Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 10344–10348, October 1995
Developmental Biology

Identification of a *Drosophila* muscle development gene with structural homology to mammalian early growth response transcription factors

JEREMY C. LEE*†, K. VIJAYRAGHAVAN‡, SUSAN E. CELNIKER§, AND MARK A. TANOUYE*¶

*Department of Molecular and Cell Biology, Division of Neurobiology, and ¶Department of Environmental Science, Policy, and Management, Division of Environmental Biology, University of California, Berkeley, CA 94720; ‡National Center for Biological Sciences, Tata Institute of Fundamental Research, P.O. Box 1234, Bangalore, 560012, India; and §Division of Biology, California Institute of Technology, Pasadena, CA 91125

Communicated by Gunther S. Stent, University of California, Berkeley, CA, June 27, 1995 (received for review May 2, 1995)

ABSTRACT **In *Drosophila*, stripe (*sr*) gene function is required for normal muscle development. Some mutations disrupt embryonic muscle development and are lethal. Other mutations cause total loss of only a single muscle in the adult. Molecular analysis shows that *sr* encodes a predicted protein containing a zinc finger motif. This motif is homologous to the DNA binding domains encoded by members of the early growth response (egr) gene family. In mammals, expression of egr genes is induced by intercellular signals, and there is evidence for their role in many developmental events. The identification of *sr* as an egr gene and its pattern of expression suggest that it functions in muscle development via intercellular communication.**

The gene stripe (*sr*) in *Drosophila* plays a critical role in muscle development. Mutations in *sr* cause severe muscle development defects; some mutant genotypes are lethal in the embryo (1–3). In homozygous sr^{54} embryos, for example, myotubes are scattered irregularly and do not appear to form contacts with their normal attachment sites in the epidermis (3). Other alleles, or allelic combinations, are viable [e.g., sr^{1}, sr^{5}, sr^{1}/*Df(3R)sr*]. In adults, these mutations cause defects in adult muscle development (1, 2). Adult muscles affected by these mutations include the indirect flight muscles, which are composed of the dorsal longitudinal muscles (DLMs) and the dorsoventral muscles (DVMs). The tergotrochanteral muscles (TTMs), which act as the primary jump muscles for initiation of flight, are also affected by certain *sr* mutant alleles. Defects in adult muscle development appear during the pupal stage and result in significant reductions in the size of some or all of these muscles. In general, the DLM is particularly affected, with some mutations resulting in its complete absence (1, 2).

Here, we present evidence that *sr* plays a role in intercellular communication required for normal muscle development. First, molecular analysis shows that *sr* encodes a predicted protein with a high level of structural homology to proteins encoded by the mammalian egr (early growth response) gene family. A characteristic of this group of mammalian transcription factors is that they are rapidly induced by a variety of intercellular signals (4–6). This suggests that *sr* activity in muscle development is induced via intercellular signaling. Second, *sr* appears to be expressed not in muscle but in other cells, including the muscle attachment sites in the epidermis. This suggests that *sr* mediates muscle development through its expression in cells that interact with developing muscle.

MATERIALS AND METHODS

***Drosophila* Stocks.** Stocks were maintained on standard medium at 22°C. Wild-type flies were the Canton Special strain. The *sr* gene is located on chromosome 3R (3–62.0) (7), in region 90E, as indicated by the breakpoint of *In(3R)Tab* (this study). sr^{1} is a spontaneous mutation; sr^{1} homozygotes show significant reductions in DLM size (1, 2). All other *sr* mutations described fail to complement the sr^{1} DLM defect. $sr^{B14.0}$, sr^{P48}, sr^{P54}, sr^{P70}, sr^{P148}, and sr^{P160} were created by insertion of the P-lwB transposon into the 90E region (3). *Df(3R)sr* is a deletion covering 90D2-91A6 (7). *In(3R)Tab* is an inversion with breakpoints in 89E and 90E (8); this inversion also fails to complement sr^{1}.

Molecular Access to the *sr* Region. We gained molecular access to the *sr* region via plasmid rescue of genomic DNA from transposon insertion stocks. To facilitate rescue, the P-lwB transposon has a single *Sal* I restriction site and also carries the ampicillin-resistance gene as a selectable marker. Genomic DNA was isolated from ≈20 flies of each transposon stock and digested with *Sal* I. The DNA was ligated in a large volume (200 μl) to favor self-ligation and then used to transform *Escherichia coli* JS5 cells by electroporation (Bio-Rad). Plasmids isolated from ampicillin-resistant clones were digested to isolate flanking genomic DNA. Isolated fragments mapped to region 90E, as indicated by digoxigenin-labeled probes hybridized to polytene chromosomes. A chromosome walk was then initiated by using probes derived from genomic DNA fragments that flank the sr^{P160} transposon insertion.

We also gained molecular access to the *sr* region by cloning the 90E breakpoint of *In(3R)Tab*. A genomic DNA library, made from *In(3R)Tab* flies, was screened with a wild-type genomic fragment from region 89E, the proximal breakpoint of *In(3R)Tab* (8). A phage clone was isolated that spanned the *In(3R)Tab* breakpoint and contained sequences from both polytene chromosome regions 89E and 90E. *Eco*RI fragments from this clone that were composed only of DNA from the 90E region were identified by *in situ* hybridization to salivary gland chromosomes. These fragments were then used as probes to screen a Canton Special wild-type genomic library.

Northern Blot Analysis and Isolation of cDNA Clones. Total fly RNA was isolated by using RNazol (Tel-Test, Friendswood, TX). Poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography, electrophoresed on formaldehyde/agarose gels, blotted onto nylon membranes, and probed with ^{32}P-labeled genomic DNA restriction fragments from the chromosome walk. Genomic DNA fragments, which had given strong signals on Northern blots, were then used as probes to isolate a *sr* cDNA from an embryonic (0–24 h) λgt11 cDNA library (gift of Kai Zinn, California Institute of Technology).

DNA Sequence Analysis. Sequence analysis was performed directly on the *sr* cDNA cloned in pBluescript (pBS) and also on similar plasmids in which the transposon TN1000 (Gold Biotechnology, St. Louis) was inserted randomly in the cDNA.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: DLM, dorsal longitudinal muscle; DVM, dorsoventral muscle; TTM, tergotrochanteral muscle; egr, early growth response.

†To whom reprint requests should be addressed.

Sequencing was performed using the double-stranded DNA cycle sequencing system (GIBCO/BRL) with ^{32}P-labeled primers.

Patterns of Expression of *sr*. The expression pattern of the *sr* transcription unit in embryos was determined by *in situ* hybridizations of the *sr* cDNA to wild-type embryos, as described by Tautz and Pfeifle (9), using a digoxigenin-labeled cDNA probe. Also, embryos from $sr^{B14.0}$ stocks were stained for β-galactosidase activity, provided by the P-lwB transposon, by the method of Bellen *et al.* (10).

RESULTS

In normal adult *Drosophila melanogaster*, the main power for flight comes from the wing downstroke, which is driven by a pair of bilaterally symmetric DLMs (Fig. 1*A*). Mutations in the gene *sr* can severely reduce DLM mass or, with some allelic combinations, completely eliminate the muscle (Fig. 1 *B* and *C*). These mutations have a less pronounced effect on other thoracic muscles; some allelic combinations cause reductions in size and defects in fasciculation of the wing elevator muscles (DVMs) and the jump muscle (TTM) (Fig. 1 *B* and *C*) (2).

Molecular characterization of *sr* was initiated to determine its role in muscle development. A chromosome walk was initiated from genomic DNA fragments flanking the sr^{P160} transposon insertion and from genomic DNA segments at the 90E breakpoint of *In(3R)Tab* (Fig. 2). Restriction fragments covering the entire chromosome walk were used to probe Northern blots from embryonic and pupal RNA. This Northern blot analysis shows that there is a transcription unit that overlaps the $sr^{B14.0}$, sr^{P48}, sr^{P70}, sr^{P148}, and sr^{P160} transposon insertion sites. This transcription unit encodes two transcripts (Fig. 3). One is a 5.0-kb transcript expressed in early (0–12 h) embryos, late (12–24 h) embryos, and pupae. A second is a 6.5-kb transcript that has been observed only in late embryos.

Genomic DNA restriction fragments were then used as probes to isolate a *sr* cDNA (Fig. 2). This cDNA hybridizes to the 5.0- and 6.5-kb transcripts (Fig. 3). The *sr* cDNA (c5-1) was sequenced, and its predicted protein product was determined (Fig. 4). A search of the GenBank data base indicated that it has amino acid sequence homology with the egr family of mammalian transcription factors. Members of this family share a conserved DNA binding domain, which is flanked by variable domains (6). The variable domains distinguish the different

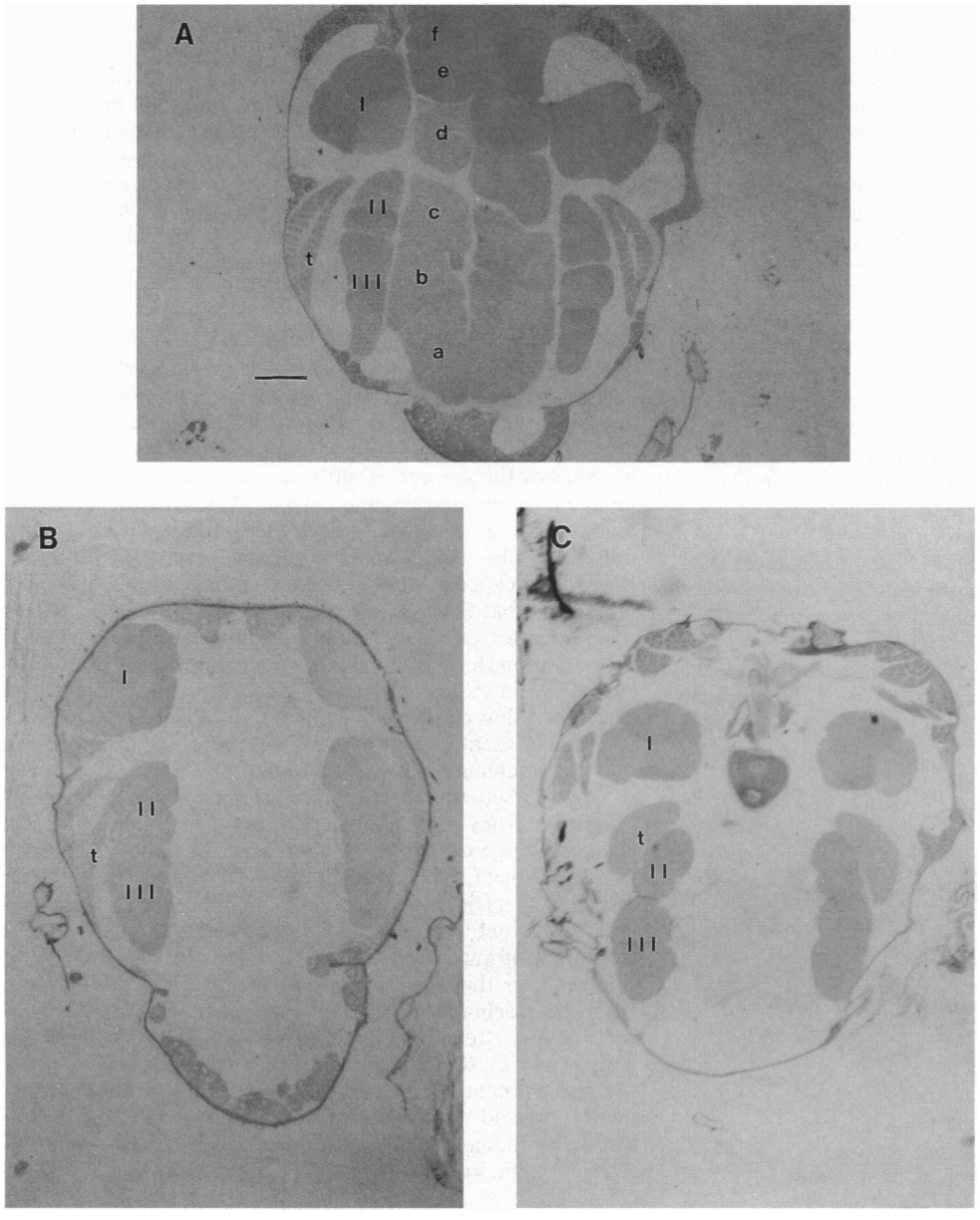

FIG. 1. Longitudinal sections through thoraxes of *D. melanogaster* wild-type and *sr* mutant adults, which were embedded in Epon 812 and propylene oxide and then sectioned and stained with 1% toluidine blue. Major flight muscles and jump muscle are indicated. DLMs, one muscle on each side composed of six fibers (a–f). Note the absence of DLMs in *sr* mutants (*B* and *C*). DVMs, three muscles on each side (I, II, and III). TTMs (t), one muscle on each side. (*A*) Wild-type Canton S fly. (*B*) $sr^1/Df(3R)sr$ mutant. (*C*) $sr^1/In(3R)Tab$ mutant. (Bar = 100 μm.)

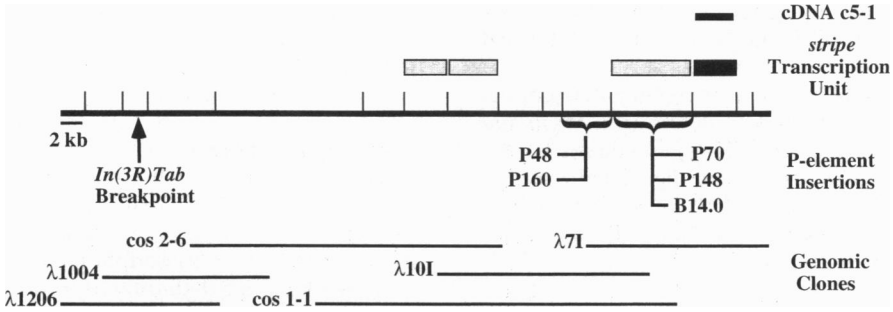

FIG. 2. Map of *sr* region. Brackets indicate genomic fragments containing *P*-element insertion sites for sr^{P48}, sr^{P160}, sr^{P70}, sr^{P148}, and $sr^{B14.0}$. Also indicated is the *In(3R)Tab* breakpoint (arrow). The 5.0- and 6.5-kb transcripts were mapped as indicated; since the precise exon boundaries have not been determined, the boxes represent genomic restriction fragments within which the exons reside. Solid box, restriction fragment that gave a strong signal in Northern blots, indicating that a large exon resides within this fragment; stippled boxes, genomic fragments with weaker signals and that presumably contain smaller exons. Genomic clones λ7I and λ10I were isolated from a *Drosophila* genomic DNA library, with a probe generated by plasmid rescue with stock $sr^{B14.0}$. Genomic clone cos 1-1 was isolated from a genomic DNA cosmid library by using a DNA probe generated by plasmid rescue of sr^{P160}, a local hop of the transposon from $sr^{B14.0}$ (3). Genomic clone cos 2-6 was isolated from the same cosmid library by using an end restriction fragment of cos 1-1 as a probe. Genomic clones λ1206 and λ1004 were isolated using stock *In(3R)Tab* as described. Map location of cDNA c5-1 is indicated. cDNA c5-1 was isolated by using the 4.5-kb fragment from λ7I as a probe.

family members. In the predicted *sr* protein, the portion predicted to encode a zinc finger DNA binding domain is highly homologous to the DNA binding domains of egr genes (Fig. 5). For example, the DNA binding domain of the human egr-1 protein is 92% identical (72/78 amino acids) to the corresponding segment in the predicted *sr* protein. Of the 6 nonidentities, 5 are conservative amino acid substitutions. This high level of homology indicates that *sr* is an egr gene family member.

Examination of the variable domain sequences indicates that the *sr* product is a unique member of the egr gene family, sharing little sequence homology with any other known egr protein. In these portions of the predicted sr protein sequence, there are segments that are glutamine-rich. Glutamine-rich motifs are common in the activation domains of transcription factors (11, 12). These motifs are thought to bind general transcription factors and/or their associated factors into a complex required for complete transcriptional activation (13). The sr sequence also has a repeat of 13 alanines; alanine-rich motifs have been shown to be associated with transcriptional repression in a number of *Drosophila* transcription factors (14). The sr sequence, therefore, strongly suggests that sr functions as a transcriptional regulator in *Drosophila*.

Although normal sr function is required for muscle development the *sr* gene is apparently not expressed in the developing muscle cells. *sr* expression was examined in wild-type embryos via *in situ* hybridization with a *sr* cDNA probe. *sr* expression was also examined in $sr^{B14.0}$ embryos. In each case, *sr* expression is not found in developing muscle (Fig. 6). Expression of *sr* in stage 14 and later embryos is found instead in the segment

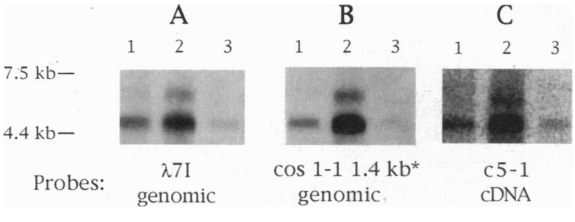

FIG. 3. Northern [poly(A)$^+$ RNA] blots from wild-type Oregon R flies. In each blot, RNA was isolated from 0- to 12-h embryos (lane 1); 12- to 24-h embryos (lane 2); and pupae (lane 3). Molecular size markers are shown on the left. (*A* and *B*) Examples of Northern blots probed with genomic fragments from the *sr* genomic walk. (*A*) RNA blot probed with genomic clone λ7I. (*B*) RNA blot probed with the distal 1.4-kb end fragment from genomic clone cos 1-1. (*C*) RNA blot probed with *sr* cDNA 5-1. *A* and *B* are x-ray autoradiographs of blots; *C* is a phosphor screen autoradiograph of a blot.

border cells, which serve as attachment sites for developing myotubes (Fig. 6) (3). At this stage of embryonic development, myotubes are disorganized in embryos homozygous for sr^{P54} and do not make contacts with their normal attachment sites in the epidermis (3). In addition, β-galactosidase staining of wing discs from $sr^{B14.0}$ third-instar larvae shows expression in cells fated to become muscle attachment sites in the adult epidermis (15). There appeared to be no staining in myoblasts in these wing discs. Finally, preliminary results with β-galactosidase staining of $sr^{B14.0}$ pupae show expression in DLM muscle insertion sites in the epidermis but not in the developing muscle fibers (J. Fernandes, S.E.C., and K.V., unpublished observations).

DISCUSSION

The defects observed in mutant embryos and adults indicate that the *sr* gene has an important function in development of many muscles. Our results suggest that this function involves intercellular communication. First, the identification of *sr* as a member of the egr gene family suggests that its expression is induced by an intercellular signal. A characteristic of the mammalian egr genes is that they are rapidly induced by many mitogenic and differentiation-inducing intercellular signals (4–6). It is expected that *sr* is induced in a similar fashion. It is possible that the signal that induces *sr* expression emanates from muscles as they develop. Second, the absence of *sr* expression in developing muscles indicates that *sr* acts via its expression in tissues with which developing muscles interact. The expression pattern of *sr* in embryos (Fig. 6) indicates that *sr* is likely to affect muscle development via expression in muscle attachment sites in the epidermis. Also, in wing discs, *sr* is expressed in cells fated to become epidermal muscle attachment sites in the adult (15). This suggests that *sr* function also acts via expression in muscle attachment sites during development of adult muscles.

The defects in muscle attachment, observed in embryos carrying lethal *sr* mutations, indicate that *sr* functions in formation of muscle attachments. Muscle attachment formation requires that developing muscles locate the appropriate sites for attachment in the epidermis. One possibility, therefore, is that *sr* regulates synthesis of a signaling molecule that is provided by the epidermis and that muscles use to locate their attachment sites. Evidence for signaling between attachment sites and developing muscle has been presented from work with the nematode *Caenorhabditis elegans* (16). Alternatively, *sr* might function in the epidermis by regulating synthesis of molecules that form the attachment structure (e.g., integ-

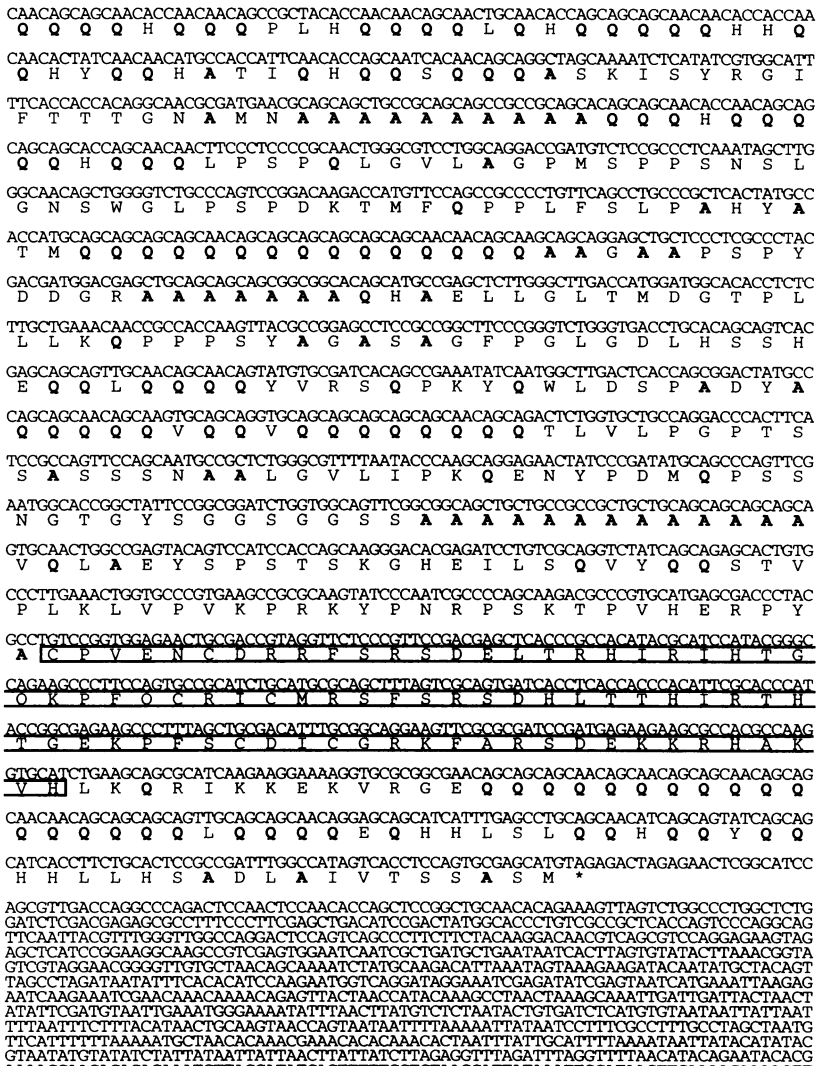

FIG. 4. Nucleotide and deduced amino acid sequence of *sr* cDNA c5-1. Box indicates zinc finger DNA binding domain; glutamine and alanine residues in variable regions are shown in boldface. Absence of an in-frame stop in the 5′ end indicates that cDNA does not contain the complete open reading frame.

rins). Finally *sr* might act by regulating a gene(s) involved in differentiation of epidermal cells so that they are competent to synthesize signaling and attachment molecules.

In adults, the defects in muscle development, observed in viable *sr* mutants, might also result from defective attachment sites. It is possible that mutations in *sr* affect synthesis of signaling molecules in the epidermis that are required for normal adult muscle assembly. Furthermore, defects in attachment structure might affect muscle development physically, through alterations in mechanical tension in the muscle; defects in muscle fiber fasciculation, which have been observed in the DVM and the TTM in *sr* mutants, might result from such changes in tension.

Whatever the precise mechanism of *sr* function in the epidermis, it appears that defects in this function do not fully explain the dramatic effect of *sr* mutations on DLM development (Fig. 1*C*). During metamorphosis, the DLMs, DVMs, and TTMs develop from myoblasts, which migrate out of the wing discs (17, 18). They must all form new attachments in the developing adult epidermis. If *sr* affected the development of these muscles only through their attachment sites, the DLM would not be specifically affected by *sr* mutations. It may be, therefore, that *sr* plays an additional and specific role in DLM development. There are differences in developmental histories of the DLM and other thoracic muscles. For example, only the DLMs utilize persistent larval muscles as developmental templates (17, 18), and they differ in the patterns

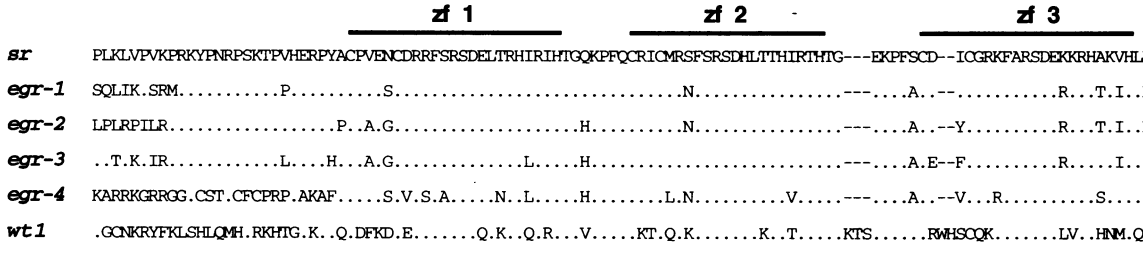

FIG. 5. Comparison of deduced amino acid sequence of the DNA binding domain of the *sr* cDNA with sequences of other members of the egr gene family. Period indicates amino acid identities; dash indicates relative gap in sequence. zf, Zinc finger.

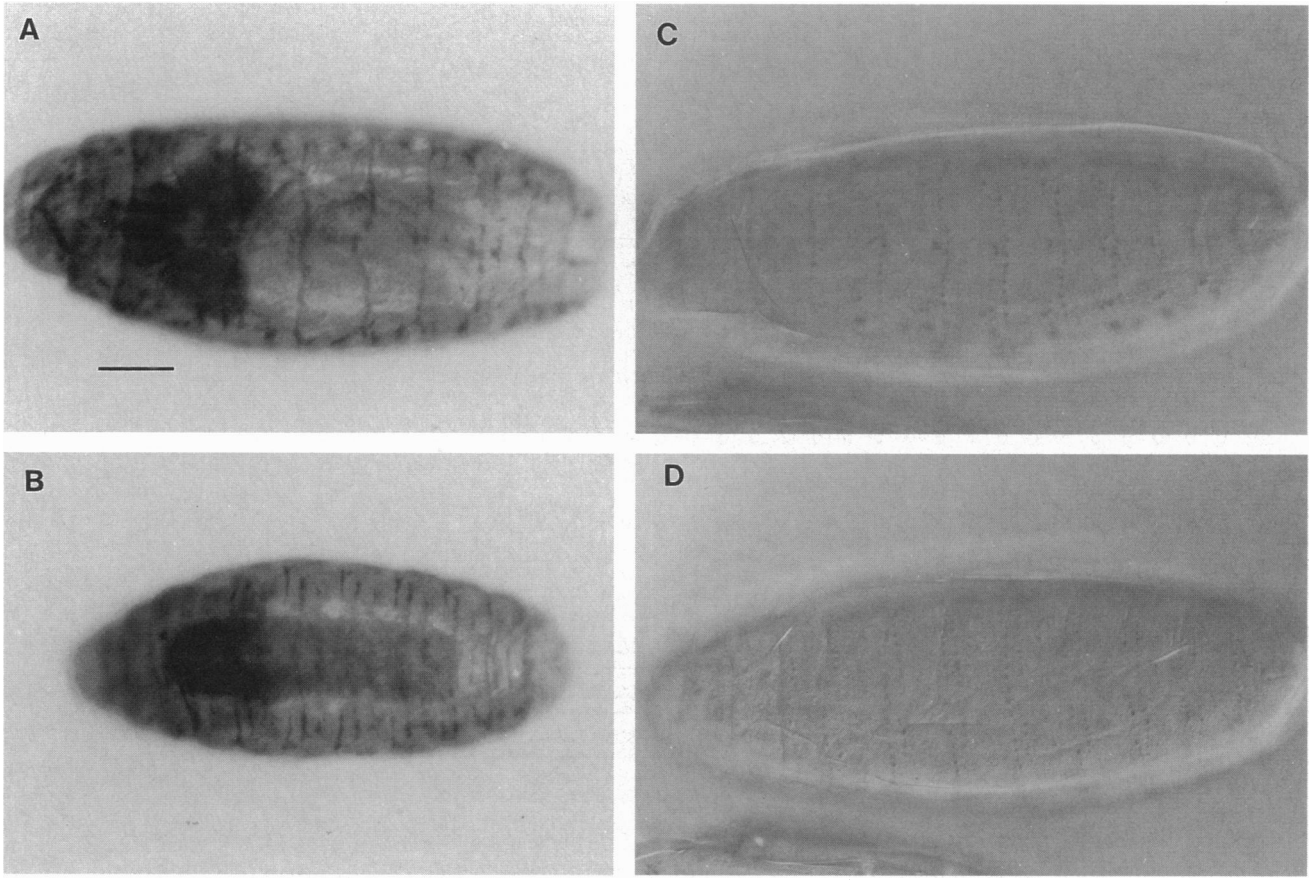

FIG. 6. Comparison of *sr* transcript expression patterns in late-stage wild-type embryos with patterns of expression of β-galactosidase in embryos of *P*-element insertion stock $sr^{B14.0}$. All photographs are with anterior on the left. (*A* and *B*) *In situ* hybridization of digoxigenin-labeled *sr* cDNA c5-1 to Canton S wild-type embryos. (*C* and *D*) Staining for β-galactosidase activity in embryos from *P*-element insertion stock $sr^{B14.0}$. (*A* and *C*) Dorsal aspect. Gene expression occurs in segment border cells. (*B* and *D*) Ventral aspect. Gene expression occurs in segment border cells and two other rows of cells per segment. (Bar = 50 μm in *A*, *C*, and *D*; bar = 70 μm in *B*.)

and timing of innervation (19). There may be a relationship between these features of DLM development and a specific role for *sr* in this process.

There are two primary examples in mammals that demonstrate the function of egr gene products *in vivo*. Mice carrying a mutation in egr2 (also known as *Krox20*) show severe developmental defects in the hindbrain and in cranial nerves (20). Second, mutations in the egr gene *wt1* are associated with Wilms tumors in children (21). In addition, *in vitro* studies suggest that egr genes play roles in the growth and differentiation of many cell types (e.g., see refs. 22 and 23). The identification of *sr* as an egr gene family member demonstrates that these genes are required for development in invertebrates as well as in mammals and indicates that they can play an important role *in vivo* in the development of muscles and formation of their attachments.

We thank A. Kolodkin for stock $sr^{B14.0}$ and genomic DNA clones; C. Oh, E. Reynolds, and P. Pavlidis for comments on the manuscript and technical advice; M. Palazzolo for technical advice; S. Faulhaber, A. Francisco, D. Chang, and R. Young for technical assistance; and S. Benzer for use of laboratory facilities. This work was supported in part by Public Health Service National Research Service Award GM15867-02 to J.C.L. and by grants from the Department of Biotechnology, Government of India, and the Rockefeller Foundation to K.V.

1. Costello, W. J. & Wyman, R. J. (1986) *Dev. Biol.* **118,** 247–258.
2. de la Pompa, J. R., Garcia, J. R. & Ferrus, A. (1989) *Dev. Biol.* **131,** 439–454.
3. Volk, T. & VijayRaghavan, K. (1994) *Development (Cambridge, U.K.)* **120,** 59–70.
4. Sukhatme, V. P. (1990) *J. Am. Soc. Nephrol.* **1,** 859–866.
5. Sheng, M. & Greenberg, M. E. (1990) *Neuron* **4,** 477–485.
6. Madden, S. L. & Rauscher, F. J., III (1993) *Ann. N.Y. Acad. Sci.* **684,** 75–84.
7. Lindsley, D. L. & Zimm, G. G. (1992) *The Genome of Drosophila melanogaster* (Academic, San Diego), p. 663.
8. Celniker, S. E. & Lewis, E. B. (1987) *Genes Dev.* **1,** 111–123.
9. Tautz, D. & Pfeifle, C. (1989) *Chromosoma* **98,** 81–85.
10. Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U., Pearson, B. K. & Gehring, W. J. (1989) *Genes Dev.* **3,** 1288–1300.
11. Courey, A. J. & Tjian, R. (1988) *Cell* **55,** 887–898.
12. Mitchell, P. J. & Tjian, R. (1989) *Science* **245,** 371–378.
13. Gill, G., Pascal, E., Tseng, Z. H. & Tjian, R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 192–196.
14. Licht, J. D., Hanna-Rose, W., Reddy, J. C., English, M. A., Ro, M., Grossel, M., Shaknovich, R. & Hansen, U. (1994) *Mol. Cell. Biol.* **14,** 4057–4066.
15. Celniker, S. E. & Lewis, E. B. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 1566–1570.
16. Hresko, M. C., Williams, B. D. & Waterston, R. H. (1994) *J. Cell Biol.* **124,** 491–506.
17. Bate, M. (1993) in *The Development of Drosophila melanogaster*, eds. Bate, M. & Martinez Arias, A. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 2, pp. 1013–1090.
18. Fernandes, J., Bate, M. & VijayRaghavan, K. (1991) *Development (Cambridge, U.K.)* **113,** 67–77.
19. Fernandes, J. & VajayRaghavan, K. (1993) *Development (Cambridge, U.K.)* **118,** 215–227.
20. Schneider-Maunoury, S., Topilko, P., Seitanidou, T., Levi, G., Cohen-Tannoudji, M., Pournin, S., Babinet, C. & Charney, P. (1993) *Cell* **75,** 1199–1214.
21. Haber, D. A., Buckler, A. J., Glaser, T., Call, K. M., Pelletier, J., Sohn, R. L., Douglass, E. C. & Housman, D. E. (1990) *Cell* **61,** 1257–1269.
22. Nguyen, H. Q., Hoffman-Liebermann, B. & Liebermann, D. A. (1993) *Cell* **72,** 197–209.
23. Rothman, A., Wolner, B., Button, D. & Taylor, P. (1994) *J. Biol. Chem.* **269,** 6399–6404.