the region, with BAC195*APZ*Δ59-54, this was weak, but more pronounced in the hindlimb, in keeping with our observations on the –53.3/–48 sequence (Fig. 2A), while the presence of the 5′ sequence in BAC195*APZ*Δ54-49 directed strong forelimb and weaker hindlimb expression, following the normal developmental gradient.

At E13.5, some trunk and limb musculature continued to be compromised with BAC195*APZ*Δ63-48; however, labelling of limb muscle masses, particularly in the forelimb was now detectable (Fig. 8B). Observations on fore- and hindlimbs from E12.5 to E14.5 are shown in Fig. 8, for deleted and wild-type BACs. In the forelimb at E12.5 (Fig. 8D,E), more proximal muscles were mainly labelled in the deleted BAC, but by E13.5 (Fig. 8F,G) most muscle masses appeared to be labelled. This was also true for the hindlimb (Fig. 8H-K) with a delay of 1 day. We therefore conclude that the –58/–48 region is essential for early expression of *Myf5* in the limb buds, but that other elements located elsewhere in the locus direct later expression of the gene in the limbs. This expression initially overlaps with that directed by the –58/–48 region, which is still active in the limbs at E13.5 (Fig. 8C).

DISCUSSION

The dissection of the region situated between –58 kb and –48 kb upstream of *Myf5* demonstrates that distinct regulatory elements direct transcription of this myogenic determination gene to the limb buds and hypoglossal cord, to sites of myogenesis in the somite and to specific regions of the central nervous system, where the endogenous gene is transcribed, in the mouse embryo (Table 2). The complete region would appear to be a composite of elements which function out of the normal context and with a heterologous *tk* promoter (Fig. 1B and results not shown). Deletions of this region, in the context of a large BAC, which recapitulates the endogenous expression of *Myf5*, show that the regulatory elements are necessary for these important aspects of the spatiotemporal regulation of *Myf5* transcription. Furthermore, this analysis shows that more than one element targets the same site, revealing multiple phases of myogenesis in the different domains where *Myf5* regulatory elements orchestrate the initiation of skeletal muscle formation.

Myf5 transcription in the central nervous system

The region between –56.6 kb and –53.7 kb specifically targets prosomeres p1 and p4 in the brain where endogenous *Myf5* transcripts are detected and deletion of this region shows that it is required for transcription of the gene at these sites. It also directs expression to the ventral neural tube. The fact that this site of transcription is not entirely coincident with that of the

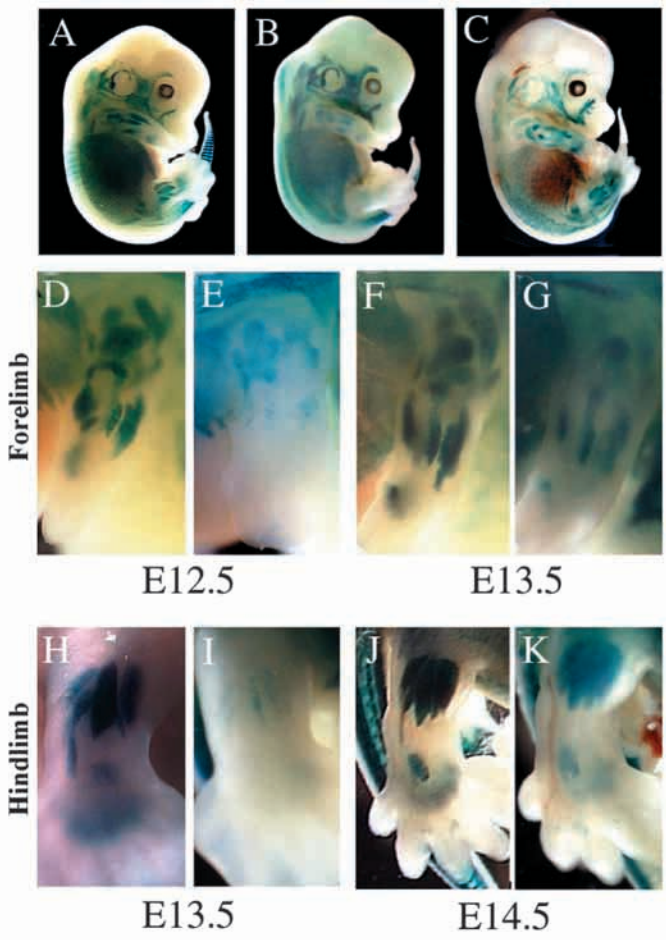

Fig. 8. Deletion of the –63 kb to –48 kb region reveals an additional element(s) directing a second phase of limb expression. Transgenic embryos at E13.5 where *nlacZ* expression is directed by BAC195*APZ* (A) or by BAC195*APZ*Δ63–48 (B). In C, an embryo with the –58/–48 *Myf5–nlacZ* transgene is shown. Comparison of the expression patterns of the limbs between embryos of BAC195*APZ* (D,F,H,J) and BAC195*APZ*Δ63-48 (E,G,I,K) transgenic lines. (E) At E12.5, all pre-muscle masses show weak X-gal staining, not seen in the hindlimb at this stage (not shown). (F,G) By E13.5, the forelimb expression pattern driven by the two constructs is similar. (I) Very weak expression is detected in hindlimb pre-muscle masses. (J,K) By E14.5, the hindlimb expression pattern driven by the two BAC constructs is indistinguishable.

endogenous gene (Tajbakhsh and Buckingham, 1995; Tajbakhsh et al., 1994) probably reflects the action of other sequences in the locus in restricting and refining the expression pattern. Indeed, the isolated fragment also tends to be transcribed more widely in prosomere p1. This effect is clearly distinguishable from ectopic expression in head mesenchyme, which is seen sporadically with many *Myf5* transgenes. In the case of the neural tube, another element located within the *Myf5* proximal region (Summerbell et al., 2000) also directs expression to this site. The upstream sequence, like this proximal sequence, can act independently, as seen in –58/–48 *tk-nlacZ* transgenic embryos. There is no apparent function associated with the activation of *Myf5* in specific neurones of the central nervous system, and based on observations that canonical Wnt signalling is implicated in *Myf5* transcription in both somites (Tajbakhsh et al., 1998) and the central nervous system (Daubas et al., 2000), we had speculated that the latter is caused by inappropriate triggering of a somite regulatory element. This is clearly not the case because the –56.6/–53.7 sequence only directs *Myf5* transcription to the specific sites in the central nervous system. It is probable that the region between –58 kb and –48 kb was put together by evolutionary tinkering (‘bricolage moléculaire’) (Jacob, 1977), which may have led to the inclusion of an unwanted regulatory element. However, it is surprising that it should have conserved its function in the absence of apparent selective pressure. It is also surprising that a second neural tube element is present in the 5′ proximal region of *Myf5*. Although the Myf5 protein is not detectable and does not appear to play a role in the embryonic central nervous system (Daubas et al., 2000), a possible role in the postnatal neural tube and brain was not examined because of the lethality of the *Myf5* mutation at birth.

Table 2. Regulatory elements within the –58/–48 kb region: a summary of transgenic results

	Sequence interval (kb)					
Sites of expression	–58/–54	–58/–57	–57.5/–57	–57.5/–56.6	–56.6/–53.7	–53.3/–48
Somites						
Early epaxial and later intercalated epaxial myotome	+	N.D.	–	N.D.	–	N.D.
Intercalated epaxial myotome	+	–	–	+	–	–
Subset of cells in central/hypaxial myotome (mainly anterior somites)	–	–	–	–	–	+
Later hypaxial somitic bud	+	+	+	(+)	–	–
Later epaxial lip	+	+	+	(+)	–	–
Limbs						
Forelimb	+	+	+	+	–	(+)
Hindlimb	+	+	+	+	–	+
Hypoglossal cord	+	+	+	+	–	–
Central nervous system						
Brain prosomeres p1, p4	+	–	–	–	+	–
Ventral neural tube	+	–	–	–	+	–

These results lead to the conclusion that, in somites, the –57/–56.6 kb fragment directs expression in the intercalated myotome and suggest that in younger embryos this element targets the early epaxial myotome. Other regulatory elements that direct expression to somites, notably to the early epaxial dermomyotome (Teboul et al., 2002) and hypaxial myotome (Summerbell et al., 2000), and to later limb muscles, are revealed by deletion of this region, as discussed in the text. N.D., not done; (), weak.

Myf5 expression in the limbs

Sequences within the –58/–48 kb region are essential for early expression of *Myf5* in limb buds. An element located between –57.5 kb and –57 kb directs robust transcription of the *nlacZ* reporter to the fore- and hindlimb buds, following the anteroposterior developmental gradient, in the same spatiotemporal pattern as the endogenous gene. This 500 bp sequence, which is conserved between human and mouse, is therefore a potential target of signalling pathways/transcription factors responsible for the activation of this myogenic determination gene once the muscle progenitor cells have migrated from the somite to the limb bud. The same sequence also directs activation of *Myf5* transcription in the hypoglossal cord. This is yet another example of multiple elements targeting the same site, because another element that also directs expression to the hypoglossal cord is present in the region between –81 kb and –63 kb (Carvajal et al., 2001b) and is necessary in the context of the locus. Cells of the hypoglossal cord also delaminate from the hypaxial dermomyotome of occipital somites, but move as a coherent mass (Noden, 1983), with activation of *Myf5* before they attain their final location around the larynx or in the tongue, where they contribute to the formation of skeletal muscle (Mackenzie et al., 1998). It is therefore surprising that this different mode of myogenesis is activated by the same element, and it remains to be seen if the same regulation is involved. It is also striking that this element directs reporter gene expression to the hypaxial somitic bud in the interlimb region, at a stage when this is a remnant of the epithelial dermomyotome, which still harbours myogenic progenitor cells that activate *Myf5*. This is in contrast to cells of the hypoglossal cord and limb buds, which express *Myf5* only when they have left the somite.

A second element in the region between –53.3 kb and –48 kb also contributes to *Myf5* expression in the limb buds. This element is more active in the hindlimb, where reporter gene expression is clearly observed in the developing muscle masses at a time when β-galactosidase activity is barely detectable in the forelimb. Deletion of the region between –59 kb and –54 kb, where the other limb element is located, confirmed this result. Forelimb expression appears to be delayed and generally weaker, but seems to extend to most of the muscle masses. This distinction between fore- and hindlimbs is particularly interesting in view of the unexpected results of mutations in the *Lbx1* and *Mox2* homeobox genes. Despite the fact that they are present in myogenic cells in fore- and hindlimbs, the absence of Lbx1 seriously compromises the formation of many forelimb muscles (Brohmann et al., 2000; Gross et al., 2000; Schafer and Braun, 1999), whereas mutations in *Mox2* mutants mainly affect hindlimb muscles (Mankoo et al., 1999). Recently, it has been shown that *Tbx5* and *Tbx4* genes are expressed in fore- and hindlimbs, respectively (Gibson-Brown et al., 1996) and required for their development (Rodriguez-Esteban et al., 1999). *Pitx1* (Logan and Tabin, 1999) is another example of a gene that plays a role in hind- but not forelimb development. These mutant phenotypes point to regulatory differences in the limb environment that the muscle progenitor cells enter, in addition to potential intrinsic differences in fore- and hindlimb progenitors themselves. Further analysis of the *Myf5* limb element at –53.3/–48 will provide more insight into the differences in the regulatory programme between fore- and hindlimbs.

Deletion of the –58/–48 region, although demonstrating that it is required for early expression of *Myf5* in limb buds, also shows that by E12.5 other regulatory sequences begin to participate in *Myf5* transcription in the muscle masses of the limbs. This is an autonomous function and therefore not simple maintenance, because it is initiated in the absence of the –63/–48 region. All muscles appear labelled eventually by the BAC transgene with this deletion, although initially this is more readily detectable in proximal muscles. It may be associated with the initiation of secondary myogenesis, which makes a major contribution to the growth of differentiated limb muscles from about E14 (Ontell and Kozeka, 1984).

Myf5 expression in the somite

Several sites of myogenesis in the somite are targeted by the region between –58 kb and –48 kb that, together with the previously identified early epaxial enhancer (Teboul et al., 2002), intragenic hypaxial enhancer (Summerbell et al., 2000), upstream hypaxial sequence (Carvajal et al., 2001a) and myotome sequence present in the –23 kb region (Hadchouel et al., 2000), constitute a set of regulatory modules that orchestrate *Myf5* transcription in the somitic cells that form the myotome. Initially, the early epaxial enhancer (Teboul et al., 2002) activates *Myf5* transcription in the epaxial dermomyotome, from which cells delaminate, and then, in the presence of Myf5, become correctly positioned to form the epaxial myotome (Tajbakhsh et al., 1996b). A second element contained in the –57.5/–56.6 region of the upstream enhancer then activates *Myf5* transcription in the epaxial myotome, which, as the dermomyotome continues to grow in an epaxial direction (Denetclaw and Ordahl, 2000; Spörle, 2001) and to produce myogenic precursors, becomes positioned more centrally, intercalated between epaxial-most and hypaxial components of the myotome. It is here that transcripts of *Mrf4* and those for differentiation markers, such as myosins (Lyons et al., 1990; Spörle, 2001), are first detected. Myf5 is required at this stage to activate *Mrf4*/myogenin transcription. The –57.5/–57 kb fragment does not direct epaxial myotome expression, indicating that the 400 bp at –57/–56.6 kb contains this transcriptional module. This region is conserved between human and mouse genomic DNA, in part because it also contains an exon of the *Ptprq* gene (Carvajal et al., 2001a). Another regulatory element contained within –23 kb of *Myf5* also targets a subdomain of the epaxial myotome (Hadchouel et al., 2000).

The region between –53.3 kb and –48 kb also displays some somitic activity in a subset of cells in a more hypaxial part of the myotome. Labelling is seen at E11.5 in the most anterior somites and later in some muscles anterior to the forelimb. At earlier stages of development, somitic labelling is only occasionally seen. More extensive transgene expression in somites on the anteroposterior axis is also sometimes seen, indicating that the –53.3/–48 region can potentially direct such transcription. In the deleted BACs, it is difficult to distinguish the contribution of this region from that of the intragenic hypaxial enhancer. The hypaxial enhancer, located within the *Myf5* gene (Summerbell et al., 2000) directs transcription in the early hypaxial myotome, and labels the caudal edge of the somite. This labelling, together with that due to the early epaxial enhancer, is evident when the –63/–48 region is deleted from BAC195*APZ*. A further hypaxial element is present in the 5′ region upstream of –88 kb (Carvajal et al., 2001a). As the somite matures, and the dermomyotome disintegrates, epithelial structures, which probably continue to be

sources of myogenic progenitor cells, are retained at the epaxial and hypaxial extremities of the somite. The 500 bp fragment at –57.5/–57 kb, which also directs *Myf5* transcription in the limb buds, targets these sites. The fact that epaxial and hypaxial extremities of the dermomyotome are targeted earlier by other regulatory sequences, illustrates the combinatorial action of elements directing the complete spatiotemporal expression of *Myf5* in the somite. At earlier stages of myogenesis, the somitic bud contributes to the hypaxial myotome and then to the formation of body wall and intercostal muscles. The behaviour and fate of cells in the later somitic bud can now be examined by means of these regulatory elements. The same is true for the later epaxial bud in relation to the contribution of the earlier epaxial dermomyotome, targeted by the early epaxial enhancer (Teboul et al., 2002).

Analysis of transgenes which direct *Myf5* transcription in the somite would suggest that at least six and possibly eight different regulatory modules are responsible for spatiotemporal aspects of this expression. Presumably, this reflects the complexity of the signals that modulate the construction of the myotome, from which all the different muscles of the trunk originate. It probably also reflects the way in which the mammalian myotome has evolved from the myotome of more primitive vertebrates, which prefigures simpler epaxial/hypaxial muscle derivatives, as in the body of fish for example. The analysis of *Myf5* regulation in the zebrafish embryo suggests that the proximal promoter region can direct somite expression (Chen et al., 2001). The –58/–48 enhancer region, together with the early epaxial and hypaxial enhancers play a major role in the expression of *Myf5* in the mouse somite, whereas the proximal promoter itself appears to have little or no regulatory capacity (Summerbell et al., 2000).

The way in which *Myf5* regulation may have evolved is intimately related to that of the *Mrf4* gene, which is linked to it in the same locus in all vertebrates examined (Braun et al., 1990; Miner and Wold, 1990). The expression profiles of these two myogenic factor genes are distinct. Although *Mrf4* appears to be regulated, in part, by sequences immediately 5′ to it (Patapoutian et al., 1993; Pin et al., 1997), it is also dependent on sequences further upstream (Carvajal et al., 2001a); it will be interesting to establish to what extent the –58/–48 kb enhancer region interacts with the *Mrf4* promoter, which lies 5′ to that of *Myf5*. Another myogenic factor gene, *MyoD*, which, like *Myf5*, plays a role in skeletal muscle determination (Rudnicki et al., 1993), has a distal enhancer that is located at –22 kb from the gene (Goldhamer et al., 1995). This directs early expression at sites of muscle formation in the embryo, whereas more proximal 5′ elements are necessary for transcription at later stages of skeletal muscle development. However, the *MyoD* enhancer, which is also active in limb buds, hypoglossal cord and myotome, can be reduced to a single 258 bp core element (Goldhamer et al., 1995; Kucharczuk et al., 1999), in contrast to *Myf5*, where the –58/–48 kb enhancer region breaks down into discrete elements that target different sites. Furthermore, as discussed here, crucial elements that direct *Myf5* expression in the embryo, as well as during later stages of development, are present elsewhere in the locus. These differences probably reflect the fact that *MyoD* lies genetically downstream of *Pax3* and *Myf5*, which govern the entry of cells into the myogenic programme. *Myf5* is therefore the target of the signalling pathways that specify myogenic cell fate and that differ between sites of muscle formation in the embryo, acting through different *Myf5* regulatory sequences. Although it was not surprising that limb versus dermomyotome/myotome expression of *Myf5* should be subject to different controls, the analysis presented here reveals the extraordinary complexity of myogenic patterning. At least three different regulatory circuits govern *Myf5* transcription in the limb, two in the hypoglossal cord, two in the epaxial and three in the hypaxial dermomyotome, and probably at least three more in the myotome. Other upstream regulatory genes, which have partially overlapping patterns of expression, may well display a similar complexity, such that small numbers of cells may read a unique code specifying muscle fate. Such codes, which would determine the myogenic body plan of a mouse, no doubt reflect the many ways in which further cell populations were co-opted into the muscle programme as vertebrates evolved and ever more sophisticated muscle functions were required. Identification of multiple *Myf5* regulatory elements now makes it possible to dissect muscle formation at any one site, both in terms of molecular regulation and of unique cellular contributions.

The authors thank Catherine Bodin for technical help. They are grateful to Ralf Spörle who contributed his insight into somite subdomains. They also thank H. Arnold and collaborators for sharing information prior to publication. This work in M.B.'s laboratory was supported by the Pasteur Institute and the CNRS and by grants from the ACI Integrative Biology Programme of the MJER, the AFM and the European Community (QLK3-CT-99/02). J.H. benefited from fellowships from ARC and the AFM, L.B. from funding from the MJER, and T.C. from fellowships from NIH and the AFM. The work in P.R.'s laboratory was supported by a grant from The Institute of Cancer Research.

REFERENCES

Bergstrom, D. A. and Tapscott, S. J. (2001). Molecular distinction between specification and differentiation in the myogenic basic helix-loop-helix transcription factor family. *Mol. Cell. Biol.* **21**, 2404-2412.

Borycki, A. G., Brunk, B., Tajbakhsh, S., Buckingham, M., Chiang, C. and Emerson, C. P., Jr (1999). Sonic hedgehog controls epaxial muscle determination through Myf5 activation. *Development* **126**, 4053-4063.

Braun, T., Bober, E., Winter, B., Rosenthal, N. and Arnold, H. H. (1990). *Myf-6*, a new member of the human gene family of myogenic determination factors: evidence for a gene cluster on chromosome 12. *EMBO J.* **9**, 821-831.

Braun, T., Rudnicki, M. A., Arnold, H. H. and Jaenisch, R. (1992). Targeted inactivation of the muscle regulatory gene Myf-5 results in abnormal rib development and perinatal death. *Cell* **71**, 369-382.

Brohmann, H., Jagla, K. and Birchmeier, C. (2000). The role of Lbx1 in migration of muscle precursor cells. *Development* **127**, 437-445.

Buckingham, M. (1994). Why myogenic factors make muscle? *Curr. Biol.* **4**, 61-63.

Carvajal, J. J., Cox, D., Summerbell, D. and Rigby, P. W. J. (2001a). A BAC transgenic analysis of the Mrf4/Myf5 locus reveals interdigitated elements that control activation and maintenance of gene expression during muscle development. *Development* **128**, 1857-1868.

Carvajal, J. J., Cox, D., Summerbell, D. and Rigby, P. W. J. (2001b). Control of the expression of Mrf5 and Myf5 genes: a BAC transgenic approach. *Int. J. Dev. Biol.* **45**, S139-S140.

Chen, Y. H., Lee, W. C., Liu, C. F. and Tsai, H. J. (2001). Molecular structure, dynamic expression, and promoter analysis of zebrafish (Danio rerio) myf-5 gene. *Genesis* **29**, 22-35.

Christ, B. and Ordahl, C. P. (1995). Early stages of chick somite development. *Anat. Embryol.* **191**, 381-396.

Christ, B., Jacob, M. and Jacob, H. J. (1983). On the origin and development of the ventrolateral abdominal muscles in the avian embryo. An experimental and ultrastructural study. *Anat. Embryol.* **166**, 87-101.

Cox, D. H., Carvajal, J. J. and Rigby, P. W. J. (2002). Enhanced efficiency of pSV1-RecA-based BAC recombineering. *Biotechniques* **33**, 1206-1208.

Daubas, P., Tajbakhsh, S., Hadchouel, J., Primig, M. and Buckingham, M. (2000). Myf5 is a novel early axonal marker in the mouse brain and is subjected to post-transcriptional regulation in neurons. *Development* **127**, 319-331.

Denetclaw, W. F. and Ordahl, C. P. (2000). The growth of the dermomyotome and formation of early myotome lineages in thoracolumbar somites of chicken embryos. *Development* **127**, 893-905.

Gerber, A. N., Klesert, T. R., Bergstrom, D. A. and Tapscott, S. J. (1997). Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. *Genes Dev.* **11**, 436-450.

Gibson-Brown, J. J., Agulnik, S. I., Chapman, D. L., Alexiou, M., Garvey, N., Silver, L. M. and Papaioannou, V. E. (1996). Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. *Mech. Dev.* **56**, 93-101.

Goldhamer, D. J., Brunk, B. P., Faerman, A., King, A., Shani, M. and Emerson, C. P., Jr (1995). Embryonic activation of the myoD gene is regulated by a highly conserved distal control element. *Development* **121**, 637-649.

Gross, M. K., Moran-Rivard, L., Velasquez, T., Nakatsu, M. N., Jagla, K. and Goulding, M. (2000). Lbx1 is required for muscle precursor migration along a lateral pathway into the limb. *Development* **127**, 413-424.

Gustafsson, M. K., Pan, H., Pinney, D. F., Liu, Y., Lewandowski, A., Epstein, D. J. and Emerson, C. P., Jr (2002). Myf5 is a direct target of long-range Shh signaling and Gli regulation for muscle specification. *Genes Dev.* **16**, 114-126.

Hadchouel, J., Tajbakhsh, S., Primig, M., Chang, T. H., Daubas, P., Rocancourt, D. and Buckingham, M. (2000). Modular long-range regulation of Myf5 reveals unexpected heterogeneity between skeletal muscles in the mouse embryo. *Development* **127**, 4455-4467.

Hogan, B., Costantini, F. and Lacy, E. (1994). *Manipulating the Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Jacob, F. (1977). Evolution and tinkering. *Science* **196**, 1161-1166.

Kelly, R., Alonso, S., Tajbakhsh, S., Cossu, G. and Buckingham, M. (1995). Myosin light chain 3F regulatory sequences confer regionalised cardiac and skeletal muscle reporter gene expression in transgenic mice. *J. Cell Biol.* **129**, 383-396.

Kucharczuk, K. L., Love, C. M., Dougherty, N. M. and Goldhamer, D. J. (1999). Fine-scale transgenic mapping of the MyoD core enhancer: MyoD is regulated by distinct but overlapping mechanisms in myotomal and non-myotomal muscle lineages. *Development* **126**, 1957-1965.

Lee, E. C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N. A. and Copeland, N. G. (2001). A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65.

Logan, M. and Tabin, C. J. (1999). Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. *Science* **283**, 1736-1739.

Luckow, B. and Schütz, G. (1987). CAT constructions with multiple unique restriction sites for the functional analysis of eukaryotic promoters and regulatory elements. *Nucleic Acids Res.* **15**, 5490.

Lyons, G. E., Ontell, M., Cox, R., Sassoon, D. and Buckingham, M. (1990). The expression of myosin genes in developing skeletal muscle in the mouse embryo. *J. Cell Biol.* **111**, 1465-1476.

Mackenzie, S., Walsh, F. S. and Graham, A. (1998). Migration of hypoglossal myoblast precursors. *Dev. Dyn.* **213**, 349-358.

Mankoo, B. S., Collins, N. S., Ashby, P., Grigorieva, E., Pevny, L. H., Candia, A., Wright, C. V., Rigby, P. W. J. and Pachnis, V. (1999). Mox2 is a component of the genetic hierarchy controlling limb muscle development. *Nature* **400**, 69-73.

Miner, J. H. and Wold, B. J. (1990). Herculin a fourth member of the MyoD family of myogenic regulatory genes. *Proc. Natl. Acad. Sci. USA* **87**, 1089-1093.

Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S. and Nonaka, I. (1993). Myogenin gene disruption results in perinatal lethality because of severe muscle defect. *Nature* **364**, 532-535.

Noden, D. M. (1983). The embryonic origins of avian cephalic and cervical muscles and associated connective tissues. *Am. J. Anat.* **168**, 257-276.

Ontell, M. and Kozeka, K. (1984). The organogenesis of murine striated muscle: a cytoarchitectural study. *Am. J. Anat.* **171**, 133-148.

Patapoutian, A., Miner, J. H., Lyons, G. E. and Wold, B. (1993). Isolated sequences from the linked Myf-5 and MRF4 genes drive distinct patterns of muscle-specific expression in transgenic mice. *Development* **118**, 61-69.

Pin, C. L., Ludolph, D. C., Cooper, S. T., Klocke, B. J., Merlie, J. P. and Konieczny, S. F. (1997). Distal regulatory elements control MRF4 gene expression in early and late myogenic cell populations. *Dev. Dyn.* **208**, 299-312.

Rawls, A., Valdez, M. R., Zhang, W., Richardson, J., Klein, W. H. and Olson, E. N. (1998). Overlapping functions of the myogenic bHLH genes MRF4 and MyoD revealed in double mutant mice. *Development* **125**, 2349-2358.

Rodriguez-Esteban, C., Tsukui, T., Yonei, S., Magallon, J., Tamura, K. and Izpisua Belmonte, J. C. (1999). The T-box genes Tbx4 and Tbx5 regulate limb outgrowth and identity. *Nature* **398**, 814-818.

Rudnicki, M. A., Schnegelsberg, P. N., Stead, R. H., Braun, T., Arnold, H. H. and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351-1359.

Schafer, K. and Braun, T. (1999). Early specification of limb muscle precursor cells by the homeobox gene Lbx1h. *Nat. Genet.* **23**, 213-216.

Spörle, R. (2001). Epaxial-adaxial-hypaxial regionalisation of the vertebrate somite: evidence for a somitic organiser and a mirror-image duplication. *Dev. Genes Evol.* **211**, 198-217.

Summerbell, D. and Rigby, P. W. J. (2000). Transcriptional regulation during somitogenesis. *Curr. Top. Dev. Biol.* **48**, 301-318.

Summerbell, D., Ashby, P. R., Coutelle, O., Cox, D., Yee, S. and Rigby, P. W. J. (2000). The expression of Myf5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors. *Development* **127**, 3745-3757.

Swaminathan, S., Ellis, H. M., Waters, L. S., Yu, D., Lee, E. C., Court, D. L. and Sharan, S. K. (2001). Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29**, 14-21.

Tajbakhsh, S. and Buckingham, M. E. (1994). Mouse limb muscle is determined in the absence of the earliest myogenic factor myf-5. *Proc. Natl. Acad. Sci. USA* **91**, 747-751.

Tajbakhsh, S. and Buckingham, M. E. (1995). Lineage restriction of the myogenic conversion factor myf-5 in the brain. *Development* **121**, 4077-4083.

Tajbakhsh, S. and Buckingham, M. (2000). The birth of muscle progenitor cells in the mouse: spatiotemporal considerations. *Curr. Top. Dev. Biol.* **48**, 225-268.

Tajbakhsh, S., Vivarelli, E., Cusella-De Angelis, G., Rocancourt, D., Buckingham, M. and Cossu, G. (1994). A population of myogenic cells derived from the mouse neural tube. *Neuron* **13**, 813-821.

Tajbakhsh, S., Bober, E., Babinet, C., Pournin, S., Arnold, H. and Buckingham, M. (1996a). Gene targeting the myf-5 locus with nlacZ reveals expression of this myogenic factor in mature skeletal muscle fibres as well as early embryonic muscle. *Dev. Dyn.* **206**, 291-300.

Tajbakhsh, S., Rocancourt, D. and Buckingham, M. (1996b). Muscle progenitor cells failing to respond to positional cues adopt non-myogenic fates in myf-5 null mice. *Nature* **384**, 266-270.

Tajbakhsh, S., Rocancourt, D., Cossu, G. and Buckingham, M. (1997). Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. *Cell* **89**, 127-138.

Tajbakhsh, S., Borello, U., Vivarelli, E., Kelly, R., Papkoff, J., Duprez, D., Buckingham, M. and Cossu, G. (1998). Differential activation of Myf5 and MyoD by different Wnts in explants of mouse paraxial mesoderm and the later activation of myogenesis in the absence of Myf5. *Development* **125**, 4155-4162.

Teboul, L., Hadchouel, J., Daubas, P., Summerbell, D., Buckingham, M. and Rigby, P. W. J. (2002). The early epaxial enhancer is essential for the initial expression of the skeletal muscle determination gene Myf5 but not for subsequent, multiple phases of somitic myogenesis. *Development* **129**, 4571-4580.

Tremblay, P., Dietrich, S., Mericskay, M., Schubert, F. R., Li, Z. and Paulin, D. (1998). A crucial role for Pax3 in the development of the hypaxial musculature and the long-range migration of muscle precursors. *Dev. Biol.* **203**, 49-61.

Venuti, J. M., Morris, J. H., Vivian, J. L., Olson, E. N. and Klein, W. H. (1995). Myogenin is required for late but not early aspects of myogenesis during mouse development. *J. Cell Biol.* **128**, 563-576.

Yang, X. W., Model, P. and Heintz, N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat. Biotechnol.* **15**, 859-865.

Zweigerdt, R., Braun, T. and Arnold, H.-H. (1997). Faithful expression of the Myf-5 gene during mouse myogenesis requires distant control regions: a transgene approach using yeast artificial chromosomes. *Dev. Biol.* **192**, 172-180.