Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 7973–7977, September 1991
Biochemistry

Drosophila has a twitchin/titin-related gene that appears to encode projectin

(muscle proteins/immunoglobulin superfamily/sequence evolution)

A. AYME-SOUTHGATE*, J. VIGOREAUX*, G. BENIAN†, AND M. L. PARDUE*

*Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and †Department of Pathology, Emory University School of Medicine, Atlanta, GA 30322

Contributed by M. L. Pardue, June 14, 1991

ABSTRACT **The sequences of twitchin and titin identify a superfamily of muscle proteins whose functions are not completely understood. In spite of their shared structural features, twitchin and titin appear to differ in function. Genetic and molecular evidence suggests that twitchin has a regulatory role in muscle contraction, whereas it has been proposed that titin has a structural function. We report here that *Drosophila* has a single-copy gene containing the two-motif amino acid sequence pattern that characterizes twitchin and titin. This gene appears to encode projectin, a muscle protein that is thought to play a structural role in asynchronous flight muscle but may have a role like that of twitchin in synchronous muscle. Thus *Drosophila* appears to be a case where the apparently diverged functions of twitchin and titin are encoded by a single gene.**

Recent studies have characterized several new proteins associated with the contractile apparatus of muscle. One notable class of these proteins is represented by the very large polypeptides titin (1000–3000 kDa) in vertebrate muscle (1, 2) and twitchin (753.570 kDa) in muscle of the nematode *Caenorhabditis elegans* (3). Both of these polypeptides contain large numbers of two conserved amino acid sequence motifs (motifs I and II) that occur in regular arrays (refs. 4 and 5; Fig. 1). Motif I is similar to fibronectin type III domains and motif II is similar to the C2 set of immunoglobulin domains (4). Twitchin, titin, and five other proteins [smooth muscle and nonmuscle myosin light chain kinases (6, 7), C-protein (8), an 86-kDa protein (9), and skelemin (10)] make up an intracellular subset (mainly muscle) of the immunoglobulin superfamily.

In spite of the structural similarities between titin and twitchin, the two proteins may have different roles in muscle. Titin extends from the Z band to the M line of the sarcomere (11, 12), and it has been proposed to act in positioning the thick filaments and in providing for resting tension in the sarcomere (13, 14). In contrast, twitchin is detected only within the myosin-containing regions of the sarcomere. Mutant animals lacking twitchin have a characteristic, nearly constant, "twitching" of the body wall muscles, suggesting that twitchin is involved in the regulation of the contraction–relaxation cycle (3).

Projectin is a *Drosophila* muscle protein, similar in size to twitchin. Projectin was first identified (15) as a component of the connecting filaments in insect flight muscle (an asynchronous muscle). These filaments connect the thick filaments to the Z band and may be structural adaptations for stretch activation, the mechanism that allows for the exceedingly rapid contractions characterizing asynchronous flight muscles (16, 17). Most *Drosophila* muscles (referred to as synchronous muscles) are not stretch-activated, yet unexpectedly projectin is also present in these muscles. In synchronous muscles projectin is localized in the region of the myosin filaments, rather than between these filaments and the Z band as it is in flight muscle (18). The difference in projectin localization in the two muscle types suggests that the protein may play different roles in the two muscles. Projectins from flight muscle, leg muscle, and supercontractile muscle differ somewhat in size, but analyses of partial proteolytic digests indicate that these proteins are, nevertheless, very similar over much of the molecule (19).

The localization of projectin within the myosin-containing regions in the synchronous muscles in *Drosophila* resembles the localization of twitchin in the muscles of the nematode. This, in addition to the similarity in size of the two proteins, suggested that projectin and twitchin might be related. The suggestion is further supported by evidence that projectin is recognized by a polyclonal antibody against twitchin. This antibody binds to a *Drosophila* muscle protein that comigrates with projectin on gels; the twitchin antibody also binds to both flight muscle and synchronous muscle of *Drosophila* in the same patterns seen with antibodies against projectin (19). We now report that the *Drosophila* gene that appears to encode projectin contains sequences with homology to both motif I and motif II of twitchin and titin.‡

MATERIALS AND METHODS

PCR and Library Screening. The PCRs were performed either on genomic DNA or on cDNA reverse-transcribed from total RNA by using random primers for the reverse transcription (20). Phage were selected from a *Drosophila* λ phage library (21) with gel-isolated PCR 186-base-pair (bp) fragments (Fig. 2) as probe, by using conventional techniques (22).

DNA Sequencing. Sequencing was done by the dideoxynucleotide method (23) using as templates either M13 single-stranded DNA or double-stranded DNA (pGEM-3 subclones using Sp6 and T7 primers; Promega).

Polytene Chromosome *in Situ* Hybridization. Hybridization to polytene chromosomes was performed as described (24).

RNA Analysis. Total RNA was prepared from adult *Drosophila* as described (25). Total RNA (10 μg) was electrophoresed on a formaldehyde/agarose gel and transferred as described (22). The probes were sense and antisense RNA from the B region (Fig. 3).

RNA *in Situ* Tissue Hybridization. Frozen sections of adults and late pupa were processed for hybridization as described (26). The probe was antisense [^{3}H]RNA (24) from the B region (Fig. 3). After hybridization at 55°C, sections were washed twice at 55°C in 2× standard saline citrate (SSC)/50% (vol/vol) formamide for 30 min, treated with RNase (20 μg/ml, 25°C, 30 min), rinsed in 2× SSC, dehydrated, and covered with emulsion (24).

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.

‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M73433–M73435).

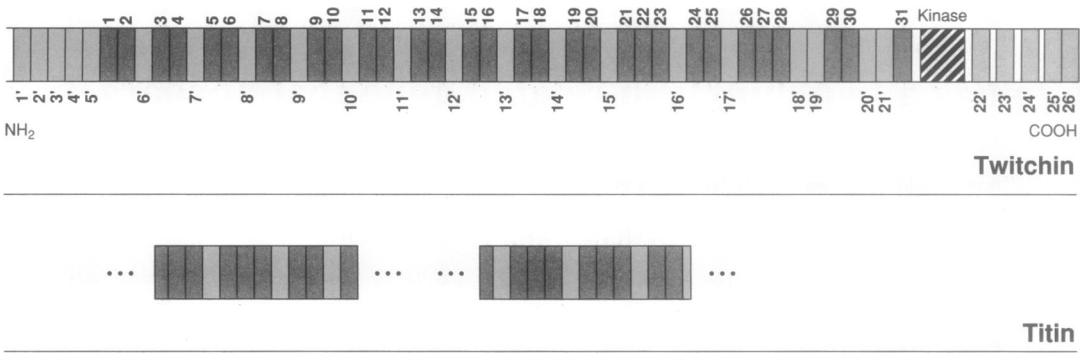

FIG. 1. Schematic representation of the arrangement of motifs I and II in the nematode protein twitchin and the vertebrate protein titin. Motif I sequences are dark gray; motif II sequences are light gray. The twitchin sequence (4) is complete except for a short region at the N terminus. The titin sequence (5) is incomplete.

RESULTS

Isolation of a *Drosophila* Gene Related to the *C. elegans* Twitchin Gene. The strong evidence that projectin is related to twitchin led us to use DNA sequences from the twitchin gene as primers to isolate and clone *Drosophila* DNA segments containing related sequences. Comparison of all the copies of twitchin motif I has produced a consensus amino acid sequence (4). Of the 100 amino acids composing motif I, one region of 12 amino acids and another of 13 amino acids are particularly well conserved. Degenerate primers based on the consensus sequence from these two regions were synthesized and used in a PCR to amplify *Drosophila* genomic DNA. As predicted from the twitchin motif I sequence, an amplified fragment of 186-bp was the major product of the PCR (Fig. 2). This 186-bp fragment was used to screen a *Drosophila* genomic library (21) and the resulting five clones all overlapped to produce a contig of ≈30 kb (Fig. 3). By *in situ* hybridization to polytene chromosomes, the contig mapped on the fourth chromosome at 102C/D.

When the PCR was performed with the twitchin primers, in addition to the 186-bp fragment, five other products were obtained. Five of the six PCR products hybridized within the cloned contig (Fig. 3). The sixth product, a 380-bp fragment, was not detected within the cloned contig. When used for *in situ* hybridization to polytene chromosomes the 380-bp probe did not hybridize to the 102C/D site with the other PCR fragments. Instead the 380-bp fragment hybridized to a single site at 63B/C and thus came from an entirely different gene. A genomic DNA probe containing the 380-bp fragment recognized a rare message on a Northern blot of adult poly(A^+) RNA that was too small to encode projectin (data not shown); however, we have not determined whether this 63B/C gene encodes a muscle-specific protein.

An initial sequencing of one of the 186-bp PCR products confirmed the perfect homology with the twitchin motif I consensus (Fig. 4*B*, sequence 8). Therefore, two restriction fragments (fragments A and B in Fig. 3) that hybridized with the other PCR products were chosen for sequencing to determine their relationships with twitchin motifs I. The deduced amino acid sequence of these fragments (Fig. 4) showed that perfect copies of motif I were present in both segments. These motifs I were interspersed with the second motif (motif II) that was also found in twitchin. The pattern of interspersion is identical to the predominant pattern in twitchin, namely, two motifs I followed by one motif II (Fig. 4*A*). The alignment between the various copies of each motif and the twitchin consensus is presented in Fig. 4 *B* and *C*. *Drosophila* motifs I are identical to twitchin motif I in their overall organization. The *Drosophila* amino acid sequences follow the twitchin consensus at least as well as most of the twitchin copies. The two regions that are the best conserved in twitchin are also the best conserved in *Drosophila* and, conversely, the regions of maximum divergence in twitchin are very variable in *Drosophila*. Motifs II are generally less conserved but still fit the consensus derived from twitchin.

An example of the degree of the conservation between the *Drosophila* and the *C. elegans* sequences is seen in comparing the sequences of the motif I; in both species these motifs can be grouped into two subsets. In Fig. 4*B*, the motif I

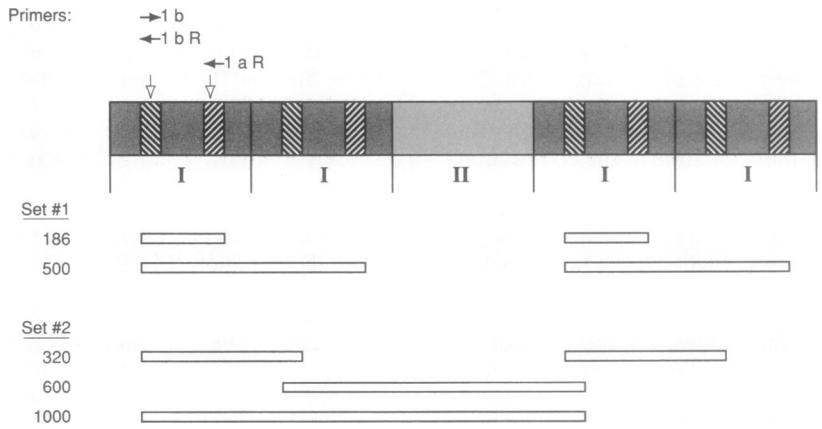

FIG. 2. Alignment of the PCR products obtained from *Drosophila* DNA using twitchin-derived primers with the most common interspersion pattern of motifs I and II found in twitchin (4). Motif I is dark gray; motif II is light gray. Hatched areas in motif I represent the two best conserved regions of 12 and 13 amino acids, respectively (used for PCR primers). Bars underneath represent PCR products obtained by two sets of primers. Primers 1b and 1aR yield set 1; primers 1b and 1bR yield set 2. Both genomic and cDNA PCRs yielded the same sets of fragments. The PCR 380-bp product, obtained in set 2, is not represented in this diagram since its length is not predictable from any combination of motifs I and II.

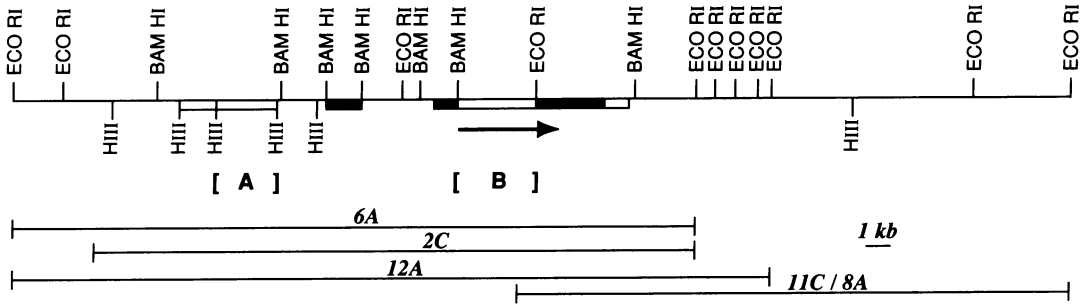

FIG. 3. Schematic map of the cloned region of DNA. Overlapping clones 2C, 6A, 8A, 11C, and 12A cover 30 kilobases (kb) of DNA. Solid boxes beneath map indicate fragments binding PCR 186-bp probes; open boxes represent fragments binding all other PCR products shown in Fig. 2. Arrow indicates direction of transcription. The two regions in brackets are fragments A and B for which the sequence is presented (Fig. 4).

sequences were numbered 1 and 2 (fragment A) and 3–7 (fragment B) based on their order within the fragments from N to C terminus. Sequences 1, 2, 4, and 6 (called set Ia) have an extra leucine at position 10, whereas in sequences 3, 5, and 7 (called set Ib) a gap was introduced at position 10 to maintain the maximum alignment. In twitchin also the odd-numbered motif I sequences are like *Drosophila* set Ia and even-numbered motif I sequences are like *Drosophila* set Ib. The order is such that the motif I sequences that are C-terminal to a motif II contain the extra leucine whereas motif I sequences that are N-terminal to a motif II do not. This pattern is lost in twitchin at the position where the regular pattern of two motifs I to one motif II is disrupted (4). The extra amino acid is always a leucine (except for isoleucine in one twitchin repeat) and, at least in *Drosophila*, even the codon is conserved (CUU) although leucine can be specified by any of six codons. In terms of evolution, the existence of these two subsets of motif I may indicate that the actual unit of duplication is motif Ia–motif II–motif Ib. The effect of the leucine on protein folding and/or function is more difficult to assess, but the conservation of this feature between *C. elegans* and *Drosophila* suggests that the extra leucine plays an important role. The argument is stronger because of the high degree of divergence observed elsewhere in the copies

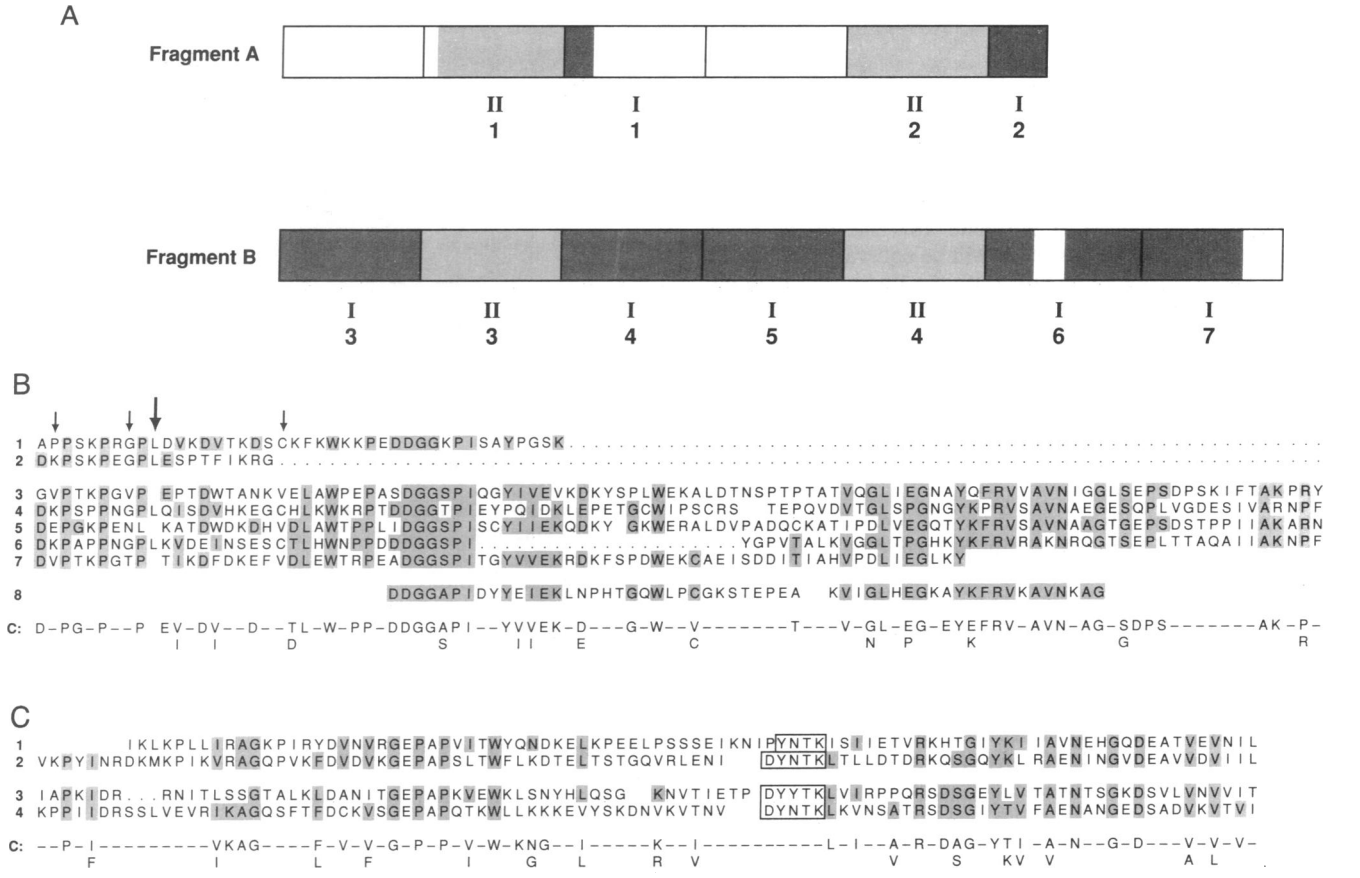

FIG. 4. Alignment of the amino acid sequence of *Drosophila* motifs I and II for comparison with the twitchin motif I and II consensus sequences. (*A*) Diagram showing location of individual motifs within the *Drosophila* sequenced regions. Motif I, dark gray; motif II, light gray. Unsequenced regions, open. (*B* and *C*) Numbers represent *Drosophila* motifs. C, consensus sequence derived from twitchin motifs (4). Shaded amino acids agree with the twitchin consensus. Dots indicate unsequenced regions. Blank spaces are gaps introduced for alignment. (*B*) Alignment of *Drosophila* motif I sequences. Arrows point to leucine 10 and the other amino acids differentiating the two subsets of motif I. Sequence 8 is the motif I sequence derived from one of the PCR 186-bp fragments. (*C*) Alignment of *Drosophila* motif II sequences. Boxed regions are additional sequences conserved between the four *Drosophila* copies but absent in twitchin.

of motif I. This distinction of two subsets that was originally recognized by the leucine residue holds true for several other amino acids along the motif I sequence, especially within the first 50 amino acids (indicated by arrows in Fig. 4*B*).

The *Drosophila* Twitchin-Related Gene Encodes Projectin. The twitchin gene encodes an mRNA of 21,614 nucleotides (G.B., unpublished results). Projectin is approximately as large as twitchin (19) and, therefore, if the *Drosophila* twitchin-related gene that we have isolated encodes projectin, we would expect the mRNA to also be on the order of 20 kb. In two *Drosophila* RNA preparations, we have detected hybridization to a probe that binds the strand that should be transcribed into mRNA (as determined from the sequence data); probes for the nontranscribed strand did not detect any RNA. In those experiments the band of hybridization was at the limit of migration, as expected of an RNA >20 kb. In other preparations we have failed to detect any hybridization. (We presume that there has been degradation of the very large RNA in these latter preparations, although smaller mRNAs, complementary to probes from other genes, were still present.) We have never seen hybridization to any band smaller than 20 kb with probes from the 102C/D gene (data not shown) and, therefore, conclude that the transcript of the 102C/D gene is large enough to encode projectin.

To determine whether the *Drosophila* 102C/D gene is transcribed in muscle, we have probed tissue sections from *Drosophila* adults and pupae with segments of RNA complementary to the predicted mRNA sequences (Fig. 5). The hybrids were detected only over muscle tissue. All types of muscle, both synchronous and asynchronous, showed hybridization as expected from the distribution of the projectin protein (19).

DISCUSSION

Our experiments to date show that *Drosophila* has a muscle-specific gene that is clearly related to twitchin and titin on the basis of the amino acid sequence. Although we have not yet completed the sequence of the gene, the size of the RNA predicts that the encoded protein will be equivalent in size to twitchin and to projectin. Thus, with the result that twitchin antibodies recognize projectin, this is strong evidence that the *Drosophila* 102C/D twitchin-related gene encodes projectin. Recently, Lakey *et al.* (27) have reported that antibodies to *Lethocerous* p800 (which appears to be projectin) bound a polypeptide encoded by a partial cDNA with sequence homology to motif II. This result also supports the hypothesis that projectin is the insect equivalent of twitchin and titin. Further evidence that projectin is a member of the twitchin/titin family is provided by the recognition of crayfish projectin by a monoclonal antibody to chicken titin (28).

When the PCR-generated 186-bp fragment was used to identify *Drosophila* genomic clones, all of the clones selected were located in one contiguous region of the genome (at 102C/D). This result argues that there is only one twitchin-equivalent gene. This conclusion is supported by results from *in situ* hybridization on polytene chromosomes that also show a single site of hybridization at 102C/D. The only fragment amplified in the PCR from the twitchin primers that is not found in the 102C/D gene defines a second gene located at 63B/C. The transcript from 63B/C is too small to encode projectin. The product of this gene has not been further defined.

Our evidence that the 102C/D gene encodes projectin raises some interesting questions. The gene is single copy yet we have found muscle-type-specific forms of projectin that differ both in size and in apparent function (19). These muscle-type-specific forms must be produced posttranscriptionally, perhaps by alternative splicing or by modification of the protein. Understanding the structures of these alternative forms and how the structures relate to the different functions of projectin in synchronous and asynchronous muscle may give insights into the apparent divergence of function between twitchin and titin. The localization of the projectin gene to 102C/D is a first step toward a genetic analysis of the protein. We note that there are two known mutations (29) in this region that have defects in wing posture, a phenotype that has been associated with defects in other known muscle genes. The mutations are bent (*bt*) and bent Dominant (bt^D). If the *bt* and bt^D mutations can be shown to be in the projectin

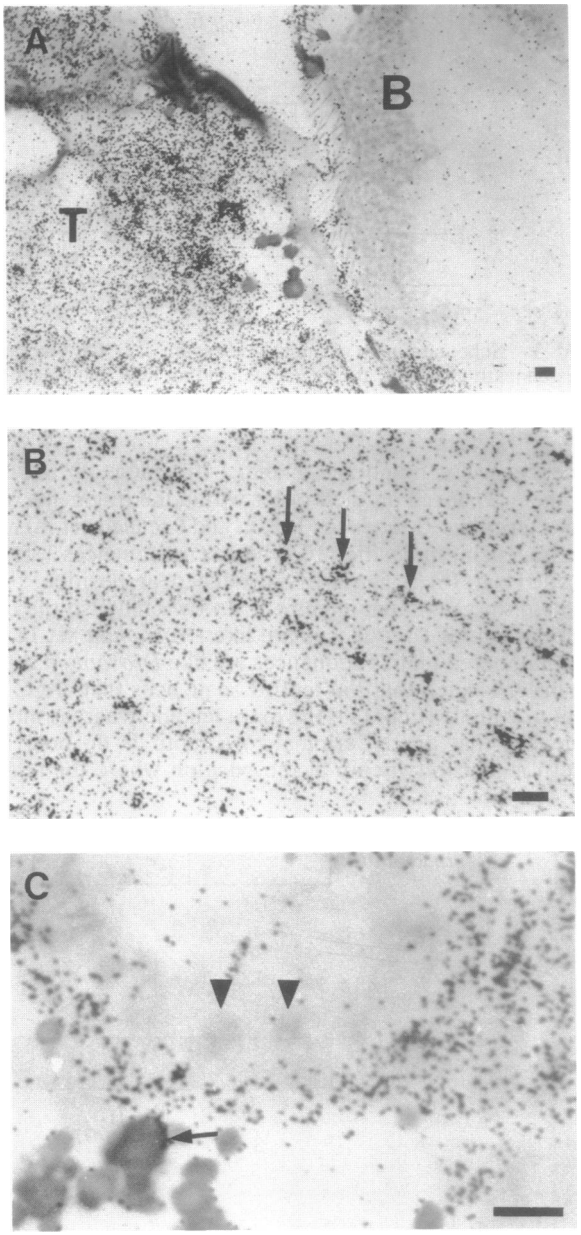

FIG. 5. Autoradiographs showing muscle-specific expression of the 102C/D gene. All sections were probed with antisense transcript of fragment B. (*A*) Pupal section through the thorax (T) and the brain (B), showing heavy hybridization on all thoracic muscles but no hybridization to the brain. (*B*) Adult section through the indirect flight muscles showing hybridization throughout the muscle with concentration of label over the linearly arranged nuclei (e.g., arrows). We assume this represents nascent transcripts. Because of the large size of the mRNA, nascent transcripts should represent a significant nuclear accumulation. (*C*) Pupal section through the abdomen. Hybridization is localized over the gut muscles. Gut parenchymal cells (arrowheads) and fat cells (arrow) do not show any hybridization. (Bars = 10 μm.)

gene, they may offer important clues to the functions of the projectin isoforms.

We thank Maryann Impeduglia for graphics/artwork. This work has been supported by grants from the National Institutes of Health (to M.L.P. and G.B.). J.V. was a postdoctoral fellow of the American Cancer Society.

1. Maruyama, K., Kimura, S., Ohashi, K. & Kuwano, Y. (1981) *J. Biochem.* **89,** 701–709.
2. Wang, K. (1982) *Methods Enzymol.* **85,** 264–282.
3. Moerman, D. G., Benian, G. M., Barstead, R. J., Schrieffer, L. A. & Waterston, R. H. (1988) *Genes Dev.* **2,** 93–105.
4. Benian, G. M., Kiff, J. E., Neckelmann, N., Moerman, D. G. & Waterston, R. H. (1989) *Nature (London)* **342,** 45–50.
5. Labeit, S., Barlow, D. P., Gautel, M., Gibson, T., Holt, J., Hsieh, C.-L., Francke, U., Leonard, K., Wardale, J., Whiting, A. & Trinick, J. (1990) *Nature (London)* **345,** 273–276.
6. Olson, N. J., Pearson, R. B., Needleman, D. S., Hurwitz, M. Y., Kemp, B. E. & Means, A. R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2284–2288.
7. Shoemaker, M. O., Lau, W., Shattuck, R. L., Kwiatkowski, A. P., Matrisian, P. E., Guerra-Santos, L., Wilson, E., Lukas, T. J., Van Eldik, L. J. & Watterson, D. M. (1990) *J. Cell Biol.* **111,** 1107–1125.
8. Einheber, S. & Fischman, D. A. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2157–2161.
9. Fischman, D. A., Vaughan, K., Weber, F. & Einheber, S. (1991) in *Frontiers of Muscle Research*, ed. Ozawa, E. (Elsevier, Amsterdam), in press.
10. Price, M. G., Brooks, C. A. & Gomer, R. H. (1990) *J. Cell Biol.* **111,** 170a (abstr.).
11. Maruyama, K., Yoshoika, T., Higuchi, H., Ohashi, K., Kimura, S. & Natori, R. (1985) *J. Cell Biol.* **101,** 2167–2172.
12. Furst, D. O., Osborn, M., Nave, R. & Weber, K. (1988) *J. Cell Biol.* **106,** 1563–1572.
13. Horowits, R., Kempner, E. S., Bisher, M. E. & Podolsky, R. J. (1986) *Nature (London)* **323,** 160–164.
14. Horowits, R. & Podolsky, R. J. (1987) *J. Cell Biol.* **105,** 2217–2223.
15. Saide, J. D. (1981) *J. Mol. Biol.* **153,** 661–679.
16. Wray, J. S. (1979) *Nature (London)* **280,** 325–326.
17. White, D. C. S. (1983) *J. Physiol. (London)* **343,** 31–57.
18. Saide, J. D., Chin-Bow, S., Hogan-Sheldon, J., Busquets-Turner, L., Vigoreaux, J. O., Valgeirsdottir, K. & Pardue, M. L. (1989) *J. Cell Biol.* **109,** 2157–2167.
19. Vigoreaux, J. O., Saide, J. D. & Pardue, M. L. (1991) *J. Muscle Res. Cell Motil.*, in press.
20. Saiki, R. K. (1989) in *PCR Technology, Principles and Applications for DNA Amplification*, ed. Erlich, H. A. (Stockton, New York), pp. 7–16.
21. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
24. Pardue, M. L. (1986) in *Drosophila: A Practical Approach*, ed. Roberts, D. B. (IRL, Oxford), pp. 111–137.
25. Ayme, A. & Tissieres, A. (1985) *EMBO J.* **4,** 2949–2954.
26. Hafen, E. & Levine, M. (1986) in *Drosophila: A Practical Approach*, ed. Roberts, D. B. (IRL, Oxford), pp. 139–157.
27. Lakey, A., Ferguson, C., Labeit, S., Reedy, M., Larkins, A., Butcher, G., Leonard, K. & Bullard, B. (1990) *EMBO J.* **9,** 3459–3467.
28. Hu, D. H., Matsuno, A., Terakado, K., Matsuura, T., Kimura, S. & Maruyama, K. (1990) *J. Muscle Res. Cell Motil.* **11,** 497–511.
29. Lindsley, D. L. & Grell, E. H. (1968) *Genetic Variations of Drosophila melanogaster* (Carnegie Institution of Washington, Washington, DC), p. 36.