The expression of Myf5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors

Dennis Summerbell*, Peter R. Ashby‡, Oliver Coutelle, David Cox*, Siu-Pok Yee§ and Peter W. J. Rigby*¶

Division of Eukaryotic Molecular Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK
*Present address: Section of Gene Function and Regulation, Institute of Cancer Research, Chester Beatty Laboratories, 237, Fulham Road, London, SW3 6JB, UK
‡Present address: Welcome Trust Building, University of Dundee, WTB/MSI Complex, Dow Street, Dundee DD1 5EH, UK
§Present address: Cancer Research Laboratories, London Regional Cancer Centre, 790 Commissioners Road East, London, Ontario, N6A 4L6, Canada
¶Author for correspondence (e-mail: p.rigby@icr.ac.uk)

Accepted 20 June; published on WWW 9 August 2000

SUMMARY

The development of skeletal muscle in vertebrate embryos is controlled by a transcriptional cascade that includes the four myogenic regulatory factors Myf5, Myogenin, MRF4 and MyoD. In the mouse embryo, Myf5 is the first of these factors to be expressed and mutational analyses suggest that this protein acts early in the process of commitment to the skeletal muscle fate. We have therefore analysed the regulation of Myf5 gene expression using transgenic technology and find that its control is markedly different from that of the other two myogenic regulatory factor genes previously analysed, Myogenin and MyoD. We show that Myf5 is regulated through a number of distinct and discrete enhancers, dispersed throughout 14 kb spanning the MRF4/Myf5 locus, each of which drives reporter gene expression in a particular subset of skeletal muscle precursors. This region includes four separate enhancers controlling expression in the epaxial muscle precursors of the body, some hypaxial precursors of the body, some facial muscles and the central nervous system. These elements separately or together are unable to drive expression in the cells that migrate to the limb buds and in some other muscle subsets and to correctly maintain expression at late times. We suggest that this complex mechanism of control has evolved because different inductive signals operate in each population of muscle precursors and thus distinct enhancers, and cognate transcription factors, are required to interpret them.

Key words: Myf5, Myogenic regulatory factor, MRF, Muscle, Branchial arch, Epaxial, Hypaxial

INTRODUCTION

In vertebrate embryos, skeletal muscle is derived from two components: a connective tissue substratum that develops in situ and provides the form and anchorage to the skeleton (Chevalier and Kiency, 1982; Köntges and Lumsden, 1996), and the myoblasts, which migrate into position, differentiate into myotubes and provide the bulk of the muscle and its characteristic histotype (Chevalier et al., 1977; Christ et al., 1977). We are interested primarily in the commitment of myoblasts, which derive from the paraxial mesoderm that forms immediately adjacent to the neural tube during gastrulation. In the trunk, the segmental units, the somites, which are located either side of the neural tube, form sequentially from the anterior end of the presomitic mesoderm such that the most cranial pair was the first to be born (review: Christ et al., 1992). The somites produce the myoblasts of the trunk, limbs and tongue. The remaining head myoblasts are derived from the anterior paraxial mesoderm, which is not overtly segmented (Noden, 1986; Trainor and Tam, 1995).

It is generally assumed that the developmental processes of somitic myogenesis in the mouse will be essentially the same as those defined using direct cell marking experiments in chick and quail embryos (review: Christ and Ordahl, 1995). The epaxial myoblasts (which migrate into the prospective muscles dorsal to the transverse processes of the vertebrae) appear first, originating at the dorsomedial lip of the dermomyotome. The cells involute from the epithelial edge and accumulate on the inner surface of the dermomyotome to form the dorsal component of the myotome. The behaviour of hypaxial myoblasts depends on their position along the craniocaudal axis. In most trunk somites, some hypaxial myoblasts involute from the ventrolateral lip of the dermomyotome and accumulate on its inner surface forming the ventral component of the myotome. However, other presumptive myoblasts remain attached to the epithelial dermomyotome forming the somitic bud, which invades the lateral body wall by blastemal growth and then differentiates in situ. At the appropriate somitic levels, cells also detach from the ventral dermomyotomal lip and migrate into their final location in the
forelimbs and hindlimbs and the thoracic and pelvic diaphragms. Head muscle formation is less well understood but experiments in both mouse and chick suggest that myoblasts from the occipital somites and head mesoderm follow a complicated program of cell migration in which myoblasts migrate first into the hypoglossal cord (Noden, 1983; Mackenzie et al., 1998) or into the branchial arches (Hacker and Guthrie, 1998) and then back out into appropriate locations in the tongue and head.

While each of these groups of cells from the somites or anterior paraxial mesoderm differentiate into myoblasts, they follow very different paths and presumably respond to different environmental signals that govern their behaviour. Manipulative embryological and explant experiments, in both chick and mouse, have shown that the environmental signals that instruct cells as to their skeletal muscle fate emanate from axial structures including the notochord (Pourquié et al., 1993) and neural tube (Teillet and Le Douarin, 1983; Rong et al., 1992; Buffinger and Stockdale, 1994; Munsterberg and Lassar, 1995), from the overlying surface ectoderm and from lateral plate mesoderm (Pourquié et al., 1995; Munsterberg et al., 1995; Dietrich et al., 1998). The effects of some of these signals can be mimicked by well-characterised molecules, for example Sonic hedgehog (Shh) and various Wnt proteins (Cossu et al., 1996a). The signals induce a cascade of transcription factors that involves the four myogenic regulatory factors (MRFs): Myf5 (Braun et al., 1989), Myogenin (Myog: Edmondson and Olson, 1990; Wright et al., 1989), MRF4 (Rhodes and Konieczny, 1989; Braun et al., 1990; Miner and Wold, 1990) and MyoD (Davis et al., 1987), which are members of the basic helix-loop-helix (bHLH) superfamily (Weintraub et al., 1989). In the mouse, Myf5 is the first such gene to be expressed at 8.0 dpc (Ott et al., 1991), followed within half a day by Myog (Sassoon et al., 1989). It has been shown that Myog expression depends on an E-box, the consensus binding site for the bHLH factors themselves and shown that expression depends on an E-box, the determination, we need to identify all of the transcription factors that interact with the Myf5 promoter and enhancers. Such knowledge would open the way to elucidation of the biochemical mechanisms by which the inductive signals regulate the activities of these factors. We have therefore used reporter gene assays in transgenic mice to identify the elements that regulate Myf5 expression. Here we provide evidence that the control of the expression of Myf5 is extremely complex, with discrete elements responsible for driving expression in different anatomical locations. Our data show that the control mechanisms for Myf5 are distinct from those that operate for either Myog or MyoD.

MATERIALS AND METHODS

Isolation of a genomic clone containing the murine Myf5 locus

A cosm id (MF5.2) was isolated by screening a genomic library derived from a T-cell clone of (CBA × B10)F1 origin in the vector cos202 (gift of D. Kioussis, NIMR) with a mouse Myf5 cDNA probe (gift of A. Buonanno, NIH, accession number: X56182). All of the reporter constructs used to produce transgenic mice were derived from this cosm id using standard recombinant DNA techniques (Sambrook et al., 1989). Full construction details and maps of all constructs are available on request.

Preparation of Myf5-lacZ fusion constructs

For further manipulations, a 14 kb KpnI-EcoRI cosmid fragment containing Myf5 was subcloned into pBluescript KS(+) (Stratagene) with a modified polylinker. PCR was used to mutate the ATG of the translational start codon of Myf5 into a BamHI restriction site (Yee and Rigby, 1993), and a BamHI cassette containing the lacZ reporter, which included a nuclear localisation signal and an SV40 polyadenylation sequence (gift of R. Krumlauf, NIMR) was inserted at this site. This construct (#1) was used to generate a deletion series (constructs #1-7). Construct #4 was prepared similarly to constructs #1-3 and 5-7 but the reporter cassette came from pD16.43 (Fire et al., 1990). Construct #4 utilises the endogenous Myf5 polyadenylation sequences. Constructs #10-13 contained the appropriate Myf5 sequences cloned upstream of a β-globin minimal promoter driving a cytoplasmic lacZ reporter with an SV40 polyadenylation sequence in plasmid BGZA, a pBluescript KS(+) based derivative of BGZ40 (Yee and Rigby, 1993). Construct #14 contained the appropriate PCR generated Myf5 fragment (primer pair: IN1 [5′-ctgagggacagttg-agaa-3′] and UTRr [5′-catgctgtataattgcacct-3′]) cloned upstream of the hsp68-lacZ-SV40poly(A) reporter gene (Whiting et al., 1991). Prior to injection, novel reporter cassettes were tested for function by transfection of 16 µg of plasmid DNA into mouse C2C12 (myoblast), 10T1/2 (fibroblast) or Neuro2a (neuroblastoma) cells using the calcium phosphate method (Sambrook et al., 1989).

Analysis of transgenic animals, in situ hybridisation and histology

Transgenic mice were produced as previously described (Yee and Rigby, 1993). Transgenic mice were diagnosed using the primers: LZD (5′-gtttttccgatttggctac-3′) and STD (5′-ggacaaaccacaactag- atgc-3′) which span the lacZ-SV40 poly(A) junction; or in the case of construct #4, Myf-140f (5′-caggactgctgatgttaaac-3′) and NLSr (5′-tgtgaaaccggtgga-3′), which span the Myf3 nuclear localisation signal boundary. Embryos for lacZ whole mounts and sections were fixed in Mirsky’s fixative (National Diagnostics) and stained using 500 µg/ml of X-gal in 5 mM K3Fe(CN)6, 5 mM K4Fe(CN)6, 2 mM MgCl2 in PBS with 0.1% NP40 at 37°C in the dark. In some cases, such stained embryos were lightly counterstained in 0.1% aqueous Acid Fuchsin so as to enhance contrast between somites and other tissues. Whole-mount in situ hybridisation was
carried out using digoxigenin-labelled riboprobes (Wilkinson, 1992) with modifications provided by D. Henrique and D. Ish-Horowicz (ICRF, Lincoln’s Inn Fields, London, UK). Riboprobes were prepared using T7 RNA polymerase on an XbaI linearised 1063 bp (BbrPI-MscI) genomic subclone of Myf5. Only 354 bp of the 504 bp riboprobe form duplexes with Myf5 mRNA. Some of the whole-mount-stained embryos were embedded in paraffin wax or agarose and sectioned.

RESULTS

The expression pattern of the mouse Myf5 gene has previously been studied using both radioactive in situ hybridisation (Ott et al., 1991) and histochemistry on embryos in which a lacZ reporter gene has been knocked-in to one of the Myf5 alleles (Tajbakhsh and Buckingham, 1994). Whilst very informative, the published data do not provide a complete control set for our experiments. We therefore reanalysed the pattern of Myf5 expression using digoxigenin whole-mount in situ hybridisation.

Myf5 in situ expression pattern

Fig. 1 shows a 9.5 dpc (25-somite) embryo that was hybridised with a digoxigenin-labelled Myf5 antisense probe and then serially sectioned. Myf5 transcripts were not detectable in the presomitic mesoderm (psm) and at this stage first appeared in cells at the dorsal lip of the dermomyotome of somite I. These cells upregulate expression of the gene as they involute to form the myotome (m) in somite II. This expression domain lies close to dorsal neural tube and ectoderm, but distant from the notochord/floorplate and lateral plate mesoderm. Approximately 4 hours later, Myf5 was strongly expressed in the ventral myotome of somite IV, a domain close to ectoderm and lateral plate but distant from neural tube or notochord. Note also some Myf5-expressing cells that do not obviously belong to either the dorsal or ventral domains (red arrows, somites II and IV). By somite VI, Myf5 was expressed throughout the myotome and the somitic bud (sb) had formed in thoracic somites. It is now widely assumed that these dorsal and ventral domains of expression equate to the presumptive epaxial and hypaxial muscle masses (Ordahl and Williams, 1998). The hypaxial cells from somite XIV (corresponding to thoracic somite 1), which migrate into the thoracic diaphragm (td) expressed Myf5 while the equivalent cells in the adjacent somite XV (cervical somite 8), which later migrate into the limb bud, did not (arrowhead). The Myf5-expressing cells previously described in the neural tube and midbrain (Ott et al., 1991; Tajbakhsh et al., 1994) appeared very faint and were normally not visible in intact whole-mount embryos.

Expression of a 14 kb Myf5 transgene

Construct #1 contains sequences from 8.6 kb upstream (KpnI) to 5.3 kb downstream (EcoRI) of the Myf5 transcription initiation site (Fig. 2). The upstream end-point of this construct is within the first exon of the adjacent MRF4 gene, so as to remove the MRF4 promoter and avoid the possibility that expression of MRF4 from the transgene altered normal muscle development.

Expression began well before 8.5 dpc in the domain of the somites (Figs 2a, 3a) which corresponds to the normal onset of Myf5 expression (Ott et al., 1991; Fig. 1). At this stage, six of the nine somites were clearly expressing. In older embryos (Fig. 2b, 9.5 dpc), the transgene was expressed in the dorsal lip of somite I as soon as it formed from the presomitic mesoderm (Fig. 2b, side panel arrow). This is precisely the
same timing and location of Myf5 expression as shown in Fig. 1.

At 10.5 dpc, expression in the hypaxial domain had begun at thoracic and lumbar levels and included the somitic bud (Figs 2c, 3c, arrow). At all times, expression in the hypaxial domain was inappropriately weak relative to that in the epaxial domain. This was particularly apparent at 11.5 dpc in the hypaxially derived intercostal muscles (Fig. 3d). Expression in the relatively immature somites of the tail appeared normal except at the extreme ventral margin (Figs 2d, side panel, 3d,e, arrow).

While the transgene in construct #1 lines activated correctly in the epaxial region of newly born somites, it was apparent, by comparison with the whole-mount in situ data, that in more mature somites the expression pattern was inexact. In the dorsal half of the somites, expression corresponded to the normal pattern, but progressively extended down the posterior rather than anterior margin of the somite (Figs 2b, 3b, arrow). The phase of inappropriate expression in the somites was transient. Transverse sections at interlimb level of a 10.5 dpc embryo (Fig. 2c, side panel XVI) showed that, in young somites, the transgene was correctly expressed in the myotome (m) but also incorrectly expressed in the dermomyotome (dm). Transverse sections through more mature rostral somites (Fig. 2c, side panel XXV) showed that expression became restricted to the myotome (m). The early, strong dermomyotomal expression was not characteristic of either whole-mount in situ hybridisation (Fig. 1) or heterozygous lacZ knock-in embryos (Tajbakhsh and Buckingham, 1994; Tajbakhsh et al., 1997), suggesting that construct #1 lacks an element that normally represses expression in the dermomyotome. By 11.5 dpc, all inappropriate somitic expression had disappeared and lacZ was restricted to the normal Myf5-positive population (Fig. 2d), except in the dermomyotome of the distal tail somites (Fig. 2d, side panel).

Expression in the branchial arches commenced at around 9.0 dpc (20 somites; Figs 2b, 3b). While expression in the second arch was constant and reliable, expression in arches 1 and 3 was often faint and variable, even between siblings of the same line. By 10.5 dpc, there was strong expression in facial muscle precursors (Figs 2c, 3c), which continued at 11.5 dpc (Figs 2d, 3d) but was downregulated by 12.5 dpc (Fig. 3e).

From 12.5 dpc onwards, expression was markedly downregulated in the epaxial and facial muscle precursors (Fig. 3e) and, by 13.5 dpc, the only muscle components still expressing were in the epaxial somitic compartment (Fig. 3f). This downregulation is not seen in heterozygous knock-in embryos (Tajbakhsh et al., 1997), indicating that construct #1 lacks an element(s) required for the maintenance of expression.

We have never observed staining that parallels the known expression pattern of Myf5 in the limbs and thus conclude that...
3749

Elements controlling Myf5 expression

construct #1 lacks at least one necessary element required for expression in the limb, confirming the data of Zweigerdt et al. (1997).

Expression in the neural tube was inconsistent and weak and in whole mounts normally masked by strong expression elsewhere. We were unable to detect neural expression in the appropriate location in the midbrain.

We routinely observed unexpected expression in head mesoderm, which was strong from 8.5 (Fig. 2a, side panel, arrow) until 12.5 dpc. Towards the end of this period the expression started to concentrate in the meninges in the region of the midbrain flexure (data not shown), presumably forming part of the dura, which is usually considered to be of mesodermal origin. It is noteworthy that this ectopic expression occurred.

**Fig. 3.** Time course of expression in a line of mice carrying construct #1. (a) 8.5 dpc (9 somite): strong expression in the epaxial domain of the somites and ectopically in head mesoderm. (b) 9.5 dpc (23 somite): expression spreads ventrally into the hypaxial domain of more mature somites, expression begins in the branchial arches. (c) 10.5 dpc: somitic expression extends into tail, in mature somites localises in myotome and in the thoracic region extends into the presumptive intercostals. (d) 11.5 dpc: expression strong in all previously positive domains except intercostals which remain weak, no indication of expression in limbs. (e) 12.5 dpc: expression in both intercostals and face downregulated. (f) 13.5 dpc: expression only in presumptive back muscles of epaxial domain. Note the lack of limb expression at all times.

---

**Table 1. Percentage of expressing embryos with β-galactosidase-positive cells in identified anatomical domains**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Transient expressors</th>
<th>Line expressors</th>
<th>Non-expressing</th>
<th>Epaxial domain (%)</th>
<th>Hypaxial domain (%)</th>
<th>Branchial arches (%)</th>
<th>Neural tube (%)</th>
<th>Head mesoderm (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletion series</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#1</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>83</td>
<td>100</td>
<td>83</td>
<td>17</td>
<td>72</td>
</tr>
<tr>
<td>#2</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>83</td>
<td>100</td>
<td>83</td>
<td>17</td>
<td>56</td>
</tr>
<tr>
<td>#3, 4</td>
<td>17</td>
<td>4</td>
<td>11</td>
<td>78</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#5</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#6, 7</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>91</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#8</td>
<td>14</td>
<td>2</td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#9</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#12, 13, 14</td>
<td>32</td>
<td>6</td>
<td>18</td>
<td>18</td>
<td>100</td>
<td>100</td>
<td></td>
<td>24</td>
</tr>
</tbody>
</table>

1Cord expression was always difficult to diagnose as it was almost always weak. It was therefore difficult to see in weakly expressing embryos, and easily masked by more superficial staining in strongly expressing embryos.

2A few placentae were not tested for transgenesis.

3Five of the seven blue embryos had only light inconsistent ectopic expression patterns. Most sites of expression were outside the normal Myf5 expression domains.
Fig. 4. Deletion mapping of Myf5 regulatory elements. Details of all constructs used are shown on the right; constructs in grey are not illustrated. (a) Construct #1 (10.0 dpc): starting point of the deletion series. Immature somites initially express correctly along the dorsal lip (dl) of the dermomyotome but later have ectopic expression in the posterior part of the dermomyotome (dm). Correct myotomal (m) expression can be seen in the mature somites. Expression is also faintly visible in the neural tube (nt) beneath the more cranial somites. The consistently strong expression in the head mesoderm is ectopic. (b) Construct #2 (9.5 dpc): the expression pattern is unchanged. (c) Construct #3 (9.75 dpc): expression in the epaxial domain of the somites and in the head mesoderm is abolished; the same pattern of expression is observed with construct #4 (data not shown). (d) Construct #5 (10.0 dpc): the pattern of expression is very similar to the previous two constructs but consistently gave stronger somitic and weaker arch expression. (e) Construct #7 (11.0 dpc): the remaining somitic expression is abolished. The same pattern of expression is observed with construct #6 (data not shown). (f) Construct #8 (10.5 dpc): expression in the arches is abolished leaving only neural expression, seen as a double column from the dorsal surface. In 4 cases out of 14, there was strong ectopic expression at the caudal end of the embryo. (Not shown: construct #9, 294 bp upstream of the translation start site is insufficient to drive neural tube expression.)
Elements controlling Myf5 expression

3751

Domain is also observed in transgenic mice carrying YAC constructs containing close to 600 kb of DNA (Zweigerdt et al., 1997). Nonetheless we have no evidence from our in situ hybridisation studies that this head mesoderm is a normal site of Myf5 expression.

Deletion mapping of Myf5 regulatory elements

To map control elements within construct #1, we generated a deletion series. The numbers of transgenic embryos and the percentage expressing in different anatomical domains are shown in Table 1 and key members of the series are illustrated in Fig. 4. Removal of 2 kb from the 5′-end deleted the main body of the MRF4 gene (construct #2, Fig. 4b) but produced no obvious changes in the expression pattern. Removal of a further 3 kb (construct #3, Fig. 4c) abolished expression in the epaxial domain of the somites and the ectopic expression in the head mesoderm. Following deletion of 1.4 kb from the 5′-end and 1.8 kb from the 3′-end (construct #5, Fig. 4d), the pattern remained unchanged except that construct #3 produced consistent and strong expression in the branchial arches and weak expression in the somites, while construct #5 showed the converse tendency. In constructs #3-5, where there is no epaxial somitic expression, it is clear (Fig. 4c,d) that in cervical, brachial and tail somites there is little or no hypaxial expression, while in the somites between the limbs such expression is more extensive. Removal of the remainder of the DNA 3′ to the reporter deleted most of the Myf5 gene (construct #6 or #7, Fig. 4e) and abolished expression in the hypaxial domain leaving only branchial arch and neural tube expression. When seen in transverse section, the latter coincided precisely with the location described by Tajbakhsh et al. (1995, Fig. 5b; our data not shown). Removal of a further 1 kb from the 5′ intergenic region (construct #8, Fig. 4f) abolished arch expression and left expression only in the neural tube. This neural element on the homologous promoter driving lacZ had a high incidence (28%) of strong ectopic expression, particularly at the caudal end of the embryo (data not shown). Finally, removal of a further 420 bp down to a 294 bp minimal promoter (construct #9, data not shown) abolished the remaining expression. These results delineated four separate elements which, in the context of

Fig. 5. Enhancer test for sufficiency. (a,d) Construct #10 (9.5 dpc, epaxial element): a 485 bp XmnI-BamHI fragment. The initial expression along the dorsal lip (dl) is correct. In more mature somites there is correct expression in the dorsal myotome but also ectopic expression throughout the dermomyotome (dm) which persists until this structure disperses to become the dermis. The transverse section through a mature somite of the same embryo shows appropriate expression in the dorsal part (epaxial domain) of the myotome (m) but widespread ectopic expression in the dermomyotome (dm). b) Construct #11 (10.5 dpc, arch element): a 1111 bp Nhel-BsaBI fragment driving the β-globin promoter expresses in the branchial arch domain of the somites. c, e, f, g) Construct #14 (9.5 dpc, hypaxial element) a 2.9 kb fragment comprising the Myf5 introns and 3′ UTR on the hsp68 promoter drives expression only in the ventral half of the somites. The close-up (which has been counterstained) shows clear expression in ventral somite IV (arrow), coinciding with the normal onset of hypaxial expression as shown in Fig. 1. The timing of onset of expression is correct even though labelled cells appear first in the dermomyotome rather than the myotome. The coronal section (f) through immature somites of the embryo in panel c) shows that the initial expression pattern is mainly ectopic along the posterior edge of the dermomyotome (dm). The transverse section (g) through a more mature somite of the same embryo shows that expression subsequently appears in the ventral myotome (m) but ectopic expression persists in the dermomyotome (dm). Expression in the notochord is ectopic. Details of all constructs used are shown below; constructs in grey are not illustrated.
construct #1, are required for expression in the epaxial and hypaxial domains of the somites, the branchial arches and the neural tube.

**The somite and arch elements act as enhancers**

Comparison between constructs #2 and #3 (Fig. 4b,c) delineated a putative element required for expression in the epaxial domain of the somite. Analysis of subfragments of this region on the β-globin promoter driving lacZ identified a 484 bp XmnI-BamHI fragment sufficient for expression of the reporter in the epaxial domain (construct #10, Fig. 5a). In early somites (I to VII), the expression pattern appeared normal, but later the dermomyotomal expression persisted and spread ventrally. Transverse sections (Fig. 5d) showed a transitional phase in which there was expression in both dorsal myotome and dermomyotome. This expression domain coincided with the location of the precursors of the epaxial muscles and thus the enhancer may represent the epaxial control element. There was also consistent expression in the head mesoderm. Strong ectopic staining in the dermomyotome (dm) again suggested a missing element that acts negatively in the dermomyotome.

This enhancer never gave expression in the hypaxial myotome, in the arches or in the neural tube.

Comparison between constructs #7 and #8 (Fig. 4e,f) delineated a putative element required for branchial arch expression. Analysis of subfragments of this region on the β-globin promoter driving lacZ identified a 1111 bp NheI-BsaBI fragment (construct #11) that was sufficient to give the expression pattern (Fig. 5b) previously described by Patapoutian et al. (1993). This expression domain coincided with the location of the branchial arch muscle precursors and thus the enhancer may represent an arch control element.

Comparison between constructs #5 and #6 (Fig. 4d,e) delineated a putative enhancer element required for expression in the ventral half of the somites. Analysis of subfragments of this region on the β-globin or hsp68 promoters driving lacZ showed that the introns and 3’ UTR of Myf5 were sufficient to give this pattern (construct #12-14, Fig. 5c). The expression pattern was anomalous in that it appeared in the ventral-posterior quadrant of the dermomyotome. While the timing of expression in the ventral half of the somite was correct, the location was inaccurate, spreading along the posterior rather
than the anterior margin from somite IV (Fig. 5e). Coronal sections through immature somites showed that this posterior expression domain was located primarily in the dermomyotome (Fig. 5f). In more mature somites, transgene expression converged on the normal Myf5 pattern, appearing in both myotome and somitic bud (Fig. 5g). It later extended into the intercostal muscles of thoracic somites (data not shown).

In some cases, these constructs also gave weak transient expression in the branchial arches (see below). They also tended to express ectopically in dorsal root ganglia. While this enhancer activated expression in the ventral part of the somite and is complementary to the epaxial control element described above, it was insufficient to drive the full hypaxial expression domain.

**Late maintenance activity**

We have already drawn attention to a number of differences between the pattern of expression produced by our largest construct and the pattern as shown by the lacZ knock-in (Tajbakhsh and Buckingham, 1994; Tajbakhsh et al., 1997). These are (1) posterior rather than anterior expression in ventral somites, (2) missing limb expression, (3) ectopic expression in head mesoderm, (4) persistent expression in dermomyotome, and (5) premature downregulation in several populations of muscle precursors. We do not know how many additional elements will be required to correct these defects. While the premature loss of expression, most obvious in the ventral somites, indicates a requirement for a distal element(s) that maintains expression at late times (Figs 2d, 3d-f), late maintenance of expression is also influenced by elements proximal to Myf5.

Construct #11 was sufficient to drive expression in the branchial arches (5 out of 5 transient embryos, Table 1; Fig. 5b). We have no lines from this construct; however, all 8 lines from constructs #3 and #5 (which contain the arch control elements defined by construct #11) expressed from 9.5-12.5 dpc (Fig. 6a-d, arrows). This pattern of expression was also consistent with other transient data (e.g. Fig. 4e, construct #7) and with the description of Patapoutian et al. (1993). In contrast, construct #1 downregulated arch expression from 11.5 dpc (8/8 lines, compare Fig. 2d and e with Fig. 6c and d). This suggests that there are sequences within construct #1 but not in construct #3 which block the late maintenance of arch expression. This block is itself suppressed in the wider context of the lacZ knock-in (Tajbakhsh and Buckingham, 1994; Tajbakhsh et al., 1997). Note that the late arch maintenance element in construct #11 was required from 10.0 dpc (see below) but was suppressed in construct #1 from 11.0 dpc.

Constructs #12-14 were sufficient to drive expression in ventral somites (32/32 transient embryos and 6/6 lines) but were also able to drive expression in the branchial arches (5/32 embryos and 4/6 lines; Table 1; Fig. 5c). In isolation, the intragenic sequences downregulated their expression in the hypaxial domain from 10.5 dpc and in the arch domain from before 10.0 dpc (Fig. 6e,f; arrows). In the transient embryos, the five that expressed within the arches all lay within the 9.0-10.0 dpc time window indicated by the lines, while most of the non-expressing transient embryos were older (10.5 dpc). In both domains, downregulation was earlier than in other constructs. In the context of constructs #1-#5 the characteristic modified hypaxial expression was maintained until 11.5 dpc (Figs 3d, 6c; black arrow). In the context of constructs #1-#7
Our data show that the transcription of Myf5 is controlled by a number of elements dispersed throughout the MRF4/Myf5 locus (Fig. 7). Each of the three elements that we have examined in detail functions as a classical enhancer; it is capable of imposing on a heterologous promoter a discrete subset of the overall Myf5 expression pattern, and there is, thus far, no evidence of overlap or synergy between these enhancers. It appears that the transcription of Myf5 in the epaxial, hypaxial and facial muscle precursors is independently controlled and, furthermore, the data indicate that there are separate regulatory circuits operating in subsets of the hypaxial and facial precursors. The facts that even our largest construct gives no expression in the limbs, fails to maintain expression properly after 12.5 dpc and allows ectopic expression in the dermomyotome, indicate that further control elements must exist either 5' of MRF4 or 3' of Myf5. It is thus clear that the regulation of Myf5 is markedly more complex than that of MyoD or Myog, genes that are known to be downstream of Myf5 in the MRF transcriptional hierarchy. We conclude that the regulation of Myf5 is tailored to its role as the determination gene for skeletal muscle. The initiation of the transcriptional cascade that leads to terminally differentiated skeletal muscle cells requires that Myf5 transcription be initiated in a variety of distinct progenitor cell populations; each population is located in a different signalling environment and the gene (or the locus) has evolved to have a distinct enhancer capable of interpreting each signalling environment.

The epaxial enhancer that we have characterised, located close to the MRF4 gene, is capable of activating transcription in the dorsomedial lip of the dermomyotome of newborn somites and is thus likely to be responsible for the initial activation of Myf5 transcription. We cannot exclude the possibility that there are other, more remote enhancers that also function in the epaxial domain, and it is clear from the present data that this enhancer is normally constrained by a negatively acting element that prevents expression in the dermomyotome. In the enhancer tests, this element downregulated sharply between 11.5 dpc and 12.5 dpc but, in the context of our largest construct, retained levels of expression at later stages that were similar to the Myf5 heterozygous lacZ knock-in mice (Tajbakhsh et al., 1997). It may therefore require a late maintenance element, not yet identified, which lies within our largest construct.

In agreement with the previous data of Patapoutian et al. (1993), we find that sequences from the intergenic region act to initiate transcription in the cells that have migrated from the anterior paraxial mesoderm into the branchial arches, and which subsequently give rise to facial muscles. However, examination of a large number of transgenic embryos leads us to the conclusion that this intergenic enhancer does not operate with equal efficiency in all facial muscle precursors. It is most efficient in arch 2 where expression is constant and reliable. It is less reliable in arch 1 and even less so in arch 3; however, expression is very variable even within the same line. Despite this variability, where there is expression, it appears on time at 9.5 dpc and is maintained through to 12.5 dpc. However in the context of our largest construct, late expression driven by this element is blocked by a second element lying in or near the epaxial enhancer region, which is itself blocked by a missing distal element in the context of the entire locus. This blocking action affects maintenance but not initiation.

The intragenic enhancer is particularly complex. It clearly functions in the hypaxial, as opposed to the epaxial, domain but it is not sufficient to direct all aspects of the hypaxial expression pattern. In isolation, it drives reporter gene expression in only the posterior half of each somite and it does not appear to operate in the most ventral regions of the somitic bud. Like the epaxial enhancer, this element is normally constrained by a negatively acting element that functions in the dermomyotome and, like the intergenic arch element, it has a context-dependent requirement for maintenance. In isolation, the hypaxial enhancer downregulates in the somites from 10.5 dpc. In all constructs containing intergenic sequence (including both the arch element and the Myf5 promoter), expression persists in the somites until 11.5 dpc and then downregulates. In the Myf5 heterozygous lacZ knock-in mice, hypaxial somitic expression persists until at least 14.5 dpc (Tajbakhsh et al., 1997). The intragenic enhancer also drives weak transient expression in the arches but, because of the strength of expression, it is not possible to tell if the pattern mirrors that of the intergenic arch element. Only the strongest expressing lines showed this activity and it downregulated from 10.0 dpc. The onset of downregulation in the arches significantly preceded downregulation in the somites.

We have further shown that the expression in the neural tube, which has been extensively documented by Tajbakhsh et al. (1994), is controlled by an element close to the Myf5 minimal promoter. This minimal promoter, defined in this analysis as the 294 bp upstream of the transcriptional start site, is, in isolation, incapable of directing any discrete aspect of the overall Myf5 expression pattern.

In summary, our data show that the regulation of the mouse Myf5 gene is much more complex than might have been expected from comparison to other MRFs, and that expression of Myf5 requires multiple discrete enhancer elements that are responsible for driving or suppressing expression in at least five different anatomical expression domains.

Myf5 regulation is different from that of Myog and MyoD

The complex and highly dispersed arrangement of the Myf5 transcriptional control elements contrasts with the relative simplicity of the elements that control the transcription of the closely related genes Myog and MyoD, both of which act downstream of Myf5 in the transcriptional hierarchy. In the case of MyoD, there are two identified enhancers. A 258 bp element that lies 20 kb upstream of the transcriptional start site appears to recapitulate the full expression pattern (Goldhamer et al., 1992, 1995) while a proximal enhancer approx. 5 kb upstream of the promoter, although widely expressed, gives delayed expression in limb buds and branchial arches (Tapscott et al., 1992; Asakura et al., 1995). Liner scanning mutagenesis of the distal element identified sequences that are not required in head muscle precursors or in the migratory precursors of the limb and diaphragm but are essential in most body muscle precursors (Kucharczuk et al., 1999).
In the case of Myog, all of the cis-acting sequences required for correct spatial and temporal expression in the embryo lie within 133 bp upstream of the transcriptional start site, although there are additional elements within 1 kb of the promoter that increase the level of expression (Cheng et al., 1993; Yee and Rigby, 1993). Mutation of the MEF2-binding site within the minimal control sequences specifically affects expression in the epaxial domain, while mutation of the adjacent MEF3 binding-site affects expression in the epaxial domain and in the muscle progenitors that migrate to the limb and head (Spitz et al., 1998; P. R. A. and P. W. J. R., unpublished data).

Thus, for both MyoD and Myog, there are different upstream regulatory circuits in different muscle precursor cell populations. However, in neither of these cases are there discrete enhancer elements that are specific for particular precursor cell populations arising at distinct locations in the embryo. We believe that the more complex arrangement of the Myf5 control sequences reflects the fact that it alone initiates the myogenic cascade by responding to diverse sets of inductive signals that vary depending on the location of the particular population of myoblasts: epaxial somite, hypaxial somite, face and limb. It has been argued that there is functional redundancy between Myf5 and MyoD (Rudnicki et al., 1993; Braun et al., 1994). However, while MyoD can initiate myogenesis in the absence of Myf5, it does not normally do so. At least in mice, Myf5 initiates the myogenic cascade before MyoD first appears and, indeed, in Myf5 null knock-ins, the expression of MyoD and Myog is significantly delayed (Tajbakhsh et al., 1997). In normal mice both genes lie downstream of Myf5: the fact that Myf5 has to respond to diverse inductive signals necessitates the complex control system that we have revealed.

**Signals regulating Myf5 expression**

The identification of signalling molecules that might act in the process of myogenesis has generally depended on explant culture systems that are not capable of distinguishing between signals that operate in particular domains of the somite. However, the diversity of signals and their suggested sites of synthesis (Fig. 7) suggest complicated graded distributions within the responding tissues: e.g. Shh in the notochord and floorplate (Fan and Tessier-Lavigne, 1994); Wnts in the dorsal neural tube (Munsterberg et al., 1995; Stern and Hauschka, 1995), Wnts and BMPs in the surface ectoderm and BMP4 in the lateral plate mesoderm (Pourquie et al., 1996; Dietrich et al., 1998). Along the dorsoventral axis, different concentrations of signals secreted ventrally from the notochord, the floorplate and the lateral plate mesoderm and dorsally from the surface ectoderm and dorsal neural tube could potentially induce hypaxial and epaxial muscle precursors, respectively. Shh for example has short- and long-range signalling functions (Lee et al., 1994; Marti et al., 1995; Roelink et al., 1995). A long-range function of Shh is to induce competence in chick somites I-III to respond to Wnt signals from the dorsal neural tube. Indeed presomitic mesoderm requires temporary exposure to Shh to enable it to subsequently respond to Wnt signals and initiate myogenesis (Munsterberg et al., 1995). BMP4 has been implicated as a repressor of lateral Myf5 expression (Cossu et al., 1996b) and the BMP4 antagonist Noggin is expressed in dorsal somites (Ikeya and Takada, 1998) which could permit activation of Myf5 epaxially but not hypaxially.

While it is entirely possible that there are unidentified myogenic signals, combinatorial and dose-dependent signalling are sufficient to allow those molecules already implicated in myogenic induction to uniquely specify the various populations of muscle precursor cells. We do not wish to imply that the signalling molecules that act on each myoblast population are necessarily distinct, only that each population is exposed to a particular combination of signals acting at a particular concentration. According to this view, the initiation of Myf5 transcription in each cell population would depend on an enhancer that binds a particular set of transcription factors and thus has distinct sequence motifs. Indeed, we have found no evidence for recurrent sequence motifs in the various enhancer regions examined (D. S., C. Halai and P. W. J. R., unpublished data).

In the mouse, direct evidence on the relationship between putative signalling molecules upstream of Myf5 is becoming available as knock-outs of more of these genes are produced. In 9.5 dpc Wnt1 and Wnt3a double null embryos, the dorsomedial lip of the dermomyotome is lost and thus Myf5 expression in the epaxial domain is abolished (Ikeya and Takada, 1998). In shh-null embryos, which fail to express epaxial Myf5 properly, the dorsal lip of the dermomyotome appears to form initially in young somites but soon disintegrates (Borycki et al., 1999). These data support a role for Shh and Wnt signalling in myogenesis but cannot say whether such signals act directly on Myf5 transcription. There are thus numerous candidates in the different anatomical domains that may be involved in the regulation of Myf5. The next step in our research will be the identification of required binding sites within the enhancer elements.

**Know your neighbours**

Three independent null mutations of the MRF4 gene have been generated in mice (Braun and Arnold, 1995; Patapoutian et al., 1995; Zhang et al., 1995). Although the three alleles were all obtained by disruption of the MRF4 gene through insertion of a neomycin cassette involving different deletions of the locus, the phenotypes were surprisingly variable. Despite this variation, all three alleles have resemblances to the Myf5 knockout. We have previously suggested that the two most likely, but not mutually exclusive, possibilities were that transcription of the inserted neomycin selection cassette in the same direction as Myf5 interferes with the expression of the latter, or that the insertion/deletion removed essential regulatory elements lying close to the MRF4 gene (Olson et al., 1996).

In the Olson allele (Zhang et al., 1995), neomycin is inserted in the opposite transcriptional orientation to Myf5 and the deletion is more extensive than in the Arnold (Braun and Arnold, 1995) or Wold (Patapoutian et al., 1995) mutations. The Olson deletion extended from the Psrl site in the first exon of MRF4 to the BamHI site in the intergenic region, spanning noncoding regions within MRF4 and the adjacent intergenic region. In this paper, we show that this allele deletes the entire epaxial control element of Myf5 (construct #10, Fig. 5) and this is the most likely cause for the apparent Myf5 phenotype. Nevertheless, it is possible that insertion of neomycin might also interfere with Myf5 transcription.
Unlike the Olson allele, the Arnold and Wold alleles carry the neomycin gene inserted in the same transcriptional orientation as Myf5 and, while both delete part of the first exon of MRF4 and some adjacent sequences, most of the introns and untranslated region remain intact. Our data show that removal of MRF4 sequences deleted in the Wold allele does not produce an obvious change in the expression pattern of Myf5 in transgenic mice (compare constructs #1 and #2, Fig. 3a,b). Therefore, the most-likely explanation for the Myf5 phenotypes is interference of neomycin transcription with Myf5 expression. However, in the case of the Arnold allele, we cannot exclude the possibility that deletion of an additional 102 bp of 5′ non-coding sequences could also interfere with Myf5 transcription. It is plausible that such elements exist since synteny between MRF4 and Myf5 is conserved in one of the earliest vertebrates, the teleost Fugu rubripes, suggesting a functional requirement for the two genes to remain linked as a pair throughout vertebrate evolution (O. C., D. S. and P. W. J. R., unpublished data). We are currently investigating whether any of the elements that we have identified regulates MRF4 as well as Myf5.

Conclusion
Myf5 is regulated very differently to the other MRFs in the myogenic cascade. We have defined discrete enhancers for the epaxial somite, hypaxial somite and branchial arches. In addition, we have mapped a neural element and provided indirect evidence for a missing dermomyotomal repressor, for an element required for correct hypaxial expression and for late maintenance elements. We speculate that this discrete but dispersed arrangement is due to Myf5 having evolved so as to respond to the disparate developmental signals or combinations of signals in different anatomical regions of the embryo and to initiate the myogenic cascade appropriately in these regions.

We are grateful to the staff of the Biological Services Division, particularly Hannah Boyes, Jane Seabury and Zoe Webster for expert animal husbandry; to Chandrika Halai for expert technical assistance and to the members of the Division of Embryonic Molecular Genetics, past and present, for their critical reading of the manuscript. During part of this project, P. R. A. was supported by a Fellowship from the International Human Frontier Science Program Organisation. O. C. held a Graduate Studentship from the Medical Research Council, which also paid for this work.

REFERENCES
Edmondson, D. G. and Olson, E. N. (1990). A gene with homology to the myc similarity region of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev. 4. 1450.