genesis 29:22–35 (2001)

ARTICLE

Molecular Structure, Dynamic Expression, and Promoter Analysis of Zebrafish (*Danio rerio*) *myf*-5 Gene

Yau-Hung Chen, Wen-Chih Lee, Chia-Feng Liu, and Huai-Jen Tsai*

Institute of Fisheries Science, National Taiwan University, Taipei, Taiwan, Republic of China

Received 12 August 2000; Accepted 24 October 2000

Summary: We isolated a 1,438 bp cDNA fragment that encoded Myf-5 myogenic factor of zebrafish. The deduced amino acid contained 237 residues, including the basic helix-loop-helix domain that is conserved in all known Myf-5. The zebrafish *myf*-5 transcripts were first detectable at 7.5 hpf, increased substantially until 16 hpf, and then declined gradually to an undetectable level by 26 hpf. During somitogenesis, zebrafish *myf*-5 transcripts were distributed mainly in the somites and segmental plates. Prominent signals occurred transiently in adaxial cells in two parallel rows but did not extend beyond the positive-signal somites. Various lengths of upstream region of zebrafish *myf*-5 fused with EGFP gene were used to carry out transgenic analysis. Results showed that a small, 82 bp (nucleotide positions from -82 to -1), regulatory cassette is sufficient to control the somite- and stage-specific expression of zebrafish *myf*-5 during early development. genesis 29:22–35, 2001. © 2001 Wiley-Liss, Inc.

Key words: fish; microinjection; molecular structure; muscle-specificity; somitogenesis; transgenic fish

INTRODUCTION

Skeletal muscle formation during myogenesis is controlled by four basic helix-loop-helix (bHLH) transcription factors: MyoD (Davis et al., 1987), myogenin (Braun *et al.*, 1989a; Edmondson and Olson, 1989; Wright *et al.*, 1989), Myf-5 (Braun *et al.*, 1989b), and MRF4/herculin/Myf-6 (Rhodes and Konieczny, 1989; Braun *et al.*, 1990; Miner and Wold, 1990). These bHLH factors are nuclear proteins that transactivate the expression of muscle-specific genes, such as the muscle creatine kinase gene (Jaynes *et al.*, 1988) and myosin light chain gene (Braun and Arnold, 1991; Faerman and Shani, 1993), that contain one or more E-box motifs and a DNA-binding site with the general consensus sequence CANNTG.

Each bHLH factor may play a different role during myogenesis. In situ hybridization in mouse embryos demonstrate that *myf*-5 is the first myogenic bHLH gene to be expressed in the somites at 8.0 days postcoitum (dpc) (Ott *et al.*, 1991), followed by *myogenin* at 8.5 dpc (Sassoon *et al.*, 1989), and *mrf*4 at 9.0 dpc. Expression of *myo*D begins at 10.5 dpc in myotomes, but by this stage *myf*-5 expression is decreasing (Sassoon *et al.*, 1989; Lyons and Buckingham, 1992). In contrast to the sequential order of bHLH gene transcripts in mice, *Xenopus* and quail *myo*D are expressed in the somites before other myogenic HLH genes (Rupp and Weintraub, 1991; Pownall and Emerson, 1992). In addition, *Xenopus myo*D mRNA is maternally inherited (Rupp and Weintraub, 1991). Gene knockout experiments with mice showed that normal muscle development occurred when either the *myf*-5 or *myo*D gene was inactivated. However, transgenic mice in which both *myf*-5 and *myo*D genes were disrupted did not form skeletal muscle (Rudnicki *et al.*, 1993). Thus, *myf*-5 and *myo*D may compensate for each other to regulate skeletal muscle development.

In vertebrates, muscle fibers in the trunk and limbs are formed by myogenic cells that originate from somites (Chevallier *et al.*, 1977; Christ *et al.*, 1977). Somites form as epithelial balls but soon reorganize into three regions: dermatome, myotome, and sclerotome. Unlike the somites of mouse and chick, zebrafish (*Danio rerio*) somite gives rise to sclerotome and myotome (Kimmel *et al.*, 1995). Somite formation of zebrafish initiates at 10.5 h postfertilization (hpf) and forms sequentially from anterior to posterior. One pair of new somitic furrow takes approximately 30 min to complete, and about 26–30 pairs of somite are formed (Kimmel *et al.*, 1995; van Eeden *et al.*, 1996). Embryonic expression in the skeletal muscle of zebrafish is easily observed due to the transparency, rapid development, and definite stages of the embryo. These advantages make zebrafish an excellent animal model for studying myogenesis.

Myf-5 cDNA has been reported in some vertebrates, including humans (Braun *et al.*, 1989b), mice (Buonanno *et al.*, 1992), bovine (Barth *et al.*, 1993), chickens (Saitoh

Nucleotide sequence data are in the GenBank databases under the accession numbers AF184166 and AF270789.

* Correspondence to: Huai-Jen Tsai, 1, Roosevelt Road, Sec. 4, Taipei, Taiwan 106. E-mail: hjtsai@ccms.ntu.edu.tw

Contract grant sponsor: National Science Council; Contract grant number: NSC 89-2313-B002-227.

et al., 1993), *Xenopus* (Hopwood *et al.*, 1991), and carp (Kobiyama *et al.*, 1998). But the upstream regulatory regions of *myf*-5 gene have been reported only in mice (Hadchouel *et al.*, 2000; Summerbell *et al.*, 2000). Zebrafish *myf*-5 gene is not clearly and fully characterized so far. In addition, the transcriptional pattern of zebrafish *myf*-5 during the early development is also uncovered. In this study, we characterize the genomic structure and upstream sequence of zebrafish *myf*-5 gene. We perform in situ hybridization for whole embryos to observe the initiation stage of zebrafish *myf*-5 transcription and the spatial distribution of zebrafish *myf*-5 transcripts during embryonic development. Furthermore, we analyze the proximal *cis*-regulatory elements of zebrafish *myf*-5 gene using transgenic analysis. This information should help us understand much more about the molecular structure and regulatory mechanisms of the *myf*-5 gene.

RESULTS

Sequences of Nucleotide and Deduced Amino Acid of Zebrafish *myf*-5 cDNA

A 252-bp fragment was amplified by the Myf5-67F and Myf5-150R primers. A 439-bp fragment corresponding to the 5′-end, noncoding regions, was amplified with RAAPC and Myf5-150R primers. For 3′ rapid amplification of cDNA ends (RACE), primers Myf5-67F and RAAPT were used for the first PCR, and primers Myf5-96F and RAUAP were used for the second PCR, resulting in an 1,105-bp fragment. The full-length cDNA of zebrafish *myf*-5 was 1,433 bp and contained a 714-bp open reading frame with 43- and 676-bp flanking regions at the 5′- and 3′-ends, respectively (Fig. 1). The deduced amino acid sequence of zebrafish *myf*-5 revealed a 237-amino-acid polypeptide that contained a bHLH domain located at amino acid positions 66 to 124.

The sequence identity of the zebrafish *myf*-5 polypeptide bHLH domain was 82.8% with zebrafish MyoD (Weinberg *et al.*, 1996) and 74.1% with zebrafish myogenin (Chen *et al.*, 2000) (Fig. 2a). However, the entire zebrafish *myf*-5 polypeptide sequence shared only 54.3% and 40.8% amino acid identity with those of zebrafish MyoD and myogenin, respectively. The bHLH domain of the zebrafish *myf*-5 polypeptide exhibited 100, 84.5, 86.2, 82.8, 84.5, and 84.5% identity, with the Myf-5 bHLH domain from carp (Kobiyama *et al.*, 1998), *Xenopus* (Hopwood *et al.*, 1991), chickens (Saitoh *et al.*, 1993), bovine (Barth *et al.*, 1993), mice (Buonanno *et al.*, 1992), and human (Braun *et al.*, 1989b), respectively (Fig. 2b).

Dynamic Expression and Spatial Distribution of Zebrafish *myf*-5 Transcripts

Whole-mount in situ hybridization showed that zebrafish *myf*-5 transcripts were first observed in the segmental plates 7.5 hpf (Fig. 3a) and extended about two pairs of somites every hour (Figs. 3b–d). At 16-hpf embryos, about twelve pairs of somites gave positive signals for zebrafish *myf*-5 transcripts (Fig. 3d). After 16 hpf, the signals in the somites became weaker and weaker. The number of somites exhibiting gene expression gradually decreased from 16 to 24 hpf (data not shown). By 24 hpf, zebrafish *myf*-5 transcripts were present only in the last three pairs of somites close to the tail bud (Figs. 3i and 3l). The zebrafish *myf*-5 gene was expressed during segmentation. Scarcely any zebrafish *myf*-5 transcripts were detected in the tail bud after 26 hpf (Fig. 3j). Signals for zebrafish *myf*-5 transcripts were found in the somites, the lateral presomitic cells, and adaxial cells (Fig. 3k). The prominent bands, which occurred in adaxial cells in two parallel rows, did not extend beyond the positive-signal somites. We found that zebrafish *myf*-5 transcripts were expressed in a somite- and stage-specific manner.

Molecular Structure of the Zebrafish *myf*-5 Gene

To elucidate the mechanism underlying the regulation of zebrafish *myf*-5 expression, we isolated the upstream, regulatory sequence of zebrafish *myf*-5 from the zebrafish genomic library. Approximately 2×10^5 plaques were screened and eight were purified. Southern blot analysis revealed that one of recombinant phages contained a ≈14 kb insert in which a ≈6.6 kb *Sac*I fragment encompassed the 5′-flanking region, and a ≈2.4 kb *Sac*I fragment encompassed 3 exons and 2 introns of the partial zebrafish *myf*-5 genomic sequence (Figs. 4 and 5). The boundary sequences of the introns all followed to GT-AG rule.

The primer extension experiment showed that zebrafish *myf*-5 transcription started at the G located at position -48 relative to translation start site (data not shown). The first nucleotide (nt) obtained from the 5′-RACE product was the 6th nt of the sequence obtained from primer extension.

Expression Patterns of EGFP cDNA Driven by Zebrafish *myf*-5 Upstream Regions in Transgenic Embryos

An exogenous DNA (pZMYP-759Et) concentration of 25 ng/μl was chosen for microinjection because, unlike DNA concentrations of 10 and 100 ng/μl, it was sufficient for transgenic EGFP expression but did not cause lethal effects. To mimic the endogenous zebrafish *myf*-5 expression pattern, EGFP cDNA fused with -3 (pZMYP-2937E) and -6.3 (pZMYP-6212E) kb of zebrafish *myf*-5 upstream sequences were microinjected into 85 and 55 fertilized eggs, respectively. The EGFP expression patterns of these two constructs in transgenic fish were identical, except for how brightly each fluoresced. The initiation of EGFP expression occurred at the 90% epiboly stage (about 9 hpf; Fig. 6c) at the edge of a shield in a faint, but well-defined spot. During the early segmental stage (about 10–16 hpf), EGFP signals were observed in the somites and segmental plate (Fig. 6d), but the EGFP expression pattern began to change from spot to bar shaped. During the pharyngula stage (24–48 hpf), EGFP was expressed in the somites and the peripheral layer of

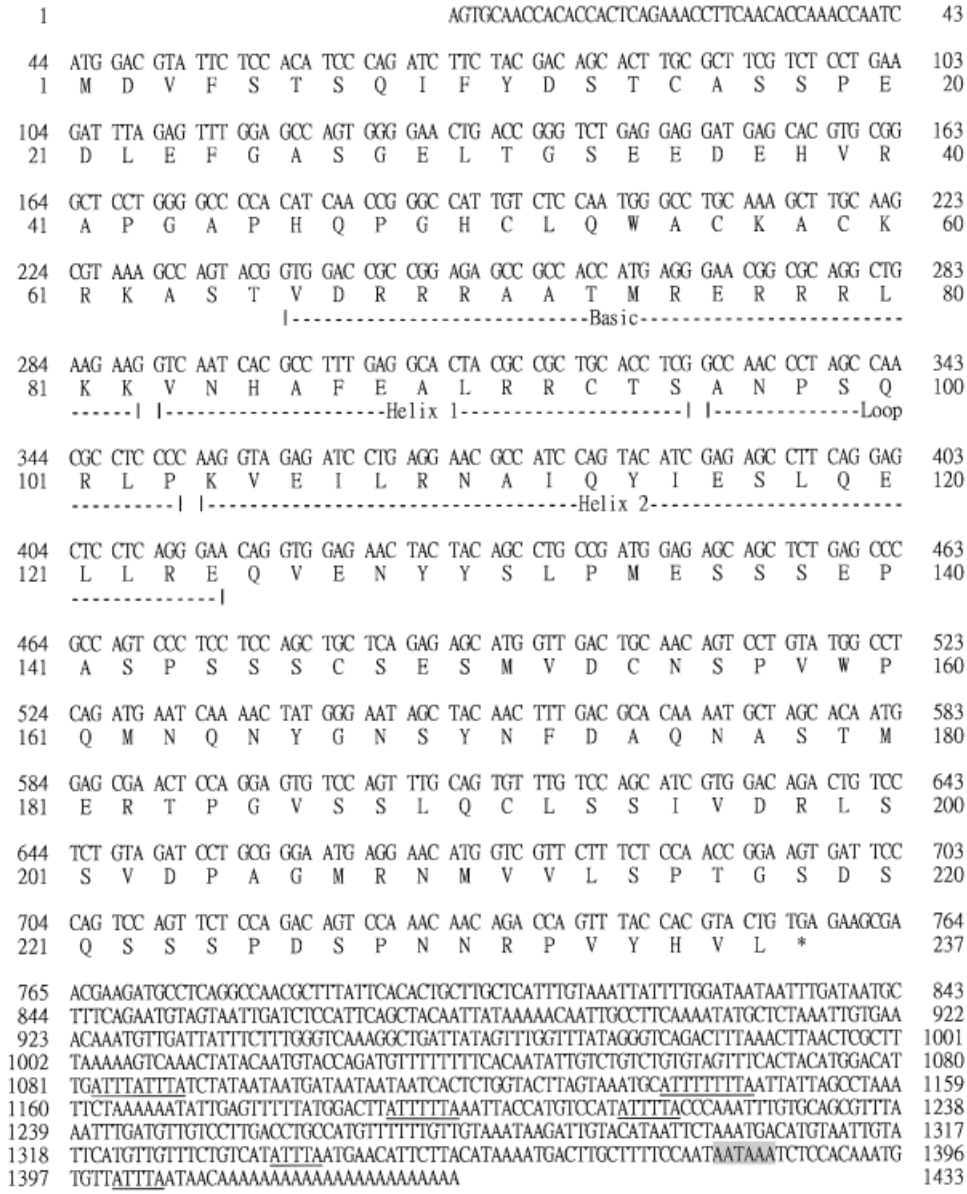

FIG. 1. Nucleotide and deduced amino acid sequence of zebrafish *myf*-5 cDNA. The nucleotides were numbered beginning with the first nucleotide of 5′-RACE product (+1). Numbers on the second line of each row indicate the amino acid sequences, and the notation on the third line indicates the structural motif. The polyadenylation signal (AATAAA) is shown in a gray box, and the stop codon is marked with an asterisk. The "destabilizer" sequence motifs are underlined. This nucleotide sequence is in the GenBank database under the accession number AF270789.

the neural tube (Figs. 6e, g), and the expression pattern began to change from bar shaped to linear. We also observed that the EGFP fluorescent signals could translocate and changed their shapes in the somites of transgenic embryos. These linear signals lasted until the hatching and early larval period (48–72 hpf).

Minimal Tissue-Specific Regulatory Sequence

To identify the *cis*-acting elements in the proximal region of zebrafish *myf*-5 promoter, we also isolated seven deletion clones that encompassed different lengths of the upstream regulatory sequence. Deletion clones of pZMYP-22E, -62E, -82E, -154E, -290E, -526E, and -702E were linearized with *Stu*I, and each was microinjected at a concentration of 25 ng/μl. Embryos injected with DNA fragments containing the -82, -154, -290, -526, and -702 bp upstream sequences (pZMYP-82E, -154E, -290E, -526E, and –702E, respectively) displayed GFP-positive signals in the somites at 10 hpf. As the total expression rates, the somite-specific rates of above groups were high (from 47% to 96.6% of survival embryos), and the nonspecific expression rates were extremely low (from 0% to 3.4%) (Fig. 7). The -82 bp upstream sequence was capable of driving somite-specific expression of zebrafish *myf*-5. On the other hand, embryos injected with DNA fragments containing less than 82 bp (pZMYP-62E and -22E) did not give off GFP-positive signals, even when the concentration of injected DNA was increased to 50 ng/μl. Based on this evidence, we suggest that the proximal sequence from

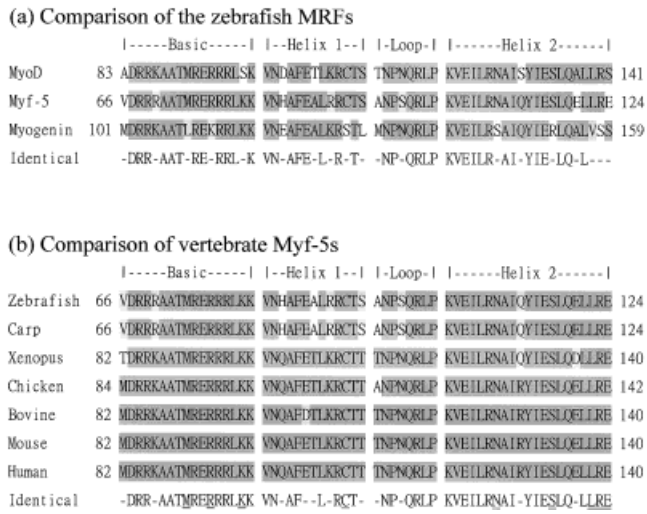

FIG. 2. Comparison of the basic helix-loop-helix domains of zebrafish MRFs, and the Myf-5 of other vertebrates. Dark and light gray boxes represent identical and similar amino acids with consensus sequence, respectively. Residues conserved in Myf-5, but not in other MRFs, are underlined. Data were obtained from GenBank: zebrafish MyoD (Z36945), zebrafish myogenin (AF202639), carp (AB012883), *Xenopus* (X56738), chicken (X63250), bovine (M95684), mouse (X56182), and human (X14894).

nt -82 to -1 of zebrafish *myf*-5 is a minimal *cis*-element for controlling specific expression. Interestingly, we observed that the proportion of embryos that not only had EGFP expression in somites but also appeared in the translocation of green signals declined sharply for pZMYP-154E- and pZMYP-82E-injected groups (Fig. 7).

DISCUSSION

Polypeptide Structure Analysis

In the MRF family of proteins, the bHLH region is the most important functional domain for mediating DNA binding and protein-protein interaction. Zebrafish bHLH domains, MyoD, myogenin, and Myf-5 are highly conserved. In addition, the bHLH domain of zebrafish Myf-5 is exactly identical to that of carp Myf-5 and exhibits 80% or greater identity with the Myf-5 of other vertebrates such as *Xenopus*, chickens, bovine, mice, and humans. However, the entire zebrafish Myf-5 polypeptide sequence exhibited only 54.3% and 40.8% identity with those of zebrafish MyoD (Weinberg *et al.*, 1996) and zebrafish myogenin (Chen *et al.*, 2000), respectively. Zebrafish Myf-5 exhibited 92.0, 62.7, 63.3, 61.0, 62.3, and 61.9% amino acid identity, respectively, with the Myf-5 of carp (Kobiyama *et al.*, 1998), *Xenopus* (Hopwood *et al.*, 1991), chickens (Saitoh *et al.*, 1993), bovine (Barth *et al.*, 1993), mice (Buonanno *et al.*, 1992), and humans (Braun *et al.*, 1989b). Thus, the bHLH domain is conserved and unique among MRF proteins, but biological function is an attribute of the entire polypeptide. This is consistent with domain-swapping experiments in MRFs (Schwarz *et al.*, 1992; Weintraub *et al.*, 1992; Winter *et al.*, 1992). As expected, the identity of the *myf*-5 polypeptide sequence is relatively higher among fish than other classes of vertebrates.

Spatial Distribution of Zebrafish *myf*-5 Transcripts

Although somite- and stage-specific transcription patterns of mouse *myf*-5 have been reported (Ott *et al.*, 1991; Lyons and Buckingham, 1992; Faerman and Shani, 1993), the transcriptional pattern and tissue distribution of *myf*-5 during early embryonic development of the zebrafish is still unclear. In this study, we used whole-mount in situ hybridization with a probe that was specific to zebrafish *myf*-5. Results showed that this probe did not cross-hybridize with zebrafish *myo*D. Zebrafish *myf*-5 transcripts were first detected in paraxial mesoderm at the 70% epiboly stage (7.5 hpf). At the 90% epiboly stage (9 hpf), increased expression of zebrafish *myf*-5 transcripts occurred at the inner boundaries of the axial mesoderm. From 10 to 26 hpf, the signals were observed in adaxial cells, the lateral presomitic cells, and newly formed somites. Furthermore, the numbers of zebrafish *myf*-5 expressed in the somites increased from 10 to 16 hpf and then gradually decreased, coincident with somite maturation, from 16 to 26 hpf.

Adaxial cells, which develop into slow muscle cells (Devoto *et al.*, 1996), are the most medial cells in the segmental plate. They were the only cells expressing abundant levels of *myo*D and other muscle-specific genes, including *myogenin* and *tropomyosin* (Devoto *et al.*, 1996; Weinberg *et al.*, 1996). In constrast, the zebrafish *myf*-5 gene was only expressed transiently in adaxial cells after segmentation, and zebrafish *myf*-5 transcripts disappeared in older adaxial cells. Compared to the very abundant expression of zebrafish *myo*D during somite formation, expression of zebrafish *myf*-5 is relatively low. In summary, the expression patterns of zebrafish *myf*-5 transcripts differ from those of zebrafish *myo*D (as reported by Weinberg *et al.*, 1996) in several important ways. First, expression of zebrafish *myf*-5 in adaxial cells was low, while zebrafish *myo*D transcripts were abundant. Second, in somites, zebrafish *myf*-5 expression was transient whereas zebrafish *myo*D transcripts were expressed until at least 60 hpf. Third, zebrafish *myf*-5 transcripts were expressed in newly formed somites only, but zebrafish *myo*D was expressed in all somites. Fourth, zebrafish *myf*-5 transcripts were highly expressed in the lateral presomitic cells, where zebrafish *myo*D was not expressed at all.

Mice Myf-5 and MyoD have functionally redundant roles in myogenesis (Rudnicki *et al.*, 1992), a hypothesis supported by findings from Myf-5 (-/-); MyoD (-/-) mutant mice (Rudnicki *et al.*, 1993). However, skeletal muscle development of mice lacking Myf-5 is apparently delayed. We believe that Myf-5 and MyoD each have a unique function in the skeletal muscle differentiation program to recruit different populations of cells into myoblast linkage. The low and transient expression of *myf*-5 in stage- and somite-specific patterns is highly

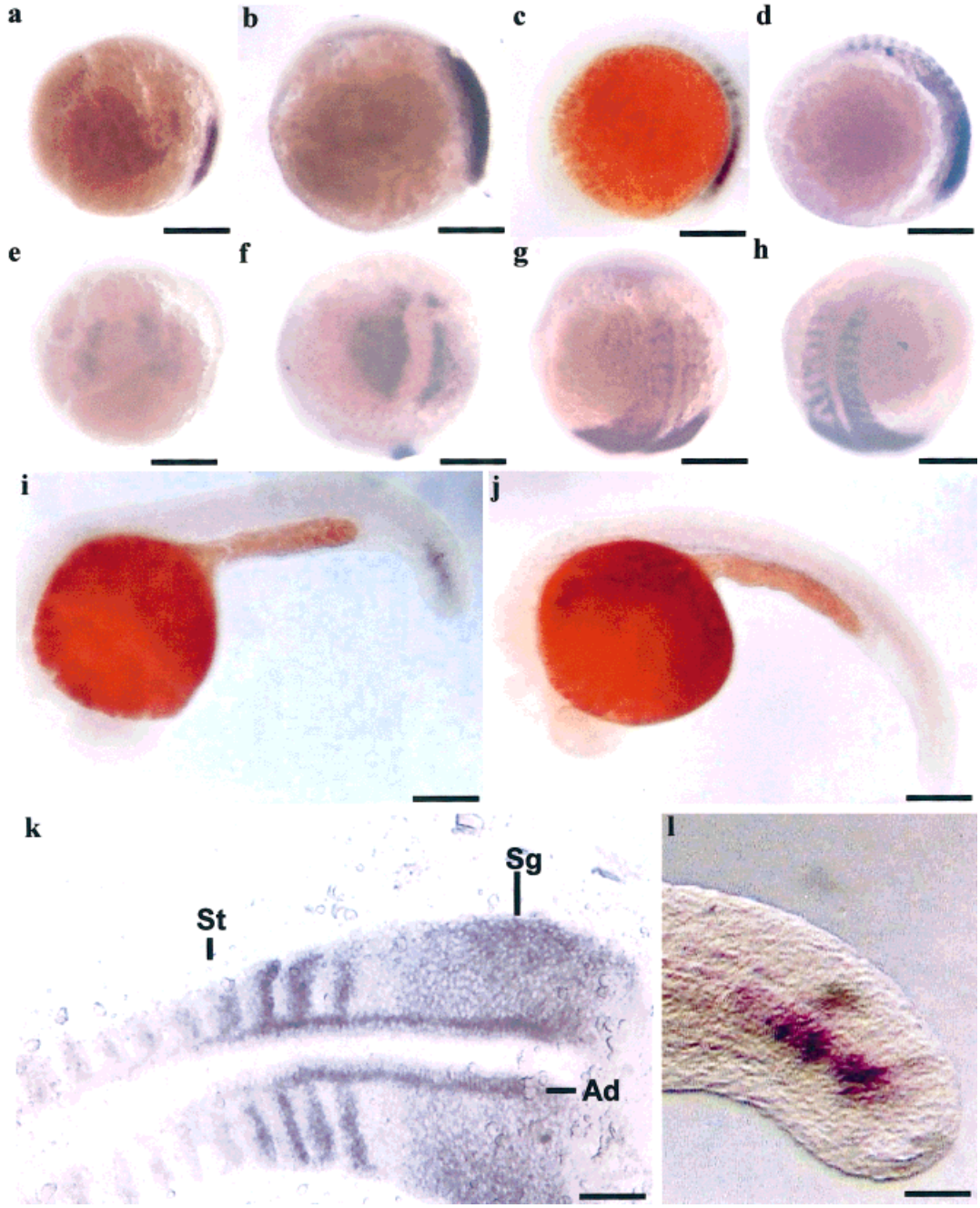

FIG. 3. Expression patterns of zebrafish *myf*-5 as determined by whole-mount in situ hybridization. Side (**a–d, i–j**) and dorsal views (**e–h**) of embryos at 7.5- (**a** and **e**), 9- (**b** and **f**), 14- (**c** and **g**), 16- (**d** and **h**), 24- (**i**), 26- (**j**) hpf. Higher magnification of positive signals in 16- (**k**) and 24- (**l**) hpf embryos. Zebrafish *myf*-5 transcripts were first detected in the adaxial mesoderm of embryos 7.5 hpf (**a** and **e**), accumulated in the inner adaxial cells by 9 hpf (**b** and **f**), and also appeared in the segmental plates (Sg), adaxial cells (Ad), and somites (St) in 14 and 16 hpf embryos (**c**, **d**, **g** and **h**). Zebrafish *myf*-5 transcripts gradually decreased from 16 to 24 hpf (**i** and **l**) and hardly any zebrafish *myf*-5 transcripts were detected 26 hpf (**j**). Scale bars: **a–j**, 250 μm; **k**, 100 μm; **l**, 50 μm.

likely in controlling the delicate development and commitment of muscle fiber precursors during zebrafish embryogenesis. It is worthwhile to notice that *myo*D is not expressed in the lateral presomitic cells of the segmental plate, suggesting that lateral presomitic cells may downregulate *myo*D prior to somite formation.

Temporal Expression of Zebrafish *myf*-5 Transcripts

In transgenic experiments, EGFP fluorescent signals were first detectable at the 90% epiboly stage (about 9 hpf) and were still present at 72 hpf. This differed

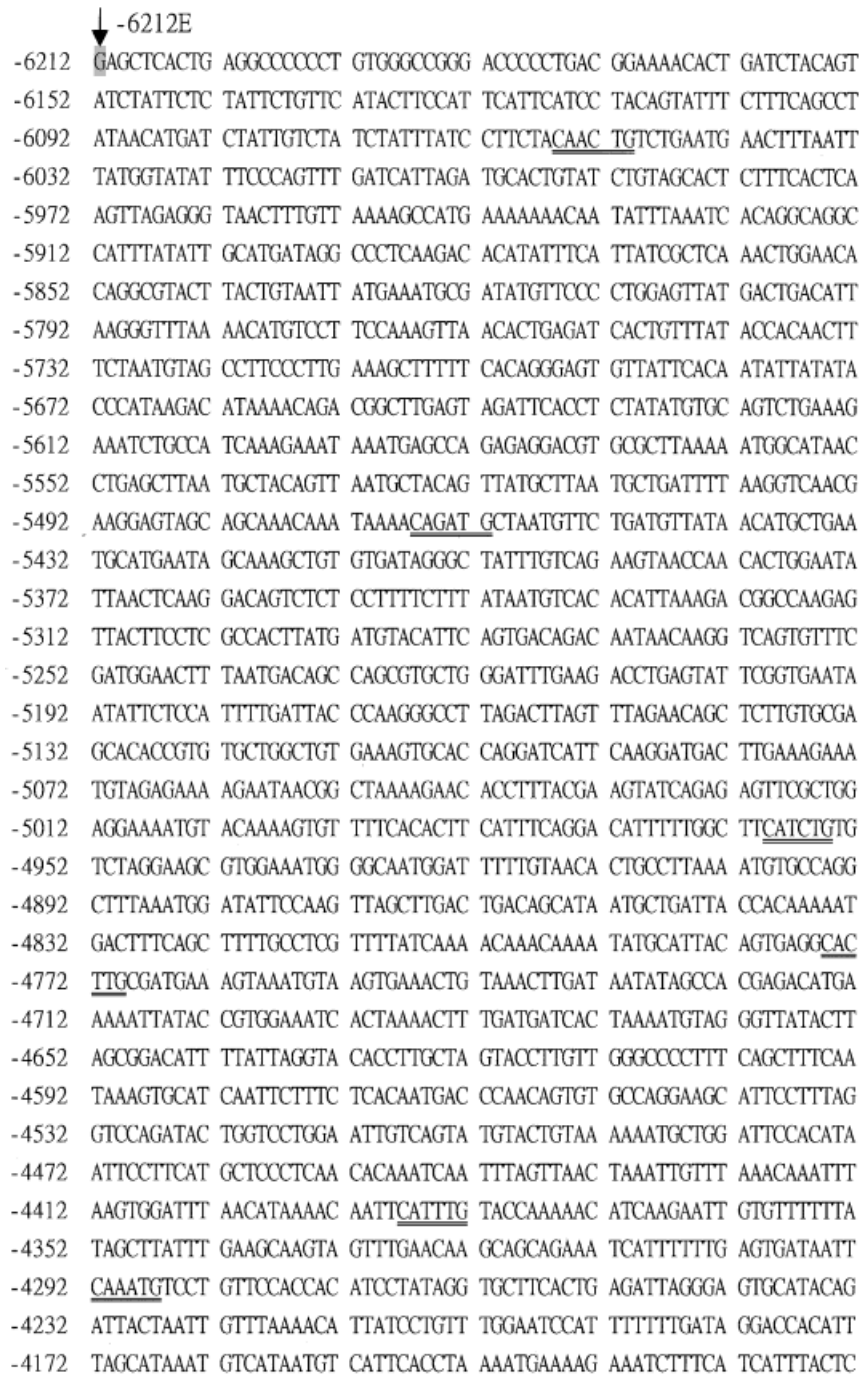

FIG. 4. Upstream sequence (-6212 to +88 bp) of the zebrafish *myf*-5 gene. Numbers indicate the nucleotide position relative to transcription start site (+1). The transcription start site and the translation start codon (ATG) are shown in the dark gray box. The putative TATA box (TATAAAT), CAAT box (GC(T/C)CAATCT), GC box (GGGCGG), E box (CANNTG), and MEF2 binding site (C/T)TA(T/A)$_4$TA(A/G)) are underlined with a single line, dashes, dots, double lines, and a solid line, respectively. The first nucleotide of each unidirectional deletion clone is indicated by an arrow, and a light gray box bearing the name of each clone. This nucleotide sequence is in the GenBank database under accession number AF184166.

slightly from in situ hybridization, in which zebrafish *myf*-5 transcripts started at 7.5 hpf (70% epiboly) but ended at 26 hpf. This discrepancy may have resulted from the high stability of EGFP in transgenic embryos. Fluorescence continued for 2 days, long after the EGFP gene was turned off. We replaced EGFP with d2EGFP (Clontech), which has a higher turnover rate (2 h). However, the expression pattern of embryos microin-

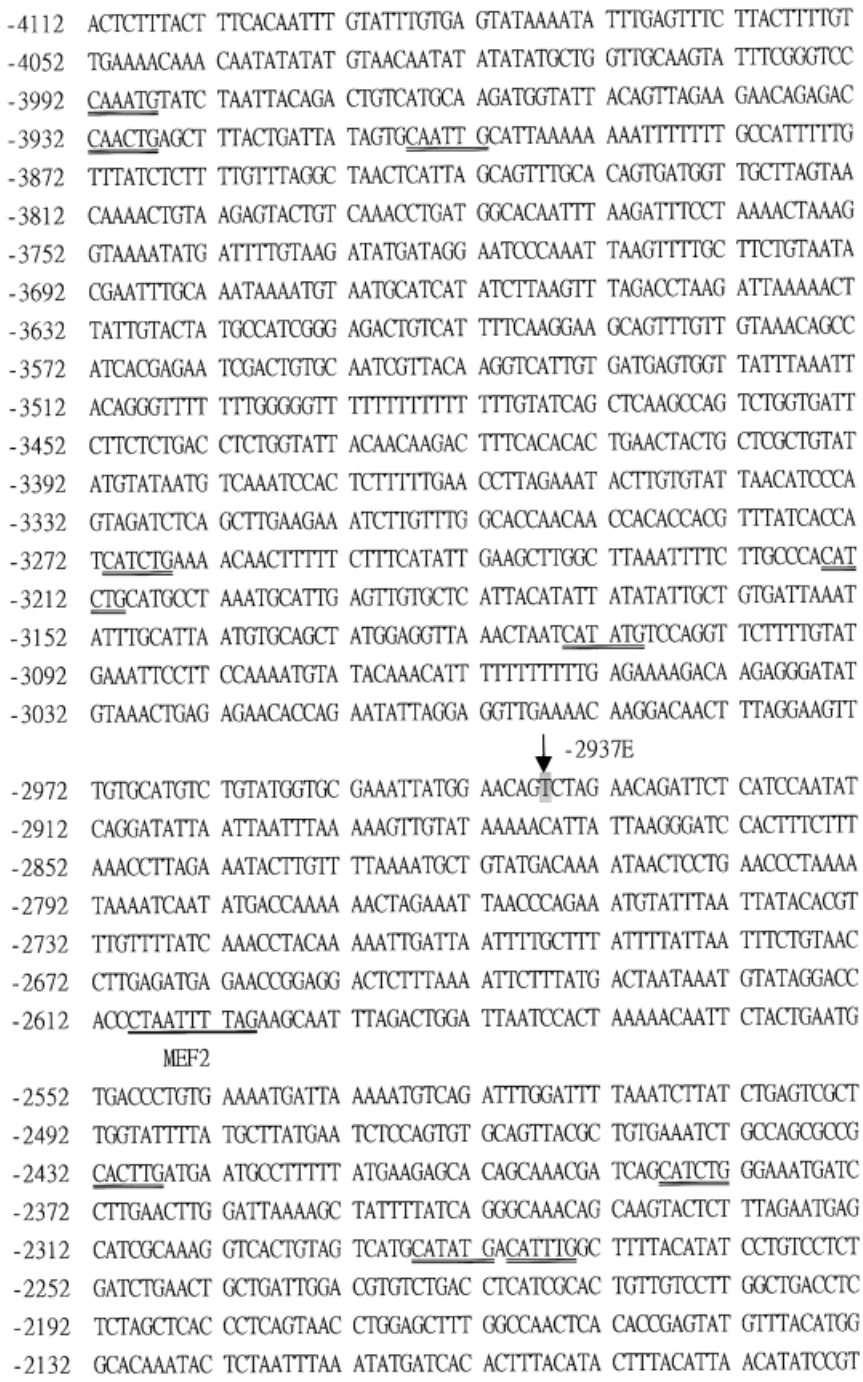

FIG. 4 (Continued)

jected with -3 kb of zebrafish *myf*-5 fused with d2EGFP cDNA was same as that of embryos injected with zebrafish *myf*-5 fused with EGFP cDNA, except that the brightness of fluorescence was reduced and the time of the first signal was delayed to 12 hpf. Detection sensitivity may be another reason for these inconsistent results because zebrafish *myf*-5 could be detected from embryos at 7.5 to 32 hpf using Reverse Transcription-Polymerase Chain Reaction (RT-PCR) (data not shown). In conclusion, we suggest that the temporal expression of zebrafish *myf*-5 transcripts starts at 7.5 hpf, increases substantially up to 16 hpf, and then gradually decreases to undetectable levels after 26 hpf. These results support that zebrafish *myf*-5 polypeptide plays an important role in somitogenesis.

Recently, Hadchouel *et al*. (2000) reported that all essential regulatory elements for controlling mouse

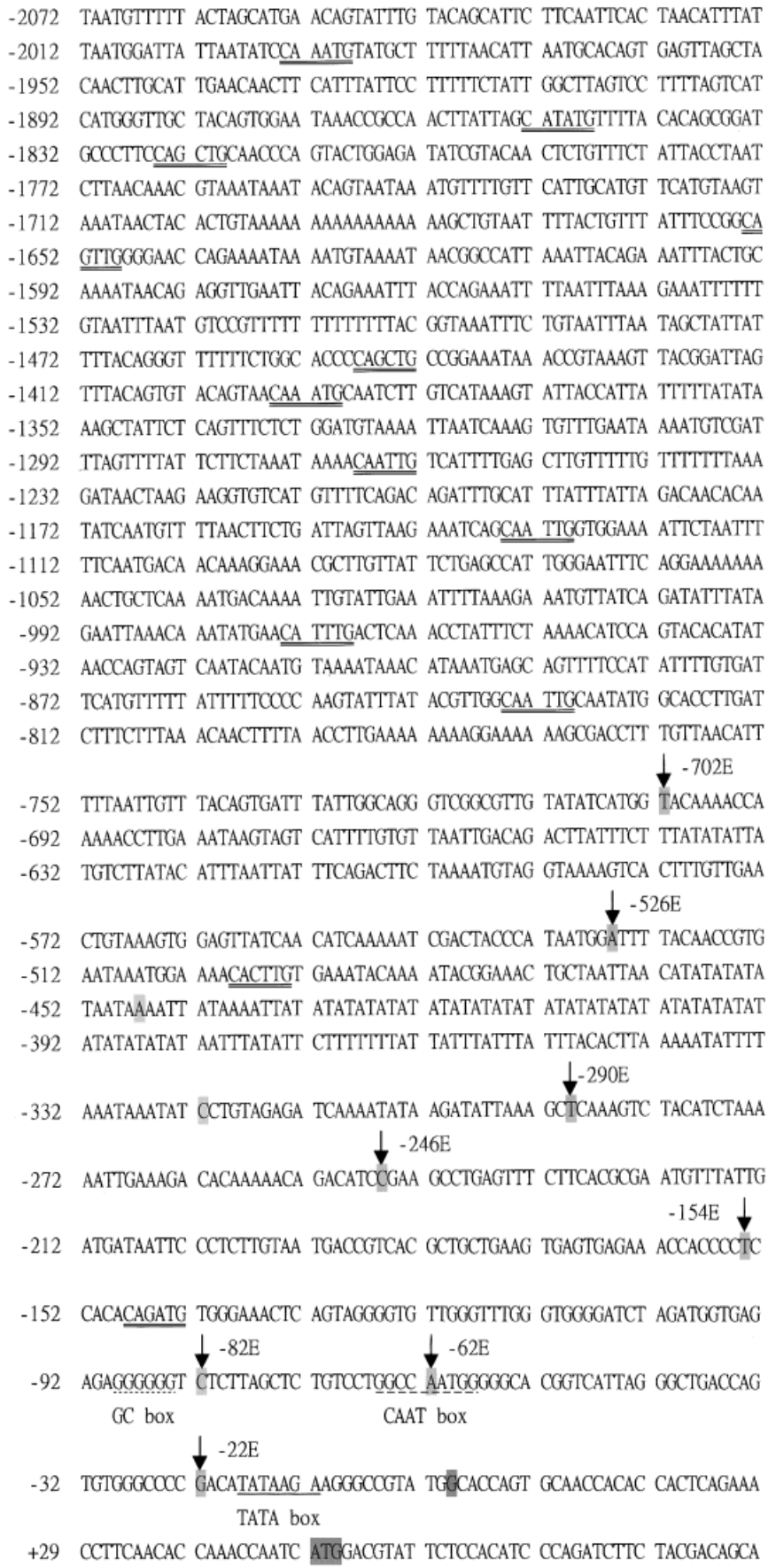

FIG. 4 (Continued)

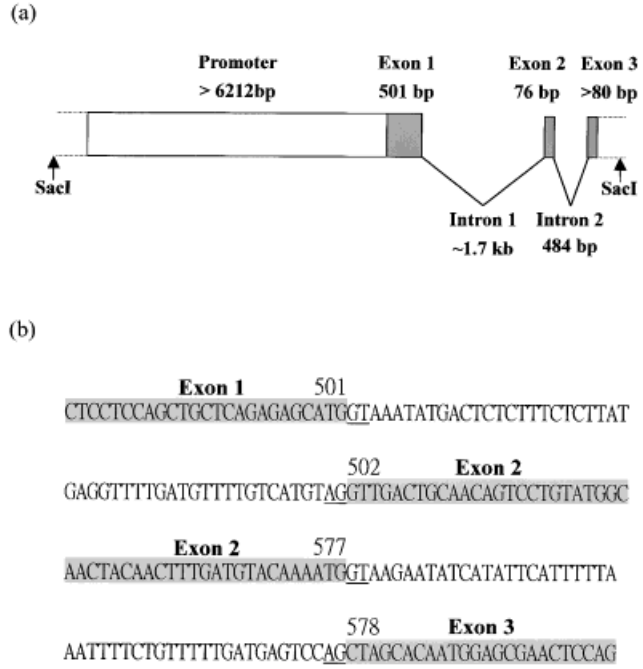

FIG. 5. Partial genomic structure of the zebrafish *myf*-5 gene. (**a**) Schematic illustration of the partial genomic structure of zebrafish *myf*-5 isolated from a recombinant phage clone. (**b**) Boundary sequences of introns 1 and 2 of zebrafish *myf*-5. Exons are shown in gray. The GT and AG sequences at exon boundarys are underlined. Numbers indicate the corresponding nucleotide positions of the zebrafish *myf*-5 cDNA.

myf-5 expression were located within -96 kb upstream segment, and a region located between -58 and -48 kb could direct *myf*-5 expression in brain. Daubas *et al*. (2000) also demonstrated that mouse *myf*-5 gene was an early axonal marker in brain, besides that *myf*-5 gene was expressed in somites. In this study, we showed that EGFP driven by -3 kb upstream regulatory region of zebrafish *myf*-5 (pZMYP-2937E construct) appeared green fluorescence in the peripheral layer of the neural tube during the pharyngula stage (24 to 48 hpf). Whether the proximal upstream regulatory region within -3 kb of zebrafish *myf*-5 gene exists an element for controlling neuron-specific expression remains to be studied.

Sequential Expression Among MRFs

Myogenic bHLH genes are transcribed sequentially during the skeletal muscle development program. Although zebrafish *myf*-5 and zebrafish *myo*D signal intensities differed, based on RT-PCR and in situ hybridization, the stages in which zebrafish *myf*-5 and zebrafish *myo*D are first expressed might be quite close. However, zebrafish *myf*-5 transcripts were detectable in paraxial mesoderm and the segmental plate, whereas zebrafish *myo*D transcripts were undetectable. Therefore, we suggest the expression sequence of MRFs during somitogenesis in zebrafish embryos was Myf-5, MyoD, and myogenin, which is consistent with their expression sequence in carp (Kobiyama *et al.*, 1998), mice (Sassoon *et al.*, 1989), and humans (Braun *et al.*, 1989b).

Cis-Element(s) Controlling the Specific Expression of the Zebrafish *myf*-5 Promoter

Despite that mosacism might occur in F0 embryos, the foreign gene driven by zebrafish promoter can be faithfully expressed in the transgenic zebrafish (Ju *et al.*, 1999; Xu *et al.*, 2000; Yoshizaki *et al.*, 2000). We used transgenesis to analyze the promoter of zebrafish *myf*-5. Results showed that the upstream regulatory sequence from nt -82 to -1 of zebrafish *myf*-5 is sufficient for controlling somite- and stage-specific expression. Using the Transcription Factors Search program and Palindrome Finding analysis (Vector NTI), an overlapping palindrome sequence, TGGCCA (-67 to -62), a SP1 site (-58 to -49) adjacent to the CCAAT box (-66 to -58), and a GGGCCC motif (-29 to -24) that precedes the TATA box (-18 to -12) were found within the – 82-bp regulatory region (Fig. 8). Tapscott *et al.* (1992) reported two regions important for regulating the muscle-specific transcription of mouse *myo*D. They are (1) a proximal regulatory region that includes a consensus SP1 binding site, a CCAAT box, and an ATAAATA sequence, adjacent to the transcription start site, and (2) a distal regulatory region that lies approximately -5 kb from the transcription start site. Our findings are consistent with the results from the mouse *myo*D proximal regulatory region. However, the upstream, -62-bp sequence of zebrafish *myf*-5 is not sufficient for controlling the specific expression of zebrafish *myf*-5. This observation gives rise to the possibility that the -82 to -62 upstream regulatory sequence may contain *cis*-element(s) required for the specificity of zebrafish *myf*-5 expression. The conserved CCAAT box, which is important for the specificity of mouse *myo*D expression (Tapscott *et al*, 1992), is not included in the pZMYP-62E construct. As a result, it cannot control zebrafish *myf*-5 specificity. Therefore, we suggest that the palindrome sequences and CCAAT-binding proteins may coordinate with each other to control the specific expression of zebrafish *myf*-5.

Interestingly, some transgenic EGFP signals shown in the somite also have a tendency for their green fluorescent signals to translocate and form a linear shape during the pharyngula stage. The proportions of embryos for which EGFP signals appeared both somite-specificity and translocation were high for pZMYP-702E-, -526E-, and -290E-injected groups. However, these translocation rates declined sharply for pZMYP-154E- and -82E-injected embryos. Therefore, it is highly possible that a *cis*-regulatory element located between nt -154 to -290 of zebrafish *myf*-5 gene may be involved in controlling the translocation ability of progenitor muscle cells.

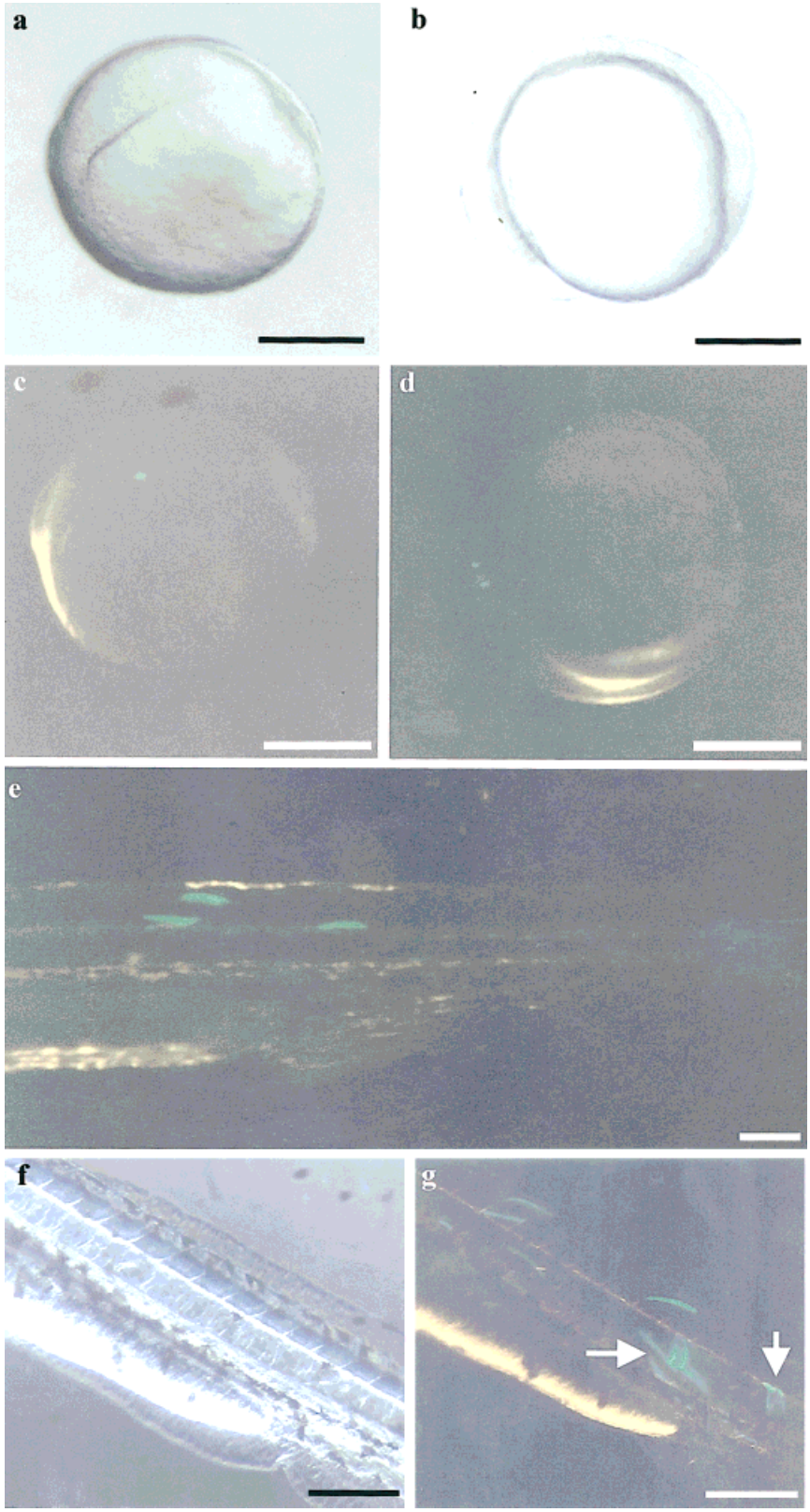

FIG. 6. EGFP expression patterns in transgenic embryos. Fertilized eggs were microinjected with pZMYP-2937E, in which the EGFP cDNA was controlled by a 3-kb, upstream region of zebrafish *myf*-5. Embryos were photographed under bright field illumination (**a**, **b**, and **f**) and in blue light (**c**, **d**, **e**, and **g**). EGFP signals were first expressed as a sharp spot at the edge of the shield during the 90% epiboly stage (**a** and **c**). During the early segmental stage, EGFP signals appeared in the somites and segmental plate (**b** and **d**) and changed in shape from spots to bars. During the pharyngula stage, the EGFP signals changed from bar shaped to linear (**e**) and remained linear through hatching and the early larval period (**f** and **g**). Arrows indicate EGFP signals displaying on the periphery layer of the neural tube. Scale bars: **a**–**d**, 250 μm; **e**–**g**, 100 μm.

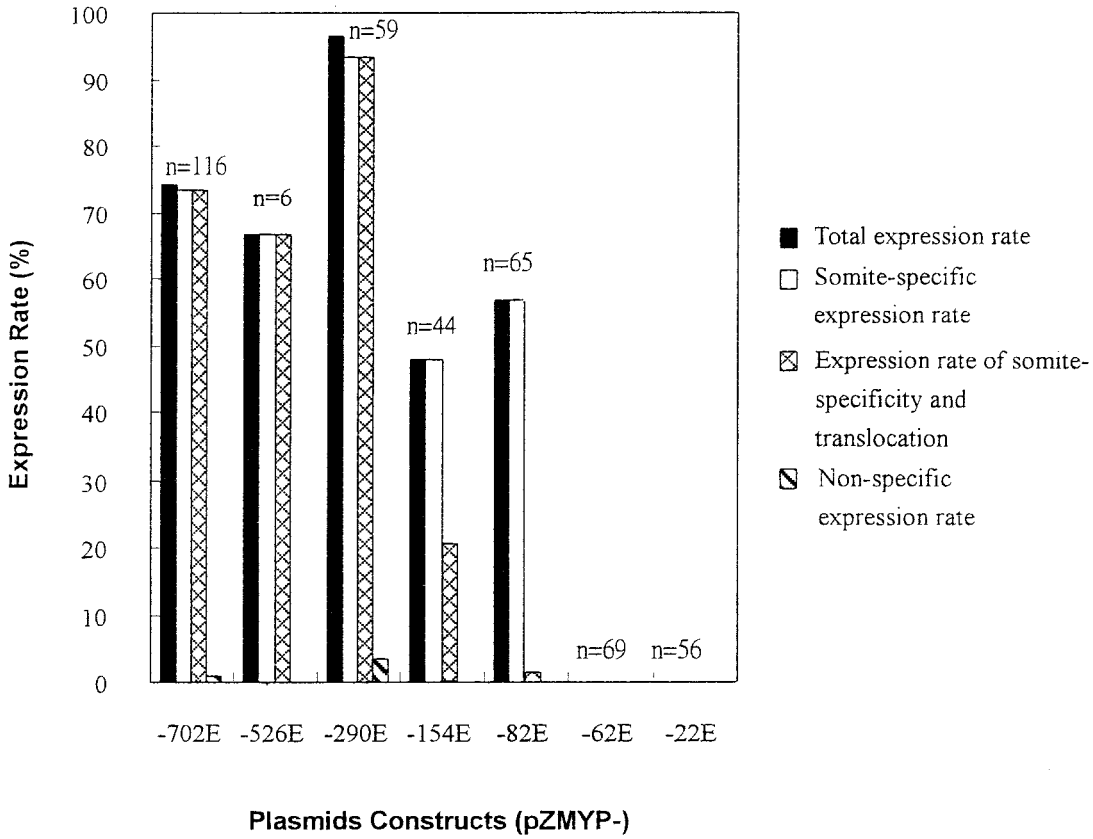

FIG. 7. Expression rates of EGFP signals in transgenic embryos. DNA fragments containing different lengths of the region upstream of zebrafish *myf*-5 were microinjected into the fertilized eggs of zebrafish at a concentration of 25 ng/μl. pZMYP-702E, -526E, -290E, -154E, -82E, -62E, and -22E represent lengths of the upstream region extending from -1 to -702, -526, -290, -154, -82, -62, and -22 bp upstream. Calculations of total expression rate (solid bars), somite-specific expression rate (empty bars), expression rate of somite-specificity and translocation (crossed bars), and nonspecific rate (hatched bars) were described in Materials and Methods. The total survival embryos after microinjection of various constructs were listed on the top of each bar (n).

MATERIALS AND METHODS

Experimental Fish

Zebrafish AB strain were kept under a 14-h light and 10-h dark photoperiod at approximately 28.5°C. After fertilization, the eggs were collected and cultured in an aquarium. Embryonic cleavage number and somite formation were observed with a light microscope to determine the developmental stages based on hpf (Kimmel *et al.*, 1995).

RNA Isolation

Embryos that had developed for 10 to 96 hpf were pooled and immediately stored in liquid nitrogen. These frozen whole embryos were homogenized with TRIzol reagent (Gibco BRL), and their RNAs were extracted according to the manufacturer's instructions (Gibco BRL).

RT-PCR

First-strand cDNA was synthesized using the SuperScript Preamplification System (Gibco BRL). Degenerated oligonucleotide primers were designed with reference to polynucleotide sequences of *myf*-5 from known vertebrates. A forward primer, Myf5-67F (GG(T/C)CACTG(T/C)CT(G/C)(A/C)T(A/G/C)TGGGCCTGCAA), and a reverse primer, Myf5-150R (GAGGCTGTAGTATTGCTCCAC(T/C)TG(T/C)TC), were synthesized. Thirty cycles of PCR amplification were performed by Taq DNA polymerase (Viogene). Each cycle consisted of denaturation for 40 s at 94°C, 1 min of annealing at 58°C, and 1 min of extension at 72°C. The last extension step was extended for 10 min at 72°C. Amplified DNA fragments were ligated with pGEM T-Easy vector (Promega) and transformed into *Escherichia coli* DH5α. DNA sequencing of both strands was carried out using a bigdye-terminator cycle sequencing ready reaction kit (Perkin-Elmer Applied Biosystems) with a DNA sequencer (Model 310, Perkin-Elmer Applied Biosystems).

RACE

First-strand cDNA used for 5′ RACE was performed as described above, then homotailed at the 5′-end using terminal transferase TdT (Boehringer Mannheim) and

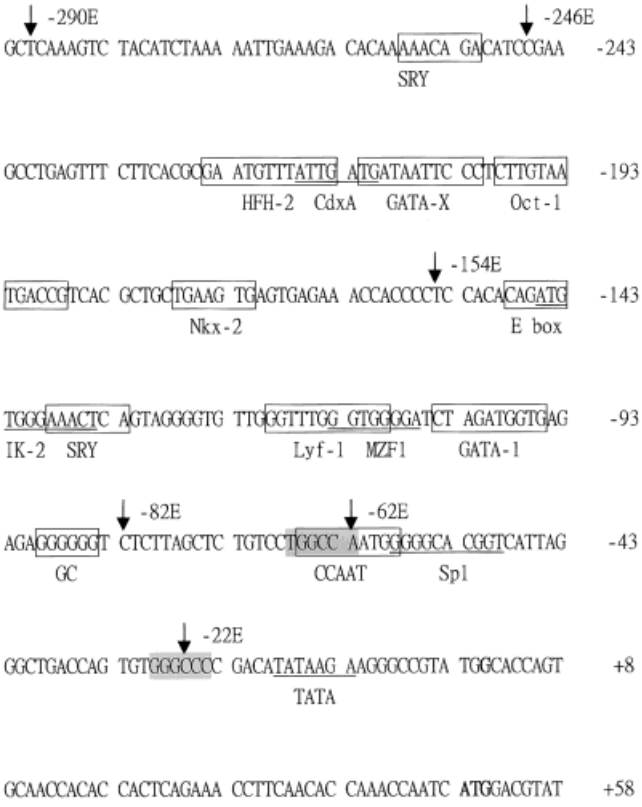

FIG. 8. Putative transcription factor binding sites in the 0.3-kb proximal regulatory region of zebrafish *myf*-5. Numbers indicate the nucleotide position relative to transcription start site (+1). The transcription start site and the translation start codon (ATG) are lettered in boldface. Putative, transcription factor binding sites are indicated. Arrows represent the first nucleotide of each deletion clone. Palindrome sequences are shown in gray boxes.

dGTP. The resulting tailed cDNA was then used to generate double-stranded cDNA by PCR amplification in the presence of a forward primer, RAAPC (GGCCACGCGTCGACTAGTACT$(C)_9$), and a reverse primer, Myf5-150R. PCR amplification was carried out as described above with the exception of annealing at 58°C. The procedure for 3′RACE was the same as that for 5′ RACE, except that (1) first-strand cDNA was synthesized by a reverse primer, RAAPT (GGCCACGCGTCGACTAGTAC$(T)_{18}$), (2) a forward primer, Myf5-67F, and a reverse primer, RAUAP (GGCCACGCGTCGACTAGTAC), were used to generate double-stranded cDNA, and the annealing temperature was 50°C, and (3) 1/50 volume of the first PCR product was used as a template for nested PCR amplification in the presence of a forward primer, Myf5-96F (GCCAACCCTAGCCAACGCCTC), and a reverse primer, RAUAP, and the annealing temperature was 56°C. Amplified DNA fragments were subcloned and sequenced as described above.

Whole-Mount In Situ Hybridization

About 5 μg of plasmids containing the insert fragment, which corresponded to nt at positions 188–439 of the zebrafish *myf*-5 cDNA, were used as template and labeled with DIG RNA Labeling Kit (Boehringer Mannheim). Normal embryos from 3 to 30 hpf were collected and then fixed with 4% paraformaldehyde for 4 h at 25°C. After fixation, embryos were dechrionized, dehydrated in absolute alcohol, and stored at -20°C. Embryos were rehydrated by immersion for 10 min each in 75%, 50%, and 25% ethanol. Rehydrated embryos were transferred to phosphate buffer saline (PBS). Prehybridization, hybridization, and detection procedures were performed with a DIG detection kit II according to the instructions in the supplier's manual (Boehriner Mannheim). All processed embryos were observed under a fluorescent stereomicroscope MZ FLIII (Leica).

Screening and Subcloning of the *zmyf*-5 Genomic Sequence

A zebrafish genomic library (5×10^8 pfu/ml) was purchased from Clonetch. Phage plating and titration were performed according to the instructions in the supplier's manuals. The DNA fragment corresponding to nt 188–439 of the zebrafish *myf*-5 cDNA was labeled with DIG. After hybridization, the positive clones were visualized with autoradiography by exposing the substrate, CDP-STAR (Tropix), for 30 min. Then, a 6.6 kb *Sac*I-digested DNA fragment from the recombinant phage, which was positive in Southern hybridization, was subcloned to pGEM-3Zf(+) (Promega). The resultant pZMYP6.6 was digested with *Xba*I to remove the 3.3kb upstream sequence to generate the plasmid pZMYP3.3.

Primer Extension

Primer, Myf5(+1)R (GGATCCGATTGGTTTGGTGTTG) was synthesized, end-labeled with [γ-^{32}P]-ATP (Amersham), and stored at -20°C. Approximately 2 ng of mRNA extracted 6 to 16 hpf were mixed with 1 μl probe (10^3cpm/μl) and hybridized for 4 h at 55°C, after which DTT, dNTP, first-strand buffer and Superscript II (GIBCO BRL) were added. The reaction was carried out at 42°C for 1 h. All samples were separated after acrylamide gel electrophoresis (containing 8M urea and 6% acrylamide), dried, and autoradiograped.

Generation of Unidirectional Deletion Clones

A forward primer, Myf5(-750)F (AGATCTGTGAACATTTTTAATTGTA), and a reverse primer, Myf5(+1)R, were used during PCR amplification in order to introduce an additional *Bam*HI site into the upstream sequence. A PCR product with a molecular mass of 809 bp was ligated to pEGFP-1 vector (Clontech) producing pZMYP759Et. Then, a 2.9-kb *Hin*dIII-*Apa*I fragment from pZMYP3.3 was cloned to pZMYP-759Et to construct pZMYP-2937E and a 3.3 kb *Xba*I fragment from pZMYP6.6 was introduced to pZMY-2937E to construct pZMYP-6212E.

pZMY-2937E, containing the upstream, ≈3kb region of zebrafish *myf*-5 and EGFP cDNA, was purified via Plasmid Miniprep Kit (Viogene). Various deletions of the

zebrafish *myf*-5 upstream region were obtained using the Double-Stranded Nested Deletion Kit (Pharmacia Biotech). Salt and DNA concentrations were adjusted to 50 mM and 100 ng, respectively, in 1× exonuclease III buffer. The reaction was carried out at 33°C with 2-min intervals. After agarose gel electrophoresis, deletions of different sizes were selected and their sequences were confirmed using a primer specific for pEGFP-1, (CCCTGATTCTGTGGATAACCGTA).

Microinjection and Green Fluorescence Detection

EGFP expression vectors fused with different lengths of the zebrafish *myf*-5 upstream sequence were constructed and linearized with suitable restriction enzymes. The linearized DNA was recovered from 0.8% low melting point agarose gel (FMC) with the Gel Extraction System (Viogene), quantified by its intensity on the agarose gel, and its absorbance measured with GeneQuant II (Pharmacia Biotech).

Linearized plasmid DNA was diluted to appropriate concentrations with 0.1% phenol red and 1×PBS. Fertilized eggs that had not undergone the first cleavage were collected and injected with about 2.3 nl of DNA solution. Cytoplasmic microinjection was performed as described by Stuart *et al.* (1988). After injection, the eggs were incubated at 28.5°C.

Transgenic embryos were observed everyday, especially from 9 to 96 hpf, under a stereo dissecting microscope (MZ12, Leica) equipped with a fluorescent module and an enhanced GFP filter cube (Kramer Scientific). Photographs were taken with a MPS 60 camera (Leica) and FUJI 400 ASA film. Expression rates were categorized into four groups: (1) total expression rate—proportion of survival embryos with EGFP expression in any cells; (2) somite-specific expression rate—proportion of survival embryos that have EGFP expression in somite exclusively; (3) somite-specificity and translocation rate—proportion of survival embryos that have EGFP expression in somite exclusively and also have translocation of green signal; and (4) nonspecific expression rate—proportion for survival embryos with somite expression that have expression in other cells.

ACKNOWLEDGMENTS

The authors wish to thank Drs. E. Chiang and B. C. Chung from the Institute of Molecular Biology, Academia Sinica, for helping with whole-mount in situ hybridization.

LITERATURE CITED

Barth JL, Worrell RA, Crawford JM, Morris J, Ivarie R. 1993. Isolation, sequence, and characterization of the bovine myogenic factor-encoding gene myf-5. Gene 127:185–191.

Braun T, Bober E, Buschhausen-Denker G, Kohtz S, Grzeschik KH, Arnold HH. 1989a. Differential expression of myogenic determination genes in muscle cells: possible autoactivation by Myf gene products. EMBO J 8:3617–3625.

Braun T, Buschhausen-Denker G, Bober E, Tannich E, Arnold HH. 1989b. A novel human muscle factor related to but distinct from MyoD1 induces myogenic conversion in 10T1/2 fibroblasts. EMBO J 8:701–709.

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

Braun T, Arnold HH. 1991. The four human muscle regulatory helix-loop-helix proteins Myf3-Myf6 exhibit similar hetero-dimerization and DNA binding properties. Nucleic Acids Res 19:5645–5651.

Buonanno A, Apone L, Morasso MI, Beers R, Brenner HR, Eftimie R. 1992. The MyoD family of myogenic factors is regulated by electrical activity: isolation and characterization of a mouse Myf-5 cDNA. Nucleic Acids Res 20:539–544.

Chen YH, Lee WC, Cheng CH, Tsai HJ. 2000. Muscle regulatory factor gene: zebrafish (*Danio rerio*) myogenin cDNA. Comp Biochem Physiol B 127:97–103.

Chevallier A, Kieny M, Mauger A. 1977. Limb-somite relationship: effect of removal of somatic mesoderm on the wing musculature. J Embryol Exp Morphol 43:263–278.

Christ B, Jacob HJ, Jacob M. 1977. Experimental analysis of the origin of the wing musculature in avian embryos. Anat Embryol (Berlin) 150:171–186.

Davis RL, Weintraub H, Lassar AB. 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51:987–1000.

Devoto SH, Melancon E, Eisen JS, Westerfield M. 1996. Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. Development 122:3371–3380.

Dubas P, Tajbakhsh S, Hadchouel J, Primig M, 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.

Edmondson DG, Olson EN. 1989. 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 3:628–640.

Faerman A, Shani M. 1993. The expression of the regulatory myosin light chain 2 gene during mouse embryogenesis. Development 118:919–929.

Hadchouel J, Tajbakhsh S, Primig M, Chang T H-T, Dubas P, Rocancourt D, Buckingham M. 2000. Modular long-range regulation of *myf*-5 reveals unexpected heterogeneity between skeletal muscles in the mouse embryo. Development 127:4455–4467.

Hopwood ND, Pluck A, Gurdon JB 1991. *Xenopus* Myf-5 marks early muscle cells and can activate muscle genes ectopically in early embryos. Development 111:551–560.

Jaynes JB, Johnson JE, Buskin JN, Gartside CL, Hauschka SD. 1988. The muscle creatine kinase gene is regulated by multiple upstream elements, including a muscle-specific enhancer. Mol Cell Biol 8:62–70.

Ju B, Xu Y, He J, Liao J, Yan T, Hew CL, Lam TJ, Gong Z. 1999. Faithful expression of green fluorescent protein (GFP) in transgenic zebrafish embryos under control of zebrafish gene promoters. Dev Genet 25:158–167.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. Dev Dyn 203: 253–310.

Kobiyama A, Nihei Y, Hirayama Y, Kikuchi K, Suetake H, Johnston IA, Watabe S. 1998. Molecular cloning and developmental expression patterns of the MyoD and MEF2 families of muscle transcription factors in the carp. J Exp Biol 201:2801–2813.

Lyons GE, Buckingham ME. 1992. Developmental regulation of myogenesis in the mouse. Semin Dev Biol 3:243–253.

Miner JH, Wold B. 1990. Herculin, a fourth member of the *MyoD* family of myogenic regulatory genes. Proc Natl Acad Sci USA 87:1089–1093.

Ott MO, Bober E, Lyons G, Arnold HH, 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.

Pownall ME, Emerson CP Jr. 1992. Sequential activation of three myo-

genic regulatory genes during somite morphogenesis in quail embryos. Dev Biol 151:67–79.

Rhodes SJ, Konieczny SF. 1989. Identification of MRF4: a new member of the muscle regulatory factor gene family. Genes Dev 3:2050–2061.

Rudnicki MA, Braun T, Hinuma S, 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 MA, Schnegelsberg PN, Stead RH, Braun T, Arnold HH, Jaenisch R. 1993. MyoD or Myf-5 is required for the formation of skletal muscle. Cell 75:1351–1359.

Rupp RA, Weintraub H. 1991. Ubiquitous MyoD transcription at the midblastula transition precedes induction-dependent MyoD expression in presumptive mesoderm of *X. laevis*. Cell 65:927–937.

Saitoh O, Fujisawa-Sehara A, Nabeshima Y, Periasamy M. 1993. Expression of myogenic factors in denervated chicken breast muscle: isolation of the chicken Myf5 gene. Nucleic Acids Res 21:2503–2509.

Sassoon D, Lyons G, Wright WE, Lin V, Lassar A, Weintraub H, Buckingham M. 1989. Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. Nature 341:303–307.

Schwarz JJ, Chakraborty T, Martin J, Zhou JM, Olson EN. 1992. The basic region of Myogenin cooperates with two transcription activation domain to induce muscle-specific transcription. Mol Cell Biol 12:266–275.

Stuart GW, McMurray JV, Westerfield M. 1988. Replication, integration and stable germ-line transmission of foreign sequences injected into early zebrafish embryos. Development 103:403–412.

Summerbell D, Ashby PR, Coutelle O, Cox D, Yee S-P, Rigby PWJ. 2000. The expression of *myf*-5 in the developing mouse embryo is controlled by discrete and dispersed enhancers specific for particular populations of skeletal muscle precursors. Development 127:3745–3757.

Tapscott SJ, Lassar AB, Weintraub H. 1992. A noval myoblat enhancer element mediates MyoD transcription. Mol Cell Biol 12:4994–5003.

van Eeden FJ, Granato M, Schach U, Brand M, Furutani-Seiki M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Kelsh RN, Mullins MC, Odenthal J, Warga RM, Allende ML, Weinberg ES, Nüsslein-Volhard C. 1996. Mutations affecting somite formation and patterning in the zebrafish, *Danio rerio*. Development 123:153–164.

Weinberg ES, Allende ML, Kelly CS, Abdelhamid A, Murakami T, Andermann P, Doerre OG, Gruwald DJ, Riggleman B. 1996. Developmental regulation of zebrafish MyoD in wild-type, *no tail* and *spadetail* embryos. Development 122:271–280

Weintraub H, Dwarki VJ, Verma I, Davis R, Hollenberg S, Snider L, Lassar AB, Tapscott ST. 1992. Muscle-specific transcriptional activation by MyoD. Genes Dev 5:1377–1386.

Winter B, Braun T, Arnold HH. 1992. Co-operativity of functional domains in the muscle-specific transcription factor Myf-5. EMBO J 11:1843–1855.

Wright WE, Sassoon DA, Lin VK. 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell 56:607–617.

Xu Y, He J, Wang X, Lim TM, Gong Z. 2000. Asynchronous activation of 10 muscle-specific protein (MSP) genes during zebrafish somitogenesis. Dev Dyn 219:201–215.

Yoshizaki G, Takeuchi Y, Sakatani S, Takeuchi T. 2000. Germ cell-specific expression of green fluorescent protein in transgenic rainbow trout under control of the rainbow trout *vasa*-like gene promoter. Int J Dev Biol 44:323–326.