

Modular long-range regulation of *Myf5* reveals unexpected heterogeneity between skeletal muscles in the mouse embryo

Juliette Hadchouel, Shahragim Tajbakhsh, Michael Primig[‡], Ted Hung-Tse Chang, Philippe Daubas, Didier Rocancourt and Margaret Buckingham*

Département de Biologie Moléculaire, CNRS URA 1947, Institut Pasteur, 25 Rue du Docteur Roux, 75724 Paris Cedex 15, France

[‡]Present address: CNRS UPR 1142, Institut de Génétique Humaine, 141 Rue de la Cardonille, 34396 Montpellier, France

*Author for correspondence (e-mail: margab@pasteur.fr)

Accepted 4 August; published on WWW 26 September 2000

SUMMARY

The myogenic factor *Myf5* plays a key role in muscle cell determination, in response to signalling cascades that lead to the specification of muscle progenitor cells. We have adopted a YAC transgenic approach to identify regulatory sequences that direct the complex spatiotemporal expression of this gene during myogenesis in the mouse embryo. Important regulatory regions with distinct properties are distributed over 96 kb upstream of the *Myf5* gene. The proximal 23 kb region directs early expression in the branchial arches, epaxial dermomyotome and in a central part of the myotome, the epaxial intercalated domain. Robust expression at most sites in the embryo where skeletal muscle forms depends on an enhancer-like sequence located between –58 and –48 kb from the *Myf5* gene. This element is active in the epaxial and hypaxial myotome, in limb muscles, in the hypoglossal chord and also at the sites of *Myf5* transcription in prosomeres p1 and

p4 of the brain. However later expression of *Myf5* depends on a more distal region between –96 and –63 kb, which does not behave as an enhancer. This element is necessary for expression in head muscles but strikingly only plays a role in a subset of trunk muscles, notably the hypaxially derived ventral body muscles and also those of the diaphragm and tongue. Transgene expression in limb muscle masses is not affected by removal of the –96/–63 region. Epaxially derived muscles and some hypaxial muscles, such as the intercostals and those of the limb girdles, are also unaffected. This region therefore reveals unexpected heterogeneity between muscle masses, which may be related to different facets of myogenesis at these sites. Such regulatory heterogeneity may underlie the observed restriction of myopathies to particular muscle subgroups.

Key words: *Myf5*, Transcriptional regulation, YACs, Transgenic mice

INTRODUCTION

The myogenic regulatory factors (MRFs), *Myf5*, *MyoD* (*Myod1*; Mouse Genome Informatics), myogenin and *Mrf4* (*Myf6*; Mouse Genome Informatics), are transcriptional activators of genes expressed in skeletal muscles. These four proteins, which form a subgroup of the basic helix-loop-helix class of transcription factors, can also induce myogenic conversion when expressed in a variety of cultured cells (for review, see Weintraub et al., 1991). Targeted mutation of the corresponding genes in the mouse demonstrates their essential role in the formation of skeletal muscle and also distinguishes their relative functions. *Myf5* and *MyoD* play a role prior to muscle cell differentiation. Although *Myf5* null (Braun et al., 1992) and *MyoD* null (Rudnicki et al., 1992) mice make muscle, *MyoD/Myf5* double mutants are devoid of myoblasts as well as muscle fibres (Rudnicki et al., 1993). Furthermore, it has been shown that cells in which these genes have been activated, and marked by β -galactosidase activity, can assume other cell fates in the absence of *Myf5* or *MyoD*, thus demonstrating their role as myogenic determination factors (Tajbakhsh et al., 1996b; Kablar et al., 1999).

The four MRFs have distinct patterns of expression, reflecting their different roles and the complex spatiotemporal organization of skeletal muscle formation, which has recently begun to be appreciated (see Tajbakhsh and Buckingham, 2000). In vertebrates, skeletal muscles of the body proper are derived from somites, which form as segmented blocks of paraxial mesoderm on either side of the neural tube and follow a rostrocaudal developmental gradient from about embryonic day (E) 8. Shortly after their formation, somites give rise to the mesenchymal sclerotome, which will contribute to the cartilage of the vertebrae and ribs, and an epithelial dermomyotome (for review, see Christ and Ordahl, 1995). Myotome formation is initiated as muscle progenitor cells (MPCs) from the epaxial (dorsal) dermomyotome lip, adjacent to the neural tube, become positioned underneath this epithelium from E8.5. From about E9.75, MPCs from the hypaxial (ventral) extremity of the dermomyotome in the interlimb region also begin to migrate underneath the epithelium to form the hypaxial component of the myotome; the hypaxial extremity of the dermomyotome and myotome defines the somitic bud. As the epaxial and hypaxial myotomes develop, they merge to form a continuous myotome layer and can no longer be distinguished

as separate domains with conventional markers. The epaxial myotome later contributes to the deep back muscles, whereas the hypaxial myotome including muscle precursors located in the somitic bud will contribute to hypaxial muscles, such as the intercostals and those of the ventral body wall (Ordahl and Le Douarin, 1992; Cinnamon et al., 1999; Denetclaw and Ordahl, 2000). Muscle progenitor cells will also migrate from the hypaxial dermomyotome to more distal sites to found muscle masses such as those of the limb, diaphragm and tongue. More recently, the notion of a third intercalated (dermo)myotome region, located between the epaxial-most and hypaxial (dermo)myotomes has emerged (Spörle and Schughart, 1998; see Tajbakhsh and Spörle, 1998). In the head, where the genetic hierarchy regulating myogenesis appears to be distinct from that in the trunk (Tajbakhsh et al., 1997), the majority of muscles arise from paraxial head mesoderm, located anterior to the most mature somite, and from prechordal mesoderm (for review, see Noden et al., 1999). The hypoglossal chord, which gives rise to tongue and pharyngeal muscles, arises from occipital/cervical somites (Noden, 1983). *Myf5* is the first myogenic regulatory factor to be expressed, initially in the epaxial dermomyotome and then in the epaxial myotome (Ott et al., 1991). It is also the first MRF to be activated in the hypaxial myotome (Tajbakhsh et al., 1997). Here, and in distal muscle masses, such as those of the limb, *MyoD* expression closely follows that of *Myf5* (Sassoon et al., 1989). This is also the case in head muscles. Unexpectedly *Myf5* is also transcribed in specific regions of the ventral neural tube (Tajbakhsh et al., 1994) and in prosomeres p1 and p4 of the mesencephalon and diencephalon (Tajbakhsh and Buckingham, 1995), although the *Myf5* protein is absent (Daubas et al., 2000).

Activation of the two myogenic determination genes depends on signals from tissues surrounding the somite, which in the mouse have been shown to differentially affect *Myf5* and *MyoD* (Cossu et al., 1996). *Wnt1*, present in the dorsal neural tube, will preferentially activate *Myf5* in explant experiments, whereas *Wnt7a*, present in surface ectoderm, preferentially activates *MyoD* (Tajbakhsh et al., 1998). Sonic hedgehog, produced by the notochord, is also required for myogenesis, but only in the epaxial domain: in its absence, *Myf5* is not expressed in the epaxial myotome, but hypaxial activation proceeds normally (Borycki et al., 1999). This illustrates the complexity of signal integration in the activation of an 'upstream' gene such as *Myf5*. Little is known in molecular terms as to how signals at different sites of myogenesis activate *Myf5* or *MyoD*, although the regulatory sequences of the MRF genes are beginning to be elucidated. In the case of *MyoD*, a 6 kb 5' flanking sequence is sufficient for muscle-specific expression in vitro but the transgene is expressed in differentiated muscles predominantly during foetal and postnatal stages in vivo (Asakura et al., 1995). A second regulatory sequence, situated at 23 kb 5' of the *MyoD* gene, is capable of reiterating the correct spatiotemporal pattern of *MyoD* expression in the mouse embryo (Faerman et al., 1995; Goldhamer et al., 1995), although this is not maintained at later stages (Asakura et al., 1995). Further dissection of the 258 bp core of this sequence, which has the properties of an enhancer, shows that it is regulated by multiple *cis*-acting elements, among which those responsible for *Myf5*-dependent activation of *MyoD* in the myotome can be distinguished (Kucharczuk et al., 1999).

Myf5, together with *Pax3*, acts upstream of *MyoD* in the genetic hierarchy that controls myogenesis (Tajbakhsh et al., 1997). Therefore, regulation of the *Myf5* gene, which plays a key role in integrating the signalling cascades leading to the acquisition of muscle cell identity, is likely to be complex. The locus is complicated by the presence of the *Mrf4* gene, situated upstream of *Myf5*, such that 8.8 kb separate their translational start codons in the mouse locus (Miner and Wold, 1990). Comparison of different *Mrf4* knockout alleles indicates that the regulatory elements of *Myf5* involved in early myotomal expression extend into the *Mrf4* gene (Olson et al., 1996; Yoon et al., 1997). Transgenic analysis also demonstrates that key regulatory elements lie outside a 5.5 kb intergenic region (Tajbakhsh et al., 1996a), which directs expression in the branchial arches and their derivatives, some head muscles, and later in some trunk muscles as well as muscles of the proximal forelimb (Patapoutian et al., 1993).

We chose a YAC transgenic approach to localise all possible *Myf5* regulatory elements required for the correct spatiotemporal expression of an *nlacZ* (where n=nuclear localisation signal) reporter targeted into the *Myf5* gene. Heterozygous embryos in which this reporter had been targeted into the endogenous *Myf5* gene (Tajbakhsh et al., 1996b) provided a control. After the initiation of this work, Zweigerdt et al. (1997) showed that an element lying between 95 kb and 45 kb upstream of the *Myf5* gene is necessary to direct expression in the limbs, whereas expression in other muscle domains is directed by sequences within -45 kb. In this paper, we report that the spatiotemporal expression of *Myf5* is regulated by multiple modules, which underlie muscle heterogeneity in the mouse embryo. An element located between -96 and -63 kb is required for later *Myf5* expression in head and many trunk muscles, but not those of the limbs. It reveals unexpected heterogeneity in hypaxial muscle derivatives. This element does not act like a classical enhancer in the embryo. A second enhancer type element located between -58 and -48 kb directs expression of the proximal *Myf5* promoter in the myotome, the limbs and the central nervous system (CNS). Lastly, within the 23 kpb region upstream of *Myf5* are sequences that direct expression in the epaxial dermomyotome, intercalated myotome and branchial arches.

MATERIALS AND METHODS

Library screening, yeast transformation and selection

The YAC library used was constructed from C57BL/6J mouse genomic DNA by Larin et al. (1991) in *Saccharomyces cerevisiae* strain AB1380 (*MATa ade2-1 can1-100 lys2-1 trp1 ura3 his5-u*) using pYAC4. This library was PCR-screened with two oligos, amplifying in exon 1 of the *Myf5* gene, leading to the isolation of Y13 and Y2.

Lithium acetate transformations of yeast cells were carried out, based on the protocol of Ito et al. (1983), using 10 µg of linearized plasmid DNA. Following transformation, YAC-containing strains were selected and maintained in synthetic dextrose (SD) medium lacking either uracil, lysine or histidine depending on the YAC. Selection for G418-resistant transformants was performed as described by Fairhead et al. (1996) Following integration of the URAbaster cassette (see below), *Ura*⁻ derivatives of *Ura*⁺ transformants were obtained by plating onto SD containing 1 g/l of 5-fluoro-orotic acid (5-FOA) and 50 mg/l of uracil (Boeke et al., 1984).

Analysis of yeast clones

Yeast clones were tested by colony PCR for the presence of *Myf5* (targeted or not), using the oligos m5KO1, m5KO2 and m5KO3, described by Tajbakhsh et al. (1997). Agarose plugs of high molecular mass DNA from positive clones were prepared as described by Huxley et al. (1991). Low molecular mass DNA was prepared as described by Philippsen et al. (1991). High and low molecular mass DNA were subjected to Southern blot analysis following enzymatic digestion and Pulse-Field Gel Electrophoresis (PFGE) or classical electrophoresis, respectively. DNA transfer and hybridisation were carried out as described by Kelly et al. (1995). Probes used to analyse the *lacZ* targeting in *Myf5* are described by Tajbakhsh et al. (1996b). Probes isolated either from genomic or YAC DNA are indicated in Fig. 1A.

YAC and vector construction

Deletion vectors

The centromeric deletion vectors pB1FCT-His and pB1RCT-His were constructed by replacing the Alu element in pCAT-H5 by the B1 element of pB1F and pB1R (McKee-Johnson and Reeves, 1996). pB1FCTUrab and pB1RCTUrab were created by replacing the *HIS5* gene with the URAbI cassette, a yeast selection marker that can be excised subsequently in the presence of 5-FOA (Alani et al., 1987). The 6 kb *PstI-SalI* fragment, subcloned from the YAC y200-Myf5-nlacZ and located upstream of the *SalI* site at -90 kb from *Myf5*, replaced the B1 element in this vector, creating pDel96-Ura. A fragment extending 1.1 kb downstream of the *Clal* site and located -63 kb from *Myf5*, subcloned from y200-Myf5-nlacZ, was inserted into pB1FCT-His replacing the B1 element to create pDel63-His. To delete sequences downstream of the *Myf5* gene in y200-Myf5-nlacZ, the B1 element in pB1F was replaced by a 5.3 kb *SmaI-BamHI* fragment containing exon 3 of *Myf5* and 3.8 kb downstream of its polyadenylation site (see Fig. 1A), to create pM5FL.

Targeting vectors

To construct y200-Myf5-nlacZ, an *nlacZ* reporter gene flanked by two *I-SceI* sites (ω; Choulika et al., 1994; Plessis et al., 1992) and a *neomycin* cassette (Tajbakhsh et al., 1996b) were introduced into a 13 kb *BamHI* fragment of the mouse *Myf5* gene at the ATG of *Myf5*, with 5.5 kb of 5' flanking sequence. The bHLH domain of *Myf5* was deleted in this construct (amino acids 14-122). The construct, for facility of cloning, differs slightly from that of Tajbakhsh et al. (1996b) where the *nlacZ* reporter was introduced 13 amino acids downstream of the *Myf5* ATG to create a fusion protein. The URAbI cassette was inserted between *nlacZ* and *neomycin*. To construct y240-Myf5-lacZ, the *nlacZ* reporter gene was replaced by *lacZ*, and *URA3* was inserted instead of the URAbI cassette. The *I-SceI* sites facilitated localisation of the *Mrf4-Myf5* locus on the YAC.

Split-marker vectors

The split-marker vectors pKA and pAN described by Fairhead et al. (1996) were used to delete from the *Clal* site located at -63 kb from the *Myf5* gene to various proximal regulatory elements. A 800 bp *Clal-EcoRI* fragment located downstream of the *Clal* site was inserted into pKA to create pKA-Δ63. A 3.4 kb *XbaI* fragment containing the *Myf5* minimal promoter and the *nlacZ* gene was inserted into pAN to obtain pAN-Myf5. These two vectors were used to create the YAC y96Δ63-Myf5-nlacZ. A 700 bp *XbaI-EcoRI* fragment located between -6.8 and -6.1 kb from *Myf5* was inserted into pAN to create pAN-Epax. Together with pKA-Δ63, this vector was used to generate the YAC y96Δ63-EpaxMyf5-nlacZ. pAN-23 was constructed by inserting a 1.4 kb *EcoRV* fragment located between -22.9 and -21.5 kb from *Myf5* into pAN. pKA-Δ63 and pAN-23 were used to generate the YAC y96Δ63-23Myf5-nlacZ.

p58/48baMyf5-nlacZ

The 5' genomic region and *nlacZ* of the knock-in construct (see

Tajbakhsh et al., 1996b) were recuperated and deleted to -2.6 kb to create pbaMyf5-nlacZ. This contains the branchial arch element located between -1706 bp (*NheI* site) and -592 bp (*BsaBI* site) (Summerbell et al., 2000) and the minimal promoter (-175 bp) from the Cap site, assumed to be at -116 bp from the *Myf5* ATG (E. Bober, personal communication). A 10 kb *XhoI-HindIII* fragment located between -58 and -48 kb of *Myf5* was inserted into the *XhoI* and *HindIII* sites of pbaMyf5-nlacZ to create p58/48baMyf5-nlacZ.

Generation of transgenic mice

YAC DNA purification was carried out as described by Manson et al. (1997). YAC integrity was verified after each step of the purification process by PFGE and the concentration of the YAC DNA was estimated on a mini-gel using as a standard a DNA mass ladder (Gibco Life Technologies). Transgenic mice were generated by microinjection of purified YAC DNA into fertilized (C57BL/6J × SJL) F₂ eggs at a concentration of approx. 1 ng/μl using standard techniques (Hogan et al., 1994). Injected eggs were reimplanted the day after the injection into pseudopregnant (C57BL/6J × CBA) F₁ foster mothers.

Identification of transgenic animals, determination of transgene copy number and analysis of YAC integrity in transgenic animals

DNA was prepared from mouse tails or, for transient transgenics, a portion of the embryo and analysed by Southern blot or PCR analysis (Tajbakhsh et al., 1997). For Southern blot analysis, 15 μg of DNA was digested with *BamHI* or *BglIII* and subjected to electrophoresis. Hybridisation probes were either from the *nlacZ* gene (1.1 kb *Clal-SacI* fragment) or, to determine copy number, *Myf5* gene fragments (either 1 kb *SmaI-ScaI* or 776 bp *KpnI-BglIII* fragments; see Fig. 1A). Transgene copy number for each transgenic line (see Table 1) was determined by analysis of hybridisation signals with a Phosphorimager (Molecular Dynamics).

YAC DNA integrity in transgenic animals was assessed by Southern blot analysis following PFGE. Agarose plugs containing intact chromosomal DNA from spleen cells were prepared as described by Manson et al. (1997). One plug (approx. 8 μg of DNA) was digested overnight with 20 Units of *Clal* or *XhoI* in a total reaction volume of 200 μl and submitted to PFGE, using 1% agarose gels in 0.5 × Tris-borate EDTA buffer at 14°C; running conditions were 6 V/cm, 3 seconds pulse interval for 22 hours. Hybridisation probes were either from *nlacZ* (ATG to *EcoRV*, 1.6 kb fragment) or from the *Mrf4-Myf5* intergenic region (Fig. 1A).

Analysis of transgene expression

Heterozygous and homozygous transgenic males were crossed with non-transgenic females ([C57BL/6J × SJL] F₁). Embryos were staged taking E0.5 as the day of the vaginal plug. Transient transgenic embryos were staged taking the day of reimplantation into the pseudopregnant foster mothers as E0.5. They were dissected in PBS, fixed in 4% paraformaldehyde (for 5-60 minutes depending on the age

Table 1. Transgene copy number for each transgenic line

Transgene	Copy number per line		
	a	b	c
y240-Myf5-lacZ	3	6	2*
y200-Myf5-nlacZ	10	6	6
y96-Myf5-nlacZ	10	3-20	
y63-Myf5-nlacZ	ND*	2	3
y23-Myf5-nlacZ	ND	8	ND

ND, not determined. For the 96-Myf5-nlacZ line, the copy number varies between 3-20 among the F₁ generation, indicative of multiple insertion sites and distinct segregation events. The asterisk indicates the lines with an incomplete expression pattern.

of the embryo), rinsed twice in PBS and stained in X-gal solution (Tajbakhsh et al., 1996a) at 37°C from 2 hours to overnight. Transgenic embryos were examined microscopically either as whole-mounts or after cryostat sectioning, as described by Kelly et al. (1995). In situ hybridisations on transgenic embryos were carried out as described by Tajbakhsh et al. (1997). Antisense probes used during this study were *nlacZ* (*Clal* to the polyadenylation site, 2.4 kb fragment from the 3' end) and *Myf5* (Ott et al., 1991).

Relative quantification of transcripts

Total RNA from E14.5 transgenic embryos was prepared as described by Daubas et al. (2000). 1 mg of total RNA was used to isolate poly(A)⁺ mRNA with Dynabeads mRNA DIRECT kit (Dyna) to eliminate contaminating genomic DNA. cDNA was synthesised as described by Daubas et al. (2000) and amplified using an ABI PRISM 7700 Sequence Detection System (PE Applied Biosystems). Sequences of specific primers for the detection of *Myf5* and *nlacZ* transcripts were as follows: *Myf5* identical (exon 1): 5'-CC-AGCCCCACCTCCAAC-3'; *Myf5* complementary (exons 2/3): 5'-CTTTTATCTGCAGCACATGCATTT-3'; *nlacZ* identical: 5'-GC-AGCCTGAATGGCGAAT-3'; *nlacZ* complementary: 5'-CGCATC-GTAACCGTGCATC-3'. TaqMan[®] probes, respectively for *Myf5* and *nlacZ*, were 5'^{Fam}-AGCCCTGTCTGGTCCCGAAAGAACA-3' (exon 2) and 5'^{Vic}-CCGATACTGTCGTCGTCCTCCCTCAAAC-3'. The PCR reactions were done in separate tubes for the amplification of *Myf5* and *nlacZ* transcripts, using the TaqMan[®] Universal PCR Master Mix (PE Biosystems), with 300 nM of each primer and 200 nM of TaqMan[®] probes. The cycling sequence consisted of an initial step at 50°C for 2 minutes and a denaturation for 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C. Contaminating genomic DNA was checked not to perturbate the detection of *nlacZ* transcripts. Relative quantitation with data from the ABI PRISM 7700 Sequence Detection System was performed using the comparative C_T method after verification that the efficiencies of target (*nlacZ*) and reference (*Myf5*) amplification were approximately equal. C_T is the threshold cycle, indicating the fractional cycle number at which the amount of amplified target reaches a fixed threshold level. The number N of molecules accumulated at C_T is calculated according to the following formula: $\log N = \log N_0 + C_T \cdot \log(1+E)$, where N₀ represents the number of molecules prior to amplification. For two different concentrations of *NMyf5* and *NnlacZ*, $\frac{NMyf5}{NnlacZ} = 2^{C_{TnlacZ} - C_{TMyf5}}$, assuming that the efficiency, E=1.

RESULTS

Construction of a series of deletions around the *Myf5* gene

In order to include potential regulatory elements located at a distance from the *Myf5* gene, a YAC library containing large fragments (100-1600 kb) of mouse (C57BL/6) genomic DNA (Larin et al., 1991) was screened by PCR with *Myf5*-specific oligonucleotides. Two YACs, Y13 (1000 kb) and Y2 (650 kb) were identified. Pulsed field gel electrophoresis (PGFE), followed by Southern analysis, showed that Y13 and Y2 contain about 450 kb and 550 kb upstream of the linked *Mrf4-Myf5* genes, respectively. All subsequent modifications were carried out on Y13.

A fragmentation vector, pM5FL, containing 3.8 kb of *Myf5* 3' flanking sequence was used for targeted deletion of Y13 beyond this downstream region. This was followed by random deletion of the upstream region using the B1 mouse repeat element as the homologous sequence (Heard et al., 1994), cloned in both orientations in the fragmentation vectors pB1F and pBIR (see Materials and Methods). The resulting YAC,

y200-*Myf5*, contains about 190 kb of upstream sequence. The *nlacZ* reporter gene was then targeted at the ATG of *Myf5* by homologous recombination to give y200-*Myf5-nlacZ* (Fig. 1B). This reporter gene, which had previously been targeted into the endogenous mouse *Myf5* gene (Tajbakhsh et al., 1996b), permits comparison between heterozygous *Myf5^{nlacZ/+}* embryos and transgenics expressing *Myf5* on YAC constructs. To make further 5' deletions, defined regions of y200-*Myf5-nlacZ*, located at 96 and 63 kb 5' of the *Myf5* Cap site, were subcloned into the fragmentation vector (see Materials and Methods). Chromosome fragmentation (Pavan et al., 1990) of y200-*Myf5-nlacZ* with these genomic sequences gave y96-*Myf5-nlacZ* and y63-*Myf5-nlacZ* (Fig. 1B).

An additional YAC, y240-*Myf5-lacZ*, was also obtained using B1 fragmentation vectors (see Materials and Methods). This contains 190 kb of sequence upstream and 50 kb of sequence downstream of the *Mrf4-Myf5* locus (Fig. 1B). In this case the *lacZ* reporter does not contain a nuclear localisation signal, making it possible to trace the axonal tracts of neurons expressing *Myf5* (Daubas et al., 2000).

All essential regulatory elements are located within 96 kb upstream of *Myf5*

Purified YAC DNA was injected into the pronucleus of fertilized mouse eggs and a number of independent transgenic lines were established for each construct. Embryos were examined from E8.5 until birth and compared with *Myf5^{nlacZ/+}* heterozygotes (Tajbakhsh et al., 1996b). Two out of three lines generated with y240-*Myf5-lacZ* gave the expected skeletal muscle expression pattern; the third line showed only very weak expression, possibly due to breakage of the YAC upon integration. All three lines generated with y200-*Myf5-nlacZ* are comparable to the *Myf5^{nlacZ/+}* heterozygotes. The first β-galactosidase positive (β-gal⁺) cells are detected in the epaxial lip of the dermomyotome at E8.5 in the recently formed somites of heterozygous knock-in and transgenic embryos (Fig. 2A,D,G). By E9.75, X-gal staining begins to be detectable in the hypaxial somitic bud (data not shown). In more mature somites, *lacZ* expression is observed in both dermomyotomes and myotomes, following a rostrocaudal gradient; by E11.5, β-gal⁺ cells extend from the epaxial (green arrowhead, Fig. 2B) to the hypaxial (red arrowhead, Fig. 2B) extremities of the myotomes. At E9, labelling is detected in the mandibular arch (results not shown). By E11.5, expression is seen in the masseter, the hyoid arch (black arrows, Fig. 2B) and faintly in the third branchial arch. At this stage, both fore and hind limb buds contain β-gal⁺ cells (Fig. 2B,E,H). Labelling is also detected in defined regions of the brain (cns, Fig. 2B; Tajbakhsh and Buckingham, 1995), where the neuronal tracts were identified as a result of cytoplasmic X-gal labelling in the 240-*Myf5-lacZ* lines (Daubas et al., 2000). All skeletal muscles in the body proper, head and limbs, are labelled at E14.5 (Fig. 2C,F,I) and this is maintained during subsequent foetal development. We therefore conclude that all essential regulatory elements required for correct *Myf5* expression are located upstream of the locus within a 190 kb region.

The y96-*Myf5-nlacZ* construct was then analysed for expression. In the two lines obtained, X-gal labelling is again comparable to that of *Myf5^{nlacZ/+}* heterozygous embryos (Fig. 2J-L), indicating that the essential regulatory sequences are

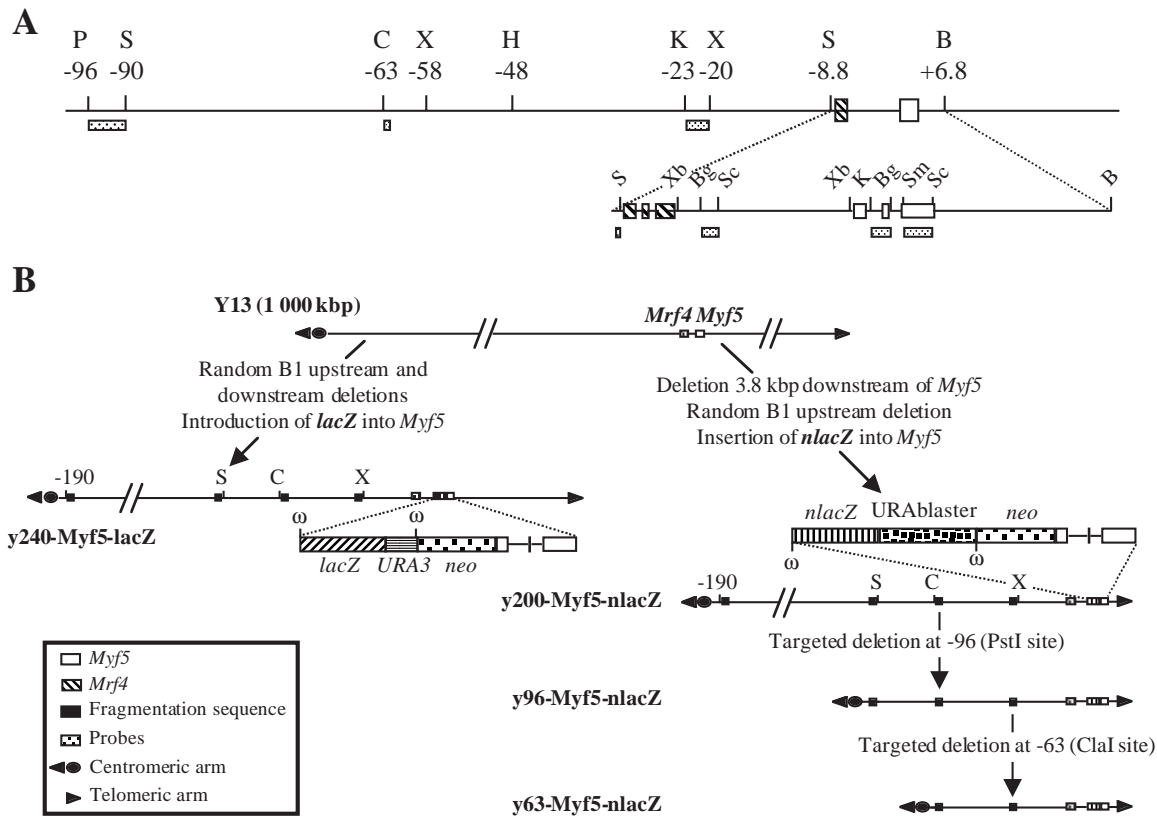


Fig. 1. Creation of a series of deletions around the *Mrf4-Myf5* locus. (A) Schematic long range map of the locus. Dotted boxes represent the different probes sub-cloned from either genomic or YAC DNA. (B) Schematic representation of the modifications carried out on Y13 (1000 kb). In two independent sets of experiments, chromosome fragmentation using either the repeated B1 element or defined sequences (black boxes) led to the generation of 4 different YACs carrying various amounts of sequences upstream and downstream of the locus. The reporter gene *lacZ*, with or without a nuclear localisation signal (n), was inserted into *Myf5* by homologous recombination in yeast. B, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sal*I; Sc, *Scal*I; Sm, *Sma*I; X, *Xho*I; Xb, *Xba*I; ω , *I-Sce*I.

within the 96 kb upstream region. It is notable, however, that a weaker expression is observed in the muscles of the most anterior portion of the face, compared to other facial muscles at E14.5 (Fig. 2L). Additional ectopic expression is seen in one of the 96-Myf5-nlacZ transgenic lines in the distal limb mesenchyme (Fig. 2K,L). It is interesting to note that most of the transgenic lines show some degree of reporter gene expression in cells located dorsally to the somites, most notably in the interlimb region (e.g. Fig. 2H,K). In *Myf5^{nlacZ/+}* embryos, only a few cells express *nlacZ* transiently in this region, along the anterior-posterior axis (Tajbakhsh et al., 1996b). The nature of these cells is unclear.

Regulatory elements directing later *nMyf5* expression in head and ventral trunk muscles are located between -96 and -63 kb

Deletion to 63 kb upstream of *Myf5* resulted in an incomplete expression pattern. Three mouse lines were established with y63-Myf5-nlacZ, one of which exhibited weak and partial X-gal labelling, again suggesting transgene breakage upon integration. In the two other lines, *nlacZ* expression initiates correctly in somites at E8.5 (data not shown). However, at E11.5 (Fig. 3A) the expression pattern begins to differ from that of the *Myf5^{nlacZ/+}* heterozygotes (Fig. 2B) or embryos expressing the y96-Myf5-nlacZ transgene (Fig. 3B),

particularly in the branchial arches and their derivatives. By E12.5, a decrease in X-gal staining is observed in the rostral somites as well as in other differentiating muscle masses of the head and trunk (data not shown). By E14.5 (Fig. 3C), in contrast with limb muscles, the transgene is not expressed in the head and ventral trunk muscles, such as the musculus (m)-rectus abdominis and the diaphragm. The m-latissimus dorsi is stained only at the attachment site behind the shoulder. In contrast, muscles of the girdle, such as the m-pectoralis (Fig. 3G), and intercostal muscles (Fig. 3C,E,G) continue to express the transgene, as do the deep back muscles (Fig. 3E). The complete expression pattern seen with the y96-Myf5-nlacZ transgene is shown for comparison (Fig. 3D,F,H).

The integrity of y63-Myf5-nlacZ, which had been checked prior to injection, was verified by PFGE and Southern blotting once integrated into the mouse genome. The two lines with the complete pattern of early muscle expression show no indication of breakage or rearrangement of this YAC transgene (data not shown). To verify that the differential staining observed among distinct muscle masses is not due to a sensitivity problem in the detection of β -galactosidase activity, 63-Myf5-nlacZ and 96-Myf5-nlacZ embryos were stained for different times (overnight and 6 hours, respectively) to give similar staining intensities in the deep back and limb muscles. This result, shown in Fig. 3C-H, clearly demonstrates that, at

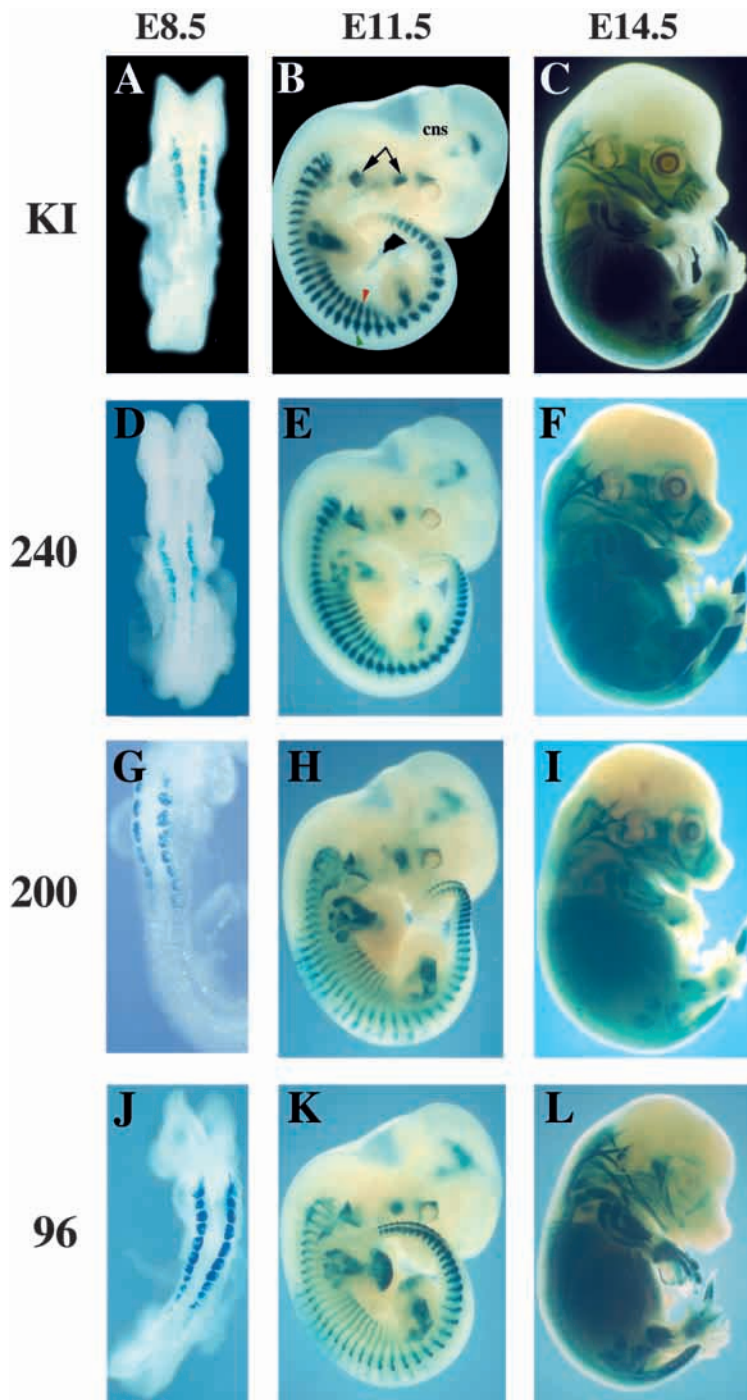


Fig. 2. Essential regulatory elements for *Myf5* expression are located within 96 kb upstream of the gene. The expression pattern of y240-*Myf5*-lacZ (D-F), y200-*Myf5*-lacZ (G-I) and y96-*Myf5*-lacZ (J-L) was revealed by X-gal staining in toto and compared with that of *Myf5*^{nlacZ+} heterozygous embryos (A-C). At about E8.5 (A,D,G,J), expression is observed in the epaxial dermomyotomal lip of the somites, adjacent to the neural tube. At about E11.5 (B, E, H and K), β -gal⁺ cells extend from the epaxial (green arrowhead in B) to the hypaxial (red arrowhead in B) limits of the myotome and are present in the branchial arches and derivatives (black arrows in B), in the limb buds and defined regions of the brain (cns in B). At E14.5, extensive staining is observed in all skeletal muscles (C,F,I,L). Ectopic expression in the distal limb mesenchyme is seen in K and L. Ectopic labelling is also seen between the epaxial dermomyotome and the neural tube in H and K.

E14.5, many muscles in the trunk and head are negative for y63-*Myf5*-nlacZ expression, whereas such a difference is not noted for limb muscles. This pattern is maintained until birth (data not shown). It is notable that, apart from this differential staining, the β -galactosidase activity appears weaker in 63-*Myf5*-nlacZ embryos throughout development, possibly due to the lower transgene copy number. Relative quantification of transcripts (Fig. 4) shows that *nlacZ* transcripts are indeed less abundant in 63-*Myf5*-nlacZ than in 96-*Myf5*-nlacZ embryos, when normalized to transgene copy number. A two-fold decrease in the level of *nlacZ* transcripts, normalized to those of the endogenous *Myf5* gene, is observed in the trunk versus the limbs of the 63-*Myf5*-nlacZ embryos, compared to 96-*Myf5*-nlacZ embryos. Given that the whole trunk was used to prepare mRNA and that some trunk muscles continue to express y63-*Myf5*-nlacZ (Fig. 3), this difference is an underestimate.

The -96/-63 region does not interact with identified proximal *Myf5* regulatory elements

To test the properties of the -96/-63 region, we first placed it directly in front of a minimal *Myf5* promoter, including the presumptive TATA box. Using the split-marker vectors (Fairhead et al., 1996; see Materials and Methods), we created a deletion in y96-*Myf5*-nlacZ between the *Cla*I site at -63 kb and the *Xba*I site at -175 bp from the Cap site of *Myf5*, to yield y96 Δ 63-*Myf5*-nlacZ (Fig. 5A). The pattern of expression of this YAC was examined at E14.5 in transient transgenic embryos (Fig. 5B). The absence of any muscle expression suggests that the -96/-63 element does not act as an enhancer on the minimal *Myf5* promoter to direct transcription in those head and trunk muscles where the endogenous gene is expressed. We therefore investigated whether it requires more extensive proximal regulatory sequences. A proximal region that drives *Myf5* expression in the branchial arches has previously been identified (Patapoutian et al., 1993) and its importance confirmed more recently (Summerbell et al., 2000). This 1168 bp region located between -1760 and -592 bp from the *Myf5* Cap site directs early expression in cells that will contribute to head musculature. A second 600 bp region, located between -6.1 and -5.5 kb upstream of *Myf5*, has been found to direct early *Myf5* expression in the epaxial somite (Summerbell et al., 2000). We therefore created a second deletion in y96-*Myf5*-nlacZ extending from the *Cla*I site at -63 kb to the *Xba*I site 5' of the epaxial element (Fig. 5A). The expression pattern of the resulting YAC, y96 Δ 63-Epax*Myf5*-nlacZ, was again analysed in transient transgenic embryos at E14.5. Only ectopic expression was observed. At E10.75, transgenic embryos showed expression in the branchial arches and in a small number of myotomal cells (Fig. 5B). These experiments confirm the early regulation of *Myf5* by the epaxial and branchial arch elements and indicate that the -96/-63 region has no *cis*-acting effect at E14.5 in the context of these proximal sequences.

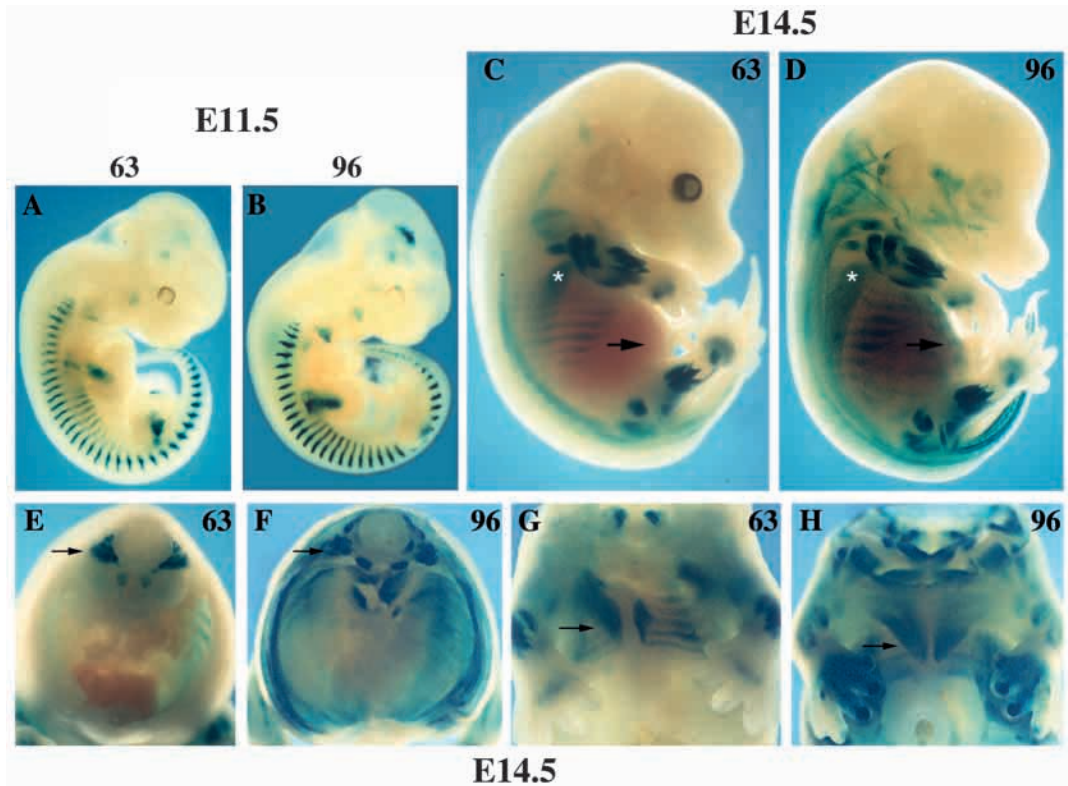


Fig. 3. Elements required for later *Myf5* expression in the head and many trunk muscles are located between -96 and -63 kb. Expression of *y63-Myf5-nlacZ* (A) and *y96-Myf5-nlacZ* (B) at about E11.5. X-gal staining of E14.5 *63-Myf5-nlacZ* (C, E, G) and *96-Myf5-nlacZ* (D, F, H) embryos. Note no staining in the m-lattissimus dorsi, beyond the shoulder level (asterisk in C and D) and the m-rectus abdominis (arrow in C and D) in the *63-Myf5-nlacZ* embryos. Staining of the deep back muscles (arrow in E and F) and m-pectoralis (arrow in G and H) is similar in both lines. In G, the m-pectoralis on one side has been removed to expose the intercostals. The embryo in D was stained for 6 hours (0.4 mg/ml of X-gal), whereas those shown in C, E and G were stained overnight in 0.6 mg/ml of X-gal. The *96-Myf5-nlacZ* embryos are from a different line from those presented in Fig. 2.

We have identified (T. C. and M. B., unpublished) an additional myogenic regulatory sequence located at -17 kb from the *Myf5* gene (A17, Fig. 5A). To test the potential interaction of this element with the remote $-96/-63$ region, we constructed a third YAC, *y96 Δ 63-23Myf5-nlacZ*, by deleting the region from the *ClaI* site at -63 kb to the *KpnI* site at -23 kb from *Myf5* (Figs 1A, 5A). Of 19 embryos dissected at E14.5, 1/19 showed only ectopic β -galactosidase activity, while 1/19 had some β -gal⁺ myotomal cells posterior to the hind limb, and in the caudal deep back muscles (Fig. 5C, D) in addition to ectopic expression. An additional 3/19 embryos were positive by PCR, but showed no X-gal staining.

We conclude from these experiments that the $-96/-63$ region does not act as a classical enhancer-type element since it shows no activity with the branchial arch element, or the early epaxial element, or other potential elements within 23 kb of the *Myf5* gene. It appears therefore that the $-96/-63$ region, if it can act out of context, may require other regulatory sequences, potentially located between -63 and -23 kb from *Myf5*.

The region between $-58/-48$ kb contains important regulatory elements for *Myf5* expression

To examine earlier expression directed by the 23 kb sequence upstream of the *Myf5* gene, we derived transgenic lines with the

YAC *y23-Myf5-nlacZ*. This YAC was created by chromosome fragmentation of *y63-Myf5-nlacZ* with a genomic sequence located between -23 and -20 kb (Fig. 6A). Three independent lines were generated, showing variable intensities of X-gal staining and different ectopic sites of expression. Nevertheless, in all three lines, the transgene exhibits a reproducible muscle expression pattern. From E8.5, β -gal⁺ cells are present in the epaxial dermomyotome of the somite (Fig. 6B) and later in the epaxial myotome (Fig. 6C). *nlacZ* expression decreases from E10.75 following a rostrocaudal gradient, such that only very weak staining is observed in deep back muscles by E14.5 (Fig. 6D, see also Fig. 5C). Also, as expected, early branchial arch expression is seen with this YAC; later expression in head muscles is not detected. Furthermore, no expression is seen in the limbs or in the hypaxial region of the myotome and its derivatives at all stages examined, as opposed to the *63-Myf5-nlacZ* embryos. Some X-gal staining is detectable in a subset of cells of the rostral intercostal muscles in *23-Myf5-nlacZ* transgenics between E11.5 and E12.5 (Fig. 6C), whereas intercostal muscles are fully labelled at E14.5 in the *63-Myf5-nlacZ* embryos (Fig. 3). No expression is detected in the central nervous system with *y23-Myf5-nlacZ*. Head expression seen in Fig. 6C corresponds to ectopic labelling of head mesenchyme. These results suggest that the $-63/-23$ kb region contains regulatory sequences responsible for *Myf5* expression

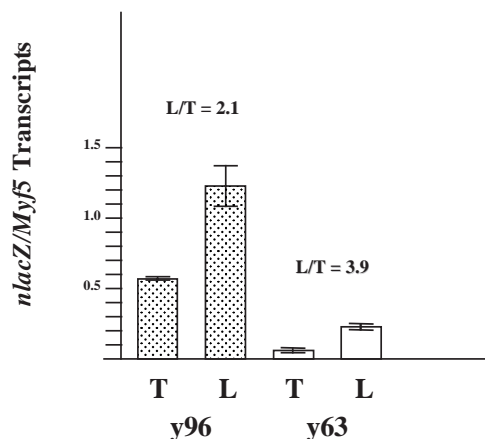


Fig. 4. Relative quantification of *nlacZ* transcripts by RT-PCR. Poly(A)⁺ mRNA was prepared from trunk and limbs of 96-Myf5-*nlacZ* and 63-Myf5-*nlacZ* E14.5 embryos. The level of *nlacZ* transcripts was determined relative to that of the endogenous *Myf5* gene and normalized to transgene copy number (e.g. 10 for 96-Myf5-*nlacZ* and 3 for 63-Myf5-*nlacZ*). The L/T value represents the ratio between the relative level of *nlacZ* to *Myf5* transcripts in the limbs (L) and trunk (T). RT-PCR was performed in duplicate with four different cDNA concentrations. Values are means \pm s.e.m.

in the limbs, the hypaxial myotome and derivatives, and the brain. From previous studies (Zweigerdt et al., 1997) and those

described here, sequences driving limb expression are probably located just 3' to -63 kb.

In fact we found that, in p58/48baMyf5-*nlacZ* transgenic embryos, the -58/-48 kb region, linked to a fragment containing the *Myf5* minimal promoter with the branchial arch region as a positive control (see Materials and Methods), is active in the branchial arches, the limb buds (Fig. 7D), the brain (Fig. 7C), the hypoglossal chord (Fig. 7B) and the myotome from its epaxial to hypaxial extent (Fig. 7E). *Myf5* expression is absent from the epaxial dermomyotome lip (Fig. 7B,E). This is particularly evident in the caudal somites (c/f Fig. 2). We therefore conclude that key regulatory elements for *Myf5* expression at different sites in the embryo, including limb muscles, are localised in this distal 10 kb fragment.

The distal 10 kb fragment and the -23 kb region direct distinct aspects of *Myf5* expression in the somite

It is notable that the epaxial myotome is positive for β -galactosidase activity with both y23-Myf5-*nlacZ* and p58/48baMyf5-*nlacZ* transgenes. However, a decrease in the epaxial staining in more rostral somites is observed only in the 23-Myf5-*nlacZ* embryos (Fig. 6C, compared with Fig. 7B). Given the stability of the β -galactosidase protein, it is possible that the X-gal staining in the epaxial myotome of the 23-Myf5-*nlacZ* embryos is due to a persistence of β -galactosidase activity rather than a continuous expression of the transgene. In situ hybridisation with an *nlacZ* probe performed on

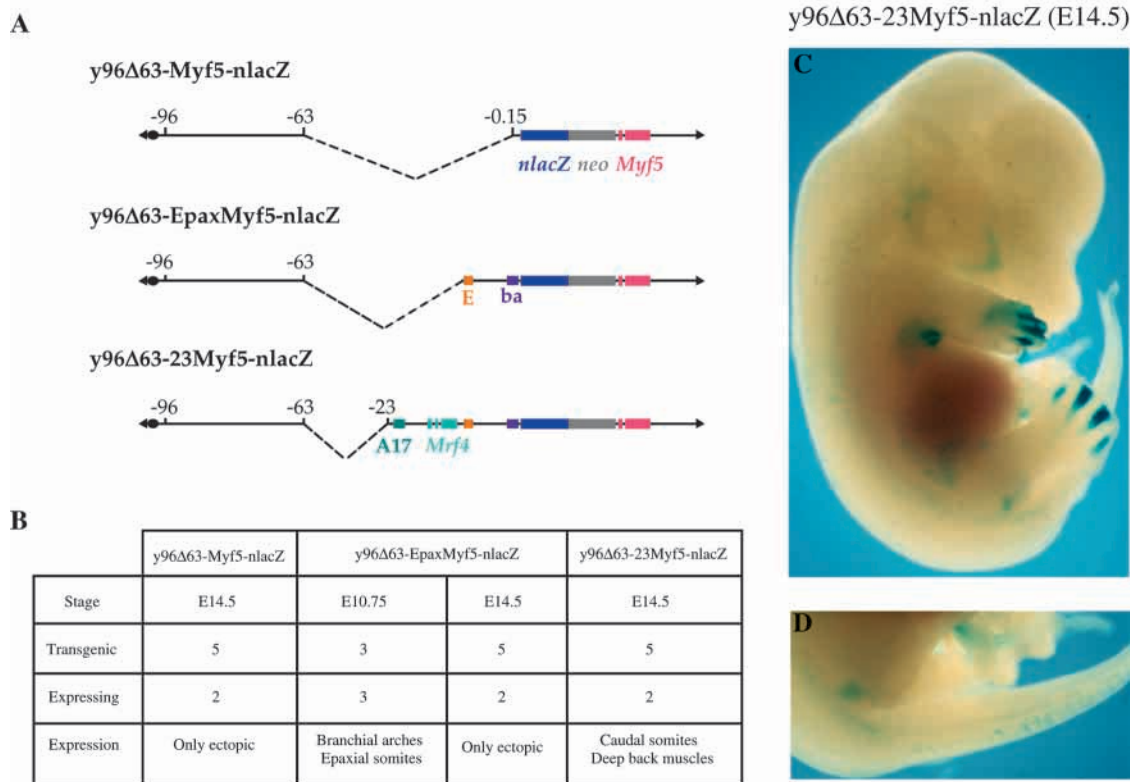


Fig. 5. The -96/-63 region does not interact with putative early *Myf5* regulatory elements. (A) Schematic representation of different deletions (see text). ba, branchial arch element; E, epaxial element. (B) Results using these constructs in transient transgenic experiments are summarized in the table. (C) 96Δ63-23Myf5-*nlacZ* transgenic embryo at E14.5 showing faint expression in caudal myotomes and back musculature. (D) Magnification of caudal region in C. The hind limb has been removed.

transgenic embryos shows that this is the case. At E9.75 (Fig. 6H), *nlacZ* transcripts are restricted to the dorsal lip of the epaxial dermomyotome in 23-Myf5-*nlacZ* transgenic embryos; the epaxial dermomyotome is also labelled in caudal somites

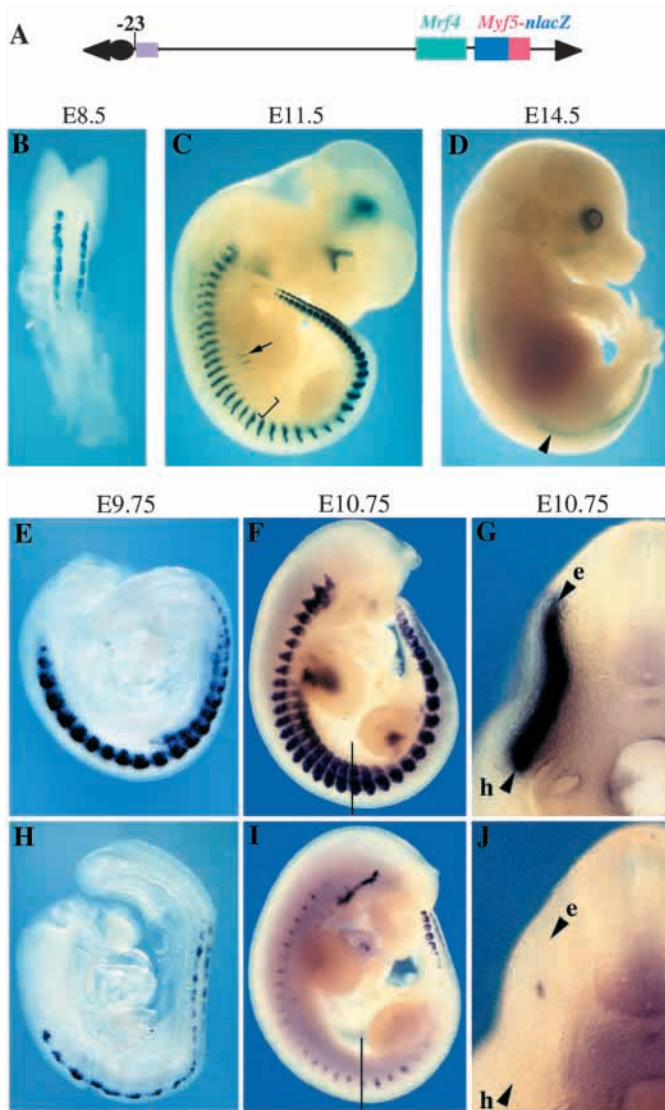


Fig. 6. The 23 kb region upstream of *Myf5* directs expression in the epaxial dermomyotome and the intercalated myotome. (A) Schematic representation of y23-Myf5-*nlacZ*, which carries 23 kb upstream of *Myf5*. (B-D) Expression of y23-Myf5-*nlacZ* is observed at E8.5 in the epaxial dermomyotome of the somites (B). At E11.5 (C), expression is observed clearly in the epaxial dermomyotome of the caudal somites and in the epaxial myotome along the anteroposterior axis, but not in the hypaxial domain (bracket). Low expression is detected in some of the intercostal muscles (black arrow). At E14.5 (D), a weak staining is observed in the caudal deep back muscles (black arrowhead) and in the epaxial domain of mature caudal somites. (E-J) Whole-mount in situ hybridisations of E9.75 (E,H) and E10.75 (F,G,I,J) transgenic embryos with an *Myf5* (E-G) or a *nlacZ* (H-J) probe. Interlimb sections (see line in F and I) presented in G and J show that *nlacZ* transcripts are observed in the epaxial intercalated domain of the myotome, but not in the epaxial-most (e in G and J) or hypaxial (h in G and J) myotome, delimited by arrowheads in G and J. Labelling in the anterior neck region in I is non-specific.

of older embryos (Fig. 6I). At E10.75, a second site of transcription is seen in the central region of the myotome (Fig. 6I), which corresponds to part of the intercalated epaxial domain (see Tajbakhsh and Spörle, 1998), as shown also on a cross section (Fig. 6J). Transcripts are not detected in the hypaxial myotome (c/f Fig. 6F,G). We therefore conclude that *Myf5* expression in the MPCs of the epaxial dermomyotome and in the differentiating epaxial-most myotome is regulated by two separable elements. Furthermore, this analysis identifies the intercalated domain of the epaxial myotome as a distinct site of *Myf5* myotomal transcription.

DISCUSSION

Myf5 expression is regulated by multiple modules

Using a YAC and plasmid transgenic approach, we show that correct spatiotemporal expression of *Myf5* in skeletal muscle

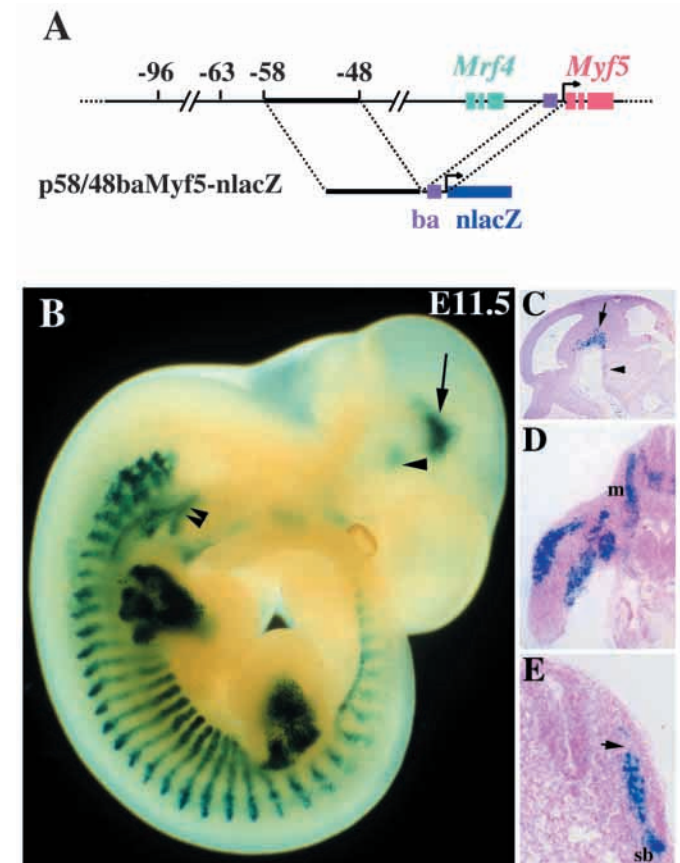


Fig. 7. The -58/-48 kb region contains regulatory elements driving *Myf5* expression in the myotome, the limb buds and the CNS. (A) Schematic representation of the transgene p58/48baMyf5-*nlacZ*. Whole-mount staining (B) and sections (C-E) of an E11.5 transgenic embryo. (C) Sagittal section through the head. (D-E) Cross sections at the level of the forelimb (D) and the tail (E). p58/48baMyf5*nlacZ* is expressed in prosomeres p1 (arrow in B and C) and p4 (arrowhead in B and C), in the hypoglossal chord (double arrowhead in B), the muscle masses of the limbs (D), the dorsoventral extent of the myotome (m in D), and the somitic bud (sb in E), but not in the epaxial dermomyotome (arrow in E). This is particularly evident in the caudal somites in B.

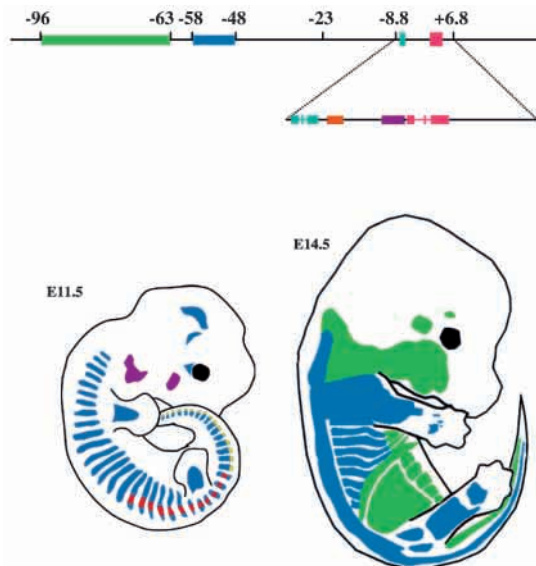


Fig. 8. *Myf5* expression is regulated by multiple modules dispersed between -96 and $+4$ kb. The region between -96 and -63 kb (green box) does not act like a classical enhancer, whereas the region located between -58 and -48 kb (blue box) functions with a proximal *Myf5* promoter. Location of the epaxial (Summerbell et al., 2000) and branchial arch elements (Patapoutian et al., 1993; Summerbell et al., 2000) are shown by orange and purple boxes respectively. Expression domains that depend on these regions are shown schematically at E11.5 and E14.5 (graphic of embryo modified from Tremblay et al., 1998). Epaxial dermomyotome (yellow) and intercalated myotome (red) expression, observed with the proximal 23 kb transgene (see Fig. 6) may be driven by the epaxial element (orange).

and the central nervous system depends on multiple proximal and far distal regulatory elements, dispersed over -96 kb from the gene (Fig. 8). In the proximal 23 kb sequence upstream of *Myf5*, lie elements directing early expression of *Myf5* in the epaxial dermomyotome, the epaxial intercalated myotome and the branchial arches. Previous studies showed that a region lying immediately upstream of the *Myf5* promoter (purple box in Fig. 8) directs robust early expression in the branchial arches (Patapoutian et al., 1993; Summerbell et al., 2000). A second sequence, located at -6.1 kb, directs expression in the epaxial dermomyotome (Summerbell et al., 2000; orange box in Fig. 8) and may also be implicated in the intercalated myotomal expression reported here. Studies carried out in other laboratories (Pin et al., 1997; Yoon et al., 1997) and our own (T. C. and M. B., unpublished) suggest that other *Myf5* and *Mrf4* elements are located within the 23 kb sequence. However, our analysis shows that these elements are not sufficient to confer overall robust expression with a *Myf5* minimal promoter. A region between -58 and -48 kb (blue box in Fig. 8) contains important regulatory elements that direct expression from a proximal *Myf5* promoter in the complete extent of the epaxial and hypaxial myotome, in the developing limb muscles and at other sites to which muscle precursors migrate from the somite, such as the hypoglossal chord. This 10 kb fragment also directs correct expression of *Myf5* in discrete regions of the brain where the endogenous gene is expressed (Tajbakhsh and Buckingham, 1995; Daubas et al.,

2000). A more distal region between -96 and -63 kb, which has an effect on the overall level of expression, is also required for late *Myf5* expression in a specific subset of skeletal muscles such as those of the head and ventral trunk muscles, but not the deep back, intercostal and limb muscles (Fig. 3). This region does not function as an enhancer. The uncoupling of early and late expression of *Myf5* is reminiscent of *MyoD* regulation, where proximal and distal elements confer late and early activity respectively. Both regions, however, function as enhancers in the context of a minimal promoter and the "late" elements are required in all foetal and postnatal muscles (Asakura et al., 1995). The core enhancer for early *MyoD* expression, situated at -23 kb, is 258 bp in size (Goldhamer et al., 1995). It remains to be seen whether the activity of the 10 kb fragment at $-58/-48$ kb from *Myf5* can be reduced to such an extent. Preliminary data suggest that its structure is more complex.

In a previous study of chimaeric embryos, using ES cells containing YAC transgenes, Zweigerdt et al. (1997) suggested that all muscle regulatory elements of *Myf5* are present in a region spanning from -95 kb to $+500$ kb of the gene, and that, whereas limb muscle regulatory elements are located between -45 and -95 , early muscle expression is seen with the -45 kb sequence. In contrast, we find that the $-58/-48$ fragment is essential for correct myotomal expression. This discrepancy may be due in part to the fact that Zweigerdt et al. (1997) analysed chimaeric embryos. Given the recent more detailed understanding of the complex spatiotemporal expression of *Myf5* (see Tajbakhsh and Buckingham, 2000), we would suggest that some domains of the (dermo)myotome do not express their YAC-45 construct (most evident in the epaxial somite of the cervical region, and hypaxial somite in the tail of E12.5 embryos; Fig. 2 in Zweigerdt et al., 1997) and that sequences located between -45 and -96 kb are indeed necessary for faithful expression of *Myf5* in trunk and tail muscles. Our analysis shows that this region can be subdivided into two domains with different characteristics (Fig. 8), where the most distal regulatory region is required for later expression of *Myf5* in the head and some trunk muscles. A previous report on an *nlacZ* transgene with 5.5 kb of *Myf5* 5' flanking sequence described expression in some facial muscles, in deep back muscles and in the proximal limb at E13.5 (Patapoutian et al., 1993). These observations probably reflect the presence of further regulatory elements within this 5.5 kb region, which direct low level expression, only detectable perhaps with high copy numbers, or different genetic backgrounds. Zweigerdt et al. (1997) also observed labelling in the head, but not in limbs, with a 45 kb 5' flanking sequence. Again without a quantitative comparison of transcript levels, it is difficult to assess whether this is lower than that seen with larger constructs. There are also negative regulatory elements for expression in head muscles lying 5' to the 5.5 kb region (Summerbell et al., 2000). With the 23 kb construct, we see relatively weak expression in the arches at E11.5 and head muscles derived from them are negative at E14.5.

Comparison of expression between the YAC transgenes and the endogenous *Myf5* gene

The sequences that lie within 58 kb upstream of the *Myf5* gene direct correct spatiotemporal transcription initially in the embryo, but not at later stages (Fig. 3). In addition to a specific

effect on certain muscle masses, the absence of more upstream sequences results in a general reduction in the overall level of expression, confirmed by quantification of *n lacZ/Myf5* transcripts by RT-PCR with the y63-Myf5-nlacZ transgene (Fig. 4). With an additional 30 kb of 5' upstream sequence, the level of expression per transgene copy compared to *Myf5* expression at E14.5 increases to a ratio that approaches 1. It is striking, however, that the level of *n lacZ* expression from YAC transgenes is lower than that of the *Myf5^{n lacZ/+}* allele. Even for multicopy transgenes, X-gal labelling is reduced and this phenomenon is confirmed by quantitative RT-PCR (data not shown). It is possible that this may be due to a minor difference in the construction (see Materials and Methods). Alternatively, since haploinsufficiency has been observed with one allele of *Myf5* (Rudnicki et al., 1993), the fact that less *Myf5* is being produced in the *Myf5^{n lacZ/+}* embryos may have repercussions on the regulation of the gene, although direct positive autoregulation of *Myf5* has been ruled out since *Myf5^{n lacZ/n lacZ}* embryos are more intensely stained than *Myf5^{n lacZ/+}* embryos (Tajbakhsh et al., 1996b).

Although the y96-Myf5-nlacZ transgene recapitulates almost all aspects of *Myf5* expression, oddly, X-gal staining in the region of the snout is lacking, suggesting that sequences required for this are present in the larger y200-Myf5-nlacZ transgene. Some ectopic expression is seen even with large flanking sequences (e.g. 96 kb; Fig. 2), suggesting that insulator sequences may be distant from the locus. One unexpected site of transgene expression, which is probably not ectopic, seen clearly in some transgenic lines at about E11.5, extends dorsally in a segmented pattern from the interlimb somites. Although much less prominent, a similar phenomenon is also detectable in *Myf5^{n lacZ/+}* heterozygous embryos, where occasional β -gal⁺ cells are seen dorsal to the cervical somites (Tajbakhsh et al., 1996b). These cells may be precursors of subcutaneous muscles. However, the transgene pattern resembles that seen in *Msx1^{n lacZ/n lacZ}* homozygous embryos (Houzelstein et al., 2000) and may therefore initially mark dermal precursors, derived from multipotent *Myf5* expressing cells in the dermomyotome.

The regulation of *Myf5* expression in the somite

It is becoming apparent from the analysis of different marker genes and mouse mutants that the (dermo)myotome is regionalised along the dorsoventral axis (see Tajbakhsh and Buckingham, 2000) suggesting that different regulatory elements may direct expression either in the epaxial-most, intercalated or hypaxial (dermo)myotome (Spörle and Schughart, 1998). Our observations are consistent with this notion. Analysis of the expression pattern of the y23-Myf5-nlacZ transgene by in situ hybridisation clearly identifies the intercalated myotome as a transcriptional domain distinct from the epaxial-most myotome. This transcriptional domain had been distinguished on the basis of *frizzled9* expression (Wang et al., 1999). The intercalated domains of the dermomyotome and myotome are marked by the expression of *Sim1* and *En1* (see Tajbakhsh and Spörle, 1998). Expression of the latter is also a characteristic of adaxial cells in the zebrafish embryo, which are specified by sonic hedgehog (see Currie and Ingham, 1998). Further analysis of the 23 kb region is required to establish whether sonic hedgehog activates *Myf5* in the intercalated myotome. The isolation of a sequence that

specifically targets this region will also permit its manipulation and hence investigation of its significance.

The -23 kb region also directs transgene expression to the dorsal epaxial dermomyotome of the somite. Consistent with this, epaxial myotome expression is initiated correctly by the -58/-48 region, whereas expression in the epaxial dermomyotome is lacking. In *sonic hedgehog* mutant mice, *Myf5* activation and epaxial myogenesis are compromised (Borycki et al., 1999). It is possible that this is due to an effect on the initial dermomyotome expression of *Myf5* and that both this and the intercalated myotome expression depend on sonic hedgehog. Expression of *Myf5* in the somite is also influenced by Wnts, produced by the dorsal neural tube and surface ectoderm; in explants, the presence of Wnt1 leads to preferential expression of *Myf5* (Tajbakhsh et al., 1998).

Nature of the distal regulatory elements

Whereas the -58/-48 regulatory region has the property of an enhancer element in that it functions with a proximal *Myf5* promoter, this is not the case for the -96/-63 regulatory region, which is inactive with the minimal promoter or sequences up to 23 kb 5' of *Myf5*, including the epaxial and branchial arch elements (Summerbell et al., 2000). Despite the presence of the latter, expression is not maintained in head muscles that derive from the branchial arches. The -96/-63 element does not appear to act efficiently as an insulator since early epaxial expression due to the sequence at -6.1/-5.5 is not observed in all founder 96 Δ 63-EpaxMyf5-nlacZ embryos examined at E10.75. Silencing, presumably due to the integration site, is seen, as in the case of the transgene with the epaxial element alone (Summerbell et al., 2000). The -96/-63 region may interact with elements present in the 10 kb fragment located between -58 and -48 kb, which direct early expression of *Myf5*. However, preliminary experiments suggest that this is not the case. It appears therefore that the regulatory elements present in the -96/-63 region require the specific configuration of the endogenous gene. The importance of distance between promoters and regulatory elements has been highlighted notably in the case of gene clusters, such as the β -globin locus or the *HoxD* complex, where changing the position of a given gene in the complex affects the timing and/or pattern of its expression (Dillon et al., 1997; van der Hoeven et al., 1996).

Dissection of *Myf5* regulatory regions reveals unexpected heterogeneity in hypaxial muscles

The most distal regulatory region, situated between -96 and -63 kb, is required for *Myf5* expression in a subset of skeletal muscles. The first differences between 63-Myf5-nlacZ and 96-Myf5-nlacZ embryos are observed from about E11.5. By E14.5, head muscles do not express the y63-Myf5-nlacZ transgene, whereas branchial arches, from which many head muscles derive, do at earlier stages, in keeping with the presence of a branchial arch element in the 23 kb upstream of *Myf5*. The existence of a distinct regulatory pathway for myogenesis in the head was shown by analysis of *splotch/Myf5* double mutant embryos (Tajbakhsh et al., 1997). The expression profile of y63-Myf5-nlacZ reveals an unexpected regulatory heterogeneity in body muscles: by E14.5, the differences in expression in many trunk muscles are striking when compared with the limbs, whereas others retain the y96-Myf5-nlacZ expression profile. The expressing and the

non-expressing muscles cannot be simply distinguished by their epaxial or hypaxial origin. The deep back muscles, formed by elongation and in situ from the epaxial myotome (Denetclaw and Ordahl, 2000), express the transgene. *y63-Myf5-nlacZ* is also expressed in the intercostal muscles. The medial-most (adjacent to the sclerotome) part of these muscles has an epaxial contribution, but they are mainly formed by elongation of the somitic bud including the hypaxial myotome, ventrally (Christ et al., 1983; Cinnamon et al., 1999). Ventral body wall musculature, which is of hypaxial origin, does not express the transgene. The exact location of the precursors of ventral body wall muscles in the hypaxial somite is unknown. They may derive from the somitic bud, either from the hypaxial dermomyotome or the hypaxial myotome, or both. Christ et al. (1983) described a dispersion of the cells of the somitic bud, subsequent to its extension into the somatopleure. Indeed some dispersion leading to migration of precursors is suggested by the fact that ventral body wall musculature is perturbed in *splotch* mice (Tajbakhsh et al., 1997; Tremblay et al., 1998). This is also the case for diaphragm and tongue muscles where *Myf5* transgene expression is also lacking in the absence of the -96/-63 region. However, it is clear that *Myf5* expression in limb muscles and those of the limb girdle does not depend on this region. The migration of myogenic progenitor cells to the limb is Pax3 dependent and limb muscles are absent in *splotch* mice (see Tajbakhsh and Buckingham, 2000). However, other regulatory genes of importance in limb muscle development are not expressed at other sites where muscle will form as a result of migration of hypaxial somite derivatives. This is the case for *Mox2* (Mankoo et al., 1999), which has been implicated in the regulation of *Myf5* specifically in the limbs. Another example is *Lbx1*, which is not expressed in the interlimb region (Jagla et al., 1995). The fact that distinct regulatory sequences direct *Myf5* expression in ventral body wall or limb muscles suggests that these elements respond to different regulators of the 'migratory' network which is implicated in myogenesis.

Spatiotemporal heterogeneity in the transcriptional regulation of expression as demonstrated for the gene encoding the myogenic determination factor, *Myf5*, is not a common property of downstream muscle genes. Apart from later fibre type differences, their expression tends to be directed in all skeletal muscles by the same regulatory elements. However, heterogeneity between skeletal muscles with similar characteristics is well known to clinicians, who are confronted by dystrophies affecting only a subset of muscles, such as those of the face (Mathews and Mills, 1996) or limb girdle (Bushby, 1999). The further analysis of factors regulating *Myf5* expression in muscle subdomains may throw light on this phenomenon and also open the way to targeted gene expression and manipulation in specific muscles.

The authors thank R. Kelly, F. Relaix and R. Spörle for helpful discussions and critical reading of the manuscript and C. Cimper and C. Bodin for technical assistance. We also thank C. Fairhead for the gift of the pKA and pAN vectors and her and E. Heard for useful advice on yeast manipulations, and C. Huxley for advice about YAC transgenesis. This work was supported by grants from the Pasteur Institute, the CNRS (Centre National de la Recherche Scientifique) and the AFM (Association Française contre les Myopathies). J. H. was a recipient of a fellowship from the MRES (Ministère de la Recherche et de l'Enseignement Supérieur), the ARC (Association pour la

Recherche sur le Cancer) and the AFM. T. C. was a recipient of a NATO (North Atlantic Treaty Organisation) and now a NIH (National Institutes of Health) fellowship (HD08570). M. P. was supported by fellowships from EMBO (European Molecular Biology Organisation), the Austrian Academy of Sciences and the AFM.

REFERENCES

- Alani, E., Cao, L. and Kleckner, N. (1987). A method for gene disruption that allows repeated use of URA3 selection in the construction of multiply disrupted yeast strains. *Genetics* **116**, 541-545.
- Asakura, A., Lyons, G. E. and Tapscott, S. J. (1995). The regulation of *MyoD* gene expression: conserved elements mediate expression in embryonic axial muscle. *Dev. Biol.* **171**, 386-398.
- Boeke, J. D., LaCrute, F. and Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197**, 345-346.
- 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., 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.
- Bushby, K. M. (1999). The limb-girdle muscular dystrophies-multiple genes, multiple mechanisms. *Hum Mol Genet* **8**, 1875-1882.
- Choulika, A., Perrin, A., Dujon, B. and Nicolas, J. F. (1994). The yeast I-Sce I meganuclease induces site-directed chromosomal recombination in mammalian cells. *Comptes Rendus de l'Académie des Sciences Série III, Sciences de la Vie* **317**, 1013-1019.
- Christ, B., Jacob, M. and Jacob, H. J. (1983). On the origin and development of the ventrolateral abdominal muscles in the avian embryo. *Anat. Embryol.* **166**, 87-101.
- Christ, B. and Ordahl, C. P. (1995). Early stages of chick somite development. *Anat. Embryol.* **191**, 381-396.
- Cinnamon, Y., Kahane, N. and Kalcheim, C. (1999). Characterization of the early development of specific hypaxial muscles from the ventrolateral myotome. *Development* **126**, 4305-4315.
- Cossu, G., Kelly, R., Tajbakhsh, S., Di Donna, S., Vivarelli, E. and Buckingham, M. (1996). Activation of different myogenic pathways: *myf-5* is induced by the neural tube and *MyoD* by the dorsal ectoderm in mouse paraxial mesoderm. *Development* **122**, 429-437.
- Currie, P. D. and Ingham, P. W. (1998). The generation and interpretation of positional information within the vertebrate myotome. *Mech. Dev.* **73**, 3-21.
- 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.
- Dillon, N., Trimborn, T., Strouboulis, J., Fraser, P. and Grosveld, F. (1997). The effect of distance on long-range chromatin interactions. *Mol. Cell* **1**, 131-139.
- Faerman, A., Goldhamer, D. J., Puzis, R. and Emerson, C. P., Jr. (1995). The distal human *myoD* enhancer sequences direct unique muscle-specific patterns of *lacZ* expression during mouse development. *Dev. Biol.* **171**, 27-38.
- Fairhead, C., Llorente, B., Denis, F., Soler, M. and Dujon, B. (1996). New vectors for combinatorial deletions in yeast chromosomes and for gap-repair cloning using 'split-marker' recombination. *Yeast* **12**, 1439-1457.
- Goldhamer, D. J., Brunk, B. P., Faerman, A., King, A., Shani, M. and Emerson, C. P. (1995). Embryonic activation of the *MyoD* gene is regulated by a highly conserved distal control element. *Development* **121**, 637-649.
- Heard, E., Avner, P. and Rothstein, R. (1994). Creation of a deletion series of mouse YACs covering a 500 kb region around Xist. *Nucl. Acids Res.* **22**, 1830-1837.
- Hogan, B., Beddington, R., Costantini, F. and Lacy, E. (1994). *Manipulating the Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Houzelstein, D., Cheraud, Y., Auda-Boucher, G., Fontaine-Perus, J. and Robert, B. (2000). The expression of the homeobox gene *Msx1* reveals two

- populations of dermal progenitor cells originating from the somites. *Development* **127**, 2155-2164.
- Huxley, C., Hagino, Y., Schlessinger, D. and Olson, M. V. (1991). The human HPRT gene on a yeast artificial chromosome is functional when transferred to mouse cells by cell fusion. *Genomics* **9**, 742-750.
- Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983). Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163-168.
- Jagla, K., Dolle, P., Mattei, M. G., Jagla, T., Schuhbauer, B., Dretzen, G., Bellard, F. and Bellard, M. (1995). Mouse *Lbx1* and human *LBX1* define a novel mammalian homeobox gene family related to the *Drosophila* lady bird genes. *Mech. Dev.* **53**, 345-356.
- Kablar, B., Krastel, K., Ying, C., Tapscott, S. J., Goldhamer, D. J. and Rudnicki, M. A. (1999). Myogenic determination occurs independently in somites and limb buds. *Dev. Biol.* **206**, 219-231.
- Kelly, R., Alonso, S., Tajbakhsh, S., Cossu, G. and Buckingham, M. (1995). Myosin light chain 3F regulatory sequences confer regionalized cardiac and skeletal muscle 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.
- Larin, Z., Monaco, A. P. and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl Acad. Sci. USA* **88**, 4123-4127.
- Mankoo, B. S., Collins, N. S., Ashby, P., Grigorieva, E., Pevny, L. H., Candia, A., Wright, C. V., Rigby, P. W. and Pachnis, V. (1999). *Mox2* is a component of the genetic hierarchy controlling limb muscle development. *Nature* **400**, 69-73.
- Manson, A. L., Trezise, A. E., MacVinish, L. J., Kasschau, K. D., Birchall, N., Episkopou, V., Vassaux, G., Evans, M. J., Colledge, W. H., Cuthbert, A. W. et al. (1997). Complementation of null CF mice with a human CFTR YAC transgene. *EMBO J.* **16**, 4238-49.
- Mathews, K. D. and Mills, K. A. (1996). The molecular genetics of human facioscapulohumeral muscular dystrophy and the myodystrophy mouse model. *Curr Opin Neurol* **9**, 394-399.
- McKee-Johnson, J. W. and Reeves, R. H. (1996). Fragmentation and integrative modification of YACs. In *YAC protocols*, vol. 54 (ed. D. Markie), pp. 167-186. Totowa, NJ: Humana Press, Inc.
- Miner, J. H. and Wold, B. J. (1990). Herculin: a fourth member of the MyoD family of myogenic regulator genes. *Proc. Natl Acad. Sci. USA* **87**, 1089-1093.
- Noden, D. M. (1983). The embryonic origins of avian cephalic and cervical muscles and associated connective tissues. *Am. J. Anat.* **168**, 257-76.
- Noden, D. M., Marcucio, R., Borycki, A. G. and Emerson, C. P., Jr. (1999). Differentiation of avian craniofacial muscles: I. Patterns of early regulatory gene expression and myosin heavy chain synthesis. *Dev. Dyn.* **216**, 96-112.
- Olson, E. N., Arnold, H. H., Rigby, P. W. and Wold, B. J. (1996). Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene *MRF4*. *Cell* **85**, 1-4.
- Ordahl, C. P. and Le-Douarin, N. M. (1992). Two myogenic lineages within the developing somite. *Development* **114**, 339-353.
- Ott, M. O., Bober, E., Lyons, G., Arnold, H. and Buckingham, M. (1991). Early expression of the myogenic regulatory gene, *myf-5*, in precursor cells of skeletal muscle in the mouse embryo. *Development* **111**, 1097-1107.
- Patapoutian, A., Miner, J. H., Lyons, G. E. and Wold, B. (1993). Isolated sequences from the linked *myf5* and *MRF4* genes drive distinct patterns of muscle-specific expression in transgenic mice. *Development* **118**, 61-69.
- Pavan, W. J., Hieter, P. and Reeves, R. H. (1990). Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. *Proc. Natl Acad. Sci. USA* **87**, 1300-1304.
- Philippsen, P., Stotz, A. and Scherf, C. (1991). DNA of *Saccharomyces cerevisiae*. In *Guide to yeast genetics and molecular biology*, vol. 194 (ed. G. R. Guthrie C. and Fink), pp. 169-174: Academic press, NY.
- 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.
- Plessis, A., Perrin, A., Haber, J. E. and Dujon, B. (1992). Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics* **130**, 451-60.
- Rudnicki, M. A., Braun, T., Hinuma, S. and Jaenisch, R. (1992). Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene *Myf-5* and results in apparently normal muscle development. *Cell* **71**, 383-390.
- Rudnicki, M. A., Schnegelsberg, P. N. J., 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.
- Sassoon, D., Lyons, G., Wright, W. E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M. (1989). Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. *Nature* **341**, 303-3907.
- Spörle, R. and Schughart, K. (1998). Paradox segmentation along inter and intrasomitic borderlines is followed by dysmorphology of the axial skeleton in the *open brain (opb)* mouse mutant. *Developmental genetics* **22**, 359-373.
- Summerbell, D., Ashby, P. R., Coutelle, O., Cox, D., Yee, S-P. 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.
- Tajbakhsh, S. and Buckingham, M. E. (1995). Lineage restriction of the myogenic conversion factor *myf5* in the brain. *Development* **121**, 4077-4083.
- Tajbakhsh, S. and Spörle, R. (1998). Somite development: constructing the vertebrate body. *Cell* **92**, 9-16.
- Tajbakhsh, S. and Buckingham, M. (2000). The birth of muscle progenitor cells in the mouse: spatiotemporal considerations. In *Somitogenesis*, vol. 48 (ed. C. Ordahl), pp. 225-268. San Diego: Academic Press.
- 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 fibers as well as early embryonic muscle. *Developmental Dynamics* **206**, 291-300.
- Tajbakhsh, S., Rocancourt, D. and Buckingham, M. (1996b). Muscle progenitor 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.
- 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.
- van der Hoeven, F., Zakany, J. and Duboule, D. (1996). Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls. *Cell* **85**, 1025-35.
- Wang Y-K, Spörle, R., Paperna, T., Schughart, K. and Francke, U. (1999). Characterization and expression pattern of the *frizzled* gene *Fzd9*, the mouse homolog of *FZD9* which is deleted in Williams-Beuren syndrome. *Genomics* **57**, 235-248.
- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benzera, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S. et al. (1991). The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* **251**, 761-766.
- Yoon, J. K., Olson, E. N., Arnold, H.-H. and Wold, B. J. (1997). Different *MRF4* knockout alleles differentially disrupt *Myf-5* expression: cis-regulatory interactions at the *MRF4/Myf-5* locus. *Dev. Biol.* **188**, 349-362.
- 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.