STRUCTURE AND EXPRESSION OF THE ACTIN GENE FAMILY OF *DROSOPHILA MELANOGASTER*

Thesis by

Beverley Bond Matthews

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1987

(Submitted August 20, 1986)

Acknowledgements

There are many people that I'd like to thank for making my life at Caltech pleasant and productive both in and out of the lab.

Norman Davidson was an outstanding advisor. He was always willing to listen and give valuable advice but also gave me enough freedom to learn to accomplish things on my own.

All of the members of Norman's lab and especially Eric Fyrberg, Dave Hershey, Chuck Rozek, Nevis Fregien, Lei Yu, Mickey Hu, and Sandy Sharp gave me advice and encouragement in the lab and Sandy and Lei, especially, kept me smiling.

Arthur Roach, Sue Celniker, Mark Garfinkel, and Lynn Crosby were also respected sources of scientific advice and good friends.

Lei Yu, Kath Mayne, Tina Kramer, and Joe Rushanan created a very pleasant home environment at the J. Alfred Prufrock House. The friendly folks in the Eating Group provided a welcome diversion each evening.

Music has always been very important in my life. John Dick, Sue Grether, and Dick Wingell were instrumental in keeping me happy and helping me to grow musically.

I am grateful to have graduated despite the efforts of Dave Hershey, Vann Parker, and Jeff Mayne who introduced me to some marvelous distractions available in the California mountains, deserts and seashores.

Bill Mattox was my general right hand man and a great friend. He was always there when I needed him to give scientific advice, an encouraging word, or Lido cookies.

My husband, Greg, is a virtually inexhaustable fount of intellectual curiosity and an unquenchable source of conversation who reminds me how much fun learning is even when I get bogged down in mundane procedures. He is also an

extremely kind person whose caring and backrubs helped get me through graduate school. Finally, I am very grateful to my family for their incredible love and support and to my father in particular for stressing the value of education.

Most of my support was provided by the National Institutes of Health. The Jean Weigle Fund made a contribution toward the preparation of this thesis.

ABSTRACT

We have isolated the six actin genes of *Drosophila melanogaster* from a *Drosophila* genomic DNA library and have compared structural features of the genes by restriction mapping, electron microscopy and DNA sequencing. We found that at least two of the actin genes contain intervening sequences which interrupt the genes at different positions. Several of the genes were shown to be lacking intervening sequences in the analogous positions. This nonconservation of intron position is in striking contrast to the strong conservation of intron positions seen in other gene families. The DNA sequences of the protein coding regions of the genes are highly conserved while the intron and untranslated sequences are not. The primary sequences of all the *Drosophila* actins resemble mammalian cytoplasmic actins more than mammalian muscle actins.

We studied the distribution of actin mRNAs in different developmental stages and in different dissected body parts with the use of gene specific hybridization probes which we isolated from the 3' untranslated portions of the genes. We found that the genes fall into three main categories with respect to their patterns of expression in *Drosophila*. Trancripts from two of the genes are found throughout *Drosophila* development. They are expressed at higher levels in ovaries and embryonic cultured cells than in muscle containing tissue and are thought to be cytoplasmic actins. Two others encode thoracic muscle actins. Their transcripts accumulate predominantly in the thoracic regions of the adult where the flight and jump muscles are found. The other two genes are most active in larvae and in adult abdomens. They are thought to encode actins used in the larval, pupal, and adult intersegmental muscles.

We studied the structure of the cytoplasmic actin gene, act5C, in detail and found that it encodes at least six different mRNAs. At the 5' end there are two nonhomologous leader exons which are alternately spliced to the remainder of the gene. At the 3' end of the gene, three sites of polyadenylation are used. The 3' variation is the principal cause of the transcript length heterogeneity observed in the transcripts. In whole animal RNA, the two leader exons are expressed with the same pattern through development and with all three polyadenylation sites. There is some developmental variability in the use of the three polyadenylation sites.

In order to determine if each exon is preceded by a functional promoter and to identify sequences important for transcription initiation from each exon, we made fusions between act5C promoter fragments and the bacterial chloramphenicol acetyltransferase (CAT) gene and tested these for promoter activity in transient assays in Kc cells. We found that each exon is preceded by a separate, functional promoter. At least two regions of DNA sequences are necessary for optimal expression from exon 1. One of these lies greater than 1.9 kb upstream from the exon 1 cap site. All of the sequences required for exon 2 transcription lie within 450 bases of its cap site. There is evidence from some constructions that transcription initiation from exon 1 may inhibit transcription initiation from exon 2.

TABLE OF CONTENTS

CHAPTER 1

Introduction

Multigene families are valuable systems in which to study the differential expression and the evolution of genes. In some cases, as with the globin genes (for review, see 25), a closely related set of genes is turned on sequentially with developmental stage. In other cases, as with the actins (12) and tubulins (7), the individual members of the gene families are expressed with unique time dependent and tissue specific programs. There are two possible reasons for the amplification of an ancestral gene and subsequent diversification in sequence into a multigene family. The small number of differences in amino acid sequence may be significant for the different functions of the isoforms in different tissues. Alternatively, the formation of a multigene family may be a convenient mechanism for evolution of different regulatory sequences for tissue specific expression, with the amino acid differences of no functional significance. In the case of the immunoglobulins (8), it is very clear that having multiple closely related genes is important for protein diversity. In either case multigene families are good systems in which to study how gene structure and primary amino acid sequences change over evolutionarily significant time periods and to compare sequences which may be important for regulated gene expression.

The actin genes are good candidates for such a study. Actin is a 42,000 dalton protein which is present in all eucaryotic cells where it performs a wide variety of functions. In muscle cells it is a major component of the thin filaments of myofibrils. Actin is also present in large quantities in nonmuscle cells where it comprises the microfilaments. These "cytoplasmic" actins are implicated in a diversity of cellular processes such as cytoplasmic streaming, cytokinesis, and phagacytosis and are generally associated with areas of cell movement (26,45). There is also evidence to suggest that actin may be a transcription factor for polymerase II genes. (9,40).

Given the diversity of roles played by this protein, it is not surprising that many organisms have multiple actin proteins. In mammals, three different actin protein species, designated alpha, beta and gamma, of nearly identical molecular weight but different isoelectric points can be resolved on two dimensional gels. Vandekerckhove and Weber showed that there are at least six different actin protein isoforms in mammals some of which comigrate on gels (47). They are the alpha skeletal muscle, alpha cardiac muscle, alpha smooth muscle, gamma smooth muscle, beta cytoplasmic and gamma cytoplasmic forms. Recently, a seventh actin protein, which appears to be another cytoplasmic form, has been identified in chicken by nucleic acid studies (3).

The primary amino acid sequences of these proteins show a high degree of conservation both within and between species. The highest density of amino acid replacements that are present are found at the extreme amino terminus of the protein. The various mammalian muscle forms differ by the order of 4-6 amino acids (out of 375) as do the cytoplasmic proteins. Muscle and cytoplasmic sequences differ from each other by roughly 25 residues (34). Actin proteins from several lower eucaryotes (*Physarum* and *Dictyostelium*) resemble the mammalian cytoplasmic actins which is not surprising given that these organisms do not have muscle tissue (12).

Because of their relative abundance and the high degree of homology between species, actin genes from a large number of species have been isolated. For examples see 3,5,6,11,12,14,15,17,22,23,29,32,33,41, and 46. The number of genes per haploid genome ranges from 1 in yeast to 17 in Dictyostelium and an as yet undefined large number in mammals (>20 in humans). Many of the mammalian genes are known to be pseudogenes but it is still quite possible that new protein isoforms will be discovered as more genes are characterized. In general, there tend to be many more cytoplasmic actin

related sequences than muscle genes corresponding to the multitude of functions carried out by cytoplasmic actins. For example, humans have greater than 20 each of beta and gamma related sequences but only one each of the alpha skeletal and alpha cardiac genes (10,11,38). Sea urchins have five cytoplasmic genes and only one muscle gene (24).

The accumulation of actin mRNAs has been studied in many organisms. Actin mRNAs generally are very abundant and in some cases may represent as much as 10% of the cellular poly A+ RNA. In many species several actin mRNAS of different sizes are made. Generally, the muscle actin mRNAs have been found to be smaller than the cytoplasmic RNAs (21).

In mammalian and avian systems, many studies on actin gene expression have been carried out in primary muscle cells and in established muscle cell lines. The latter can be maintained in a proliferative state indefinitely or induced to differentiate into muscle by mitogen starvation and/or growth to confluence. In replicating prefusion myoblasts, the gamma and beta actins are the predominant species expressed. As the cells stop replicating and undergo terminal differentiation, the cytoplasmic actins are turned off and the muscle specific genes are induced (1,42).

These cell culture systems have proven to be very useful for the study of the developmental switch in expression from the beta to the alpha actin genes. They have recently been used to study sequences required in *cis* for the induction of the alpha genes and the down regulation of beta actin in differentiating muscle. Several groups have demonstrated induction of alpha actins (chick skeletal 4,35, rat skeletal 28, human cardiac 30,31). Proper gene regulation has been demonstrated for the chick alpha gene with as little as 200 base pairs of 5' flanking sequences present (4). This implies that all the sequences required for induction of some muscle genes lie in the 5' untranslated

regions. Sharp *et al.* and Paterson *et al.* have shown that beta actin genes which have been transfected into muscle cell lines will down regulate when the cells start to differentiate (37,43,44). There is evidence from this work that some of the sequences which are important for deinduction lie 3' to the gene.

The 3' and 5' untranslated sequences of various mammalian and avian actin genes have been searched for DNA homologies. A comparison of the 5' untranslated sequences of the chicken actin genes showed that there is very little homology in these sequences between the different genes. This result also has held true in other mammalian species. There is very little homology in the untranslated sequences of actins of different subtypes even in the same species. In contrast, 3' UTRs of genes of similar function in different species show a remarkable degree of homology (18,20,27,36,38,39). For example, there is a 200 base stretch that is 76% homologous between the chick cardiac and the human cardiac genes in the 3' untranslated sequences. There is also significant homology in the 5' flanking regions of the chicken and rat skeletal alpha genes (36). Finally, there are short repeated conserved sequences that are found in the 5' flanking sequences of alpha genes (20,30). It seems quite likely that those conserved sequences play a role in the developmental regulation of those isoforms. The conservation of these sequences over large evolutionary distances suggests that they may be promoter elements associated with muscle specific expression of the genes. Another more general sequence motif CC(Arich)GG was found in many mammalian actins (30).

In *Drosophila*, there are several different actin protein isoforms which are expressed in a tissue specific manner (19). As in mammals, there are three major actin protein bands that can be resolved on two dimensional gels. Type I (the most acidic) is found in primary myogenic cultures postfusion (13), and in supercontractile muscles such as the larval body wall, adult abdominal and

larval and adult visceral muscles (19). Isoactins II and III are generalized cellular actins found in all cell types and in most cases type III is an unstable precursor to type II (2, 19). There is also a stable form of actin III which is found in adult thorax and is probably unique to flight muscle tissue.

One of the *Drosophila* actin genes was isolated by Fyrberg *et al.* (14). They showed that the *Drosophila* actin gene family has six members which are dispersed throughout the three major chromosomes.

We were interested in further study of this gene family for several reasons. When we began this study, actin was one of the most highly conserved proteins known over such large evolutionary distances. We were interested in studying how the structures and sequences of the genes compared with other known actins from different species. We were also interested in determining how the different protein isoforms related to those characterized in mammals. Furthermore, we knew that *Drosophila* had several different actin protein isoforms that are found at particular developmental times and in specific tissues. We presumed that the separate genes which encode the different isoforms would have different patterns of expression in the organism. These would be good candidates for studies on the regulation of gene expression.

In the work described in Chapter 2, we investigated general structural features of the gene family and compared them to other known actin gene structures. We were very surprised to find that unlike other gene families which had been studied at the time, the actin family does not show conservation of intron positions. At least three of the genes were shown to have introns which were all in different positions from each other and from the single intron which had been found in yeast. Furthermore, the primary amino acid sequences of the amino terminal ends of the proteins as deduced from the

DNA sequences showed that all of the genes encode proteins that, in this respect, resemble mammalian cytoplasmic genes. This is surprising given that *Drosophila* has muscles that are very similar in appearance to mammalian striated muscle. Because they all resembled cytoplasmic actins, a different criterion other than protein sequence was needed to determine the functions of each of the six genes.

In Chapter 3 we investigate the functional identity of each gene by determining the times in development and the tissues in which each is expressed. We describe the generation of gene specific hybridization probes which are taken from the 3' untranslated regions of the genes. These are used in *in vitro* translation assays to show that each of the genes codes for an actin protein of the expected size and isoelectric point. This implies that each of the genes contains the sequences necessary to encode a functional actin protein. The probes are also used to study RNAs from different developmental stages and tissues. We find that two of the genes are expressed in early embryos, early pupae, and ovaries. These are all conditions under which a cytoplasmic gene would be expressed. Two of the genes are expressed only in late pupae and adults. One is expressed in thorax only and the other in thorax and leg. The former is thought to be the major actin expressed in the indirect flight muscle while the other is the major actin of the jump muscle. The final two genes are expressed strongly in larvae and in adult abdomens and are probably the actins expressed in the supercontracting muscles.

Having determined when each of the genes was expressed, we were interested in studying *cis*-acting sequences involved in the regulation of gene expression. We decided to use a transient transformation system in *Drosophila* cultured cells because many constructions could be tested in a relatively short period of time. Before a detailed study of this kind could be initiated, more

precise information was needed about the actin transcriptional units. We decided to concentrate on the cytoplasmic act5C gene for several reasons. Based on RNA blots, the act5C gene appeared to have the most complex transcriptional unit. Three different sized bands were observed in contrast to the single species seen for the other actin genes. Furthermore, these transcripts appeared to be differentially developmentally regulated.

In Chapter 4 the structure of the act5C gene is studied in more detail. We find that at least six different transcripts are synthesized from this gene. We show that unlike other actin genes characterized to date, this gene has two separate transcription start sites which are associated with two leader exons. These are alternately spliced to the body of the gene. Three sites of polyadenylation are also used, accounting for the six mRNA species. We find that transcripts containing each of the two leader exons exhibit the same pattern of developmental regulation. The use of the three polyadenylation sites in contrast is developmentally regulated.

In Chapter 5 we examine whether both leader exons are preceded by functional promoters and we begin to define *cis*-acting sequences important for the expression from each of the transcription start sites. We find that each leader exon is preceded by a separate, functional promoter. By deletion analyses, we find that at least two regions of 5' flanking sequences are necessary for optimal expression of exon 1, one of which is greater than 1.9 kb upstream of its cap site. In contrast, 450 bases of sequences 5' to the exon 2 cap site contain all the sequences required for its expression.

REFERENCES

1. Bains, W., P. Ponte, H. Blau, and L. Kedes. 1984. Cardiac actin is the major actin gene product in skeletal muscle cell differentiation *in vitro*. Mol. Cell. Biol. 4:1449-1453.

2. Berger, E. and G. Cox. 1979. A precursor of cytoplasmic actin in cultured *Drosophila* cells. J. Cell Biol. 81: 680-683.

3. Bergsma, D.J., K.S. Chang, and R.J. Schwartz. 1985. Novel Chicken Actin Gene: Third cytoplasmic isoform. Mol. Cell. Biol. 5:1151-1162.

4. Bergsma, D.J., J.M. Grichnik, L.M. Gossett, and R.J. Schwartz. 1986. Delimitation and characterization of *cis*-acting sequences required for the regulated expression and transcriptional control of the chicken skeletal α-actin gene. Mol. Cell. Biol. 6:2462-2475.

5. Chang, K.S., W.E. Zimmer Jr., D.J. Bergsma, J.B. Dodgson, and R.J. Schwartz. 1984. Isolation and characterization of six different chicken actin genes. Mol. Cell. Biol. 4:2498-2508.

6. Cleveland, D.W., M.A. Lopata, R.J. MacDonald, N.J. Cowan, W.J. Rutter, and M.W. Kirschner. 1980. Number and evolutionary conservation of α- and β-tubulin and cytoplasmic β- and γ-actin genes using specific cloned cDNA probes. Cell 20:95-105.

7. Cleveland, D.W. 1983. The tubulins: From DNA to RNA to protein and back again. Cell 34:330-332.

8. Cushley, W. and A.R. Williamson. 1982. Expression of immunoglobulin genes. Essays in Biochemistry 18:1-39.

9. Egly, J.M., N.G. Miyamoto, V. Moncollin, and P. Chambon. 1984. Is actin a transcription initiation factor for RNA polymerase B?. EMBO 3:2363-2371.

10. Engel, J., P. Gunning, and L. Kedes. 1982. Human cytoplasmic actin proteins are encoded by a multigene family. Mol. Cell. Biol. 2:674-684.

11. Engel, J.N., P. W. Gunning, and L. Kedes. 1981. Isolation and characterization of human actin genes. Proc. Natl. Acad. Sci. USA 78:4674-4678

12. Firtel, R.A. 1981. Multigene families encoding actin and tubulin. Cell 24:6-7.

13. Fyrberg, E.A. and J.J. Donady. 1979. Actin heterogeneity in primary embryonic culture cells from *Drosophila melanogaster*. Dev. Biol. 68:487-502.

14. Fyrberg, E.A., K.L. Kindle, N. Davidson and A. Sodja. 1980. The actin genes of *Drosophila*: A dispersed multigene family. Cell 19:365-378.

15. Fyrberg, E.A., B.J. Bond, N.D. Hershey, K.S. Mixter, and N. Davidson. 1981. The actin genes of *Drosophila*: Protein coding regions are highly conserved but intron positions are not. Cell 24:107-116.

16. Fyrberg, E.A., J.W. Mahaffey, B.J. Bond and N. Davidson. 1983. Transcripts of the six *Drosophila* actin genes accumulate in a stage and tissue specific manner. Cell 33:115-123.

17. Gallwitz, D. and I. Sures. 1980. Structure of a split yeast gene: Complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U.S.A. 77:2546-2550.

18. Gunning, P., T. Mohun, S.-Y. Ng, P. Ponte, and L. Kedes. 1984. Evolution of the human sarcomeric-actin genes: Evidence for units of selection within the 3' untranslated regions of the mRNAs. J. Mol. Evol. 20:202-214.

19. Horovitch, S.J., R.V. Storti, A. Rich, and M.L. Pardue. 1979. Multiple actins in *Drosophila melanogaster*. J. Cell. Biol. 82:86-92.

20. Hu, M.C.-T., S.B. Sharp, and N. Davidson. 1986. The complete sequence of the mouse skeletal α-actin gene reveals several conserved and inverted repeat sequences outside of the protein-coding region. Mol. Cell. Biol. 6:15-25.

21. Hunter, T. and J.I. Garrels. 1977. Characterization of the mRNAs for alpha-, beta-, and gamma-actin. Cell 12:767-781.

22. Katcoff, D., U. Nudel, D. Zevin-Sonkin, Y. Carmon, M. Shani, H. Lehrach, A.M. Frischauf, and D. Yaffe. 1980. Construction of recombinant plasmids containing rat muscle actin and myosin light chain DNA sequences. Proc. Natl. Acad. Sci. USA 77:960-964.

23. Kindle, K.L. and R.A. Firtel. 1978. Identification and analysis of Dictyostelium actin genes, a family of moderately repeated genes. Cell 15:763-778.

24. Lee, J.J., R.J. Shott, S.J. Rose III, T.L. Thomas, R.J. Britten, and E.H. Davidson. 1984. Sea urchin actin gene subtypes; gene number, linkage and evolution. J. Mol. Biol. 172:149-176.

25. Maniatis, T., E.F. Fritsch, J. Lauer, and R.M. Lawn. 1980. The molecular genetics of human hemoglobins. Ann. Rev. Gen. 14:145-178.

26. Marx, J.L. 1975. Actin and myosin: Role in nonmuscle cells. Science 189:34-37.

27. Mayer, Y., H. Czosnek, P.E. Zeelon, D. Yaffe, and U. Nudel. 1984. Expression of genes coding for the skeletal muscle and cardiac actins in the heart. Nucleic Acids Res. 12:1087-1100.

28. Melloul, D., B. Aloni, J. Calvo, D. Yaffe, and U. Nudel. 1984. Developmentally regulated expression of chimeric genes containing muscle actin DNA sequences in transfected myogenic cells. EMBO 3:983-990.

29. Minty, A.J., S. Alonso, J.L. Guenet, and M.E. Buckingham. 1983. Number and organization of actin-related sequences in the mouse genome. J. Mol. Biol. 167:77-101.

30. Minty, A. and L. Kedes. 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. Mol. Cell. Biol. 6:2125-2136.

31. Minty, A., H. Blau, and L. Kedes. 1986. Two-level regulation of cardiac actin gene transcription: Muscle-specific modulating factors can accumulate before gene activation. Mol. Cell. Biol. 6:2137-2148.

32. Nagao, R.T., D.M. Shah, V.K. Eckenrode, and R.B. Meagher. 1981. Multigene family of actin-related sequences isolated from a soybean genomic library. DNA 2:1-9.

33. Nudel, U., D. Katcoff, R. Zakut, M. Shani, Y. Carmon, M. Finer, H. Czosnek, I. Ginsburg, and D. Yaffe. 1980. Isolation and characterization of rat skeletal muscle and cytoplasmic actin genes. Proc. Natl. Acad. Sci. USA 79:2763-2767.

34. Nudel, U., R. Zakut, M. Shani, S. Neuman, Z. Levy, and D. Yaffe. 1983. The nucleotide sequence of the rat cytoplasmic beta-actin gene. Nucleic Acids Res. 11:1759-1771.

35. Nudel, U., D. Greenberg, C.P. Ordahl, O. Saxel, S. Neuman, and D. Yaffe. 1985. Developmentally regulated expression of a chicken muscle-specific gene in stably transfected rat myogenic cells. Proc. Natl. Acad. Sci. USA 82:3106-3109.

36. Ordahl, C.P. and T.A. Cooper. 1983. Strong homology in promoter and 3'-untranslated regions of chick and rat α-actin genes. Nature 303:348-349.

37. Paterson, D., J. Eldridge, and A. Seiler-Tuyns. 1986. Regulation of actin gene expression: analysis of potential regulatory regions. In Emerson, C., D. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui (eds): "Molecular Biology of Muscle Development," New York: Alan R. Liss, Inc., 521-529.

38. Ponte, P., P. Gunning, H. Blau, and L. Kedes. 1983. Human actin genes are single copy for α-skeletal and α-cardiac actin but multicopy for β- and γ-cytoskeletal genes: 3' untranslated regions are isotype specific but are conserved in evolution. Mol. Cell. Biol. 3:1783-1791.

39. Ponte, P., S.-Y. Ng, J. Engel, P. Gunning, and L. Kedes. 1984. Evolutionary conservation in the untranslated regions of actin mRNAs: DNA sequence of a human beta-actin cDNA. Nucleic Acids Res. 12:1687-1697.

40. Scheer, U., H. Hinssen, W.W. Franke, and B.M. Jockusch. 1984. Microinjection of actin-binding proteins and actin antibodies demonstrates involvement of nuclear actin in transcription of lampbrush chromosomes. Cell 39:111-122.

41. Scheller, R.H., L.B. McAllister, W.R. Crain, Jr., D.S. Durica, J.W. Posakony, T.L. Thomas, R.J. Britten, and E.H. Davidson. 1981. Organization and expression of multiple actin genes in the sea urchin. Mol. Cell. Biol. 1:609-628.

42. Schwartz, R.J. and K.N. Rothblum. 1981. Gene switching in myogenesis: Differential expression of the chicken actin multigene family. Biochemistry 20:4122-4129.

43. Seiler-Tuyns, A., Eldridge, J.D., and B.M. Paterson. 1984. Expression and regulation of chicken actin genes introduced into mouse myogenic and nonmyogenic cells. Proc. Natl. Acad. Sci. USA 81:2980-2984.

44. Sharp, S.B., T.A. Kost, S.H. Hughes, C.P. Ordahl, and N. Davidson. 1986. Regulation of intact and hybrid β and α actin genes inserted into myogenic cells. In R. A. Firtel and E.H. Davidson (eds): "Molecular approaches to developmental biology," New York: Alan R. Liss, Inc., in press.

45. Stossel, T.P. 1978. Contractile proteins in cell structure and function. Ann. Rev. Med. 29:427-457.

46. Tobin, S.L., E. Zulauf, F. Sanchez, E.A. Craig, and B.J. McCarthy. 1980. Multiple actin-related sequences in the *Drosophila melanogaster* genome. Cell 19:121-131.

47. Vandekerckhove, J. and K. Weber. 1978. At least six different actins are expressed in a higher mammal: An analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J. Mol. Biol. 126:783-802.

CHAPTER 2

The actin genes of *Drosophila*: Protein coding regions are highly conserved but intron positions are not

(published in Cell)

Cell, Vol. 24, 107–116, April 1981, Copyright © 1981 by MIT

The Actin Genes of Drosophila: Protein Coding Regions Are Highly Conserved but Intron Positions Are Not

Eric A. Fyrberg, Beverley J. Bond,* N. Davis Hershey, Katharine S. Mixter* and Norman Davidson
Department of Chemistry and
* Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

The entire set of six closely related Drosophila actin genes was isolated using recombinant DNA methodology, and the structures of the respective coding regions were characterized by gene mapping techniques and by nucleotide sequencing of selected portions. Structural comparisons of these genes have resulted in several unexpected findings. Most striking is the nonconservation of the positions of intervening sequences within the protein-encoding regions of these genes. One of the Drosophila actin genes, DmA4, is split within a glycine codon at position 13; none of the remaining five genes is interrupted in the analogous position. Another gene, DmA6, is split within a glycine codon at position 307; at least two of the Drosophila actin genes are not split in the analogous position. Additionally, none of the Drosophila actin genes is split within codon four, where the yeast actin gene is interrupted. The six Drosophila actin genes encode several different proteins, but the amino acid sequence of each is similar to that of vertebrate cytoplasmic actins. None of the genes encodes a protein comparable in primary sequence to vertebrate skeletal muscle actin. Surprisingly, in each of these derived actin amino acid sequences the initiator methionine is directly followed by a cysteine residue, which in turn precedes the string of three acidic amino acids characteristic of the amino termini of mature vertebrate cytoplasmic actins. We discuss these findings in the context of actin gene evolution and function.

Introduction

Much remains to be learned about the expression of eucaryotic genes, and in particular how the activities of thousands of such genes are coordinated to mediate the ordered sequence of events leading to metazoan ontogeny. Increasingly, eucaryotic structural genes, or portions thereof, are being found not to be uniquely represented in animal genomes, but rather to be members of small gene families (reviewed by Long and Dawid, 1980). In some cases the members of such families are expressed differentially, either in temporal sequence or tissue specific distribution (Maniatis, et al., 1980; Hagenbüchle et al., 1980). Continued study of such gene families will increase our understanding of gene evolution and of the regulatory elements that control gene expression.

The actin genes are well suited for such a study. Because actin is a highly conserved and ubiquitous protein (Pollard and Weihing, 1974), the structure of these genes can be examined across very large evolutionary distances. Furthermore, actin synthesis is highly regulated during metazoan development. As various types of muscle cells differentiate they begin synthesis of specific actin isoforms (Whalen et al., 1976; Garrels and Gibson, 1976; Saborio et al., 1979), each of which is encoded by one or more members of a small family of closely related genes (Vandekerckhove and Weber, 1978a, 1978c). Further investigations of structural and functional aspects of these genes will help elucidate the mechanisms by which genes are selectively expressed.

We have compared the structural features of the actin genes of Drosophila melanogaster. We report several unexpected results which are relevant to the evolution and functioning of the actin gene family.

Results

Isolation of All Members of the Drosophila Actin Gene Family

To isolate all members of the Drosophila actin gene family we screened a library of Drosophila genomic DNA using the technique of Benton and Davis (1977). The probe for this screen was derived from λDmA2, a recombinant phage that contains a Drosophila actin gene, and that we have characterized previously (Fyrberg et al., 1980). Of 40,000 plaques screened (approximately four Drosophila genome equivalents), 30 phages that hybridized strongly were selected. The Eco RI restriction patterns of the purified DNA from these phages indicated that they were members of six distinct classes. Furthermore, blot hybridization experiments have demonstrated that DNA pooled from representatives of these six phage classes contains all of the actin-gene-containing fragments observed in digests of Drosophila genomic DNA (data not shown). We therefore conclude that these six phages contain all of the Drosophila actin genes.

We have denoted the six actin-gene-containing phages as λDmA1–λDmA6 in order of the decreasing sizes of their actin-gene-containing Eco RI fragments. By in situ hybridization to polytene chromosomes we have determined the chromosomal location of the Drosophila DNA segment inserted within each phage genome: λDmA1, 88F; λDmA2, 5C; λDmA3, 42A; λDmA4, 57A; λDmA5, 87E; λDmA6, 79B (data not shown). These results confirm and extend those of Tobin et al. (1980) and Fyrberg et al. (1980) who reported that actin gene probes hybridized to these six locations.

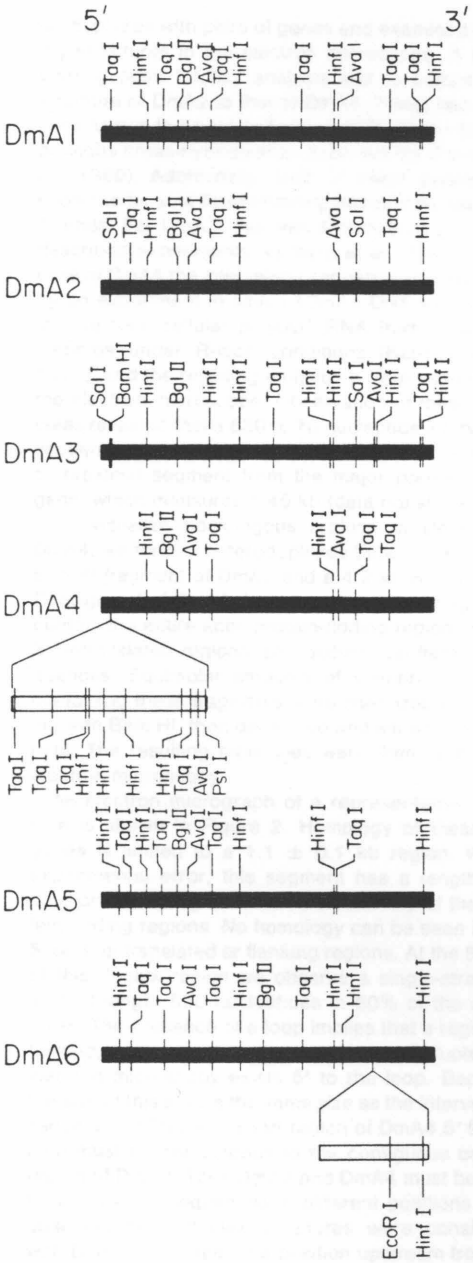


Figure 1. High Resolution Restriction Maps of the Protein Coding Regions of DmA1–DmA6

Solid blocks represent the 1.1 kb protein coding regions of the actin genes, which are aligned to facilitate comparison of restriction sites. The 5′ to 3′ orientation is left to right as indicated. Open blocks represent intervening sequences, which are inserted at the positions indicated. The intervening sequence of DmA4 measures approximately 630 nucleotides while that of DmA6 is 357 nucleotides in length. A small intervening sequence within codon 307 of DmA1 which has recently been discovered by F. Sánchez, S. L. Tobin and B. J. McCarthy (personal communication) is not shown. Intervening sequences within untranslated regions are not shown, since these regions have been analyzed only in DmA2 and DmA4. DmA2 is known to have an intervening sequence in its 5′ untranslated region, however (see text). There is an Ava1 site at codon 333 of DmA1 which is not shown in the figures.

High Resolution Restriction Mapping of the Protein-Coding Region of Each Actin Gene Reveals Many Conserved Sites

Low resolution restriction mapping and blot hybridization experiments using the chimeric phage DNAs revealed that each contained a single actin structural gene (data not shown; however, restriction maps of λDmA1—DmA6 are available on request). After localizing each gene within a reasonably small (2–6 kb) restriction fragment, subclones were prepared by inserting fragments into the plasmid vector pBR322. Subsequently the protein coding region of each of the structural genes was localized within the subcloned fragments and mapped precisely with frequently cutting restriction enzymes. Appropriate fragments were mapped with Ava I, Bgl II, Hinf I and Taq I using the technique of Smith and Birnstiel (1976). These results are presented in Figure 1.

As expected for a family of closely related genes, these maps reveal considerable conservation of restriction sites. The conserved sites allow maps of the Drosophila actin genes to be aligned with those of Dictyostelium (Firtel et al., 1979) and yeast (Ng and Abelson, 1980). This alignment revealed the tentative positions of the 5′ and 3′ ends of the protein-coding portion of each Drosophila actin gene and thus determined their transcriptional polarities. Interestingly there are four well conserved Hinf I sites within codons 51, 154, 259 and 363 of the coding regions. Hinf I recognizes the nucleotide sequence GANTC. In each of these four cases this sequence is generated by consecutive (Asp or Glu)–Ser codons. The extremely regular spacing of these (Asp or Glu)–Ser sequences in the actin proteins (104 ± 1 amino acids) may be of functional significance.

These restriction maps allowed us to identify tentatively an intervening sequence within DmA6. Alignment of restriction enzyme sites, some of which are not shown in Figure 1 (Ava II, Alu I, Hpa II, Hae III) revealed that an approximately 400 nucleotide insertion split this gene into segments of 900 and 225 nucleotides. No such insertion is indicated by maps of the other five genes. A larger intervening sequence within DmA4 was overlooked in this analysis, however, because it lies very close to the 5′ end of the coding region (refer to Figure 1).

Direct Visualization of Nonconserved Interruptions within DmA4 and DmA6

To map more precisely the sequence relationships among the Drosophila actin genes, we prepared het-

eroduplexes with pairs of genes and examined resulting structures in the electron microscope. A logical starting point for this analysis was to compare the structure of DmA2 to that of DmA4. These two genes were known to be related very closely on the basis of previous cross-hybridization experiments (Fyrberg et al., 1980). Additionally, both of these genes were known to be split by intervening sequences near their 5′ ends. For DmA2, this was known from previously described experiments (Fyrberg et al., 1980). In the case of DmA4 the intervening sequence was identified by an experiment in which λDmA4 DNA was hybridized to total cellular poly(A)$^+$ RNA from Drosophila embryos under R-loop conditions (Kaback et al., 1979), and the resulting molecules were examined in the electron microscope. Observation of these molecules revealed that a 630 ± 70 nucleotide intervening sequence separates an approximately 100 nucleotide 5′ proximal segment from the major portion of the gene, which measures 1.45 kb (data not shown).

To visualize homologous regions of DmA2 and DmA4, we formed heteroduplexes between an 8.7 kb Eco RI fragment of DmA2 and a 4.2 kb Bam HI–Eco RI fragment of DmA4. In both cases these fragments contain the entire actin protein-coding region, 5′ and 3′ untranslated regions, plus substantial flanking sequences. Equimolar amounts of chimeric plasmids containing these fragments were linearized by cleaving with Bam HI, then denatured and allowed to reanneal. The resulting molecules were then spread for electron microscopy.

An electron micrograph of a representative molecule is shown in Figure 2. Homology of these two genes is limited to a 1.1 ± 0.1 kb region. Within experimental error, this segment has a length and position indicating that it consists entirely of the protein-coding regions. No homology can be seen in the 5′ or 3′ untranslated or flanking regions. At the 5′ end of this duplex region we observe a single-stranded loop of length 630 nucleotides in 60% of the molecules. The presence of a loop implies that a region of homology too short to be recognized as duplex by electron microscopy exists 5′ to the loop. Because the size of this loop is the same size as the intervening sequence of DmA4, a short region of DmA4 5′ to the loop must be homologous to the contiguous coding region of DmA2. Thus DmA2 and DmA4 must be split by intervening sequences in different positions. The observed heteroduplex structures were consistent with DmA2 being split at a position upstream from its coding region in the 5′ untranslated region, and with DmA4 being split within the protein-coding region extremely close to the 5′ end. This interpretation is fully confirmed by the sequence data presented below.

Heteroduplexes of DmA1 with DmA5, DmA5 with DmA6, and DmA3 with DmA6 have also been prepared. In all cases a duplex segment corresponding

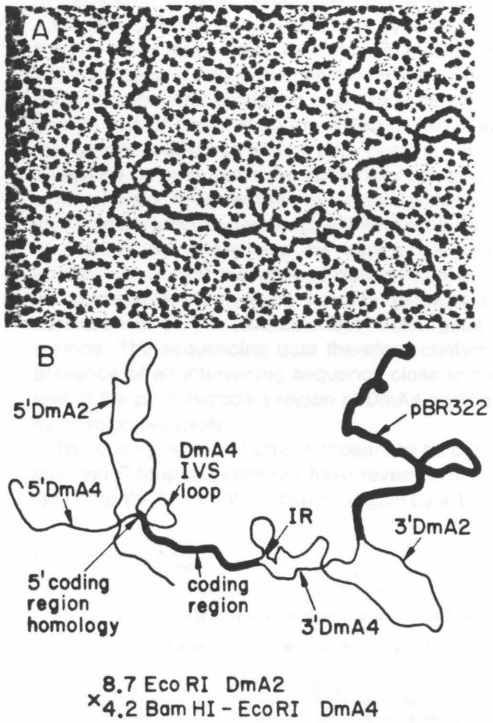


Figure 2. Heteroduplex of the DmA2 and DmA4 Actin Genes

Two chimeric plasmids, one containing an 8.7 kb Eco RI fragment of DmA2 and one a 4.2 kb Bam HI–EcoRI fragment of DmA4 were linearized with Bam HI and used to prepare heteroduplexes. A representative molecule is shown in (A) and depicted schematically in (B). The long duplex region on the right is the pBR322 vector sequence. The shorter duplex region in the lower center corresponds in size and position to the protein coding regions of the two genes. The 3′ untranslated and flanking regions are nonhomologous. Short inverted repeat segments separated by approximately 600 nucleotides form a stem-loop structure immediately 3′ to the DmA2 coding region. At the 5′ end of the coding regions a loop of the same size as the DmA4 intervening sequence (630 nucleotides) can be seen. The loop forms because a short segment 5′ to the DmA4 intervening sequence hybridizes to the 5′ end of the DmA2 coding region. The 5′ untranslated and flanking regions of the two genes are completely nonhomologous.

to the protein-coding regions terminates in forked single strands. The interpretation again is that the protein-coding regions of the several genes are homologous, but that the untranslated and flanking sequences are not. Figure 3A shows a heteroduplex of Drosophila actin genes DmA3 and DmA6. A 360 ± 40 nucleotide loop of single-stranded DNA can be seen near the 3′ end of the DmA6 coding region. This sequence splits the gene into segments of approximately 910 and 210 nucleotides in agreement with the restriction mapping data presented in Figure 1. As shown below, the nonhomologous segment within the coding region of DmA6 is an intervening sequence.

Heteroduplex analysis of DmA1 and DmA5 (Figure 3B) has failed to reveal intervening sequences within either coding region.

In summary, these experiments allowed tentative identification of a 630 nucleotide intervening sequence near the 5′ end of the DmA4 protein coding region and one of 360 nucleotides near codon 300 of DmA6. No other introns were seen within the coding regions of any of the remaining actin genes. These experiments would not allow visualization of introns less than 100 nucleotides in length. Finally, the results of the DmA2/DmA4 heteroduplex experiment indicated that the previously identified intervening sequence of DmA2 is located upstream from the protein-coding region and lies within the 5′ untranslated region.

Precise Mapping of the Introns of DmA2, DmA4 and DmA6

The precise positions of the introns of DmA2, A4, and A6 deduced from the DNA sequence data are shown in Figure 4. In panel A we show nucleotide sequence data for the regions that correspond to the beginning of translation in DmA2 and DmA4. In the former, the ATG initiator codon is followed by a cysteine codon, then by a sequence which by comparison with known actin amino acid sequences positively identifies it as an actin gene. The sequence of DmA4 is nearly identical to that of DmA2 from the amino terminus up to the glycine codon at amino acid position 13. The sequence beyond that point does not encode amino acids of actin; however there is an acceptable exon–intron splice junction sequence at this point (Lerner et al., 1980). By sequencing the appropriate region, approximately 600 nucleotides downstream we find an acceptable intron–exon junction followed by a continuation of the characteristic actin gene sequence. The sequencing data therefore confirm the presence of an intervening sequence close to the 5′ end of the protein-coding region of DmA4 and define its position precisely.

The coding region of DmA2 appears to be continuous, yet R-loop experiments have revealed that it is split near the 5′ end of the coding region by a 1.65 kb

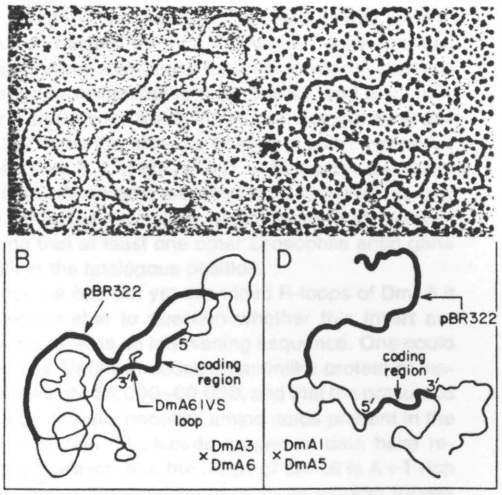


Figure 3. DmA3/DmA6 and DmA1/DmA5 Heteroduplexes

The DmA3/DmA6 heteroduplex was prepared by hybridizing a plasmid containing a 6.0 kb Eco RI fragment of DmA3 with one containing a 5.0 kb Bam HI fragment of DmA6. Plasmids were separately digested with Bgl II, which cleaves the plasmids in their respective actin coding regions, but in different positions (refer to Figure 1). Resulting heteroduplexes have long staggered ends which anneal to form circular structures. (A) shows such a molecule and (B) its schematic representation. One of the actin structural genes is interrupted by a 360 nucleotide intervening sequence, which appears as a small single-stranded loop. Since the same loop is apparent in DmA5/DmA6 heteroduplexes, we were able to assign it unambiguously to DmA6. The DmA1/DmA5 heteroduplex shown in (C) and (D) was prepared by hybridizing a plasmid containing a 3.6 kb Bam HI fragment of DmA1 to one containing a 5.2 kb Eco RI fragment of DmA5. This heteroduplex failed to reveal intervening sequences within either actin coding region.

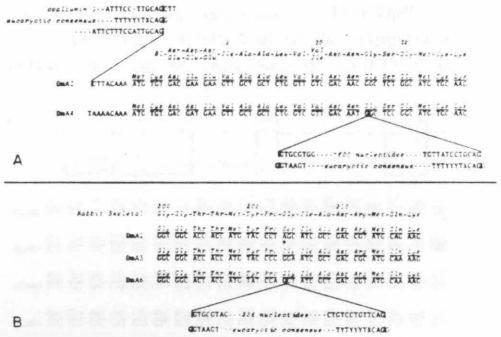


Figure 4. Nucleotide Sequences Surrounding Interruptions within Drosophila Actin Genes

(A) shows the nucleotide sequences of the 5′ ends of DmA2 and DmA4. The sequence of the amino terminal tryptic peptide encoded by each gene is aligned with that of vertebrate cytoplasmic actin (Bl represents the acetyl group which blocks the N terminus). Eight nucleotides upstream from the ATG initiation codon of DmA2 is a sequence which appears to be a functional intron–exon junction sequence based on previous R-loop experiments and on the homology of this sequence to the eucaryotic consensus sequence and to the ovalbumin G intron–exon junction sequence. (Y denotes a position where either pyrimidine can occur and X a position where either a purine or a pyrimidine is acceptable.) The lower portion of (A) illustrates that DmA4 is interrupted within the glycine codon at position 13 by an approximately 630 nucleotide intron. Inspection of the sequences of the 5′ and 3′ ends of this interruption reveals that they share homology with the eucaryotic consensus exon–intron and intron–exon sequences. (B) shows the sequence of codons 300–314 for DmA2, DmA3 and DmA6. The sequence of each encoded protein is identical to the corresponding sequences of vertebrate actins except for the serine at position 307 in DmA2 (however, we are unsure of this amino acid assignment due to the sequencing ambiguity denoted by the asterisk). Comparison of the sequences of the three Drosophila genes reveals that DmA6 is interrupted within the glycine codon at position 307. Inspection of the sequence of the 5′ and 3′ ends of this interruption again reveals that they share homology with the eucaryotic consensus exon–intron and intron–exon sequences.

intron (Fyrberg et al., 1980). From results of the heteroduplex experiment shown in Figure 5 we suspected that the gene was split in the 5′ untranslated region. Figure 4 shows that 8 nucleotides upstream from the ATG initiation codon of DmA2 there is a sequence homologous to an intron–exon junction. Fortuitously, we presume, the proposed DmA2 intron–exon junction almost exactly matches that of ovalbumin G (Breathnach et al., 1978). No other reasonable intron–exon splice junction was found for the 200 nucleotides upstream from this position. We therefore are reasonably certain that this junction is the intron–exon junction of DmA2.

The nonconservation of the intron within DmA6 is illustrated in Figure 4B. Comparison of the nucleotide sequence through the region near codon 300 of DmA6 with that of other actin genes shows that the interruption is within a glycine codon at position 307, and that an acceptable exon–intron junction exists at this point. Exactly 357 nucleotides downstream we find an intron–exon junction and the continuation of the actin gene sequence. Sequencing of DmA2 and DmA3 through the corresponding region shows that neither of these genes has an interruption in the same position. Although we have not yet sequenced DmA1, DmA4 and DmA5 through this region, none of these genes displayed evidence for an intron at this position in heteroduplex mapping experiments, and any intervening sequences which interrupt these genes at codon 307 must therefore be extremely short. However, F. Sánchez, S. L. Tobin and B. J. McCarthy (personal communication) have recently found that DmA1 is split by a 60 nucleotide intron within codon 307, demonstrating that at least one other Drosophila actin gene is split in the analogous position.

Since we had not yet examined R-loops of DmA6 it was reasonable to question whether this insert actually represents an intervening sequence. One could argue that DmA6 encodes an actinlike protein of molecular weight 55,000–60,000, and that the presumed insertion actually encodes amino acids present in the mature protein. Nucleotide sequence data have revealed, however, that the insert of DmA6 is A+T rich and encodes stop codons in all three reading frames (data not shown). This evidence argues very strongly that the insert of DmA6 is an intervening sequence.

The Amino Terminal Sequences of All Drosophila Actins Are Similar to Those of Vertebrate Cytoplasmic Actins

The amino terminal tryptic peptides of several vertebrate actins have been sequenced by Vandekerckhove and Weber (1978c). This analysis has revealed that a relatively high number of amino acid replacements occurs within these peptides and that sequence divergences tend to be tissue-specific, rather than species-specific. Thus, skeletal muscle actins from several vertebrate species are identical in amino acid sequence, but differ in sequence from the cytoplasmic actins (Vandekerckhove and Weber, 1978a, 1978c). In mammals these muscle specific versus cytoplasmic actin amino acid replacements involve charged amino acids; and, as a result, the isoelectric point of skeletal muscle actin differs from that of the cytoplasmic species (Vandekerckhove and Weber, 1978a). Because these charged amino acid exchanges are limited to the amino terminal tryptic peptide, determination of its primary sequence should allow a reasonably accurate prediction of the isoelectric points of respective actin proteins and thereby facilitate assignment of biological functions to each. We therefore derived the Drosophila actin amino terminal sequences from the nucleotide sequence of the corresponding region of each gene.

A comparison of the amino terminal tryptic peptides of each Drosophila actin with those of vertebrate skeletal muscle and cytoplasmic actin isoforms is shown in Figure 5. Within the boxes are the amino acid replacements characteristic of vertebrate skeletal and cytoplasmic actins. The sarcomeric actin sequence is Asp[1]Glu[2]Asp[3]Glu[4](Ser or Thr)[5]Thr[6] . . . Cys[10] . . . Leu[16]Val[17] While that of its cytoplasmic counterpart is (Asp or Glu)[2](Asp or

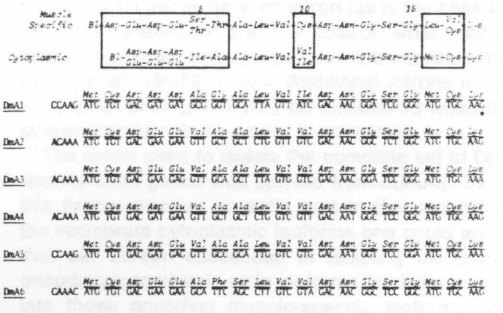


Figure 5. Sequences of the Amino Terminal Tryptic Peptides Encoded by Each Drosophila Actin Gene

The sequence of each encoded protein is aligned with the composite sequence of vertebrate muscle-specific and cytoplasmic actins. Within the boxed regions are the replacements which distinguish the vertebrate-muscle-specific actins from their cytoplasmic counterparts (Vandekerckhove and Weber, 1978a, 1978b, 1978c). Each of the Drosophila sequences resembles the vertebrate cytoplasmic actin sequences, partial exceptions being DmA1 and DmA6. Each of the derived Drosophila sequences encodes a cysteine that immediately follows the initiator methionine; however, cysteine does not occur at this position in the mature vertebrate actins. The sequence CAAPu precedes each initiator codon, except in DmA6 where the sequence is CAAAC. The asterisk in the DmA1 sequence denotes an ambiguous nucleotide. This base may be A; however, this would not alter our amino acid assignment. The unusual phenylalanine-serine sequence occurring at positions 6 and 7 in DmA6 would become serine-alanine if either T of the phenylalanine codon were inserted as the result of a sequencing error. However, then to return the sequence to the proper reading frame a nucleotide must be inserted in the leucine or valine codon at positions 8 and 9. We have failed to detect any evidence for such a nucleotide in our sequence determination through this region and we therefore believe that the sequence as shown is correct.

Glu)^{4}Ile^{5}Ala^{6} . . . (Val or Ile)^{10} . . . Met^{16}Cys^{17}. Inspection of the Drosophila actin gene sequences reveals first that only DmA4 is split within the glycine codon at position 13, second, that none of the Drosophila genes is split within codon four, as is the yeast actin gene, and third, that each of the derived actin amino acid sequences closely resembles the vertebrate cytoplasmic sequences, partial exceptions being DmA1 and DmA6. Examination of other portions of the Drosophila actin amino acid sequences shows that for the most part they also conform to the vertebrate cytoplasmic actin sequences (refer to Vandekerckhove and Weber, 1978a, 1978b, 1978c and our Appendix). Therefore, Drosophila apparently does not synthesize actin forms comparable in amino acid sequence to that of vertebrate skeletal muscle.

Figure 5 reveals another surprising result, namely the presence of a cysteine codon following the initiator methionine codons of each of the Drosophila actin genes. All vertebrate actins thus far characterized begin with an aspartic or glutamic acid residue that is acetylated. If only the methionine is cleaved from the Drosophila actins each would begin with cysteine followed by the three acidic residues characteristic of vertebrate cytoplasmic actins. Neither the Dictyostelium nor the yeast actin genes encode a cysteine at this position (Firtel et al., 1979; Gallwitz and Sures, 1980; Ng and Abelson, 1980). It is of course possible that the cysteine residue is cleaved by in vivo processing, and that the mature Drosophila actins begin with the string of acidic residues as do those of vertebrates.

A conserved sequence in the 5′ untranslated region of each Drosophila actin gene is apparent in Figure 5. In DmA1–DmA5 the sequence CAAPu precedes each ATG initiation codon. In DmA6 a cytidine has been inserted to give the sequence CAAAC. Immediately upstream from this region the sequences of the genes become completely divergent (data not shown). Thus selective pressure may preserve this four nucleotide sequence.

The Six Drosophila Actin Genes Encode Several Different Polypeptides

An interesting question is whether the six Drosophila actin genes encode proteins that differ in their primary amino acid sequences. Knowledge of such sequence differences may lead to a better understanding of the relationship of the structure of actin isoforms to their particular developmental roles. In the course of our studies we have obtained some, but not all, of the data needed for a decisive answer. These incomplete data are presented in the Appendix, in which we show Drosophila actin gene sequences not presented in Figures 4 or 5. Comparison of all of the available sequence information reveals that the six genes encode at least five different proteins. Only the comparison of DmA4 with DmA5 has as yet failed to reveal a difference in the sequence of the encoded proteins.

Far more striking than the differences in the sequences of these proteins are their similarities. Sequences of Drosophila actins are apparently extremely conserved. Most of the amino acid replacements are very conservative, such as replacement of valine with isoleucine, or threonine with serine. Therefore the observed amino acid replacements may not be functionally significant. Rather, we think it equally probable that for the most part they represent neutral mutations which have accumulated and resulted in slight divergences of the respective sequences.

Discussion

In this communication we have reported the isolation and structural characterization of the six Drosophila actin genes. We have shown that the protein-coding regions of these genes are highly homologous and that each encodes an actin protein with an amino acid sequence closely resembling those of the vertebrate cytoplasmic isoforms. We find no homology of untranslated or flanking regions of these genes. Additionally, we have discovered two introns within protein-coding regions (one in DmA4, one in DmA6) which are either in a position where no other Drosophila actin genes have an intron (as is the case for the DmA4 intron), or are in a position where some, but not all, other actin genes have introns (as is the case for the DmA6 intron). Additional introns in the Drosophila actin genes may be discovered when the characterization of the genes is completed.

The probe used to isolate the complete set of Drosophila actin genes was derived from λDmA2. Since this probe encodes an actin protein that resembles the vertebrate cytoplasmic isoforms one could argue that our screen succeeded in isolating the genes encoding cytoplasmic actin isoforms but failed to isolate those encoding muscle-specific isoforms. We think this improbable. Mammalian skeletal muscle actins differ from the cytoplasmic isoforms in approximately 25 of 375 amino acid residues (Vandekerckhove and Weber, 1978a, 1978b). Analysis of these amino acid replacements shows that they would increase nucleotide mismatch of these two types of actin genes by at most 3% above the value due to codon degeneracy. The moderately stringent hybridization conditions employed in our library screening and blot hybridization protocols do not distinguish such a small increase in nucleotide mismatch. Therefore unless the amino acid sequences of Drosophila muscle-specific actins are considerably more divergent from their cytoplasmic counterparts than is the case for vertebrates, our hybridization experiments would have detected them.

The variability in the positions of the introns in the Drosophila actin genes is in sharp contrast to the situation observed in globin genes (Maniatis et al.,

1980), vitellogenin genes (Wahli et al., 1980) and the ovalbumin-X-Y gene series (Royal et al., 1979). In each of these cases the placement of intervening sequences is exactly the same for all members of the gene family. For the globins this property extends across rather large evolutionary distances; for example, globin genes of Xenopus and humans are interrupted in exactly the same positions (Patient et al., 1980) despite the fact that their ancestral lines diverged approximately 350 million years ago (Young, 1962). In fact, there are only two cases where members of vertebrate multigene families differ in placement of introns, the first being one of the nonallelic rat insulin genes, which has lost an intron (Lomedico, et al., 1979), and the second being a nonfunctional mouse α-globin-like gene which has lost both introns (Nishioka et al., 1980).

Placement of intervening sequences is much more divergent in the Drosophila actin gene family than in the multigene families discussed above. At least three interruptions occur within the protein-coding regions of Drosophila actin genes, one within codon 13 of DmA4, one within codon 307 of DmA6, and a shorter intron at the same position of DmA1 (F. Sánchez, S. L. Tobin and B. J. McCarthy, personal communication). The interruption within codon 13 is not present in any other Drosophila actin genes, while that within codon 307 is not present in at least two other genes. Furthermore, none of the Drosophila actin genes is interrupted within codon 4 where the yeast actin gene is split (Gallwitz and Sures, 1980; Ng and Abelson, 1980) and at least two (DmA3 and DmA6) are not split between codons 121 and 122 where one of the sea urchin actin genes is split (refer to Durica et al., 1980; Schuler and Keller, 1981; and our Appendix). Finally, none of the several Dictyostelium actin genes sequenced by Firtel et al. (1979) appears to have interruptions within its protein-coding regions.

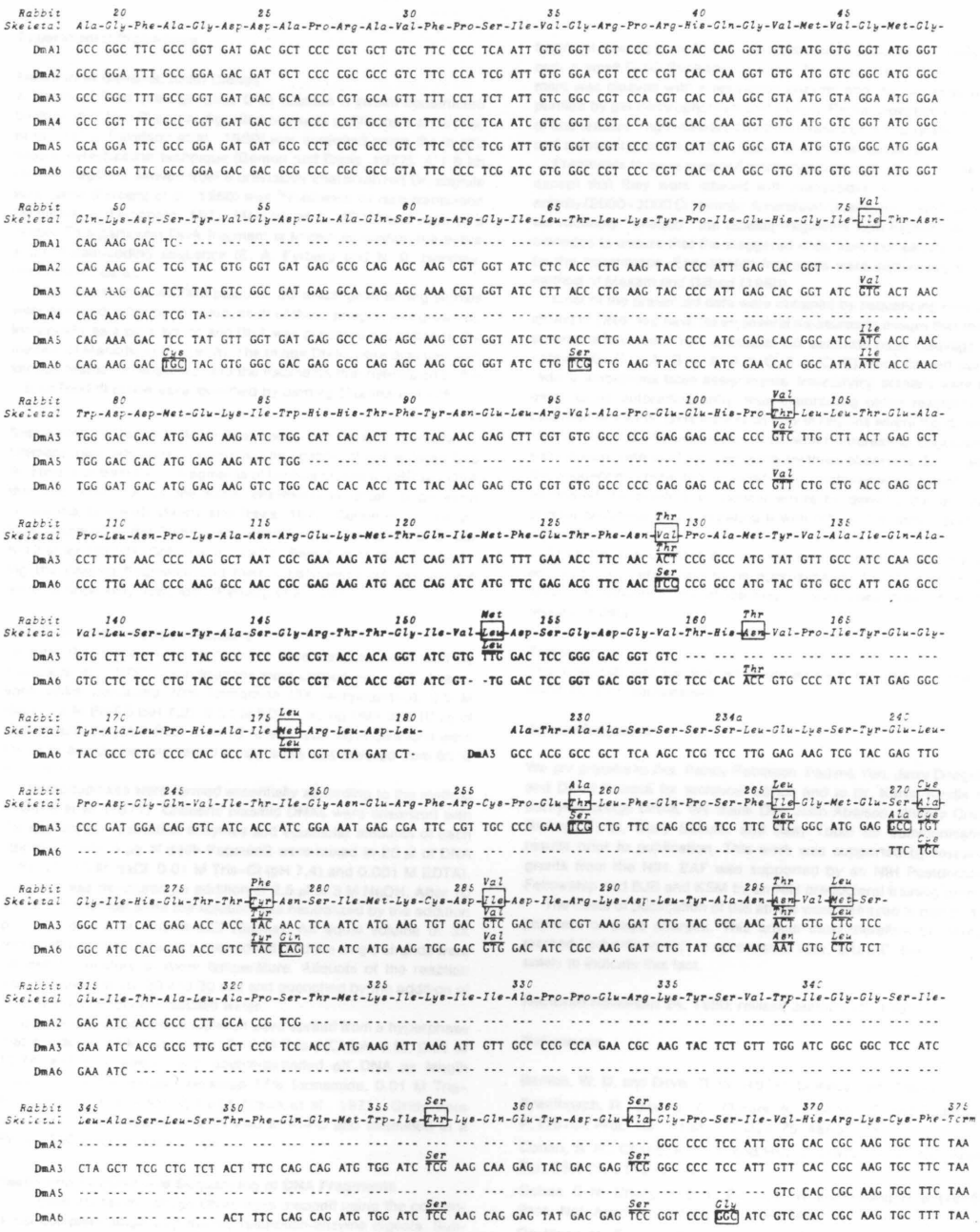
These results may help to elucidate the functional and/or evolutionary roles of intervening sequences. Gilbert (1979) has proposed that the primary function of introns is to facilitate gene evolution and as a corollary that the older form of a gene will have more introns. According to a strict form of this hypothesis, the primordial metazoan actin gene had at least four introns: one in the 5′ untranslated region (as does DmA2), one in codon 13 (as does DmA4), one after codon 121 (as does sea urchin) and one in codon 307 (as do DmA6 and DmA1). If this were the case, then many introns have been deleted from individual Drosophila actin genes during evolution. This could be for any or all of several reasons. First, because the actin genes are probably of ancient evolutionary origin there has been a long period in which to excise introns. Second, there may be strong selective pressure to delete introns from genes encoding proteins which evolve slowly, such as actin. Third, Drosophila may have more rapidly deleted introns during its evolution than have typical eucaryotes. Consistent with this notion is the observation that Drosophila has one of the smallest eucaryotic genomes.

Further comparisons of the positions of introns within the actin genes of a variety of metazoans will be most informative. If positions of introns are conserved even between protostomes and deuterostomes, this will support the hypothesis that the primordial metazoan actin gene contained many introns, several of which have been deleted from Drosophila actin genes. Conversely, if no conservation of intron positions within actin genes is found, then perhaps at least some introns are the vestiges of transposonlike elements which have inserted into genes, become fixed, and subsequently diverged in nucleotide sequence (see Doolittle and Sapienza, 1980; Orgel and Crick, 1980).

The observation that all of the Drosophila actins are similar in sequence to the vertebrate cytoplasmic isoforms has implications for both the evolution and functioning of actin proteins. Lower eucaryotes such as Physarum and Dictyostelium synthesize only actins with amino acid sequences closely resembling the vertebrate cytoplasmic isoforms (Vandekerckhove and Weber, 1978d, 1980), suggesting that the vertebrate cytoplasmic actins predate the vertebrate muscle-specific isoforms. Our results indicate that vertebrate muscle-specific actin isoforms evolved after insects separated from the phylogenetic line that gave rise to the vertebrates. Furthermore, our data suggest that the amino acid replacements found in vertebrate muscle-specific actin isoforms are not required for striated muscle function per se. Many types of insect muscle bear an anatomical resemblance to vertebrate striated muscle (Smith, 1968), and Drosophila muscle-specific actin isoforms have been documented (Storti et al., 1978; Fyrberg and Donady, 1979; Horovitch et al., 1979). However, our data clearly show that the amino acid sequences of Drosophila muscle-specific isoforms most closely resemble those of vertebrate cytoplasmic actins. In accordance with this conclusion, Lubit and Schwartz (1980) have found that rabbit antibodies elicited by actin extracted from the body wall musculature of the marine mollusc Aplysia will crossreact with vertebrate cytoplasmic actins but not with vertebrate skeletal muscle actin. Myofibrillar actin in Aplysia may therefore also have an amino acid sequence like vertebrate cytoplasmic actins.

Our results and those of Tobin et al. (1980) and Fyrberg et al. (1980) have revealed that Drosophila actin genes comprise a highly conserved family that encodes a nearly identical set of proteins. Examination of the constellation of tissue types and developmental intervals in which each gene is expressed should enhance our knowledge of the factors that regulate eucaryotic gene expression and additionally those factors that determine the cellular roles of newly synthesized proteins.

Cell
114



Appendix. Available Sequences of the Six Drosophila Actin Genes Not Presented in Figures 4 and 5

Codons of the Drosophila genes are aligned with the amino acid sequence of rabbit skeletal muscle actin. Amino acid residues of the skeletal muscle actin sequence that are replaced in vertebrate cytoplasmic actin isoforms are boxed, and their replacements appear directly above (replacements within the amino terminal tryptic peptides of actins are presented in Figure 5). In most cases the sequences of the encoded Drosophila actins conform to those of the vertebrate cytoplasmic isoforms. Amino acid replacements among the Drosophila actins occur at residues 3, 4, 5, 6, 7, 10, 52, 66, 76, 129, 270, 279, 296, 307 and 367 (refer also to Figures 4 and 5), accounting for five slightly divergent proteins. The amino acids encoded at these positions by the Drosophila genes appear above the appropriate codons. Codons that are not assigned an amino acid specify the same residue as is found in the rabbit skeletal muscle actin sequence. Codons that specify amino acids not expected on the basis of the vertebrate sequence data are boxed, and the amino acid assignments are written directly above. Codon 277 of DmA6 encodes valine, but this is not indicated in the figure.

Experimental Procedures

Isolation of Genomic Actin Clones

A library of Drosophila genomic DNA (Canton S strain) constructed by ligation of Eco RI partial digestion products to Charon 4 arms (Yen et al., 1979; Davidson et al., 1980) was screened using the in situ plaque hybridization technique (Benton and Davis, 1977). A 1.8 kb Hind III fragment derived from a previously characterized Drosophila actin gene (Fyrberg et al., 1980) was ^{32}P-labeled by nick translation (Schachat and Hogness, 1973; Maniatis et al., 1975) and used as a probe. This particular DNA fragment is known to contain the entire actin protein-coding sequence (E. A. Fyrberg and N. D. Hershey, unpublished data).

From a screen of 40,000 plaques, 30 which gave strong signals were selected. After purification, each of these phages was grown up individually as a plate lysate and DNA was prepared according to the method of Maniatis et al. (1978). The phage DNAs were digested with various restriction enzymes, and the fragments that hybridized to the 1.8 kb Hind III probe were identified by blotting (Southern, 1975).

Subcloning of Genomic Actin Sequences into Plasmid Vectors

Chimeric plasmids were constructed by ligation of gel isolated Eco RI, Hind III or Bam HI-cut fragments of lambda clones to pBR322 DNA that had been cut by the same enzyme(s) in order to generate complementary ends (Mertz and Davis, 1972; Cohen et al., 1973). DNA prepared in this fashion was used for transformation of E. coli K-12 strain HB 101 (Cohen et al., 1972). Selection of clones containing the desired fragments and their subsequent DNA preparation were as previously described (Fyrberg et al., 1980).

R-Loop and Heteroduplex Mapping

To form R-loops we used a protocol essentially as described by Kaback et al. (1979). Hybridizations were carried out in 20 μl reactions which contained 70% formamide (3X recrystallized), 0.5 M NaCl, 0.1 M PIPES (pH 7.2), 0.01 M EDTA, 50 ng DNA and 10 μg of total cellular poly(A)$^+$ RNA from 18 hr embryos. Hybridizations were for 24 hr and during this period temperature was lowered from 55°C to 47.5°C.

Heteroduplexes were formed essentially according to the method of Davis et al. (1971). Chimeric plasmid DNAs were linearized with the appropriate restriction enzymes and equimolar amounts of each (approximately ½ μg of each fragment) were mixed in 25 μl of DNA buffer (0.010 M NaCl, 0.01 M Tris-Cl (pH 7.4) and 0.001 M EDTA). The DNA was denatured by addition of 2.5 μl of 3 M NaOH. After 10 min at room temperature the solution was neutralized by the addition of 10 μl of 2.5 M Trizma-HCl (Sigma). An equal volume of 3X recrystallized formamide was then added, and the single strands were allowed to renature at room temperature. Aliquots of the reaction were removed at 10, 20 and 30 min and quenched by the addition of 2 vol of ice-cold double-distilled water.

Both R-loops and heteroduplexes were spread from a hyperphase that contained 45% formamide, 0.10 M Tris-HCl (pH 8.5), 0.01 M EDTA and both single- and double-stranded ϕX DNA as length markers. The hypophase contained 17% formamide, 0.01 M Tris-HCl (pH 8.5) and 0.001 M EDTA (Davis et al., 1971). Grids were strained in uranyl acetate, shadowed in Pt-Pd and examined in a Philips 300 electron microscope.

Restriction Mapping and Sequencing of DNA Fragments

Actin-gene-containing phage DNAs were mapped using the conventional complete single and double restriction-enzyme digests. Subcloned fragments were mapped in finer detail by the method of Smith and Birnstiel (1976). DNA to be mapped was digested using enzymes that produce 3′ recessed ends, and these were filled using appropriate α-^{32}P nucleotides and the Klenow fragment of E. coli DNA polymerase (Klenow and Henningsen, 1970). DNA was resuspended in 50 μl of 0.0066 M NaCl, 0.0066 M Tris-HCl (pH 7.5), 0.0066 M $MgCl_2$, and 0.0066 M DTT. Of each labeled nucleotide 50 μC was then added (400 Ci/mmole Amersham-Searle), and the reaction was initiated by addition of 1 U of enzyme. After 15 min at room temperature the reaction was stopped by addition of EDTA, and the labeled fragment was separated from unincorporated nucleotides by passage over a small G-50 Sephadex column. The double end-labeled fragment was cleaved with a restriction enzyme and the two ends were purified by gel electrophoresis and elution. Partial digestion products of end-labeled fragment were electrophoresed on 5% acrylamide gels and autoradiogrammed after drying on a Hoefer gel dryer.

Fragments to be sequenced were prepared in an identical fashion, except that they were labeled with nucleotides of higher specific activity (2000–3000 Ci/mmole; Amersham-Searle, Inc.). Additionally, we routinely "chased" the labeled fragments with excess cold nucleotides to ensure that the staggered ends were completely filled in by the polymerase. End-labeled fragments were sequenced by the method of Maxam and Gilbert (1980).

Most of the presented data were obtained by sequencing only one strand of DNA. We have taken several measures to ensure that these determinations were very accurate. Six separate base cleavage reactions (G, G > A, G + A, A > C, T + C, C) were carried out to reduce ambiguous base assignments. Intensifying screens were not used during autoradiography, thus maintaining better resolution of cleavage products. Gels were read only in regions where bands were clearly resolved. All sequences from which interesting conclusions were drawn were read independently by three observers. Since all of the sequences presented lie within actin protein-coding regions, we eliminated the majority of clerical errors by deriving the encoded protein sequence and comparing it with other actin amino acid sequences.

From the cases where overlapping sequences were determined, we judged our error rate to be approximately 1%. This error rate would not influence any of the major conclusions drawn from the sequence data.

Biosafety

This research was carried out in accordance with NIH Guidelines, using P2/EK1 containment.

Acknowledgments

We are grateful to Drs. Randy Robinson, Pauline Yen, Jerry Dodgson and David Kaback for technical advice and to Dr. Karen Kindle for purifying phage DNAs. We thank Drs. John Abelson, William Crain, Richard Firtel, Mary Schuler and Sally Tobin for communicating results prior to publication. This work was supported by research grants from the NIH. EAF was supported by an NIH Postdoctoral Fellowship and BJB and KSM by an NIH predoctoral training grant.

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

Received November 24, 1980; revised January 23, 1981

References

Benton, W. D. and Davis, R. W. (1977). Science *196*, 180–182.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978). Proc. Nat. Acad. Sci. USA *75*, 4853–4857.

Cohen, S. N., Chang, A. C. Y. and Hsu, L. (1972). Proc. Nat. Acad. Sci. USA *69*, 2110–2114.

Cohen, S. N., Chang, A. C. Y., Boyer, H. W. and Helling, R. B. (1973). Proc. Nat. Acad. Sci. USA *70*, 3240–3244.

Davidson, N., Fyrberg, E. A., Hershey, N. D., Kindle, K., Robinson, R. R., Sodja, A. and Yen, P. (1980). In The Genetic and Evolutionary Aspects of the Transcriptional and Translational Apparatus, S. Osawa, H. Uchida and H. Ozeki, eds. (Tokyo: Kodansha Scientific Publishing Co.), pp. 233–247.

Davis, R. W., Simon, M. and Davidson, N. (1971). Meth. Enzymol. *21D*, 413–428.

Doolittle, W. F. and Sapienza, C. (1980). Nature *284*, 601–603.

Durica, D. S., Schloss, J. A., and Crain, W. R., Jr. (1980). Proc. Nat. Acad. Sci. USA *77*, 5683–5687.

Firtel, R. A., Timm, R., Kimmel, A. R. and McKeown, M. (1979). Proc. Nat. Acad. Sci. USA *76*, 6206–6210.

Fyrberg, E. A. and Donady, J. J. (1979). Dev. Biol. *68*, 487–502.

Fyrberg, E. A., Kindle, K. L., Davidson, N. and Sodja, A. (1980). Cell *19*, 365–378.

Gallwitz, D. and Sures, I. (1980). Proc. Nat. Acad. Sci. USA *77*, 2546–2550.

Garrels, J. I. and Gibson, W. (1976). Cell *9*, 793–805.

Gilbert, W. (1979). In Eucaryotic Gene Regulation, ICN-UCLA Symposia on Molecular and Cellular Biology, *14*, R. Axel, T. Maniatis and C. F. Fox, eds. (New York: Academic Press), pp. 1–12.

Hagenbuchle, O., Bovey, R. and Young, R. A. (1980). Cell *21*, 179–188.

Horovitch, S. J., Storti, R. V., Rich, A. and Pardue, M. L. (1979). J. Cell Biol. *82*, 86–92.

Kaback, D. B., Angerer, L. M. and Davidson, N. (1979). Nucl. Acids Res. *6*, 2499–2517.

Klenow, H. and Henningsen, I. (1970). Proc. Nat. Acad. Sci. USA *65*, 168–175.

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. and Steitz, J. A. (1980). Nature *283*, 220–224.

Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979). Cell *18*, 545–558.

Long, E. O. and Dawid, I. B. (1980). Ann. Rev. Biochem. *49*, 727–764.

Lubit, B. W. and Schwartz, J. H. (1980). J. Cell Biol. *86*, 891–897.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975). Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. and Efstratiadis, A. (1978). Cell *15*, 687–701.

Maniatis, T., Fritsch, E. F., Lauer, J. and Lawn, R. M. (1980). Ann. Rev. Genet. *14*, 145–178.

Maxam, A. and Gilbert, W. (1980). Meth. Enzymol. *65*, 499–560.

Mertz, J. E. and Davis, R. W. (1972). Proc. Nat. Acad. Sci. USA *69*, 3370–3374.

Ng, R. and Abelson, J. (1980). Proc. Nat. Acad. Sci. USA *77*, 3912–3916.

Nishioka, Y., Leder, A. and Leder, P. (1980). Proc. Nat. Acad. Sci. USA *77*, 2806–2809.

Orgel, L. E. and Crick, F. H. C. (1980). Nature *284*, 604–607.

Patient, R. K., Elington, J. A., Kay, R. M. and Williams, J. G. (1980). Cell *21*, 565–573.

Pollard, T. and Weihing, R. (1974). In Critical Reviews in Biochemistry, *2*, G. Fasman, ed. (Cleveland, Ohio: CRC Press), pp. 1–65.

Royal, A., Garapin, A., Cami, B., Perrin, F., Mandel, J. L., Lemeur, M., Bregegeyre, F., Gannon, F., LePennec, J. P., Chambon, P. and Kourilsky, P. (1979). Nature *279*, 125–132.

Saborio, J. L., Segura, M., Flores, M., Garcia, R. and Palmer, E. (1979). J. Biol. Chem. *254*, 11119–11125.

Schachat, F. H. and Hogness, D. S. (1973). Cold Spring Harbor Symp. Quant. Biol. *38*, 371–381.

Schuler, M. A. and Keller, E. B. (1981). Nucl. Acids Res., *9*, 591–604.

Smith, D. S. (1968). In Insect Cells: Their Structure and Function. (Edinburgh: Oliver and Boyd), pp. 31–58.

Smith, H. O. and Birnstiel, M. L. (1976). Nuc. Acids Res. *3*, 2387–2399.

Southern, E. (1975). J. Mol. Biol. *98*, 503–517.

Storti, R. V., Horovitch, S. J., Scott, M. P., Rich, A. and Pardue, M. L. (1978). Cell *13*, 589–598.

Tobin, S. L., Zulauf, E., Sánchez, F., Craig, E. A. and McCarthy, B. J. (1980). Cell *19*, 121–131.

Vandekerckhove, J. and Weber, K. (1978a). Proc. Nat. Acad. Sci. USA *75*, 1106–1110.

Vandekerckhove, J. and Weber, K. (1978b). Eur. J. Biochem. *90*, 451–462.

Vandekerckhove, J. and Weber, K. (1978c). J. Mol. Biol. *126*, 783–802.

Vandekerckhove, J. and Weber, K. (1978d). Nature *276*, 720–721.

Vandekerckhove, J. and Weber, K. (1980). Nature *284*, 475–477.

Wahli, W., Dawid, I. B., Wyler, T., Weber, R. and Ryffel, G. U. (1980). Cell *20*, 107–117.

Whalen, R. G., Butler-Browne, G. S. and Gros, F. (1976). Proc. Nat. Acad. Sci. USA *73*, 2018–2022.

Yen, P., Hershey, N. D., Robinson, R. and Davidson, N. (1979). In ICN-UCLA Symposium on Eucaryotic Gene Regulation, R. Axel and T. Maniatis, eds. (New York: Academic Press), pp. 133–142.

Young, J. Z. (1962). The Life of Vertebrates. (Oxford: Oxford University Press).

CHAPTER 3

Transcripts of the six *Drosophila* actin genes accumulate in a stage- and tissue-specific manner

(published in Cell)

Cell, Vol. 33, 115–123, May 1983, Copyright © 1983 by MIT 0092-8674/83/050115-09 $02.00/0

Transcripts of the Six Drosophila Actin Genes Accumulate in a Stage- and Tissue-Specific Manner

Eric A. Fyrberg,† James W. Mahaffey,* Beverley J. Bond,‡ and Norman Davidson‡
* Department of Biology
The Johns Hopkins University
Baltimore, Maryland 21218
‡ Division of Biology,
and Department of Chemistry
California Institute of Technology
Pasadena, California 91125

Summary

We have surveyed expression of the six Drosophila actin genes during ontogeny. Unique portions of cloned actin genes were used to monitor levels of respective mRNAs in developmentally staged whole organisms and dissected body parts. We find that each gene is transcribed to form functional mRNA, which accumulates with a distinct pattern. Two of the genes, *act5C* and *act42A*, are expressed in undifferentiated cells and probably encode cytoplasmic actins. *Act57A* and *act87E* are expressed predominantly in larval, pupal, and adult intersegmental muscles; *act88F* in muscles of the adult thorax; and *act79B* in the thorax and leg muscles. These composite data define three main patterns of actin gene expression which are correlated with changing Drosophila morphology, particularly muscle differentiation and reorganization.

Introduction

A substantial fraction of a metazoan genome is comprised of small families of imperfectly repeated genes. Such multigene families typically encode a series of protein isoforms whose distributions are temporally and spatially regulated during development. The list of proteins known to be encoded by multigene families has grown considerably in recent years, now including structural, enzymatic, and storage polypeptides.

Multigene families represent useful paradigms for studying the function and evolution of eucaryotic genes. Comparisons of 5′ and 3′ untranslated and flanking regions of various gene family members have revealed conserved sequences which may be required for their regulated transcription and processing. Additionally, comparisons of the nucleotide sequence and intron–exon arrangement of related genes facilitate tracing their evolutionary histories, thereby elucidating the mechanisms by which genetic information is preserved and changed.

In order to utilize fully such structural information, it is important to determine in vivo expression patterns of members of multigene families. Such knowledge is necessary for identification of regulatory regions in these genes. Furthermore, it allows one to correlate evolutionary changes in a particular multigene family with phylogenetic changes in metazoan body plans and developmental pathways, and thus to comprehend more fully the relationship of genotype to phenotype.

Actin genes are well suited for such functional and evolutionary studies because of their evolutionary conservation (Firtel, 1981) and their developmentally regulated expression, which has been investigated primarily by monitoring accumulation of protein isoforms (Garrels and Gibson, 1976; Whalen et al., 1976). The isolation and characterization of the six actin genes of Drosophila melanogaster have already been reported (Tobin et al., 1980; Fyrberg, et al., 1980, 1981). We report the temporal and anatomical distribution of actin mRNAs during normal Drosophila development. Our results establish that two of the genes encode cytoplasmic actins, since complementary mRNAs are found during all stages of development and in dividing Drosophila cultured cells as well. Messenger RNAs transcribed from the remaining four genes accumulate only during the differentiation of particular types of muscle cells. We discuss these results in the contexts of Drosophila development and eucaryotic gene regulation.

Results

Construction of Gene-Specific Hybridization Probes

We have previously isolated and characterized in some detail the six Drosophila actin genes. The protein coding regions are approximately 85% homologous, and cross-hybridize efficiently even under moderately stringent conditions. As shown below, mRNAs transcribed from these genes range in length from 1.6 to 2.2 kb, considerably larger than the 1131 nucleotide protein coding region. Nucleotide sequencing data (E. Fyrberg and N. D. Hershey, unpublished data; F. Sanchez et al., 1983) have revealed that most of this additional length comprises the respective 3′ untranslated regions, which have lengths of 300 to 700 nucleotides. Sequences of these regions are different for each gene. Therefore we subcloned fragments from each 3′ untranslated region to use as probes for determining levels of homologous mRNAs.

The positions and lengths of the several subclones thus prepared are illustrated in Figure 1. We have adopted the convention of Zulauf et al. (1981), and denote these subclones and the parental clones from which they are derived by their chromosomal locations, in preference to our previous notation (Fyrberg et al., 1981; see also legend to Figure 1). Sequences of several of these subclones begin at the Hae III or Ava II site which spans codons 365 and 366. These probes include approximately 30 nucleotides of the protein coding region. One probe, probe B of *act88F*, includes 131 nucleotides of the protein coding region. Additionally, each probe contains a substantial length of the 3′ untranslated region and, in some cases, 3′ flanking sequences as well.

The specificity of each probe was tested by hybridization to DNAs of the parental actin clones. Each of the eight gene-specific probes was labeled with ^{32}P and hybridized

† To whom correspondence should be addressed.

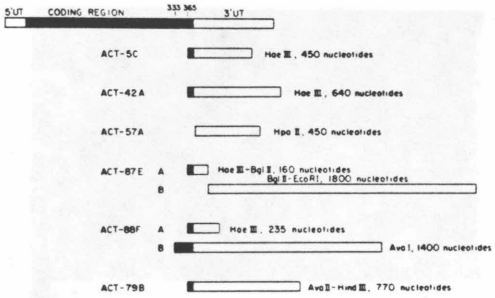


Figure 1. Actin Gene-Specific Probes

The top schematic diagram represents the transcribed region of an actin gene. The solid line represents the protein coding region, which measures 1131 nucleotides. Positions of codons 333 and 365, which form the 5′ border of particular subclones, are indicated. Open blocks represent gene regions encoding the 5′ and 3′ untranslated regions, which are of variable length. Immediately below we show the fragments subcloned from each gene. Lengths of fragments and identities of restriction enzymes used to generate them are indicated. All fragments were cloned in pBR322 (or its derivatives) as described in Experimental Procedures.

It should be noted that we follow Zulauf et al. (1981) in denoting the actin genes by their chromosomal locations, in preference to our previous notation (Fyrberg et al., 1981) which is given in parentheses: *act88F* (DmA1), *act5C* (DmA2), *act42A* (DmA3), *act57A* (DmA4), *act87E* (DmA5), and *act79B* (DmA6).

to a nitrocellulose strip containing "dots" of each of the six parental actin gene DNAs. As seen in Figure 2A, at moderately stringent conditions each probe hybridizes only to the parental actin gene DNA. Since all of the experimental hybridizations described below were performed under conditions of equal or greater stringency, we are confident that none of our results is due to cross-hybridization of specific probes with heterologous transcripts.

As illustrated by Figures 2A and 2B, when hybridization stringency is dropped sufficiently, several of these probes will hybridize weakly with nonhomologous genes, but only probe A of *act87E* will cross-hybridize significantly. To avoid erroneous results due to cross-hybridization of this probe, we have in every case performed a second hybridization using *act87E* probe B, which does not cross-hybridize at any stringency.

The Six Drosophila Actin Genes Are Functional

To determine whether each Drosophila actin gene is functional, we performed hybrid-selection experiments using the technique of Ricciardi et al., 1979. (See also Fyrberg et al., 1980.) Specific probe DNAs were hybridized to poly(A)$^+$ RNA isolated from the appropriate stage of Drosophila development (as determined by the experiments described below). Messenger RNA complementary to each probe was translated in vitro, and protein products were separated using two-dimensional gel electrophoresis. These experiments demonstrate that each gene-specific probe hybrid-selects mRNA, which translates to form one of the known Drosophila actin isoforms. Figure 3A shows results obtained using the *act5C* probe. Identical results were seen with the *act42A*, *act87E*, *act88F*, and *act79B*

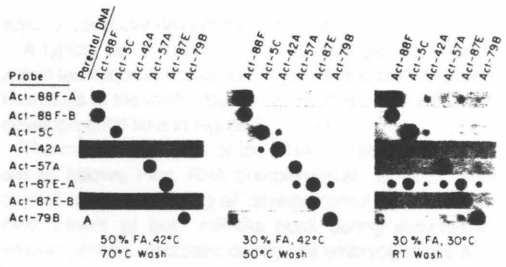


Figure 2. Dot Blot Analysis of Gene-Specific Probes

To test the specificity of the 3′ untranslated region probes, DNA of each was isolated free of vector sequences, labeled by nick-translation in the presence of four ^{32}P nucleotides, and hybridized to nitrocellulose strips containing 50 μg dots of each of the parental actin genes, which were also isolated free of vector sequences. Probes are arranged in the vertical dimension, while clones parental DNAs are arranged in the horizontal axis. Hybridizations and washes were carried out under conditions listed below each panel. (FA) formamide.

probes. In each case a major translation product having the correct molecular weight and an isoelectric point of 5.77 is seen along with a minor form of isoelectric point 5.84. Figure 3B shows results obtained using the *act57A* probe. The major translation product is more acidic (pI 5.70), as is the minor form (pI 5.77). In all of these experiments, the major translation products correspond to stable actin isoforms seen in vivo (Storti et al., 1978; Fyrberg and Donady, 1979; Horovitch et al., 1979). The minor translation forms probably represent a small percentage of actin molecules not acetylated in vitro (see Berger and Cox, 1979; Garrels and Hunter, 1979). These composite hybrid-selection data strongly indicate that each of the six Drosophila actin genes is transcribed to form stable functional messenger RNA in vivo, and that none is a "pseudogene."

Figure 3C illustrates results obtained using a non-actin gene probe for the selection. Here a clone encoding a presumed tropomyosin gene (a gift from S. Falkenthal) was used to select mRNA. Even when this autoradiograph is greatly overexposed, no actin protein can be detected, indicating the specificity of the hybrid-selection protocol.

Drosophila Actin Gene Expression Is Temporally Modulated

We performed two types of experiments to study variation in steady state levels of actin gene transcripts during the Drosophila life cycle. Poly(A)$^+$ RNA was prepared from several developmental stages, ranging from early embryo to adult, and also from several permanent Drosophila cell lines. In one set of experiments, equal amounts of RNA from different developmental stages were subjected to gel electrophoresis under denaturing conditions, transferred to nitrocellulose filters, and hybridized to nick-translated gene-specific probes (RNA gel blots) under conditions that are sufficiently stringent for specificity (see Experimental Procedures).

In a second set of experiments, denatured DNA from

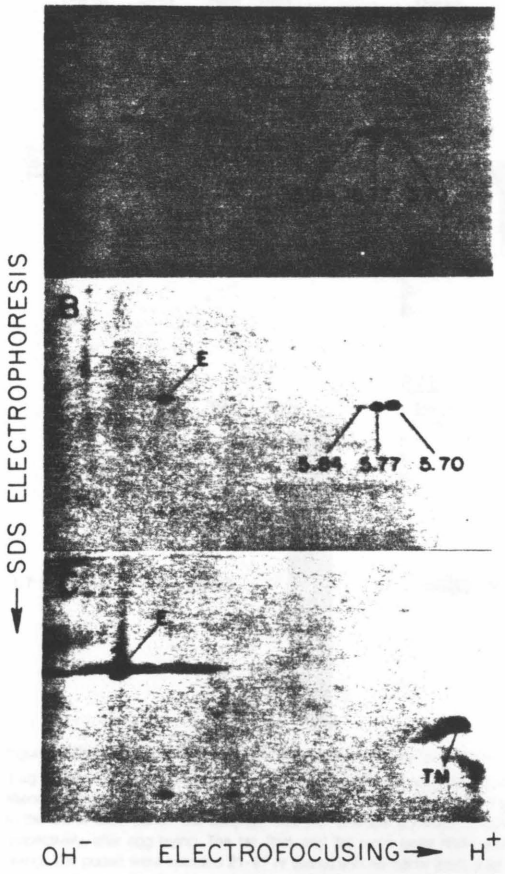


Figure 3. Selection and In Vitro Translation of mRNA Homologous to Actin Gene-Specific Probes

3A shows a fluorograph of ^{32}S-labeled proteins whose synthesis is directed by mRNA homologous to the *act5C*-specific probe. Identical results were obtained using the *act42A*, *act87E*, *act88F*, and *act79B* probes. 3B shows results of an identical experiment using the *act57A* probe for the selection. An asynchronous embryo RNA preparation (2–24 hr) was used for hybrid selections with the *5C*, *42A*, *57A*, and *87E* probes. A mixed pupal RNA sample (containing RNA from early, middle, and late pupae) was used for the hybrid selections to the *88F* and *79B* probes. 3C shows an overexposed fluorograph of proteins translated from RNA selected by a cloned Drosophila gene believed to be tropomyosin. "E" is a protein synthesized by the translation system in the absence of exogenously added mRNA. All gels contained 50,000 cpm. Exposures were for 72–96 hr. Arrows indicate the directions of SDS gel electrophoresis and isoelectric focusing.

each gene-specific probe was applied as a series of "dots" onto nitrocellulose strips (Kafatos et al., 1979). Identical dot series were individually hybridized to ^{32}P-labeled cDNA representing poly(A)$^+$ RNA isolated from a particular developmental stage (DNA dot blots). The former set of experiments provides the lengths of transcripts as well as variations in their steady state levels during development. The latter experiments conveniently demonstrate relative levels of transcripts from the six different genes within a particular developmental stage.

A typical set of results using the RNA gel blot-hybridization technique is illustrated in Figure 4, and the integrated intensities of the mRNA bands are plotted as a function of developmental time in Figure 5.

The main conclusions of the RNA gel blot experiments are as follows. First, RNA complementary to *act5C* and *act42A* is present during all developmental stages examined. Levels of both mRNAs peak during early-to-mid embryogenesis, decrease during late embryogenesis and the larval instars, and rise again during early pupal development. This pattern of expression is not correlated with Drosophila muscle differentiation, which occurs during late embryogenesis and late in pupal development as well (Poulson, 1950; Bodenstein, 1950). Furthermore, using either RNA gel blots or dot blots (described below), we find that in several undifferentiated permanent Drosophila cell lines all actin mRNA is transcribed from either *act5C* or *act42A*. Both of the aforementioned observations lead us to propose that *act5C* and *act42A* encode cytoplasmic or cytoskeletal actins. Finally, it is of interest to note that the *act5C* probe is complementary to three different size classes of mRNA, and that the relative levels of these three classes change somewhat during development.

Second, transcripts from genes *act57A* and *act87E* are most abundant during late embryogenesis. Both transcripts are present at reduced levels throughout larval growth, although the reduction is more marked for *act87E*. Rather low levels of both transcripts are found during the later larval to early pupal transition, and these subsequently rise to moderate levels during late pupal development. Periods during which these two mRNAs are abundant are correlated with the differentiation of larval musculature and its subsequent restructuring during late pupation to form adult muscles (see Poulson, 1950; Crossley, 1978).

Third, messenger RNAs encoded by the final two genes, *act79B* and *act88F*, occur at maximal levels during mid-to-late pupal development, and remain high in newly eclosed adults. Additionally, we detect very low levels of *act79B* transcripts during late embryogenesis and throughout larval growth (not visible in the exposure in Figure 4). As we show in more detail below, the timing and location of *act79B* and *act88F* mRNA accumulation during pupal development is precisely correlated with differentiation of adult leg and thoracic musculature (Crossley, 1978).

To examine further the accumulation of actin mRNAs during Drosophila development, we have used the dot blot technique described by Kafatos et al. (1979). We emphasize that we have used these experiments to assess the relative abundance of actin mRNAs within a particular developmental stage. Since exposures of the dot blots were not standardized, they cannot be reliably compared between developmental stages.

Results of temporal dot blot experiments are shown in Figure 6. As illustrated, during early embryogenesis and early pupal development, when no major Drosophila muscle types differentiate, *act5C* and *act42A* mRNAs predominate. These two transcripts are also among the most

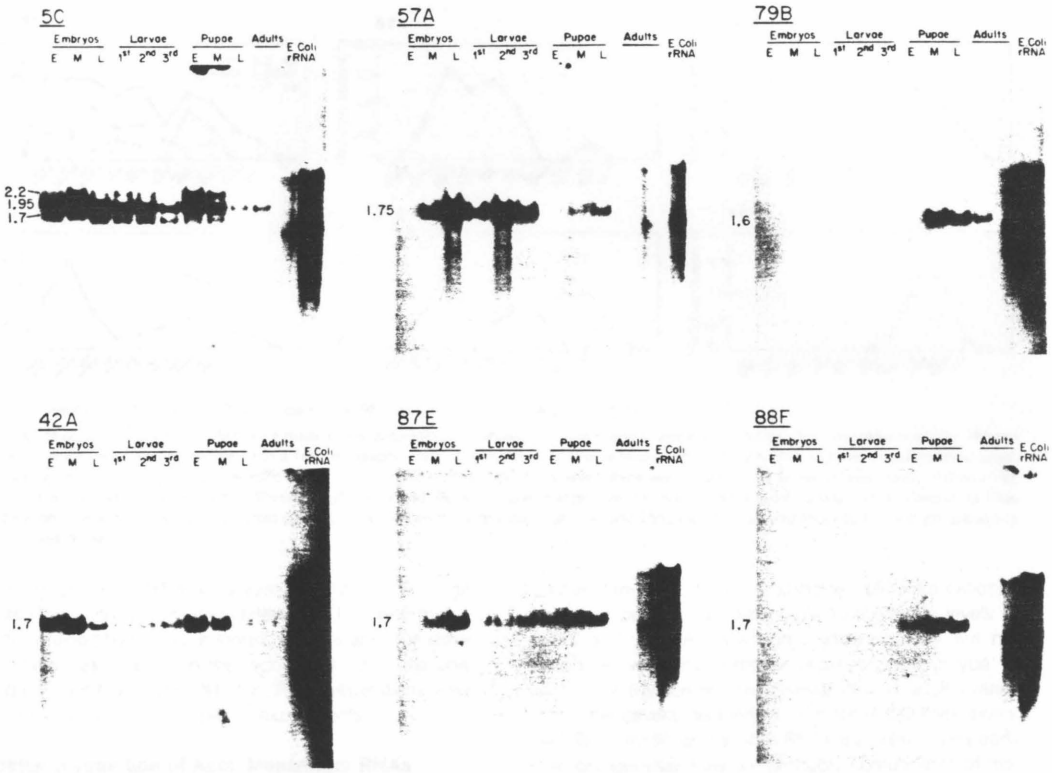


Figure 4. Developmental Patterns of Transcripts for the Six Actin Genes

1 μg of poly(A)+ RNA from different developmental stages was subjected to electrophoresis on formaldehyde–agarose gels and blotted to nitrocellulose. Six filters, identically prepared, were hybridized to the six nick-translated actin gene specific probes. The lengths of the transcripts from each gene are indicated to the left of each panel. The early (E), mid (M), and late (L) embryo RNAs were made from embryos which had been aged for 0–4, 8–12, and 6–20 hr, respectively, after egg laying. The 1st, 2nd, and 3rd instar larval RNAs were made from animals aged 32–37, 48–51, and 89–99 hr, respectively, after egg laying. The pupae were selected 21–27 hr (early) and 48–58 hr (mid) after flotation on water. The late pupal RNA sample was made from animals that had been synchronized by flotation on water to a 9 hr period. After a few animals had eclosed, RNA was made from the rest. The adult RNA was made from both males and females <5 days old. *Act88F* and *act87E* here refer to probe A only.

E. coli 16S and 23S ribosomal RNAs were used as a standard to determine the length of the actin transcripts. Nick-translated pKK2361 (a gift from Harry Noller and Mary Alice Raker) that contains the entire E. coli rrnB ribosomal RNA operon was hybridized to the lanes containing the standard.

abundant in 5-day adult flies, which have completed muscle development and would have presumably reduced levels of muscle-specific actin gene transcripts. Taken together, these results strengthen our conclusion that *act5C* and *act42A* encode cytoplasmic actins.

During late embryogenesis and throughout larval growth, *act57A* transcripts are very abundant, overshadowing *act87E*, *act79B*, and the cytoplasmic actin gene transcripts, all of which are known to be present from the RNA blots displayed in Figure 4. We conclude that *act57A* encodes the major actin isoform of larval skeletal muscle. Hybridization of cDNA to the *act87E* probe is much less intense, and we conclude that this gene encodes a minor actin isoform of larval muscle. Unexpectedly, *act87E* hybridizes detectably to early embryo cDNA. However, we have noted in the RNA blots that *act87E* probe B hybridizes to an approximately 4 kb RNA during early embryogenesis and during pupal development. This large transcript is probably not related to any of the actin genes, since it does not hybridize to an actin protein coding region probe (data not shown). Since *act87E* probe B was used in the dot blot experiments, the low level of hybridization seen during early embryogenesis is likely due to this same non-actin-related transcript.

During mid and late pupal stages, *act88F* and *act79B* transcripts are very abundant. *Act57A* and *act87E* transcripts are also present, but during this phase of development *act87E* transcripts are more abundant than those of *act57A*. These results indicate that each of these four genes is expressed in developing adult muscle. In the following section, we confirm these results by examining the anatomical distributions of the transcripts in developing pupae.

In summary, the DNA dot blots reveal the same overall

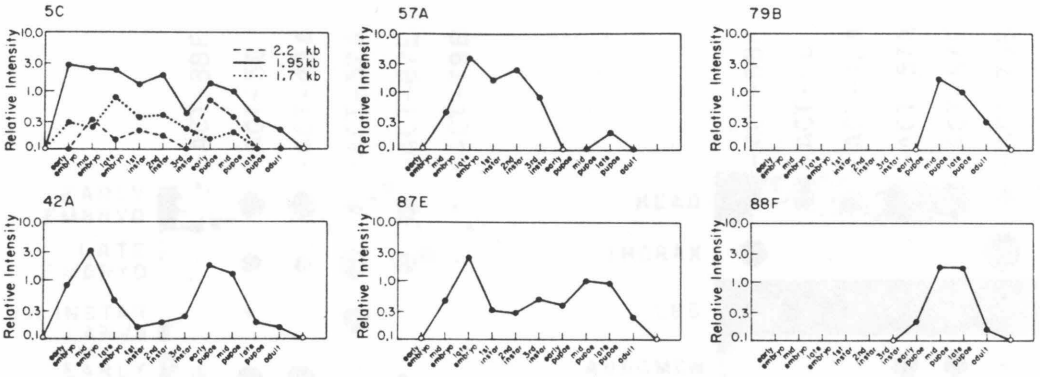


Figure 5. Relative Levels of Actin mRNA in Total Poly(A) RNA as a Function of Developmental Time

Tracings of optical density vs. position were made of the autoradiograms shown in Figure 4 with a Joyce Loebl double beam microdensitometer. Relative areas under the peaks were determined with a Tektronix graphic computing system. The relative level of actin in poly(A) RNA is plotted on a logarithmic scale. Closed circles are used to denote measured points. Open circles represent points for which there was no detectable signal. Different kinds of lines in the *act5C* panel are used to represent the different length transcripts. Refer to Figure 4 legend for the exact times at which animals were collected for RNA preparation. The above analysis was carried out on autoradiograms of blots washed at 55°C and 75°C (see Experimental Procedures), and the average of the two was plotted.

patterns of actin mRNA accumulation as do the RNA gel blots. The minor quantitative differences between the results obtained by the two methods are probably due either to differences inherent in the techniques, or to the unavoidable fact that slightly different RNA preparations were used for the respective types of experiments.

Spatial Distribution of Actin Messenger RNAs

The experiments presented above allowed us to group the six Drosophila actin genes into three pairs, with members of each pair similarly regulated and possibly functionally related. For independent tests of some of these assignments, we hybridized dot blot strips containing specific probe DNAs to ^{32}P-labeled cDNAs representing poly(A)$^+$ RNA of particular body parts. RNA was prepared from heads, thoraces, legs, and abdomens of late pupae, and from ovaries of newly eclosed adults. Results of these experiments are presented in Figure 7.

Once again, we observe similar patterns of mRNA expression for particular pairs of actin genes. *Act57A* and *act87E* transcripts are present primarily in the head and abdomen. This observation strengthens our conclusion that *act57A* and *act87E* are expressed during the differentiation and restructuring of larval muscle, since much of the cephalic and abdominal musculature of Drosophila is comprised of restructured larval muscle (Crossley, 1978). Low but significant amounts of *act87E* are also found in legs and ovaries. The significance of this observation is unknown. In the thorax, where the indirect flight muscle is the largest structure, we detect only *act88F* and *act79B* transcripts. *Act79B* transcripts are also abundant in legs. Since thoracic and leg musculature both are derived primarily from imaginal discs (Crossley, 1978; Lawrence, 1982), this observation supports the notion that both genes are activated during the differentiation of imaginal myoblasts.

When the head, thorax, or abdomen blots are exposed for longer periods, we detect low to moderate levels of *act5C* and *act42A* transcripts, indicating they are not associated with any particular organ or muscle type. In fact, these transcripts predominate only in adult ovaries and in the several permanent Drosophila cell lines examined. Since *act5C* and *act42A* RNAs are present very early in embryogenesis, they are probably components of maternal mRNA (Anderson and Lengyel, 1979). We suggest that these ovarian transcripts are transcribed in polyploid nurse cells and transported to the oocyte for use during early embryogenesis.

Discussion

We have described several major features of Drosophila actin gene expression. Specific probes for each of the six genes have been constructed and used to monitor the steady state levels of the respective gene transcripts during Drosophila development. These experiments reveal that each of the actin genes directs the synthesis of stable messenger RNA having a length sufficient to encode the 42,000 MW actin protein, and that messenger RNA from each gene can be translated in vitro to form a known Drosophila actin isoform. Qualitative analyses of the accumulation of actin gene transcripts show that no two genes display precisely identical patterns of RNA accumulation. However, as regards the main features of the temporal pattern of expression for the whole organism, the genes can be grouped into three sets of two each.

Our interpretations of the three patterns of RNA accumulation take into account the potential roles of actin as well as the timing of histogenesis during Drosophila development. The actin proteins of any metazoan are likely to be utilized either as cytoplasmic/cytoskeletal components in various tissues or as the backbone of thin filaments of

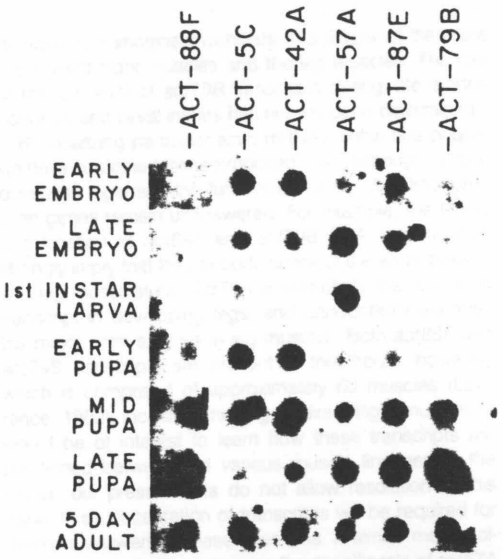


Figure 6. Dot Blots of Actin Gene-Specific Probes to Developmentally Staged ^{32}P-labeled cDNAs

In all cases, *act88F* and *act87E* refer to probe B. Early and late embryos were aged 0–2 hr and 18–24 hr after fertilization, respectively. Early, mid, and late pupae were aged 153 hr, 174, hr, and 202 hr from time of egg laying.

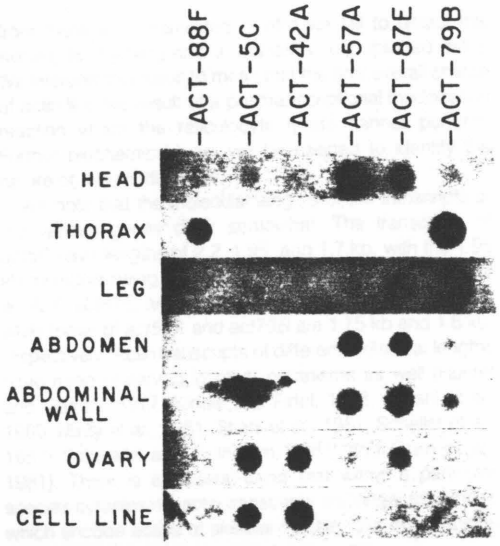


Figure 7. Dot Blots of Actin Gene-Specific Probes to Spatially Localized ^{32}P-Labeled cDNAs

Positions of specific probe DNAs are indicated. In all cases, *act88F* and *act87E* refer to probe B. Heads, thoraces, legs, and abdomens were isolated from late (191 hr) pupae. Abdominal wall RNA was prepared from abdomens which were emptied of organs using pressure from a small rolling pin. Ovaries were dissected from newly eclosed adult females.

muscle cells (Wolosewick and Porter, 1979; Huxley, 1969). Our data suggest that two Drosophila actin genes encode cytoplasmic actins, while four encode muscle-specific actin isoforms. Expression of the four muscle-specific genes can be further subdivided into two biologically meaningful patterns.

Early embryogenesis and early pupation in Drosophila are characterized by elevated mitotic activity and by widespread reorganization of cells and tissues (Poulson, 1950; Bodenstein, 1950). These activities would require large amounts of cytoplasmic/cytoskeletal actins, and presumably high levels of the corresponding actin mRNAs would also be present during these periods. Accordingly, *act5C* and *act42A* transcripts are abundant during embryogenesis, during early pupal development, and in the ovaries of the adult female, but are relatively less abundant in the whole organism during periods when Drosophila muscles differentiate. In addition, we find that transcripts from these genes, and only these genes, are found in undifferentiated permanent Drosophila cell lines, including K_c cells (Echalier and Ohanessian, 1969), D^1 cells (Schneider and Blumenthal, 1978), and a cell line derived from Shibire^{ts} embryos (A. Simcox and J. H. Sang, personal communication). These observations are consistent with the proposal that both transcripts encode cytoplasmic actins.

Muscle differentiation in Drosophila occurs during two distinct phases. During mid-to-late embryogenesis and throughout larval growth, the musculature of the larva differentiates. During metamorphosis, existing larval musculature breaks down while new sets of adult muscles differentiate. A portion of the adult musculature is fashioned from reorganized larval muscle, while many muscles differentiate de novo from myoblasts arising from imaginal discs (refer to Crossley, 1978, for a description of the Drosophila muscle system). We would therefore expect to find high levels of muscle-specific actin mRNAs during both of these phases of development.

Act57A and *act87E* transcripts are most abundant during late embryogenesis, and are present at reduced levels throughout larval growth and during late pupal development. These observations lead us to propose that both of these genes are expressed during the differentiation of larval musculature and again during its reorganization, or redifferentiation, to form adult muscle. From the actin isoform data discussed below, it is clear that *act57A* encodes the majority of actin present in larval skeletal muscle. The role of *act87E* during larval development and growth is less clear. During this phase of Drosophila development, *act87E* transcripts are not nearly as abundant as those of *act57A*. It is probable that *act87E* encodes a minor actin species of larval muscle, but this assignment is tentative. However, *Act87E* transcripts are also abundant in the abdominal wall and abdomen of late pupae.

The final two transcripts, *act88F* and *act79B*, are abundant only when the adult musculature develops. These transcripts almost certainly specify the actins found in the fibrillar and tubular muscles of the adult that differentiate

de novo from imaginal myoblasts. The largest of these are the indirect flight muscles and the leg muscles. The role of the low level of *act79B* transcripts during late embryogenesis and larval instars has not yet been determined.

By localizing particular actin mRNAs within late pupae, we have confirmed these proposed roles, although several questions regarding the functions of particular Drosophila actin genes remain unanswered. For example, the timing and location of *act88F* and *act79B* RNA accumulation strongly imply that they encode actins of the adult thoracic and leg musculature. *Act79B* is essentially the only actin transcript in developing legs, and undoubtedly encodes the major actin species in leg muscles. Both *act88F* and *act79B* transcripts are present in the thorax, however, which is comprised of approximately 80 muscles (Lawrence, 1982), including the large indirect flight muscles. It would be of interest to learn how these transcripts are partitioned between the various muscle lineages of the thorax; our present data do not allow resolution of this issue. In situ localization of transcripts will be required for definitive answers to these questions. A similar methodology will be required to determine the specific role of *act57A* and *act87E* during larval development, and to assess the relative levels of the two cytoskeletal actin transcripts in different kinds of nonmuscle cells.

Our conclusion that *act57A* encodes actin present in larval and adult muscle fibers is confirmed by considering available Drosophila actin isoform data. As shown in our hybrid-selection experiments, *act57A* encodes the most acidic Drosophila actin isoform, termed actin I. This isoform is known to be the principal actin of both larval and adult body wall musculature (Storti et al., 1978; Fyrberg and Donady, 1979; Horovitch et al., 1979). These composite results clearly demonstrate the role of the *act57A* gene during larval development. It should be mentioned, however, that a contrary result has been reported by Zulauf et al. (1981), who found an actin I isoform to be encoded by the *act79B* gene. These authors imply that *act79B* encodes the major actin isoform of larval skeletal muscle. While we *can* detect low levels of *act79B* transcripts during larval development and growth, it is obvious that these transcripts could encode only a minor component of larval skeletal muscle. The cause of the discrepancy between our results and those of Zulauf et al. is unknown. From our data, we estimate that more than 95% of *act79B* transcripts which accumulate during Drosophila development are localized within muscles of the adult thorax and leg. We therefore believe that *act57A* encodes the major actin isoform of larval skeletal muscle, while *act79B* encodes a major species of thoracic and leg muscle and possibly a minor component of larval skeletal muscle.

None of the RNAs selected by specific probes encoded a major translation product having an isoelectric point of 5.84, which is characteristic of actin III. A stable actin isoform of this charge is found in adult thoracic muscle (Horovitch et al., 1979; Mogami, Fujita, and Hotta, 1982), and we expected either (or both) *act79B* and *act88F* to encode this isoform. However, when poly(A)$^+$ RNA isolated from thoraces is translated in vitro we fail to detect this isoform (E. Fyrberg and J. Mahaffey, unpublished data). We interpret this result to mean that the final overall charge of actin III is the result of a posttranscriptional modification reaction which the reticulocyte lysate cannot perform. Further biochemical work will be needed to identify the nature of this modification.

We note that the molecular lengths of the transcripts of the various genes differ somewhat. The transcripts of *act5C* have lengths of 2.2, 1.95, and 1.7 kb, with the 1.95 kb molecule being the most abundant. The transcripts for *act42A*, *act87E*, and *act88F* are 1.7 kb for each in length, while those of *act57A* and *act79B* are 1.75 kb and 1.6 kb, respectively. Actin transcripts of different molecular lengths have been observed in other organisms as well (Hunter and Garrels, 1977; Kindle and Firtel, 1978; Ordahl et al., 1980; Minty et al., 1981; Shani et al., 1981; Scheller et al., 1981; Schwartz and Rothblum, 1981; and Crain et al., 1981). There is a general trend that within a particular species cytoplasmic actin transcripts are longer than those which encode actins of skeletal muscle. Our observations for Drosophila are in only partial agreement.

A final topic to consider is the selective value of the six Drosophila actin genes. We have shown that Drosophila actin gene transcripts display three distinct patterns of accumulation during development. Presumably, the respective actin genes are being transcribed in response to rather different regulatory cues. *Act5C* and *act42A* may respond primarily to levels of mitogenic compounds. The remaining four genes probably respond to regulatory molecules which are synthesized early during muscle cell differentiation. It may be advantageous for metazoans to have several copies of genes that encode proteins synthesized in a variety of cell types, as are contractile proteins. Each related gene could then be regulated independently and, presumably, more precisely. In this model the primary selective value of multiple actin genes would be to ensure more precise regulation of actin synthesis during metazoan ontogeny. An alternative, and by no means exclusive, hypothesis is that multigene families facilitate the accumulation of particular amino acid replacements in the encoded protein isoforms, which concurrently become specialized for particular developmental roles. Further resolution of this issue will come only when individual actin protein isoforms can be tested in new in vivo contexts. Gene transplantation techniques have recently made such experiments possible (Spradling and Rubin, 1982; Rubin and Spradling, 1982).

Experimental Procedures

Drosophila Culture

Synchronously developing cultures of the Canton-Special Drosophila strain were prepared by placing eggs on standard cornmeal–agar medium in plastic boxes and incubating at 25°C. In some cases cultures were further synchronized by floating puparia on water just prior to pupation.

Construction of Plasmids Containing Unique Regions of Drosophila Actin Genes

We have previously cloned the six Drosophila actin genes in pBR322 and characterized each in some detail (Fyrberg et al., 1981). The region of each

gene which adjoins and includes the 3′ untranslated region of the complementary messenger RNA was mapped in finer detail using restriction endonucleases, and partially sequenced using the technique of Maxam and Gilbert (1980). Once identified, the appropriate restriction fragment was isolated from a preparative agarose gel by the technique of Tabak and Flavell (1978). Ends of the isolated fragments were made flush by incubation with nucleotide triphosphates and the Klenow fragment of E. coli DNA polymerase (Klenow and Henningsen, 1970), and molecular linkers containing the sequence cleaved by Hind III were added to the ends by blunt-end ligation. In several cases, fragments were cloned by using alternative strategies. *Act88F* probe B was inserted directly into pBR322 which had been cleaved with Ava I; *act87E* probe B was inserted into the Bgl II and Eco RI sites of plasmid pKC7 (Rao and Rogers, 1979); and *act57A* was cloned directly into the Cla I site of pBR322.

Chimeric plasmids produced in this fashion were used to transform E. coli strain HB101 to ampicillin resistance. Plasmid DNA was prepared by centrifugation in ethidium bromide–cesium chloride gradients (Radloff, Bauer, and Vinograd, 1968).

Preparation of Poly(A)$^+$ RNA

Poly(A)$^+$ was prepared by either of two methodologies. RNA to be used for blot-hybridization experiments was extracted using the technique of Chirgwin et al. (1979), as previously described (Fyrberg et al., 1980). RNA to be used as template for cDNA synthesis was extracted using the SDS–phenol technique (Spradling and Mahowold, 1979).

Drosophila tissues and body parts were dissected in the saline solution described by Ephrussi and Beadle (1936). After dissection, they were transferred to buffer containing 0.5% w/v SDS and homogenized.

RNA Electrophoresis and Blotting

The RNA electrophoresis and blotting protocol was a composite of an unpublished procedure of D. Goldberg and B. Seed, and Thomas (1980). RNAs to be separated were denatured by heating for 15 min at 65°C in an electrophoresis buffer which contained 50% formamide, 17% formaldehyde. One microgram/lane was loaded on 4 mm 1.5% agarose gels containing 2.2 M formaldehyde. Electrophoresis was performed for approximately 600 volt-hr. The electrophoresis buffer consisted of 20 mM Na-MOPS (Sigma), 5 mM NaOAc, 1 mM EDTA.

After electrophoresis the gel was soaked for 30 min in 50 mM NaOH, 100 mM NaCl; neutralized for 40 min in 100 mM Tris (pH 7.5); and blotted to nitrocellulose in 20× SSC. After blotting for 12–24 hr, filters were air dried, then baked for 2 hr in a vacuum oven.

Filters were pre-hybridized at 50°C for 12–20 hr in 50% v/v formamide, 0.8 M NaCl, 0.1 M PIPES, 0.01% Sarkosyl, 5× Denhardts solution and 1 mg/ml denatured calf thymus DNA. They were then hybridized to 1×10^6 cpm/ml of nick-translated probes of specific activity $>5 \times 10^7$ cpm/μg at 50°C in the buffer above with 10% w/v dextran sulfate added and 500 μg/ml denatured calf thymus DNA rather than the 1 mg/ml above. Hybridizations were for 40–48 hr with gentle agitation. After hybridization, filters were washed once at room temperature in 2× SSC, 0.5% Sarkosyl, and 0.02% sodium pyrophosphate, and then several times at 50°C or 55°C in 0.2× SSC, 0.05% Sarkosyl, 0.01% sodium pyrophosphate, or at 75°C in 2× SSC, 0.1% sodium pyrophosphate, 0.1% SDS, and then exposed to Kodak XAR-5 x-ray film using Du Pont Lightning Plus intensifying screens.

When allowance is made for a difference of ~10° between RNA–DNA hybrids and DNA–DNA duplexes in 50% formamide (Casey and Davidson, 1977), and of ~5°C in aqueous electrolyte (Kallenbach, 1969), we estimate that the stringency of hybridization and washing in these experiments is about the same as for the DNA–DNA dot blots described below.

Preparation of DNA "Dots"

Chimeric plasmids containing the inserted actin gene-specific sequence were linearized by digestion with Eco RI, then denatured by addition of NaOH to 0.3 M. After 5 min at room temperature, the solution was neutralized by addition of Trizma–HCl (Sigma), then quenched at 0°C. After addition of an equal volume of 20× SSC, the DNA solution was applied to nitrocellulose filters (Schleicher and Schuell) using a commercial manifold (Bethesda Research Laboratories). Filters were washed 3 times in 10× SSC, then once in 3× SSC. After air drying, filters were baked in a vacuum oven at 80°C.

Hybridizations were performed at 42°C in 50% formamide, 5× SSC, 5× Denhardt's solution, 50 mM sodium phosphate (pH 6.8), which contained 40 μg/ml of denatured calf thymus DNA. Filters were washed 2 times at room temperature and 3 times at 68°C in 2× SSC, 0.1% sodium pyrophosphate, 0.1% SDS. To ensure that filters contained equimolar amounts of each DNA, a representative set was hybridized to nick-translated pBR322 DNA.

Synthesis of cDNA and Nick-Translated DNA Probes

Synthesis of ^{32}P-labeled cDNA probes using d(T) primers was performed by the technique of Efstratiadis et al. (1975). Double-stranded DNA fragments were labeled using the technique of nick-translation (Schachat and Hogness, 1973; Maniatis, Jeffrey, and Kleid, 1975).

Drosophila Permanent Cell Lines

K_c cells (Echalier and Ohanessian, 1969), D^1 cells (Schneider and Blumenthal, 1978) and a cell line derived from Shibre^{ts} embryos (A. Simcox and J. H. Sang, personal communication) were used as sources of RNA. These lines were independently derived from embryos of particular genotypes by growing primary explants either in vitro or vivo.

Acknowledgments

We thank Drs. Charles Ordahl and Sara Tobin for communicating results of their actin gene studies prior to publication; Drs. Robert Dottin, Gerry Rubin, Allen Shearn and Allan Spradling for providing technical advice and laboratory reagents; and Dr. Amada Simcox for providing cultured Drosophila cells. We are grateful to Dr. S. Falkenthal for the tropomyosin gene DNA, to Drs. David Price and Carl Parker for the K_c cell RNA, and to Drs. Harry Noller and Mary Alice Raker for the clone of the E. coli *rrnB* 16S + 23S rRNA operon. This work was supported by grants from the National Institutes of Health and Muscular Dystrophy Association to E. A. F. and N. D. J. W. M. and B. J. B. were supported by NIH Predoctoral Training Grants.

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

Received August 26, 1982; revised February 2, 1983

References

Anderson, K. V., and Lengyel, J. A. (1979). Rates of synthesis of major classes RNA in *Drosophila* embryos. Dev. Biol. *70*, 217–231.

Berger, E., and Cox, G. (1979). A precursor of cytoplasmic actin in cultured *Drosophila* cells. J. Cell Biol. *81*, 680–683.

Bodenstein, D. (1950). The postembryonic development of Drosophila. In Biology of Drosophila, M. Demerec, ed. (New York: John Wiley and Sons, Inc.), pp. 278–282.

Casey, J., and Davidson, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentration of formamide. Nucl. Acids Res. *4*, 1539–1552.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Crain, W. R., Jr., Durica, D. S., and VanDoren, K. (1981). Actin gene expression in developing sea urchin embryos. Mol. Cell. Biol. *1*, 711–720.

Crossley, A. C. (1978). The morphology and development of the *Drosophila* muscular system. In The Genetics and Biology of Drosophila, Vol. 2b., M. Ashburner and T. R. F. Wright, eds. (New York: Academic Press), pp. 499–560.

Echalier, G., and Ohanessian, A. (1969). Isolement, en cultures *in vitro* de lignees cellulaires diploides de *Drosophila melanogaster*. CR Acad. Sci. (Paris), Ser. B. *268*, 1771–1773.

Efstratiadis, A., Maniatis, T., Kafatos, F. C., Jeffrey, A., and Vournakis, J. N. (1975). Full length and discrete partial reverse transcripts of globin and chorion mRNAs. Cell *4*, 367–378.

Ephrussi, B., and Beadle, G. W. (1936). A technique of transplantation for

Drosophila. Am. Nat. *70*, 218–225.

Firtel, R. A. (1981). Multigene families encoding actin and tubulin. Cell *24*, 6–7.

Fyrberg, E. A., and Donady, J. J. (1979). Actin heterogeneity in primary embryonic culture cells from Drosophila melanogaster. Dev. Biol. *68*: 487–502.

Fyrberg, E. A., Kindle, K. L., Davidson, N., and Sodja, A. (1980). The actin genes of Drosophila: a dispersed multigene family. Cell *19*, 365–378.

Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S., and Davidson, N. (1981). The actin genes of Drosophila: protein coding regions are highly conserved but intron positions are not. Cell *24*, 107–116.

Garrels, J. I., and Gibson, W. (1976). Identification and characterization of multiple forms of actin. Cell *9*, 793–805.

Garrels, J. I., and Hunter, T. (1979). Post-translational modification of actins synthesized *in vitro*. Biochim. Biophys. Acta *564*, 517–525.

Horovitch, S. J., Storti, R. V., Rich, A., and Pardue, M. L. (1979). Multiple actins in *Drosophila melanogaster*. J. Cell Biol. *82*, 86–92.

Hunter, T., and Garrels, J. I. (1977). Characterization of the mRNAs for α-, β-, γ-Actin. Cell *12*, 767–781.

Huxley, H. E. (1969). The mechanism of muscular contraction. Science *164*, 1356–1366.

Kafatos, F. C., Jones, C. W., and Efstratiadis, A. (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucl. Acids Res. *6*, 1541–1552.

Kallenbach, N. R. (1968). Theory of thermal transitions in low molecular weight RNA chains. J. Mol. Biol. *37*, 445–466.

Kindle, K. L., and Firtel, R. A. (1978). Identification and analysis of *Dictyostelium* actin genes, a family of moderately repeated genes. Cell *15*, 763–778.

Klenow, H., and Henningsen, I. (1970). Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from Escherichia coli B by limited proteolysis. Proc. Nat. Acad. Sci. USA *65*, 168–175.

Lawrence, P. A. (1982). Cell lineage of the thoracic muscles of Drosophila. Cell *29*, 493–503.

Maniatis, T., Jeffrey, A., and Kleid, D. G. (1975). Nucleotide sequence of the rightward operator of phage lambda. Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Maxam, A., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65*, 499–560.

Minty, A. J., Carvatti, M., Robert, B., Cohen, A., Daubas, P., Weydert, A., Gros, F., and Buckingham, M. E. (1981). Mouse actin messenger RNAs: construction and characterization of a recombinant plasmid molecule containing a complementary transcript of mouse α-actin mRNA. J. Biol. Chem. *256*, 1008–1014.

Mogami, K., Fujita, S. C., and Hotta, Y. (1982). Identification of Drosophila indirect flight muscle myofibrillar proteins by means of two-dimensional electrophoresis. J. Biol. Chem. (Japan) *91*, 643–650.

Poulson, D. F. (1950). Histogenesis, organogenesis, and differentiation in the embryo of *Drosophila melanogaster* meigen. In Biology of Drosophila, M. Demerec, ed. (New York: John Wiley and Sons, Inc.), pp. 168–181.

Ordahl, C. P., Tilghman, F. M., Ovitt, C., Fornwald, J., and Largen, M. T. (1980). Structure and Developmental Expression of the chick α-actin gene. Nucl. Acids Res. *8*, 4948–5005.

Radloff, R., Bauer, W., and Vinograd, J. (1968) A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in the HeLa cells. Proc. Nat. Acad. Sci. USA *59*, 838–845.

Rao, R. N., and Rogers, S. G. (1979). Ptasmid pKC7: a vector containing ten restriction endonuclease sites suitable for cloning DNA segments. Gene *7*, 79–82.

Ricciardi, R. P., Miller, J. S., and Roberts, B. E. (1979). Purification and mapping of specific mRNAs by hybridization–selection and cell-free translation. Proc. Nat. Acad. Sci. USA *76*, 4927–4931.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348–353.

Sanchez, F., Tobin, S. L., Rodest, U., Zulauf, E., and McCarthy, B. J. (1983). Two Drosophila actin genes in detail: gene structure, protein structure, and transcription during development. J. Mol. Biol. *163*, 533–551.

Schachat, F. E., and Hogness, D. S. (1973). Repetitive sequences in isolated Thomas circles from *Drosophila melanogaster*. Cold Spring Harbor Symp. Quant. Biol. *38*, 371–381.

Scheller, R. H., McAllister, L. B., Crain, W. J., Jr., Durica, D. S., Posakony, J. W., Thomas, T. L., Britten, R. J., and Davidson, E. H. (1981). Organization and expression of multiple actin genes in the sea urchin. Mol. Cell Biol. *1*, 609–628.

Schneider, I., and Blumenthal, A. B. (1978). Drosophila cell and tissue culture. In The Genetics and Biology of Drosophila, Vol. 2a. M. Ashburner and T. R. F. Wright, ed. (New York: Academic Press), pp. 266–316.

Schwartz, R. J., and Rothblum, K. N. (1981). Gene switching in myogenesis: differential expression of the chicken actin multigene family. Biochemistry *20*, 4122–4129.

Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D., Carmon, Y., Reiter, J., Frischauf, A. N., and Yaffe, D. (1981). Skeletal muscle actin mRNA. Characterization of the three prime untranslated region. Nucl. Acids Res. *9*, 579–589.

Spradling, A. C., and Mahowald, A. P. (1979). Identification and genetic localization of mRNAs from ovarian follicle cells of Drosophila melanogaster. Cell *16*, 589–598.

Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science *218*, 341–347.

Storti, R. V., Horovitch, S. J., Scott, M. P., Rich, A., and Pardue, M. L. (1978). Myogenesis in primary cell cultures from Drosophila melanogaster: protein synthesis and actin heterogeneity during development. Cell *13*, 589–598.

Tabak, H. F., and Flavell, R. A. (1978). A method for the recovery of DNA from agarose gels. Nucl. Acids. Res. *5*, 2321–2332.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Nat. Acad. Sci. USA *77*, 5201–5205.

Tobin, S. L., Zulauf, E., Sanchez, F., Craig, E. A., and McCarthy, B. J. (1980). Multiple actin-related sequences in the Drosophila melanogaster genome. Cell *19*, 121–131.

Whalen, R. G., Butler-Browne, G. S., and Gros, F. (1976). Protein synthesis and actin heterogeneity in calf muscle cells in culture. Proc. Nat. Acad. Sci. USA *73*, 2018–2022.

Wolosewick, J. J., and Porter, K. R. (1979). Microtrabecular lattice of the cytoplasmic ground substance: artifact or reality. J. Cell Biol. *82*, 114–139.

Zulauf, E., Sanchez, F., Tobin, S. L., Rodest, U., and McCarthy, B. J. (1981). Developmental expression of a *Drosophila* actin gene encoding actin I. Nature *292*, 556–558.

CHAPTER 4

The *Drosophila melanogaster* actin 5C gene uses two transcription initiation sites and three polyadenylation sites to express multiple mRNA species

(published in Molecular and Cellular Biology)

MOLECULAR AND CELLULAR BIOLOGY, June 1986, p. 2080–2088
0270-7306/86/062080-09$02.00/0

Vol. 6, No. 6

The *Drosophilia melanogaster* Actin 5C Gene Uses Two Transcription Initiation Sites and Three Polyadenylation Sites To Express Multiple mRNA Species

BEVERLEY J. BOND[1] AND NORMAN DAVIDSON[2]*

Division of Biology,[1] and Department of Chemistry,[2] California Institute of Technology, Pasadena, California 91125

Received 23 December 1985/Accepted 10 March 1986

At least six mRNAs are made from the *Drosophila melanogaster act5C* gene. We investigated the structures of these RNAs in detail and determined that they are heterogeneous at both their 5′ and 3′ ends. At the 5′ end there were two nonhomologous leader exons which were alternately spliced to the remainder of the gene. These leader exons mapped to 1.7 and 0.7 kilobases, respectively, upstream of a common splice acceptor site which was eight base pairs 5′ to the translation initiator AUG. Exon 1 is 147 bases in length, while exon 2 is 111 bases. A consensus TATA sequence was found roughly 30 base pairs upstream from exon 1, but none was found in the analogous position upstream of exon 2. The transcript length diversity arose principally from the use of three polyadenylation sites. This gave rise to RNA molecules with 3′-untranslated regions of roughly 375, 655, and 945 base pairs. With two start sites and three termination sites, this gene has the potential to produce six different transcripts. All six possible transcripts were present in whole fly mRNA. Transcripts containing the two different leader exons were found in roughly the same relative quantities through development. In contrast, the various 3′ ends were differentially represented through development.

There are six members of the *Drosophila melanogaster* actin multigene family (18, 19, 53). They map to six widely dispersed chromosomal sites. The genes can be placed into three classes that show different developmental patterns of expression (20). Two of the genes (*act79B* and *act88F*) are expressed in the late pupae stage and in adults and encode the major fibrillar and tubular adult-specific muscle isoforms, including those for the flight and jump muscles. *act57A* and *act87E* are expressed in larval and late pupal-adult stages. They are believed to encode actins for larval musculature and abdominal muscles of the adult. *act5C* and *act42A* encode the cytoskeletal actins of nonmuscle cells. They are expressed in the whole animal in most stages of development and are the only actins expressed in early embryos.

Insofar as they have been mapped, intron positions and sequences of the six genes are not conserved, with the exception of the intervening sequence at codon 307 which is present in the *act88F* and *act79B* genes (18, 46). However, the protein-coding regions are rather highly conserved, and there is 85 to 95% amino acid sequence homology among the six actin isoforms.

In this study we were concerned with the structure of the *act5C* gene. The structure of the gene, as it was known prior to this study, is depicted in Fig. 1. R-loop studies indicated that there is a short exon (exon 1) about 1.7 kilobases (kb) upstream of the main exon (exon 3). Sequence data suggest that the protein-coding region is entirely contained in exon 3 and that there is a consensus splice acceptor sequence 8 nucleotides upstream of the translation initiator ATG (18). Sequence data and results of in vitro transcription studies provided a tentative identification of a cap site at the 5′ end of exon 1 as well as a TATA box and an activating sequence upstream (40; D. Price, B. Korber, J. Topol, and C. S. Parker, personal communication).

RNA gel blots showed that there are three *act5C* transcripts with molecular lengths of 2.2, 1.95, and 1.7 kb (20). Their relative intensity in whole animal RNA varies with developmental stage. In general, the 1.95-kb band is the most intense. We hypothesized that the reason for the length difference is alternative choices of polyadenylation sites in the 3′-untranslated region.

This study was undertaken to obtain decisive information about the structures of the several transcripts. The results reported below show that in addition to the mRNA of structure exon 1-exon 3, as depicted in Fig. 1, there is an additional cap site (and presumed transcription start site) between exons 1-3 and an additional mRNA of structure exon 2 and exon 3, as shown in Fig. 1. Exons 1 and 2 are not greatly different in length. The main cause of the three length classes is the existence of three alternate polyadenylation sites. Our data indicate that in whole animal RNA, exons 1 and 2 are used with approximately equal probability in all stages of development and with all three polyadenylation sites. Furthermore, there is some developmental variability in the usage of the three polyadenylation sites.

MATERIALS AND METHODS

Preparation of RNA. RNA was prepared by a modified guanidinium thiocyanate method (12, 19). Typically, about 15 ml of guanidinium solution was used for each gram of tissue. After the addition of CsCl, the mixture was centrifuged for 15 min at 10,000 × *g* in an SS34 rotor. The cuticles and other debris formed a band on top of the CsCl solution which was discarded. The solution was then spun in the ultracentrifuge as described above. Poly(A)$^+$ RNA was selected by oligo(dT) chromatography (2).

Primer extension. A synthetic 24-base oligonucleotide that was complementary to the actin 5C RNA between codons 26 and 34 was used as a primer. The sequence of this primer was 5′-TCGATGGGAAGACGGCGCGGGGAG-3′. It was synthesized by S. Horvath (California Institute of Technology) on an automated DNA synthesizer (29) and then

* Corresponding author.

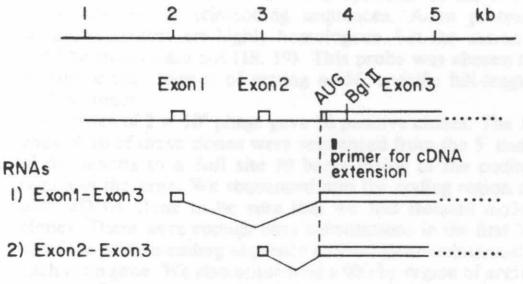


FIG. 1. Transcript map of the 5′ end of the *act5C* gene. Exon sequences are represented by open boxes, while intron and flanking sequences are solid lines. The alternate splicing pathway is depicted below. The location of the primer used for primer extension experiments is also indicated.

purified from a 20% polyacrylamide–8 M urea gel and precipitated with ethanol. It was 5′ end labeled with polynucleotide kinase and [^{32}P]ATP (~7000 Ci/mmol). Labeled primer (150 pmol) and poly(A)$^+$ RNA (1 μg) were suspended in 20 μl of 70% formamide containing 0.4 M NaCl, 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid), and 1 mM EDTA. After incubation for 2 min at 60°C and 2 h at 37°C, the reaction mixture was diluted to 200 μl. Ammonium acetate (200 μl) and isopropyl alcohol (240 μl) were added. After 15 min at room temperature, the nucleic acids were recovered by centrifugation. The RNA precipitation procedure removed most of the excess primer molecules. The ethanol-rinsed and dried pellet was suspended in 15 μl of a 2× cDNA mixture (0.1 M Tris [pH 8.3]; 12 mM magnesium acetate; 0.12 M NaCl; 20 mM dithiothreitol; 2 mM each of dATP, dGTP, and TTP). To this was added 2 μl of 10 mM dCTP, 1.9 μl of 0.8 μg of actinomycin D per ml, 15 U of reverse transcriptase, and water to 30 μl. The solution was incubated at 37°C for 45 min and precipitated with ethanol. The precipitated products were run on alkaline agarose gels (36) or, after RNase A digestion of the RNA templates, on 6% polyacrylamide–8 M urea gels.

S1 nuclease and exonuclease VII mapping of transcripts. Typically, 10 ng of ^{32}P-labeled DNA fragments was hybridized to 1 μg of poly(A)$^+$ RNA in 80% formamide at 55°C and then digested with S1 nuclease (6, 15) or exonuclease VII (7). For the exonuclease VII reactions, after the hybridization step 1 U of enzyme was added to each reaction in 100 μl of cold exonuclease VII buffer (30 mM KCl, 10 mM Tris [pH 7.4], 10 mM EDTA). After incubation at 45°C for 1 h, the reactions were precipitated with ethanol, with yeast tRNA used as a carrier. Ethanol-washed and dried samples were loaded on alkaline agarose gels (36). Protected fragments were detected by autoradiography of the dried gels.

Isolation of cDNA clones. A *D. melanogaster* embryo (8- to 20-h old) cDNA library in the vector lambda *gt*10 (kindly provided by L. Kauvar) was screened by the plaque lift technique of Benton and Davis (4) with nick-translated probes.

DNA sequencing. DNA sequencing was done by the chemical modification technique (35).

RNA blotting. Poly(A)$^+$ RNAs were electrophoresed on formaldehyde-agarose gels and blotted to nitrocellulose by a modification of the technique of Thomas (52). Blotting, hybridizations, and blot washings were performed as described previously (20). Hybridization probes were gel isolated and labeled by nick translation.

RESULTS

The *act5C* gene has two transcription start sites. The structure of the 5′ end(s) of the transcripts was first investigated by primer extension experiments. The primer was a synthethic oligonucleotide which was chosen to fulfill two criteria. First, it had to be able to prime cDNA synthesis from all *act5C* transcripts, regardless of their 5′-untranslated sequences. For this reason, the sequence was taken from the protein-coding portion of the gene. Second, it had to be specific for the *act5C* gene and not hybridize to other actin mRNAs. The actin protein-coding regions, in general, were very homologous, but a search of the protein-coding sequences revealed a 24-nucleotide stretch which contained six mismatches between the *act42A* and *act5C* genes. This sequence, which starts 80 base pairs (bp) downstream from the *act5C* gene initiator AUG, was used for the primer, and early embryo RNA was used for the experiments because only these two actin genes were expressed at that developmental stage. Hybridizations were done under conditions in which the primer would hybridize only to the *act5C* RNA.

Surprisingly, the primer extension experiments showed four bands with lengths of 258, 255, 222, and 213 nucleotides, and they were of approximately equal intensity (Fig. 2). These data indicate that transcripts of the gene have at least two and perhaps as many as four different 5′ ends but do not give any information about where they map on the genomic DNA.

To map the location of the 5′-untranslated sequences with respect to the genomic DNA, S1 nuclease and exonuclease VII experiments were done. A 4.2-kb kinase-labeled *Bgl*II fragment of genomic *act5C* DNA (Fig. 3B) was used as a hybridization probe to poly(A)$^+$ RNA from different developmental stages. This DNA fragment contains 250 bp of actin protein-coding sequence and 4 kb of upstream sequences.

Hybrids formed between early and late embryo poly(A)$^+$ RNA and the labeled DNA fragment were digested with exonuclease VII or S1 nuclease. The exonuclease VII- and S1 nuclease-protected bands are displayed on an alkaline agarose gel (Fig. 3). The two enzymes protected fragments with different lengths, indicating that splicing occurs at the 5′ end of the gene. The S1 nuclease-protected band was roughly 300 bases in length, indicating the presence of a splice acceptor site just upstream of the AUG codon in all *act5C* transcripts. Exonuclease VII digestion yielded two fragments with lengths of 0.95 and 1.98 kb which were of roughly equal intensity.

These results indicate that exon sequences terminating at their respective 5′-end cap sites are present in the genomic DNA at roughly 0.7 kb (exon 2) and 1.7 kb (exon 1) upstream from the splice acceptor site. The location of exon 1 had previously been found by R-loop analysis (17) and by in vitro transcription experiments (D. Price and C. Parker, personal communication). Exon 2 was not found by R-loop experiments.

These results do not determine the size or number of the upstream exons in the several transcripts. Furthermore, the relationship between the two exonuclease VII digestion products and the primer extension products was not resolved by this experiment.

Sequencing of the 5′ leader exons. To determine the size and number of upstream exons, we isolated and analyzed cDNA clones for the *act5C* gene. A *D. melanogaster* (8- to 20-h-old embryo cDNA library was obtained from Larry Kauvar. It was screened with a genomic probe which

contained about 4 kb of sequences upstream of the AUG codon and no protein-coding sequences. Actin protein-coding sequences are highly homologous, but the untranslated sequences are not (18, 19). This probe was chosen to maximize the chances of getting *act5C* specific full-length cDNA clones.

A screen of 2×10^4 phage gave 16 positive clones. The 5′ ends of 10 of these clones were sequenced from the 5′ ends of the inserts to a *SalI* site 30 bases inside of the coding region of the gene. We sequenced into the coding region of each cDNA clone to be sure that we had isolated *act5C* clones. There were enough base substitutions in the first 30 bp of the protein-coding sequence to identify unambiguously each actin gene. We also sequenced a 900-bp region of *act5C* genomic DNA that was shown by exonuclease VII experiments to contain exon 2. The genomic sequence data and their interpretation, as deduced from comparison with the cDNA sequences (see below), are shown in Fig. 4. A 430-bp region of genomic DNA surrounding exon 1 has been sequenced previously (40; D. Price, personal communication).

All of the cDNA clones had the same splice acceptor site which was 8 nucleotides upstream of the AUG, as previously

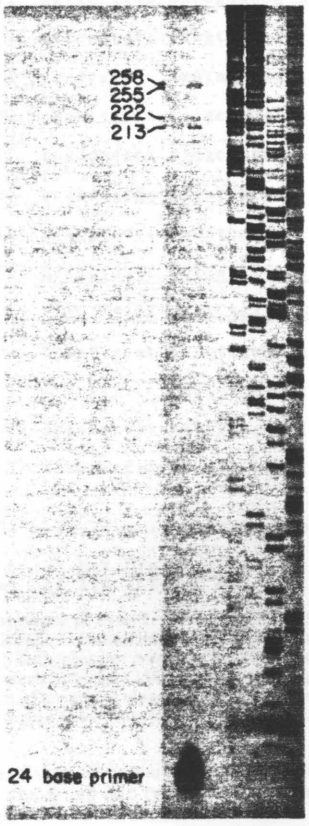


FIG. 2. Mapping of the 5′ ends of *act5C* RNAs by primer extension. A 24-base synthetic oligonucleotide homologous to the *act5C* gene between codons 26 and 33 (Fig. 1) was labeled with kinase and hybridized to 2 μg of poly(A)$^+$ RNA from 0- to 4 h-old embryos. cDNA synthesis was carried out as described in the text. The extended products were run on a 6% sequencing gel. A dideoxy-sequencing ladder is shown in the four rightmost lanes which was used for size comparison. Numbers to the left of the gel are in bases.

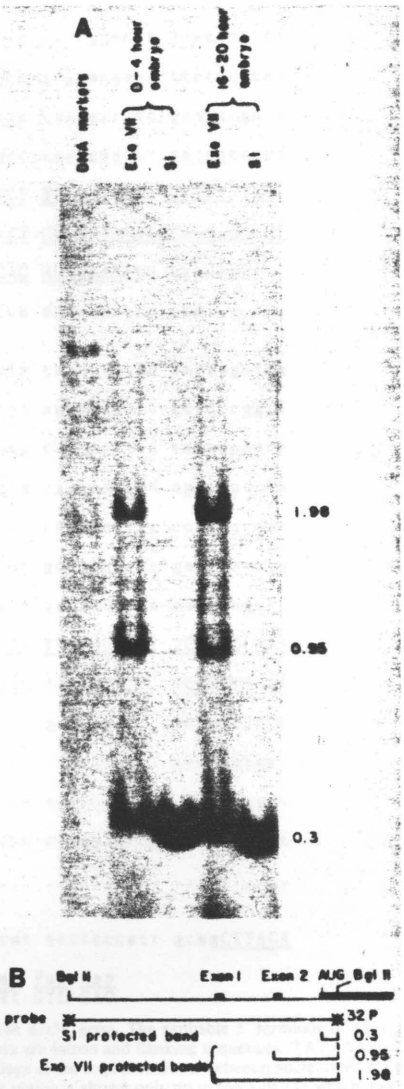


FIG. 3. (A) S1 nuclease and exonuclease VII mapping of the 5′ end of the gene. Poly(A)$^+$ RNA (1 μg) from 0- to 4-h-old or 16- to 20-h-old embryos was hybridized to a kinase-labeled *BglII* restriction fragment. The hybrid molecules were digested with either S1 nuclease or exonuclease VII. The protected products were run on a 1% alkaline agarose gel as shown. The left-hand lane contains a *Hin*dIII-*Eco*RI digest of lambda DNA which was used as a size marker. Numbers to the right of the gel are in kilobases. (B) The probe fragment and protected bands.

```
-260  .......... .......... .......... .......... ...tggaagt acactcttca
-200  tggcgatata caagacacac acaagcacga acacccagtt gcggaggaaa ttctccgtaa
-140  atgaaaaccc aatcggcgaa caattcatac ccatatatgg taaaagtttt gaacgcgact
 -80  tgagagcgga gagcattgcg gctgataagg ttttagcgct aagcgggctt tataaaacgg
 -20  gctgcgggac cagttttcat ATCACTACCG TTTGAGTTCT TGTGCTGTGT GGATACTCCT
  41  CCCGACACAA AGCCGCTCCA TCAGCCAGCA GTCGTCTAAT CCAGAGACAC CAAACCGAAA
 101  GACTTAATTT ATATTTATTT AATTAATTTT AATAAAACAC ACCAAATgta agtagctttc
 161  cccttcccaa caacaaaaca ccatcgaacc actcccacca agaaaaagca ata.......

 591  .......... aatgtacata catacagtat atgcatatta taatctgtaa aactagatca
 651  ggttcttgaa aatagtgacg taggcagccg ttttggctga agcagaaatt tttgccggtt
 711  tttcaaagtt gtagttgcaa aaatggagaa aaccttcgag cattcgttca tatacacaca
 771  ctcacgcgca aaataacgag agagagtgta tgtgtgtgtg agagagcgaa agccagacga
 831  cggtttgctt ttcgcctcga aacatgacca tatatggtca caaaacttgg ccgccgcaat
 891  tcaacacacc agcgctctcc ttcgcaccca tagcgaccat gcggcggagc gagcgagatg
 951  gcgagagcga gcgacgccta tggcgacgtc gacgcaggca gcgattgaaa aacgcagtta
1011  actggcattc aacattcacc agccactttc AGTCGGTTTA TTCCAGTCAT TCCTTTCAAA
1071  CCGTGCGGTC GCTTAGCTCA GCCTCGCCAC TTGCGTTTAC AGTAGTTTTC ACGCCTTGAA
1131  TTTGTTAAAT CGAACAAAAA Ggtaaagttt aactagcttt gaaaagtttc gtggctctta
1191  attgttaaat tttctagagt gcgtttagtg ttttttttt ttttatttt gtaatgttaa
1251  tttcgggttc caattcgagt tttaggcagc cgcactttta agggcgcata cacacaggca
1311  actgtgctct ctttgcggct ttcttttgca ccggcattcg ttaagtgtcg tctagaagct
1371  tctcccctcc .......... .......... .......... .......... ..........

1681  .......... ggtaaccaaa aactaatggg aaatccgcat tctttccatt gcagCTTACA

          Met Cys Asp Glu Glu Val Ala Ala Leu Val Val Asp
1741  AA  ATG TGT GAC GAA GAA GTT GCT GCT CTG GTT GTC GAC
```

FIG. 4. DNA sequence of exon 1, exon 2, intron, and 5′-flanking sequences of the *act5C* gene. The probable 5′ terminus of exon 1 is designated as nucleotide 1. Capital letters indicate exon sequences. Lower case letters are intron and flanking sequences. TATAA (exon 1) and TTAA (exon 2) sequences just preceding the transcription starts are underlined. Gaps in the sequence occur between nucleotides 213 and 601 and again between nucleotides 1381 and 1691. The sequence of the protein-coding region is shown only up to the *Sal*I site which was the 3′ termination point of the cDNA clone sequencing. The sequence between −217 and 213 was determined by D. Price. The rest was done by sequencing of both strands by the method of Maxam and Gilbert (35).

surmised (20), and they all contained *act5C* protein-coding sequences. The clones fell into two classes with respect to sequences further in the 5′ direction.

Of the 10 clones, 7 had roughly 147 bp of sequence joined directly to the previously mentioned splice acceptor site. A comparison of these sequences with known genomic sequences revealed that they are completely homologous, with a 147-bp stretch of sequence about 1.7 kb upstream of the initiator AUG (exon 1). A primer extension product made from an RNA of this structure would be 258 bases in length or the same as the longest of the four products seen. Five of the seven clones were full length, while one was 3 bp shorter and the final clone was 25 bp shorter. The clone with the 144-bp exon 1 sequence was of the correct size to correspond to the primer extension product with a length of 255 nucleotides. We do not know whether this is a real alternate start site or a premature stop by reverse transcriptase. We presume that the clone that was 25 bp short is an incomplete cDNA.

The remaining three clones had a different sequence spliced to the acceptor site. This sequence had no homology to a 450-bp region which included exon 1. Rather, it was identical to sequences roughly 0.7 kb upstream of the initiator AUG codon at the site predicted by the exonuclease VII analysis (exon 2). In two of these clones, this exon was 111 bp in length, while the third clone was 9 bp shorter,

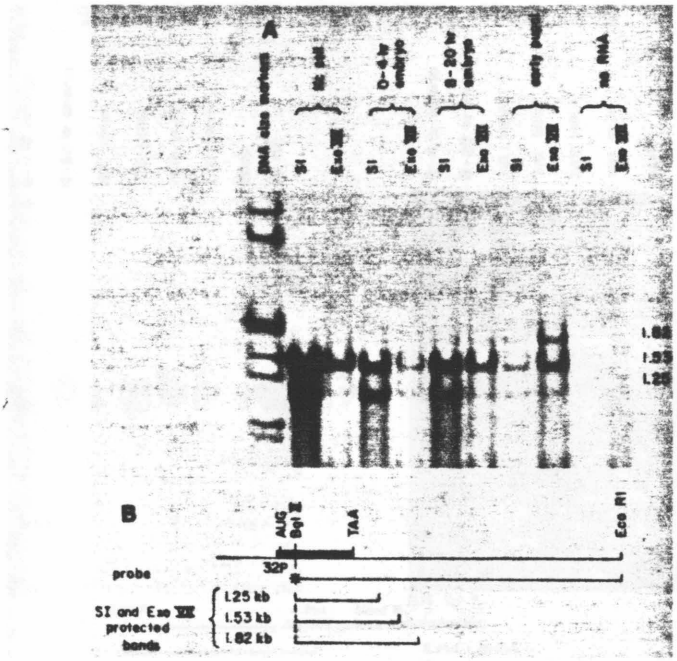


FIG. 5. (A) S1 nuclease and exonuclease VII mapping of the 3′ end of the gene. Poly(A)$^+$ RNA (1 μg) from the developmental stages indicated was hybridized to a 4-kb *Bgl*II-*Eco*RI fragment labeled with the Klenow fragment of DNA polymerase I. The hybrid molecules were digested with S1 nuclease or exonuclease VII, and the protected products were run on a 1% alkaline agarose gel. Numbers to the right of the gel are in kilobases. (B) The probe and protected fragments. The DNA markers are a lambda *Eco*RI-*Hin*dIII digest.

corresponding to the observed primer extension bands of length 222 and 213 nucleotides. Again, we do not know whether the shorter product is real or a reverse transcriptase artifact. This kind of microheterogeneity would not have been seen in our lower resolution exonuclease VII experiments.

Of the 10 clones, 4 had an extra guanine nucleotide between the poly(dC) tail used in the cDNA cloning and the transcription start that did not correspond to the genomic sequence. This may be due to reverse transcription of the cap nucleotide.

The correspondence between the cDNA clones and the primer extension products was confirmed by sequencing the latter (data not shown). These results indicate that at least two independent transcription start sites are used to give rise to two classes of actin 5C mRNA. They are alternately spliced to the body of the gene.

Three sites of polyadenylation are used by the *act5C* gene. The difference in length between the two alternate leader exons is not sufficient to be resolved on an RNA blot. The major source of the RNA size heterogeneity seen on RNA blots must lie elsewhere. Alternate splicing and different sites of polyadenylation are two possibilities.

The structure of the 3′ ends of the RNAs was investigated by S1 nuclease and exonuclease VII experiments. A 4-kb fragment originating within the protein-coding region (at the same *Bgl*II site used for the 5′ experiments) and extending 3 kb beyond the translation stop codon was used as the hybridization probe. This labeled fragment was hybridized to poly(A)$^+$ RNA from various developmental stages or from the Kc line of *D. melanogaster* cultured cells. The hybrids were treated with exonuclease VII or S1 nuclease, and the protected products were run on alkaline agarose gels (Fig. 5).

For each RNA used, the S1 nuclease- and exonuclease VII-protected products were the same length, indicating that no splicing occurs in the transcripts 3′ to the *Bgl*II site. In all cases, bands of 1.25 and 1.53 kb were seen. These corresponded to 3′-untranslated regions of 375 and 655 bp, respectively. In the 8- to 20-h-old embryo and early pupal RNAs, a larger 1.82-kb band was also seen. This corresponds to a 3′-untranslated region of 945 bp. These results indicate that different sites of polyadenylation are used to yield three size classes of RNAs. These size differences are of an appropriate size to acount for the three different *act5C* species with molecular lengths of 1.7, 1.95, and 2.2 kb, as resolved on RNA blots.

Differential expression of the *act5C* transcripts. It is also apparent from the S1 nuclease and exonuclease VII data and from RNA blots (20) that use of the different sites of polyadenylation is developmentally regulated. The intermediate site of polyadenylation was used in all stages of development tested. Transcripts of this type were generally the most abundant. Transcripts with the shortest 3′-untranslated region were also present at almost every stage tested, but they were usually present at a lower level and peaked at different times during development. Transcripts which used the last polyadenylation site and had the longest 3′-untranslated region were only found at specific times in development. They were most abundant at mid-embryo and early pupal stages.

To study the relative representation of the two different

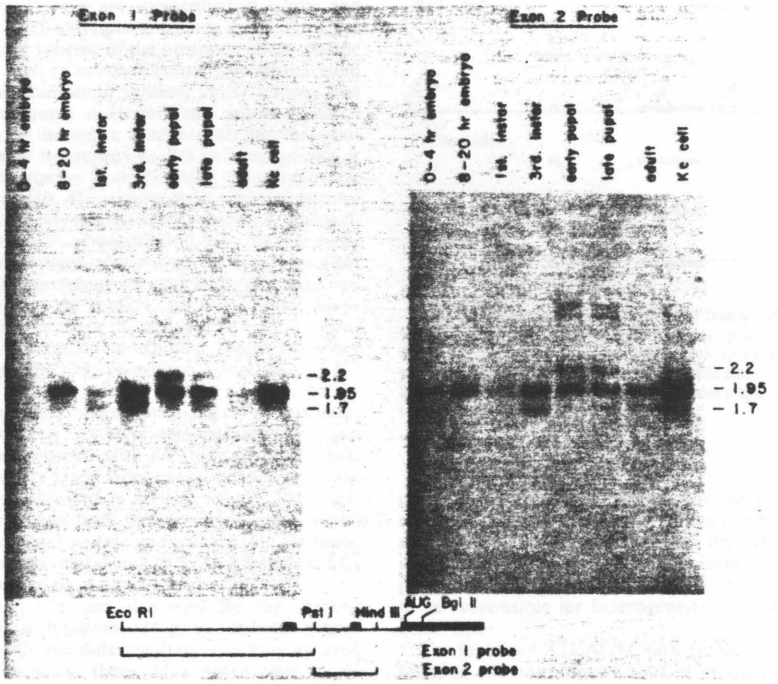


FIG. 6. Developmental RNA blots probed with exon 1- and exon 2-specific probes. Identical blots containing 1 μg of poly(A)$^+$ RNA from the developmental stages indicated, and Kc cells were probed with nick-translated exon 1- or exon 2-specific fragments, which are indicated in the diagram below. Numbers to the right of the gels are in kilobases.

leader exon sequences among *act5C* transcripts, parallel RNA blots were made which contained poly(A)$^+$ RNA from several different stages of *D. melanogaster* development and *D. melanogaster* Kc cells. These were probed with exon 1- and exon 2-specific probes (see Fig. 2; Fig. 6).

The blots that were hybridized with exon 1 and 2 probes looked very similar. Both leader exons were used in RNAs of 1.7, 1.95, and 2.2 kb, indicating that all three polyadenylation signals are used with each start site (Fig. 6). Both leader exons were represented in transcripts at each stage of development studied. The level of these transcripts appeared to rise and fall with a similar pattern. No remarkable differences in the pattern of expression of these two leader exons were seen, with the exception of two higher molecular weight species which were seen only in pupal RNA and were homologous only to the exon 2 probe. The identity of these bands is unknown.

DISCUSSION

We have shown that six different transcripts are made from the *D. melanogaster act5C* gene. These results are shown in Fig. 7 and discussed below.

The *act5C* gene has two transcription start sites. We found that two classes of RNA with respect to 5′-untranslated sequences are made from the *act5C* gene. Their 5′ sequences were encoded by two distinct leader exons which were located 1.7 and 0.7 kb upstream of the translation initiator AUG. These leader sequences were alternately spliced to a common splice acceptor site which is 8 bp 5′ of the AUG codon. We saw some microheterogeneity in the exact initiation sites of exons 1 and 2, both in primer extension experiments and cDNA clones. It is not clear whether all of these sites are actually used in vivo or if they are reverse transcriptase artifacts.

We did not see any developmental specificity in the use of the two leader exons. Transcripts containing the alternate exons displayed approximately the same pattern of expression through development. It is possible that transcription from the two start sites is regulated but that we would not have detected the difference on the RNA blots that we did. For example, there may be tissue-specific expression of the two *act5C* start sites at a given developmental stage. Alternately, transcription from the two promoters could be linked to the cell cycle. Further experiments will be necessary to test these possibilities.

Many other cases of alternate transcription start sites and alternate first exons for eucaryotic genes are known. The resulting transcripts are often made in a developmental or tissue-specific manner, as in the following examples. In each case, the structural organization of the transcripts is slightly different.

The mouse α-amylase 1^a gene (25, 26, 48, 57), like *act5C*, has two leader exons which are alternately spliced to a common third exon. The distal promoter is 30 times stronger than the proximal promoter and drives transcription of a parotid gland-specific mRNA. Transcription from the proximal promoter produces an mRNA which accumulates to a level that is 100 times lower in the liver, parotid gland, and pancreas. Two classes of *D. melanogaster* ADH mRNA that

differ only at their 5′ ends are transcribed in adults and larvae (5). The adult transcripts are made from the distal promoter and require splicing of the upstream adult-specific leader exon to the next exon which contains 5′-untranslated sequences and the translation initiator AUG codon. The larval transcripts originate at the proximal promoter which lies just upstream of the splice acceptor site for the adult transcript. The larval transcripts do not require splicing at their 5′ ends. Transcription from the distal promoter of the yeast invertase gene (10, 43) is under the control of glucose and produces an RNA which encodes the secreted, glycosylated form of the enzyme. Transcription from the proximal start site, on the other hand, is constitutive. The transcripts which originate at the proximal site start within the sequence of the signal peptide of the protein coded for by the larger RNA and therefore produce a protein without a signal sequence, which is therefore intracellular. Neither product requires 5′ splicing. The chicken myosin light chain 1 (LC_1) and 3 (LC_3) gene (38) uses alternate splicing of two separate leader exons to produce different transcripts. In this case, the splice acceptor site is in the protein-coding region, and the RNAs code for proteins with different amino-terminal sequences. Transcripts made from the distal promoter encode LC_1 which is expressed in skeletal muscle and heart. LC_3 transcripts are made from the proximal promoter and are expressed in skeletal muscle and gizzard. LC_1 is made earlier in embryonic development (10 to 14 days) than LC_3 which appears in embryos at 14 to 15 days.

Multiple initiation sites are also used for the chicken ovomucoid (21, 30) and lysozyme (24) genes. In both of these cases, however, while the different start sites may be used with different efficiencies, there does not appear to be developmental or tissue-specific use of the start sites.

Many genes have been shown to have microheterogeneity around the cap sites. In these cases several mRNAs are made from a single gene, but they differ by only a few bases in length. Some examples are the human alpha actin gene (27), the *D. melanogaster* myosin LC-2 gene (41), the adenovirus type 2 EII gene (1), the simian virus 40 late genes (16, 22, 23), the polyomavirus late genes (17), the chicken ovalbumin gene (34), the chicken lysozyme gene (24), and the yeast Adh-I gene (3).

In conclusion, this review of the literature indicates that for most of the known cases of alternate transcription start sites and alternate first exons, there is some tissue or stage specificity of expression, a resulting difference in the gene product, or both. In contrast, there is no indication of such differences for the two transcription start sites of the actin 5C gene.

Exon 2 lacks a TATAA sequence. An examination of the DNA sequences immediately 5′ of the two cap sites showed that some sequences present in this region in many eucaryotic genes are lacking. We did find a TATAA sequence 30 bp upstream of the transcription start at exon 1, but no such sequence was found in the vicinity of exon 2. An AATT sequence was found 30 bp upstream of the exon 2 cap site. Neither start site was preceded by a CCAAT sequence which is generally found roughly 65 bp upstream from the transcription start.

There are several cases of eucaryotic genes with no TATAA sequence in the region 30 bases upstream of the transcription start. Some examples are the *D. melanogaster* myosin alkali LC (14), the human antithrombin III gene (44), the simian virus 40 late genes (16, 45), and the adenovirus type 2 EII gene (1). Several of the genes mentioned earlier, namely the *D. melanogaster* Adh-I adult promoter (5), the

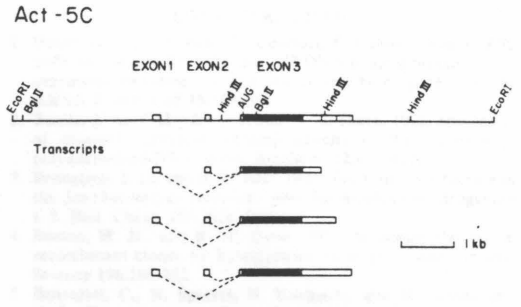


FIG. 7. Structure of the *act5C* gene transcripts. Schematic representation of the generation of transcripts from the *D. melanogaster act5C* gene. The solid boxes represent protein-coding sequences, while the open boxes represent transcribed, untranslated sequences. Intron and flanking sequences are designated by the solid lines.

chicken myosin LC (38), chicken ovomucoid (21, 30), chicken lysozyme (24), *D. melanogaster* myosin LC-2 (41), and yeast Adh-I (3), are missing the consensus TATAA sequences but have an AT-rich region in the analogous position. The lack of TATAA sequences has been postulated to be responsible for heterogeneity around the cap site in some cases.

The sequence ATCAGTC or a similar sequence has been found at the transcription start of several *D. melanogaster* mRNAs (51). The sequences ATCACTA and TTCAGTC were found at the starts of exons 1 and 2, respectively.

Do exons 1 and 2 have independent promoters? Two lines of evidence suggest that the sequences at and upstream of exon 1 are sufficient for the independent initiation of transcription. Price and co-workers (Parker et al., personal communication) have tested a DNA segment containing exon 1 and upstream sequences, but not exon 2, in an in vitro transcription system. They found a high level of transcription initiation at the same site that we mapped as the cap site for exon 1. Furthermore, we made fusions between putative actin promoter sequences and the bacterial chloramphenicol acetyltransferase gene and tested these constructs by transient assays in *D. melanogaster* Kc cells (unpublished data). We found that sequences upstream of exon 1 are sufficient to cause a high level of expression from this promoter in these cells.

However, at present, there is no evidence as to whether transcription can initiate at the cap site of exon 2 under the control of promoter sequences immediately upstream, or whether sequences around the promoter for exon 1 are necessary for production of the RNA beginning at exon 2. Experiments to test this question are in progress.

Multiple sites of polyadenylation. Three major size classes of RNA exist for the *act5C* gene. We found that the major source of this size heterogeneity lies at the 3′ end of the gene at which three sites of polyadenylation are used. The first two sites were used in all stages of development studied, but the corresponding mRNAs were not always present in the same relative amounts. The transcripts which terminated after the second polyadenylation signal were always the most abundant. The third and distant site was used only at particular times in development, especially at mid-embryo and early pupal stages. There is as yet no known functional

significance to the different developmental choices of polyadenylation sites in the *act5C* gene.

Most eucaryotic genes have a sequence AATAAA that is roughly 30 bp upstream of the site of polyadenylation in the RNA. We did not determine the precise sites of poly(A) addition for these transcripts. We estimate from S1 nuclease and exonuclease VII analyses that the sites of poly(A) addition lie rougly 375, 655, and 945 bases from the translation stop codon. There are no AATAAA sequences in the 3′-untranslated region of the *act5C* gene (D. Price, E. A. Fyrberg, and B. J. Bond, unpublished data). However, there are closely related sequences 20 to 60 bp upstream from each presumed poly(A) addition site. That is, AATATA and AATGAAA sequences are found roughly 55 and 35 bp, respectively, in front of the proximal poly(A) site. The AATATA sequence is a minor variant that has been seen in other eucaryotic genes, while the AATGAAA variant has been found by Wickens and Stephenson (56) to be an inefficient substrate for poly(A) addition. TATAAA and AATCAAA sequences lie roughly 60 and 20 bp 5′ to the middle poly(A) site. The TATAAA is another minor variant in eucaryotes, while the AATCAAA has not been seen previously. The distal site of polyadenylation is preceded by an AATTAAA sequence which lies roughly 45 bp upstream. This variant has been seen in 12% of the RNAs compiled (8).

Many eucaryotic genes generate multiple transcripts from a single gene through the use of different sites of polyadenylation. Some examples of this are the mouse α-amylase gene (54), the dihydrofolate reductase gene (49, 50), the bovine prolactin gene (47), the ovalbumin gene (33), the ovalbumin X and Y genes (28, 32), the human β-tubulin gene (31), the chicken vimentin gene (9, 58), the chicken ovomucoid gene (21), the immunoglobulin heavy-chain genes (11, 13, 37), the major late adenovirus type 2 transcription unit (39), the yeast Adh gene (3), the α-2 microglobulin gene (55), the β-2 microglobulin gene (42), and the *D. melanogaster* myosin alkali LC gene (14).

In many cases the alternative sites of polyadenylation are used with different frequencies, and in the case of the chicken vimentin gene (9) there is some tissue specificity in the use of the different sites. The significance of having multiple mRNA species transcribed from a single gene and differing only in the lengths of their 3′-untranslated sequences, however, is in general unknown.

The only case in which the alternate use of polyadenylation sites has a known functional significance is the case of the immunoglobulin heavy-chain genes. Early et al. (13) showed that the mRNAs for the secreted and membrane-bound forms of immunoglobulin M are transcribed from the same gene and differ only at their 3′ ends. A similar arrangement has been found for all of the immunoglobulin heavy-chain classes. Milcarek and Hall (37) showed that the ratio of the membrane-bound and secreted forms of the γ-2b mRNA in the cell is determined by selective use of the alternate polyadenylation sites.

ACKNOWLEDGMENTS

We are grateful to Bette Korber, David Price, Joanne Topol, and Carl Parker for communicating results prior to publication and to Willian Mattox and Eric Fryrberg for helpful discussions. We thank Larry Kauvar for providing the *D. melanogaster* (8- to 20-h old cDNA library in lambda *gt*10 and Mickey Chien-Tsung Hu for running a primer extension sample on a sequencing gel.

This work was supported by a Public Health Service research grant from the National Institutes of Health.

LITERATURE CITED

1. **Baker, C. C., J. Herisse, G. Courtois, F. Gailbert, and E. Ziff.** 1979. Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5′ end. Cell **18**:569–580.
2. **Bantle, J. A., I. H. Maxwell, and W. E. Hahn.** 1976. Specificity of oligo(dT) cellulose chromatography in the isolation of polyadenylated RNA. Anal. Biochem. **72**:413–427.
3. **Bennetzen, J. L., and B. D. Hall.** 1982. The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. J. Biol. Chem. **257**:3018–3025.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening (lambda)gt recombinant clones by hybridization of single plaques in situ. Science **196**:180–182.
5. **Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner.** 1983. The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell **33**:125–133.
6. **Berk, A., and P. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
7. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of simian virus 40. Proc. Natl. Acad. Sci. USA **75**:1274–1278.
8. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site. Cell **41**:349–359.
9. **Capetanaki, Y. G., J. Ngai, C. N. Flytzanis, and E. Lazarides.** 1983. Tissue-specific expression of two mRNA species transcribed from a single vimentin gene. Cell **35**:411–420.
10. **Carlson, M., and D. Botstein.** 1982. Two differentially regulated mRNAs with different 5′ ends encode secreted and intracellular forms of yeast invertase. Cell **28**:145–154.
11. **Cheng, H., F. R. Blattner, L. Fitzmaurice, J. F. Mushinski, and P. W. Tucker.** 1982. Structure of genes for membrane and secreted murine IgD heavy chains. Nature (London) **296**: 410–415.
12. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of Biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**: 5294–5299.
13. **Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood.** 1980. Two mRNAs can be produced from a single immunoglobulin (mu) gene by alternative RNA processing pathways. Cell **20**:313–319.
14. **Falkenthal, S., V. P. Parker, and N. Davidson.** 1985. Developmental variations in the splicing pattern of transcripts from the *Drosophila* gene encoding myosin alkali light chain result in different carboxly-terminal amino acid sequences. Proc. Natl. Acad. Sci. USA **82**:449–453.
15. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
16. **Fiers, W., R. Contreras, G. Haegeman, R. Rogiers, A. Van de Voorde, H. Van Heuverswyn, J. Van Herreweghe, G. Volckaert, and M. Ysebaert.** 1978. Complete nucleotide sequence of SV40 DNA. Nature (London) **273**:113–120.
17. **Flavell, A. J., A. Cowie, S. Legon, and R. Kamen.** 1979. Multiple 5′ terminal cap structures in late polyoma virus RNA. Cell **16**:357–371.
18. **Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixter, and N. Davidson.** 1981. The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. Cell **24**:107–116.
19. **Fyrberg, E. A., K. L. Kindle, N. Davidson, and A. Sodja.** 1980. The actin genes of *Drosophila*: a dispersed multigene family. Cell **19**:365–378.
20. **Fyrberg, E. A., J. W. Mahaffey, B. J. Bond, and N. Davidson.** 1983. Transcripts of the six *Drosophila* actin genes accumulate in a stage and tissue specific manner. Cell **33**:115–123.
21. **Gerlinger, P., A. Krust, M. LeMeur, F. Perrin, M. Cochet, F. Gannon, D. Dupret, and P. Chambon.** 1982. Multiple initiation and polyadenylation sites for the chicken ovomucoid transcrip-

tion unit. J. Mol. Biol. **162**:345–364.
22. **Ghosh, P., V. Reddy, J. Swinscoe, P. Choudary, P. Lebowitz, and S. Weissman.** 1978. The 5′ terminal leader sequences of late 16S mRNA from cells infected with simian virus 40. J. Biol. Chem. **253**:3643–3647.
23. **Ghosh, P., V. Reddy, J. Swinscoe, P. Lebowitz, and S. Weissman.** 1978. Heterogeneity and 5′-terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. **126**:813–846.
24. **Grez, M., H. Land, K. Giesecke, and G. Schutz.** 1981. Multiple mRNAs are generated from the chicken lysozyme gene. Cell **25**:743–752.
25. **Hagenbuchle, O., U. Schibler, S. Petrucco, G. C. Van Tuyle, and P. K. Wellauer.** 1985. Expression of mouse *Amy-2*^{a} alpha-amylase genes is regulated by strong pancreas-specific promoters. J. Mol. Biol. **185**:285–293.
26. **Hagenbuchle, O., M. Tosi, U. Schibler, R. Bovey, P. K. Wellauer, and R. A. Young.** 1981. Mouse liver and salivary gland alpha-amylase mRNAs differ only in 5′ nontranslated sequences. Nature (London) **289**:643–646.
27. **Hanauer, A., M. Levin, R. Heilig, D. Daegelen, A. Kahn, and J. L. Mandel.** 1983. Isolation and characterization of cDNA clones for human skeletal muscle alpha actin. Nucleic Acids Res. **11**:3503–3516.
28. **Hellig, R., F. Perrin, F. Gannon, J. L. Mandel, and P. Chambon.** 1980. The ovalbumin gene family: structure of the X gene and evolution of duplicated split genes. Cell **20**:625–637.
29. **Hunkapiller, M., S. Kent, M. Caruthers, W. Dreyer, J. Firca, C. Giffin, S. Horvath, T. Hunkapiller, P. Tempst, and L. Hood.** 1984. A microchemical facility for the analysis and synthesis of genes and proteins. Nature (London) **310**:105–111.
30. **Lai, E. C., D. R. Roop, M. Tsai, S. L. C. Woo, and B. O'Malley.** 1982. Heterogeneous initiation regions for transcription of the chicken ovomucoid gene. Nucleic Acids Res. **10**:5553–5567.
31. **Lee, M. G., S. A. Lewis, C. D. Wilde, and N. J. Cowan.** 1983. Evolutionary history of a multigene family: an expressed human beta-tubulin gene and three processed pseudogenes. Cell **33**:477–487.
32. **LeMeur, M., N. Glanville, J. L. Mandel, P. Gerlinger, R. Palmiter, and P. Chambon.** 1981. The ovalbumin gene family: hormonal control of X and Y gene transcription and mRNA accumulation. Cell **23**:561–571.
33. **LeMeur, M. A., B. Galliot, and P. Gerlinger.** 1984. Termination of the ovalbumin gene transcription. EMBO J. **3**:2779–2786.
34. **Malek, L. T., W. H. Eschenfeldt, T. W. Munns, and R. E. Rhoads.** 1981. Heterogeneity of the 5′ terminus of hen ovalbumin messenger ribonucleic acid. Nucleic Acids Res. **9**:1657–1673.
35. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
36. **McDonell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline agarose gels. J. Mol. Biol. **110**:119–146.
37. **Milcarek, C., and B. Hall.** 1985. Cell-specific expression of secreted versus membrane forms of immunoglobulin gamma 2b mRNA involves selective use of alternate polyadenylation sites. Mol. Cell. Biol. **5**:2514–2520.
38. **Nabeshima, Y., Y. Fujii-Kuriyama, M. Muramatsu, and K. Ogata.** 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. Nature (London) **308**:333–338.
39. **Nevins, J. R., and M. C. Wilson.** 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. Nature (London) **290**:113–118.
40. **Parker, C. S., and J. Topol.** 1984. A *Drosophila* RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell **36**:357–369.
41. **Parker, V. P., S. Falkenthal, and N. Davidson.** 1985. Characterization of the myosin light-chain-2 gene of *Drosophila melanogaster*. Mol. Cell. Biol. **5**:3058–3068.
42. **Parnes, J. R., R. R. Robinson, and J. G. Seidman.** 1983. Multiple mRNA species with distinct 3′ termini are transcribed from the beta$_2$-microglobulin gene. Nature (London) **302**:449–452.
43. **Perlman, D., H. O. Halvorson, and L. E. Cannon.** 1982. Presecretory and cytoplasmic invertase polypeptides encoded by distinct mRNAs derived from the same structural gene differ by a signal sequence. Proc. Natl. Acad. Sci. USA **79**:781–785.
44. **Prochownik, E. V.** 1985. Relationship between an enhancer elemenent in the human antithrombin III gene and an immunoglobulin light-chain enhancer. Nature (London) **316**:845–848.
45. **Reddy, V., B. Thimmappaya, R. Dhar, K. Subrumanian, S. Zain J. Pan, P. Ghosh, M. Celma, and S. Weissman.** 1978. The genome of simian virus 40. Science **200**:494–500.
46. **Sanchez, F., S. L. Tobin, U. Rdest, E. Zulauf, and B. J. McCarthy.** 1983. Two *Drosophila* actin genes in detail. Gene structure, protein structure, and transcription during development. J. Mol. Biol. **163**:533–551.
47. **Sasavage, N. L., M. Smith, S. Gillam, R. P. Woychik, and F. M. Rothman.** 1982. Variation in the polyadenylation site of bovine prolactin mRNA. Proc. Natl. Acad. Sci. USA **79**:223–227.
48. **Schibler, U., O. Hagenbuchle, P. K. Wellauer, and A. C. Pittet.** 1983. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene *Amy-1*^{a} in the parotid gland and the liver. Cell **33**:501–508.
49. **Setzer, D. R., M. McGrogan, J. H. Nunberg, and R. T. Schimke.** 1980. Size heterogeneity in the 3′ end of dihydrofolate reductase messenger RNAs in mouse cells. Cell **22**:361–370.
50. **Setzer, D. R., M. McGrogan, and R. T. Schimke.** 1982. Nucleotide sequence surrounding multiple polyadenylation sites in the mouse dihydrofolate reductase gene. J. Biol. Chem. **257**:5143–5147.
51. **Snyder, M., J. Hirsh, and N. Davidson.** 1981. The cuticle genes of *Drosophila*: a developmentally regulated gene cluster. Cell **19**:365–378.
52. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
53. **Tobin, S. L., E. Zulauf, F. Sanchez, E. A. Craig, and B. J. McCarthy.** 1980. Multiple actin-related sequences in the *Drosophila melanogaster* genome. Cell **19**:121–131.
54. **Tosi, M., R. A. Young, O. Hagenbuchle, and U. Schibler.** 1981. Multiple polyadenylation sites in a mouse alpha-amylase gene. Nucleic Acids Res. **9**:2313–2323.
55. **Unterman, R. D., K. R. Lynch, H. L. Nakhasi, K. P. Dolan, J. W. Hamilton, D. V. Cohn, and P. Feigelson.** 1981. Cloning and sequence of several $\alpha_{2\mu}$-globulin cDNAs. Proc. Natl. Acad. Sci. USA **78**:3478–3482.
56. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent mRNA 3′ end formation. Science **226**:1045–1051.
57. **Young, R. A., O. Hagenbuchle, and U. Schiber.** 1981. A single mouse alpha-amylase gene specifies two different tissue-specific mRNAs. Cell **23**:451–458.
58. **Zehner, Z. E., and B. M. Peterson.** 1983. Characterization of the chicken vimentin gene: single copy gene producing multiple mRNAs. Biochemistry **80**:911–915.

CHAPTER 5

Transcription from each of the *Drosophila* act5C leader exons is driven by a separate, functional promoter

ABSTRACT

The *Drosophila* act5C gene has two leader exons which are alternately spliced to the remainder of the gene. In this way two classes of transcripts are made with respect to the 5' untranslated sequences both of which are present in *Drosophila* Kc cell mRNA. In order to define the sequences necessary for transcription from each start site and to determine if each is driven by a separate promoter, 5' flanking regions from the act5C gene were inserted upstream of the bacterial chloramphenicol acetyltransferase (CAT) gene and tested for promoter activity by transient assays in *Drosophila* Kc cells. We show that both leader exons are preceded by separate, functional, promoters. The exon 1 proximal promoter contains at least two regions important for optimal expression. One is greater than 1.9 kb upstream from the exon 1 cap site while the other lies between 1.2 and 0.09 kb upstream of the cap site. The promoter elements necessary for transcription from exon 2 are within 450 bases upstream of the cap site. The data suggest that, in some constructions, transcription initiation at exon 1 inhibits transcription initiation at exon 2.

Examination of the DNA sequences reveals two identical 10 base regions of dyad symmetry one of which is 110 bases upstream of exon 1 and the other is 183 bases upstream of exon 2. These are of the same form as the CCArGG sequences found upstream of many mammalian and chicken actin genes which have been proposed to be transcription regulatory sequences.

INTRODUCTION

Drosophila melanogaster has six actin genes which are dispersed among the three major chromosomes (16,17,45). The members of the multigene family

have been shown to be expressed in a developmental stage and tissue specific manner (18,41).

Transcripts from the act5C gene are found in all stages of development studied and are especially abundant in early embryos and early pupae (18). These are stages at which there is rapid cell division but little muscle differentiation. This gene is also one of the two actin genes expressed in ovaries. For these reasons the act5C gene is thought to encode a cytoplasmic form of actin.

We have recently shown (5) that at least six messenger RNAs are made from this gene. The structural basis for the transcript heterogeneity is the presence of two transcription initiation sites and three polyadenylation sites. The two initiation sites lie 1.7 kb (exon 1) and 0.7 kb (exon 2) upstream of the common splice acceptor which is 8 base pairs 5' to the translation initiator AUG. Transcripts containing each of the alternate leader exons rise and fall with approximately the same pattern during development. Both exon 1 and exon 2 containing transcripts are found in Kc cells.

Korber, Topol, Price, and Parker (personal communication) demonstrated the existence of a promoter upstream of exon 1 of the act5C by *in vitro* transcription experiments with Kc cell extracts. A series of 5' deletions was used to study the *in vitro* sequence requirements for transcription. They found that 210 bases of sequence upstream of the cap site is enough for high levels of transcription in the assay and deletion to -85 causes a threefold drop in transcription. They used DNA footprint analyses to identify two regions in the DNA where factors present in the Kc cell extracts bind the DNA. One of these is centered around the TATA box (37) while the other is at -88 to -117 (Topol and Parker, personal communication). They have not studied transcription from exon 2 in the *in vitro* system.

Several investigators have demonstrated the possibility of carrying out transient transformation experiments in *Drosophila* cultured cells (3,13,29,33,34). This method offers a practical way to study large numbers of constructs in order to better define sequences necessary for expression of *Drosophila* genes. The bacterial chloramphenicol acetyltransferase (CAT) gene (21) is a very convenient marker for gene expression in transient assay experiments. The assay for the enzyme is very sensitive and is linear over a wide range of activities. Furthermore there is no analogous enzyme activity in eucaryotic cells, thus eliminating the problem of background activity.

We were interested in determining whether each of the two act5C transcription starts is preceded by a functional promoter. We were further interested in identifying the particular sequence elements necessary for transcription from each promoter and in comparing the *in vivo* and *in vitro* sequence requirements for transcription.

We have made several fusions of actin promoter sequences to CAT sequences and tested their ability to drive transcription of the CAT gene in the Kc cell transient assay system. We have found that exon 1 and exon 2 are each preceded by a functional promoter and that exon 2 can be transcribed in the absence of sequences upstream of exon 1. At least two regions of upstream sequences are important for optimal expression from exon 1 one of which is greater than 1.9 kb upstream of exon 1. Upstream sequences important for transcription initiation of exon 2 lie within 450 bases of its cap site.

MATERIALS:

All of the oligonucleotides used were synthesized in the laboratory of S. Horvath (California Institute of Technology) on an automated DNA synthesizer (27). *Drosophila* Kc cells were obtained from Carl Parker (California Institute of Technology). They were maintained at 22-23^{o}C in D20 media (15) from the laboratory of John D. O'Connor (U.C.L.A.).

METHODS:

Construction of puc^{PL}CAT

The vector puc^{SV0}CAT was obtained from Nevis Fregien (C.I.T.). This plasmid was made by inserting the BglII-BamHI fragment of $pA_{10}cat_2$ (28) into the BamHI site of the puc 18 polylinker and then removing the SV40 promoter containing Hind III fragment. It was linearized at the single Hind III site and phosphatased. It was then ligated to the oligonucleotide of sequence 5'AGCTTCATGATCTCGAGGGCCCTGCAGCATGCAGATCT3' and the complement 5'AGCTAGATCTGCATGCTGCAGGGCCCTCGAGATCATGA3'. These oligonucleotides hybridized to form a duplex with a multiple cloning site and with protruding Hind III sticky ends on each side. When ligated into the vector, the Hind III site is retained at only one end. In order to orient the polylinker with respect to the vector, the miniprep DNAs were cut with Hind III and a second enzyme which cut asymmetrically elsewhere in the plasmid, run on a gel, and blotted. They were hybridized to one labeled strand of the oligomer. The number (0,1 or 2) and identity of the labeled bands indicated the orientation of the inserted linker(s) directly adjacent to the puc^{SV0} sequences (see lower part of Fig. 1A). Clones that gave the desired pattern of hybridization were digested with an enzyme for which there was a unique restriction site internal to the polylinker and ligated under dilute conditions to

Figure 1A. Restriction map of puc^{PL}CAT.

The multiple cloning site shown at the bottom of the panel was inserted at the single HindIII site of puc^{SV0}CAT (see text) to make puc^{PL}CAT. This vector contains the CAT gene (open box) and a 3' untranslated region from SV40 (hatched box) which contains an intron and a polyadenylation site. The plasmid carrier is puc 18 which includes the polylinker adjacent to the SV40 sequences.

Figure 1B. Schematic map of act5C promoter - CAT fusions.

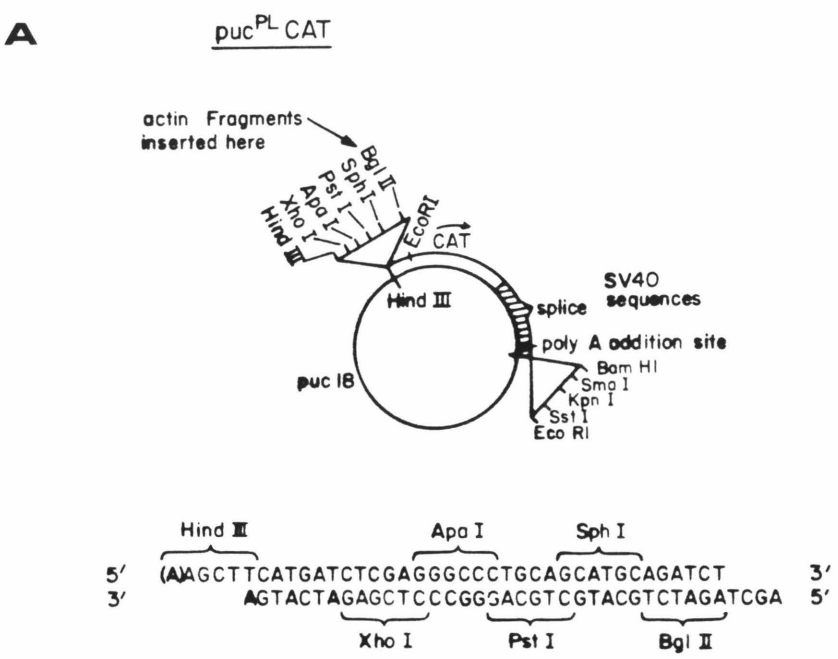
A restriction map of the act5C gene is shown at the top. Exons are represented by open boxes while introns and flanking sequences are solid lines.

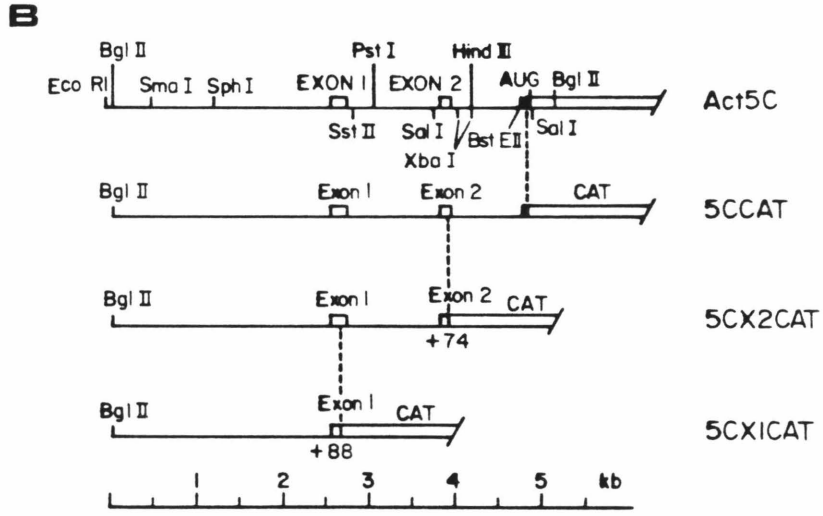
The act5C promoter-CAT gene junctions for the indicated clones are shown below. All of the cloned inserts have the same 5' endpoint at the BglII site of the actin gene, but they continue to different extents in the 3' direction. 5CX2CAT is joined to the CAT gene at nucleotide +74 of exon 2. The actin/CAT junction of 5CX1CAT is at nucleotide +88 of exon 1. The actin/CAT junction of 5CCAT is shown in more detail in Figure 1C.

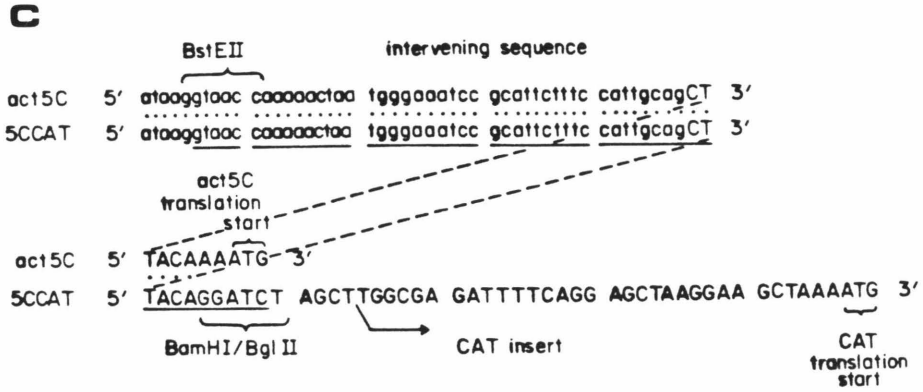
Figure 1C. The sequence of the act5C-CAT junction in 5CCAT.

The sense strand of act5C is compared to the sense strand of the 5CCAT actin/CAT junction in order to show at what point they diverge. The 5CCAT construct was made as follows: Act5C was cut at the BstEII site and ligated to the synthetic oligonucleotides whose sequence is underlined. Following digestion with BglII, the modified act5C fragment was ligated to the vector puc^{PL}CAT. The purpose of the oligonucleotide linker was to reconstruct the 3' end of the act5C intron including a splice acceptor site, and to provide the BglII cohesive end for cloning into the puc^{PL}CAT vector. Homology between 5CCAT and act5C are shown by dashes. The actin and CAT translation start sites are

indicated. Lower case and upper case letters designate intron and exon sequences, respectively.


A
pucPL CAT
actin Fragments inserted here
Bgl II
Sph I
Pst I
Apa I
Xho I
Hind III
EcoRI
CAT
Hind III
SV40 sequences
splice
poly A addition site
puc 18
Bam HI
Sma I
Kpn I
Sst I
Eco RI
Hind III
Apa I
Sph I
5′ (A)AGCTTCATGATCTCGAGGGCCCTGCAGCATGCAGATCT 3′
3′ AGTACTAGAGCTCCCGGGACGTCGTACGTCTAGATCGA 5′
Xho I
Pst I
Bgl II


B
Bgl II
Pst I
Hind III
Eco RI
Sma I
Sph I
EXON 1
EXON 2
AUG
Bgl II
Act5C
Sst II
Sal I
Bst EII
Sal I
Xba I
Bgl II
Exon 1
Exon 2
CAT
5CCAT
Bgl II
Exon 1
Exon 2
CAT
+74
5CX2CAT
Bgl II
Exon 1
CAT
+88
5CX1CAT
1
2
3
4
5
kb


C
BstEII
intervening sequence
act5C 5′ ataaggtaac caaaaactaa tgggaaatcc gcattctttc cattgcagCT 3′
5CCAT 5′ ataaggtaac caaaaactaa tgggaaatcc gcattctttc cattgcagCT 3′
act5C translation start
act5C 5′ TACAAAATG 3′
5CCAT 5′ TACAGGATCT AGCTTGGCGA GATTTTCAGG AGCTAAGGAA GCTAAAATG 3′
BamHI/Bgl II
CAT insert
CAT translation start

eliminate extra copies of the polylinker. The selected clone was sequenced by the method of Maxam and Gilbert (30) across the polylinker and was shown to have a single linker insert of the expected sequence in the desired orientation.

Construction of actin CAT fusions

The plasmid 5CCAT (Fig. 1B) was constructed as follows: The act5C gene was cut at the single BstEII site which cleaves at the 3' end of the intron and was ligated to oligonucleotides of sequences

1. 5'GTAACCAAAAACTAATGGGAAATC3'
2. 5'CGCATTCTTTCCATTGCAGCTTACAG3'
3. 5'GAATGCGGATTTCCCATTAGTTTTTG3'
4. 5'GATCCTGTAAGCTGCAATGGAAA3'

one of which was kinase labeled with ^{32}P ATP. These oligonucleotides together comprise a 50 base pair stretch with a BstEII overhang at one end and a BamHI overhang at the other which serve to add back the 3' end of the act5C intervening sequence and six of eight bases of the untranslated sequences 5' to the translation initiator AUG (see Fig. 1C). This sequence includes a splice acceptor site and a proposed lariat site, CTAAT. The DNA was digested with BamHI to remove excess linker and with BglII which cuts 2.5 kb upstream of exon 1. The appropriate BglII-BamHI fragment was gel isolated and ligated to phosphatased puc^{PL}CAT that had been linearized with BglII. The oligonucleotide insertion was checked by DNA sequencing (30) in the clone finally selected.

In order to make 5CX1CAT, a BglII-BamHI fragment from the clone A29 (a gift of David Price) was cloned into the BglII site of puc^{PL}CAT. A29 is a pBR322 subclone of the Act5C upstream region from the EcoRI site to a BamHI site which had been added by linker insertion in the middle of exon 1 at base +88.

To make 5CX2CAT, Act 5C was cut with XbaI and then treated with Bal 31 for various lengths of time. Aliquots from each time point were tested for extent of digestion. DNA from the appropriate timepoint was ligated to BamHI linkers and then digested with BamHI and BglII. The desired fragment was gel isolated and ligated to BglII cut, phosphatased puc^{PL}CAT. The extent of Bal 31 digestion of correctly oriented inserts was determined precisely by DNA sequencing (30). The clone, 5CX2CAT, that was used contains 74 base pairs of exon 2 before the BamHI linker.

Construction of Deletion Clones

The PstI and SphI deletion clones were made by digesting each of the above actin CAT fusion clones with PstI or SphI, and then religating under very dilute conditions to eliminate the small actin fragment. Series of unidirectional deletions were made from 5CCAT and 5CX1CAT with Exonuclease III by the method of Henikoff (26). In each case the starting plasmids were digested with ApaI and BglII. ApaI leaves a 3' overhang and was protected from digestion with Exo III. The deletions occurred from the BglII site in the direction of the actin sequences.

Kc Cell Transfections

The Kc cells were plated at a density of $2x10^7$ cells/T75 flask in 10 ml of D20 media. 26-30 hours later, 10 µg of calcium phosphate precipitated (22) supercoiled plasmid DNA was added per flask. The medium was not changed after the addition of DNA. The cells were collected 48 hours later. They were rinsed once with 1 ml of 0.25M Tris pH 7.8, resuspended in 100 µl of the same buffer, and frozen in dry ice/ethanol. Hsp CAT 1 (a gift of P. Di Nocera and I. Dawid, 13) was used initially in the transient assay system to optimize conditions. Protein concentrations of cell extracts were determined by the Biorad protein assay which is taken from the method of Bradford (7).

CAT Assays

CAT assays (21) were carried out in a final reaction volume of 100 µl with 0.2 µCi ^{14}C chloramphenicol per reaction and 0.8 mM acetyl coenzyme A. Reactions were incubated at 37°C for two hours and were stopped by the addition of 0.9 ml of ethyl acetate. The organic layer was dried in a speed vac and resuspended in 10 µl of ethyl acetate. Reaction products were separated by thin layer chromatography as described (21). Activities are expressed as percentage of ^{14}C chloramphenicol acetylated. Each plasmid was transfected into cells at least three times. The activities were normalized to 5CCAT or 5CX1CAT and averaged. Assays were performed such that CAT activity would be in the linear range. Activities were generally below 50% acetylation.

RESULTS

Design of Vector

In order to study the promoter activity of *Drosophila* act5C sequences, we have made fusions between various actin promoter sequences and the bacterial chloramphenicol acetyltransferase (CAT) gene. These were to be assayed by transient transformation in *Drosophila* Kc cells. We needed a vector which had no eucaryotic promoter sequences upstream of the CAT coding region and which would allow easy insertion of act5C fragments in this region. A vector which satisfied the first criterion, puc^{SV0}CAT, was obtained from Dr. Nevis Fregien. Puc^{SV0}CAT contains the CAT protein coding sequences preceded by a single HindIII site and followed by a 3' untranslated region from SV40 (including splice sites and a polyadenylation signal) all in puc 18. This plasmid was modified by the addition of a polylinker at the 5' end of the CAT gene which was designed to facilitate the insertion of actin promoter fragments and the subsequent deletion of promoter sequences. The new vector puc^{PL}CAT

is shown in Figure 1A and the sequence of the inserted polylinker is shown below it.

Construction of act5C promoter CAT fusions

Three act5C promoter fragments were cloned into the BglII site of puc^{PL}CAT, as illustrated in Figure 1B. Each has the same BglII site as a 5' endpoint but they continue to different extents in the 3' direction.

5CCAT contains nearly the whole subcloned 5' region of the act5C gene upstream of the translation initiator AUG. This includes 2.5 kb of 5' flanking sequences, exon 1, exon 2, and the entire upstream intron. It therefore contains an intact splice acceptor signal and also has six of the eight bases of the 5' untranslated region normally found between the splice acceptor site and the translation initiator AUG. Transcription from either start site of this construct should yield a functional CAT mRNA.

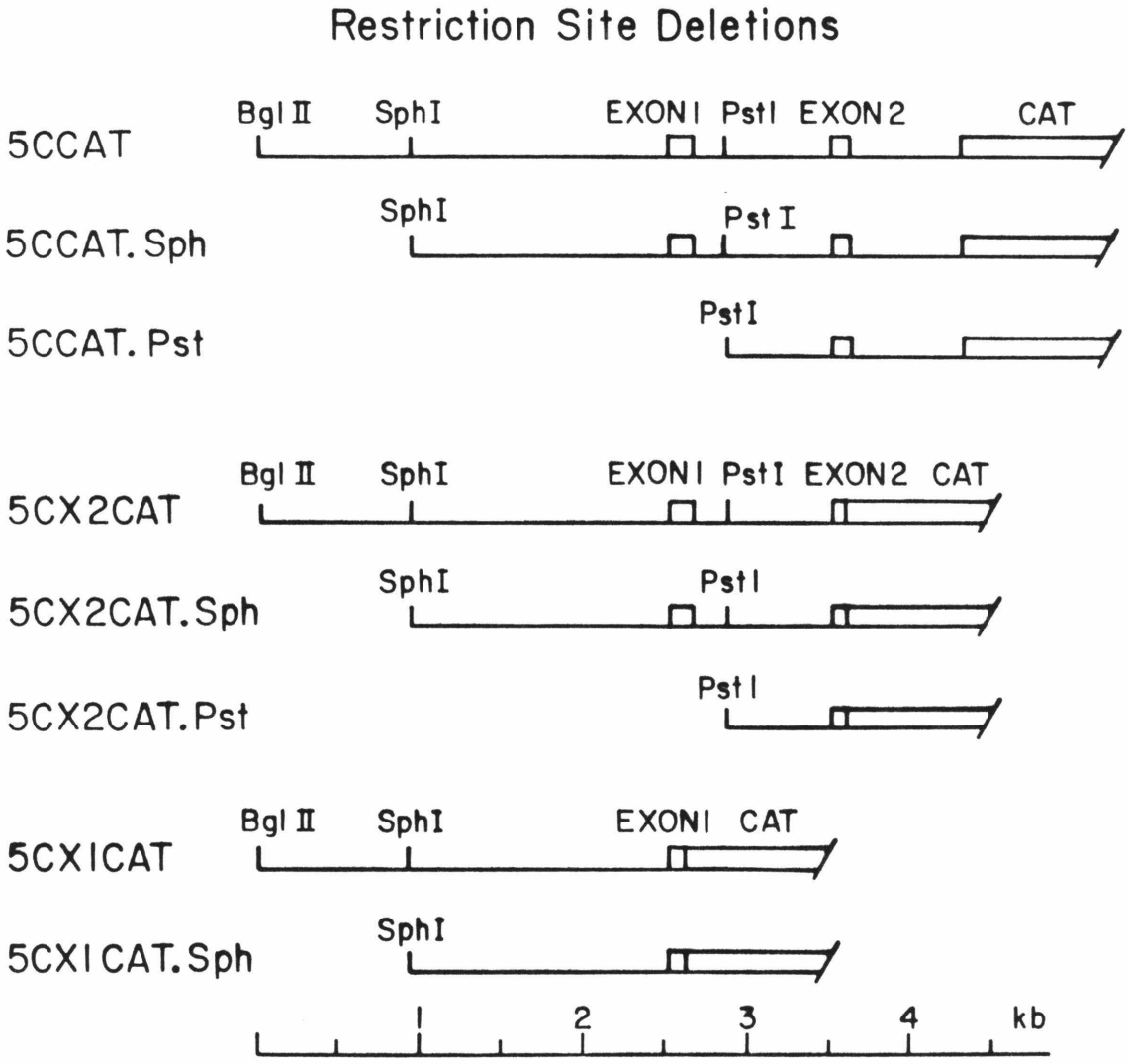
5CX1CAT contains the same upstream sequences but is joined to the CAT gene at nucleotide +88 of exon 1. This construct provides a means to study transcription from exon 1 alone in the absence of exon 2 or any introns.

5CX2CAT is joined to the CAT gene at nucleotide +74 of exon 2. It is not clear whether transcripts from exon 1 in this construct will produce functional CAT mRNA. Cleavage may occur at the splice donor 3' to exon 1, but there are no natural splice acceptors (except for the one in the SV40 3' end) at which to complete the splicing reaction. It is possible that a cryptic splice acceptor site in the intron would be used. Deletion of exon 1 or deletion past exon 1 in this construct provides a means to study transcription from exon 2 alone.

Two restriction sites in the 5' end of the act5C gene were used to make a set of 5' deletions (Fig 2). The SphI site lies 0.95 kb from the 5' BglII site or 1.55 kb upstream of exon 1. The PstI site is 2.75 kb downstream of the BglII

Figure 2. Maps of the PstI and SphI 5' deletion clones.

Schematic maps show the remaining actin sequences in each of the deletion clones. The deletion clones are shown underneath the parent plasmid from which they were made. The open boxes are exons and the solid lines are introns and untranslated sequences.


Restriction Site Deletions
5CCAT
Bgl II
SphI
EXON1
PstI
EXON2
CAT
5CCAT. Sph
SphI
PstI
5CCAT. Pst
PstI
5CX2CAT
Bgl II
SphI
EXON1
PstI
EXON2
CAT
5CX2CAT. Sph
SphI
PstI
5CX2CAT. Pst
PstI
5CX1CAT
Bgl II
SphI
EXON1
CAT
5CX1CAT. Sph
SphI
1
2
3
4
kb

site and lies between exon 1 and exon 2. SphI and PstI sites were included in the polylinker so that these upstream fragments could be easily deleted. The SphI deletions enabled us to determine the effect of deleting far upstream sequences. The PstI deletions made it possible to test whether transcription could occur from exon 2 in the absence of exon 1.

Transient Expression Analysis of act5C-CAT Fusions and SphI and PstI deletions

10 µg of each of the act5C-CAT plasmids and the restriction site deletion plasmids was transfected into Kc cells. The cells were collected two days later and cell extracts prepared. The protein concentrations of the extracts were measured and protein-equivalent amounts of extract were used to carry out CAT assays. An autoradiograph of a typical CAT assay is shown in Figure 3.

Each construct was transfected into cells and assayed at least three times. For each sample, the CAT activity was calculated as the percentage of ^{14}C chloramphenicol acetylated. The results from experiments performed at different times were normalized to 5CCAT activity and averaged. The activities observed for this series of plasmids as normalized to 5CCAT activity of 1.0 are shown in Figure 4. The vertical bars indicate the standard errors of the mean.

As can be seen in the figure, 5CCAT which contains both of the transcription initiation sites and all of the upstream sequences gives the highest level of expression. The 5CCAT.Sph deletion on average decreases the activity by about 15%. The activity of 5CCAT.Pst, a construct which is deleted in all of exon 1 and sequences further upstream, is down another 15% from 5CCAT.Sph but still has a relatively high level of activity. This result indicates that exon 2 has its own promoter which can be expressed independently of exon 1 proximal sequences.

Figure 3. CAT assays on SphI and PstI deletion clones.

An autoradiograph of a typical CAT assay is shown. The plasmids used to transfect Kc cells are shown along the bottom. Protein equivalent amounts of cell extract were used to perform CAT assays as described in the text. The spots at the bottom of the figure are unreacted ^{14}C chloramphenicol. The upper spots are the monoacetylated chloramophenicol products of the CAT reaction.

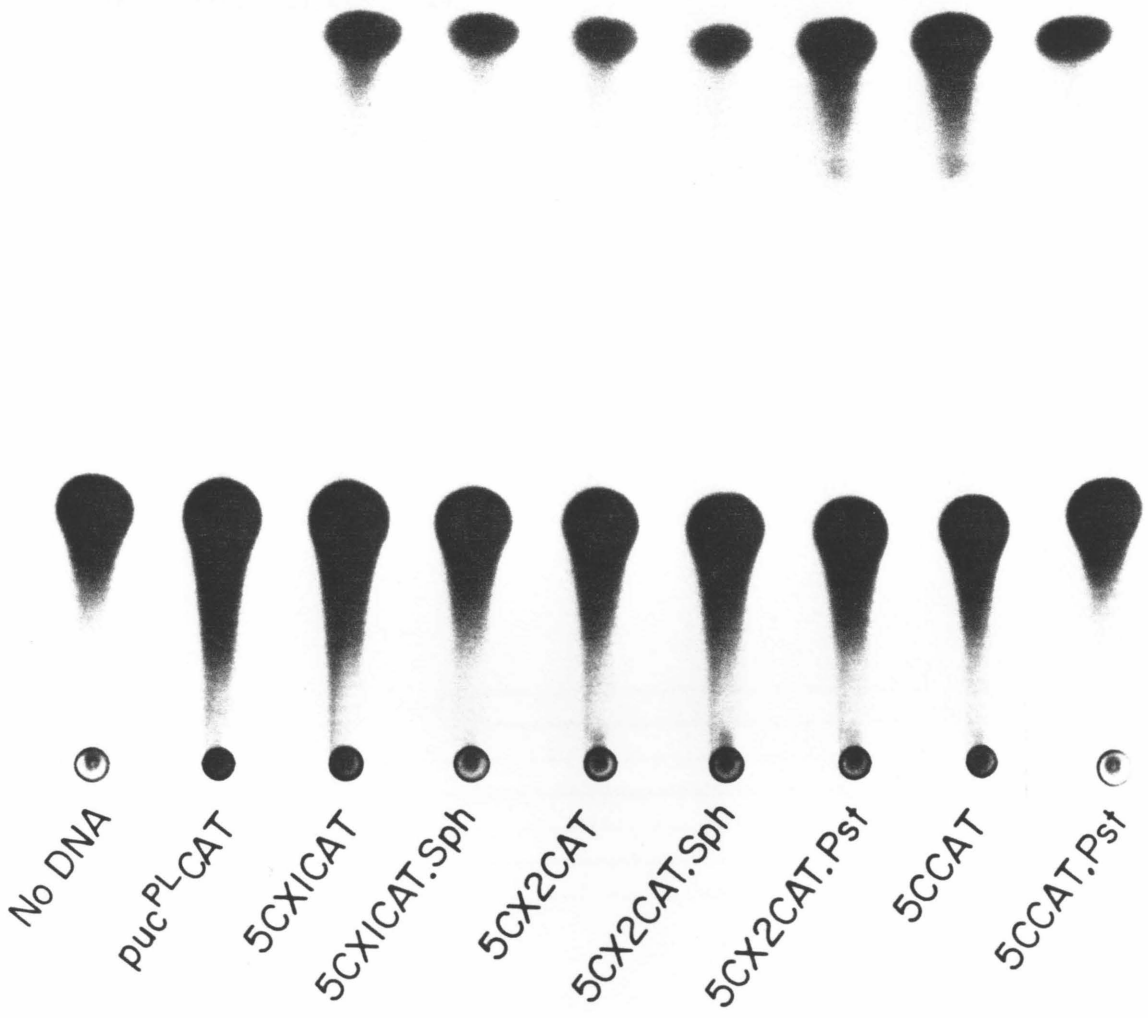
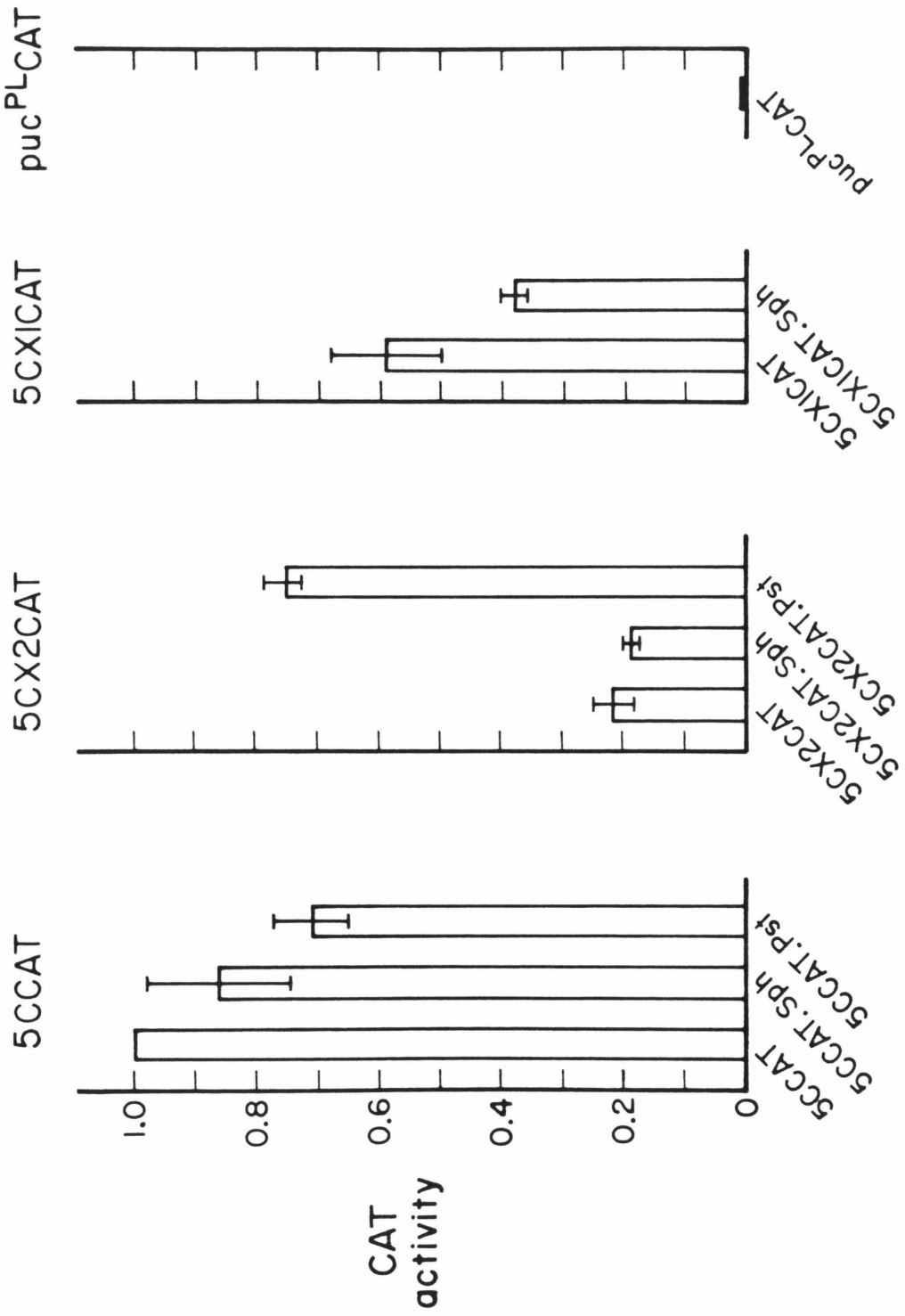

No DNA
pucPLCAT
5CXICAT
5CXICAT.Sph
5CX2CAT
5CX2CAT.Sph
5CX2CAT.Pst
5CCAT
5CCAT.Pst

Figure 4. CAT activity of PstI and SphI deletion clones.

The CAT activities of the clones indicated along the bottom of the figure are plotted in the bar graph. Each was calculated as the percentage of ^{14}C chloramphenicol acetylated and represents the average of three or more experiments. The results of cell transfections and CAT assays carried out at different times were normalized using 5CCAT activity, which is set to 1.0. On this scale the puc^{PL}CAT activity is 0.003. The vertical bars indicate the standard errors of the mean.


5CCAT
5CX2CAT
5CXICAT
pucPLCAT
5CCAT
5CCAT.Sph
5CCAT.Pst
5CX2CAT
5CX2CAT.Sph
5CX2CAT.Pst
5CXICAT
5CXICAT.Sph
pucPLCAT
CAT activity
0
0.2
0.4
0.6
0.8
1.0

5CX1CAT which contains only the exon 1 promoter and no upstream splice sequences gives about 40% less activity than 5CCAT. The activity of 5CX1CAT.Sph drops off 35% from 5CX1CAT suggesting the presence of sequences important for the optimal expression of exon 1 in the far upstream region from -1.5 kb to -2.5 kb.

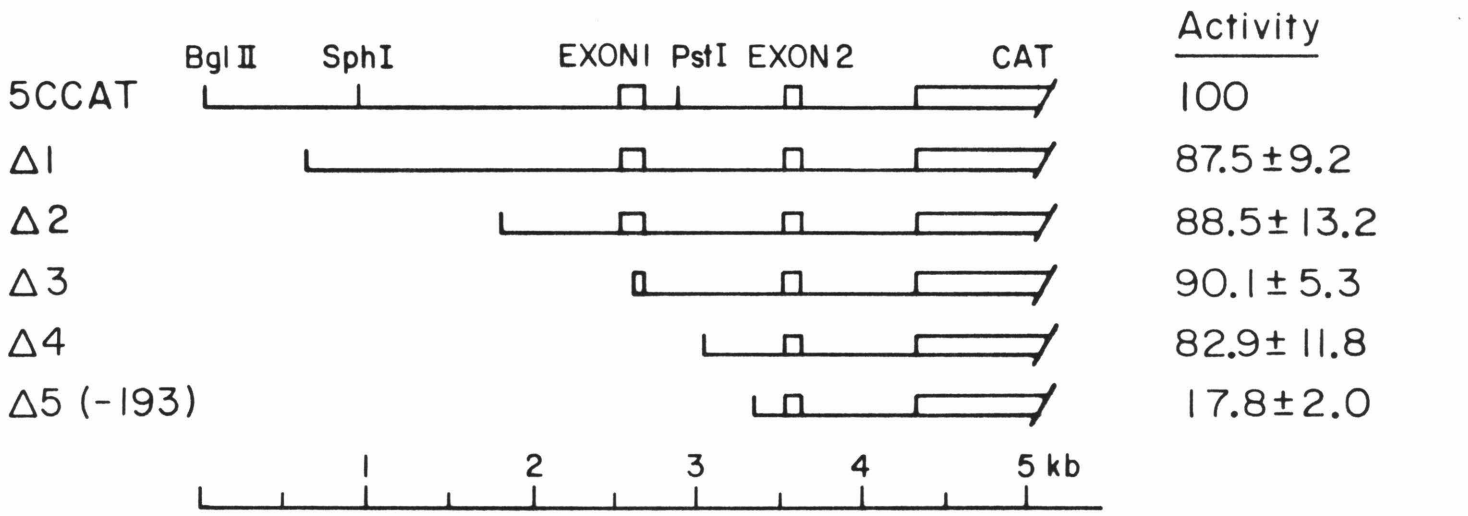
5CX2CAT and 5CX2CAT.Sph are expressed at 22% and 19%, respectively, of 5CCAT. This low level of activity is probably related to the previously mentioned potential RNA splicing problem. In each of these constructs exon 1 and the exon 1 splice donor are present, but there is no splice acceptor 5' to the CAT gene. 5CX2CAT.Pst which has all of the exon 1 sequences deleted shows a sharp rise in activity at a level 75% that of 5CCAT or up almost fourfold from 5CX2CAT.Sph. 5CX2CAT.Pst contains all the sequences necessary for efficient expression from exon 2. 5CX2CAT and 5CX2CAT.Sph contain the same exon 2 adjacent sequences yet the activities of these plasmids in the transient assays are significantly reduced. For some reason, the presence of exon 1 is causing a decrease in transcription from exon 2. Possible explanations will be addressed in the discussion section.

Analysis of 5CCAT Exonuclease III Deletions

It is clear from the above experiments that both exon 1 and exon 2 are preceded by functional promoters. In order to begin to localize sequences necessary for transcription from each promoter, we have made sets of deletions at the 5' end of the gene. Deletions upstream of exon 1 were made in the 5CCAT construct by the use of Exonuclease III. They initiated at the BglII site and continued in the direction of the act5C promoters. These deletions are illustrated in Figure 5. The normalized, averaged results of three or more assays for CAT activity in Kc cells are also shown in Figure 5. In this case the activity drops off very little in deletions 1, 2, 3, and 4 which take out 0.6, 1.8,

Figure 5. Maps and CAT activities of the 5CCAT exonuclease III deletion clones.

The deletion clones are illustrated underneath the parent plasmid from which they were made. Deletions 1, 2, 3, and 4 are missing 0.6, 1.8, 2.6, and 3.04 kb, respectively, of 5' sequences. Deletion 5 has 193 remaining bases of sequence upstream of the exon 2 cap site. The associated CAT activities measured as previously described are normalized to 5CCAT activity of 100 and are reported with standard errors of the mean.


5CCAT Deletions
Activity
BglII
SphI
EXON1
PstI
EXON2
CAT
5CCAT
100
Δ1
87.5±9.2
Δ2
88.5±13.2
Δ3
90.1±5.3
Δ4
82.9±11.8
Δ5 (-193)
17.8±2.0
1
2
3
4
5 kb

2.6, and 3.04 kb, respectively, from the 5' end. A significant drop in activity occurs finally with deletion 5 which leaves only 193 base pairs of sequence upstream of exon 2. This construct shows a decrease in activity of 82% from 5CCAT and 81% from 5CCATΔ4. Thus, a fivefold drop in activity is associated with deleting the sequences between -193 and -450 with respect to exon 2. Deletion 5 falls just short of removing a CCArGG sequence and region of dyad symmetry between -171 and -186 (see Discussion).

Analysis of 5CX1CAT Exonuclease III Deletions

A series of exonuclease III deletions was made from the BglII site in 5CX1CAT as with 5CCAT. These deletions are shown in Figure 6. Note that the starting plasmid for this series, 5CX1CAT, contains no exon 2 sequences, over half of exon 1, and all of its upstream flanking sequences. It expressed 60% of the CAT activity of the complete 5CCAT construction. The averaged results of the CAT assays as normalized to 5CX1CAT activity are also shown in Figure 6.

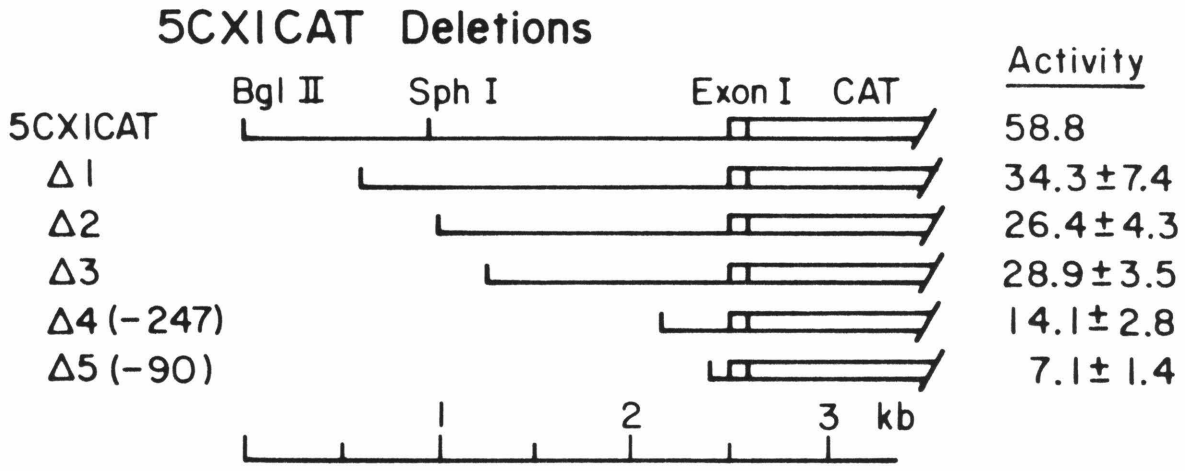
Deletions 1, 2, and 3 which take out 0.6, 1.0, and 1.3 kb, respectively, cause a drop in activity analogous to that seen with the 5CX1CAT.Sph plasmid which is to a level that is 45-60% of 5CX1CAT. Deletion of 2.25 kb (Δ4, to a point -247 from exon 1) causes another drop in activity to 24% and deletion to -90, Δ5, gives a level of activity of 12% of 5CX1CAT which is an eightfold drop in activity. Thus, there appear to be at least two regions that contain sequences that are important for the optimal expression of exon 1.

DISCUSSION

The act5C gene of *Drosophila melanogaster* has two alternate leader exons both of which are expressed in transcripts present in Kc cells (5). A series of act5C promoter - CAT fusions were tested in a *Drosophila* Kc cell

Figure 6. Maps and CAT activities of the 5CX1CAT exonuclease III deletion clones.

The deletion clones are illustrated underneath the parent plasmid from which they were made. Deletions 1, 2, and 3 have lost 0.6, 1.0 and 1.3 kb of upstream sequences, respectively. Deletions 4 and 5 are deleted to 247 bases and 90 bases upstream of the exon 1 cap site, respectively. The activities of the clones in CAT assays as normalized to 5CX1CAT activity are indicated as well as the standard errors of the mean.


5CXICAT Deletions
Activity
Bgl II
Sph I
Exon I
CAT
5CXICAT
58.8
Δ1
34.3±7.4
Δ2
26.4±4.3
Δ3
28.9±3.5
Δ4 (-247)
14.1±2.8
Δ5 (-90)
7.1±1.4
1
2
3
kb

transient assay system to determine if transcription from these exons is driven by separate promoters and to begin to delimit sequences necessary for their transcription.

Exon 1 and Exon 2 are each preceded by separate functional promoters

We find that sequences adjacent to exon 1 will drive expression of the CAT gene in the absence of downstream sequences near exon 2. It was expected that there is a promoter upstream of exon 1 for two reasons. First, it is generally found that sequences important for transcription and regulation of polymerase II genes lie upstream of the cap site (8) though there are exceptions (1,10,19). Secondly, Parker and coworkers showed that sequences upstream of exon 1 are sufficient for independent initiation of transcription in their *in vitro* transcription system (37).

Likewise, exon 2 adjacent sequences will drive CAT expression in the absence of exon 1 and sequences further upstream indicating the existence of a separate promoter for exon 2. This result was not necessarily expected. Exon 2 has no TATA or CAAT sequences or any recognizable generalized eucaryotic promoter elements. It is expressed with the same pattern through development as exon 1 (5). It was therefore not clear that it was driven by a separate promoter.

There are several other known cases of eucaryotic genes that have multiple transcription starts driven by separate promoters. They include the *Drosophila* Adh gene (2), the mouse α-amylase gene (24,25,42,47), and the chicken (35), rat (38,44), and mouse (40) myosin light chain I/III genes. In all of these cases, the separate promoters specify unique tissue specific or developmentally timed programs of expression from the alternate start sites.

It is not clear why there are separate promoters for the act5C gene. It is possible that they direct tissue specific patterns of expression that would not

have been seen in our RNA blot analyses on whole animal RNA. There are precedents in other organisms for different cytoplasmic genes being expressed in a time or tissue specific way. In *Dictyostelium*, all of the actins characterized are cytoplasmic. Four of the genes have been shown to have different patterns of expression during development though three of these encode the same protein (31). In sea urchin there are five cytoplasmic actin genes with only minor variations in the amino acid sequences of the encoded proteins. These fall into three categories with respect to the time and sites of expression in embryogenesis (11). In *Drosophila*, the act42A and act5C genes are both thought to encode cytoplasmic actins. There is evidence that only the act5C gene is expressed in *Drosophila* Schneider line 2 cells (20) and that the act42A gene is preferentially induced by ecdysone in the S3 cell line (46). It may be that these examples reflect the need for cytoplasmic actins of slightly different amino acid sequence in different cellular environments. Given that the differences in sequence are very slight, it seems more likely that these cases demonstrate the need for complex regulatory machinery to allow the cytoplasmic actins to be expressed as required to perform a multitude of roles. This latter explanation could account for the existence of six transcripts from the single *Drosophila* act5C gene some of which are developmentally regulated (5).

Two regions of upstream sequence contribute to transcription of exon 1

Our results indicate that there are at least two regions containing sequences important for optimal expression from exon 1. One is far upstream and the other is closer to the cap site. Loss of sequences between 2.5 and 1.9 kb upstream of the cap site is associated with a 35-40% drop in CAT activity. We have not tested sequences further upstream. It is possible that we would achieve a greater starting level of activity with additional 5' sequence

present. Nor have we tested the orientation dependence of these upstream sequences.

Most eucaryotic promoter elements characterized lie within a few hundred bases upstream of the associated cap site. There are a few known examples of promoter elements that are naturally located further upstream. It has also been shown that many enhancer elements will act over a large distance. One example of a far upstream promoter element is part of the *Drosophila* Adh proximal promoter. This is a bipartite promoter. One element which lies greater than 2 kb upstream of the cap site controls the level of expression from the promoter. The other lies within 400 base pairs of the transcription start and is necessary for tissue specific expression (39).

Similarly, the *Drosophila* sgs-3 gene contains two regions of control. Sequences within 130 bases are sufficient for correct tissue specific expression of the gene while sequences further upstream between -130 and -2270 are required for a tenfold greater level of expression (12). There is evidence to suggest that this upstream element lies greater than 1.4 kb upstream (6).

Another example of a gene with distant control sequences is the yeast *HO* gene for which three different control regions have been identified. An upstream activating sequence between -1250 and -1400 and the TATA box region at -90 are necessary for transcription. Another region between -150 and -900 is necessary for mother/daughter control, i.e., for expression in a or α haploid mother cells but not daughter cells (36).

The second region important for transcription from exon 1 of act5C lies within 1.2 kb of the cap site. Loss of sequences between -1200 and -247 results in a twofold drop in activity and deletion to -90 results in another twofold drop to a level that is 12% of the starting level with X1CAT.

Within this region there are two sequences with homology to other known or suspected promoter elements. As may be seen in Figure 7, the first is the TATA box which is present at position -30 in most eucaryotic genes. In higher eucaryotes it is thought to be necessary for the precise positioning of the mRNA start site (8).

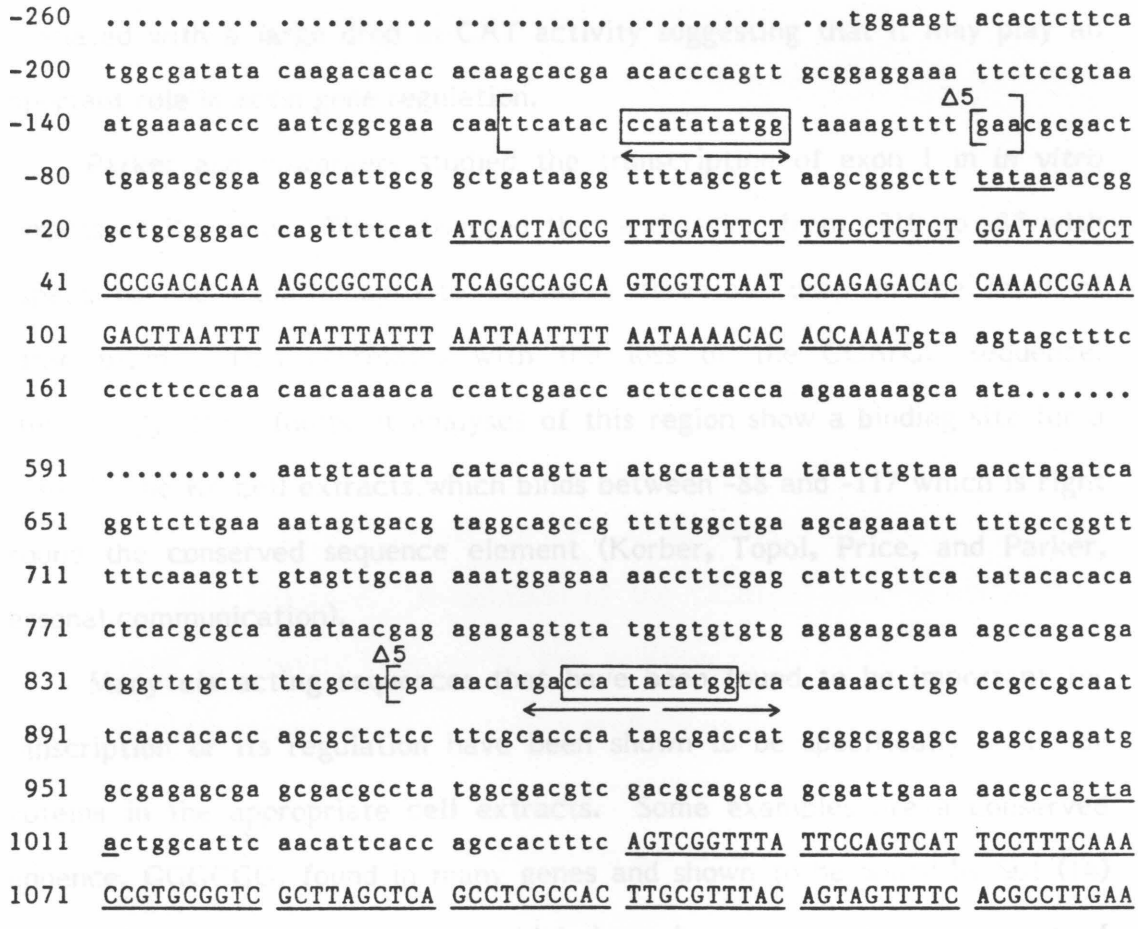
The second is a 10 base sequence similar to that found in many mammalian and chicken actin genes (32). In the act5C gene and in the human gamma actin gene it has the sequence CCATATATGG. In other actin genes it has the general form $CC(A/T)_6GG$ where one G/C base pair may be present in the A/T rich core and has been designated CC-A rich-GG (designated CCArGG below). The particular A rich sequence is often conserved between actin subtypes so, for example, beta actins will have a CCTTTTATGG sequence. The position and number of these sequences varies somewhat but they are generally within 220 bases of the cap site.

The CCArGG sequence upstream of exon 1 in the *Drosophila* act5C gene is centered at -105. In the series of 5CX1CAT deletions that we tested, deletion 5 removed the CCArGG sequence whereas deletion 4 did not. This change was associated with a twofold drop in activity though we don't know if the loss of this sequence element and the drop in CAT activity are causally related. Deletion 4 was already down to 25% of the activity of 5CX1CAT. If the CCArGG sequence is affecting expression it is clearly only one of several important sequence elements. It is also conceivable that this sequence is important for the specific timing of expression of the gene. We were unable to test for this in the Kc cell transient assay system.

Minty and Kedes (32) found that there are four CCArGG sequences in the 5' region of the human cardiac alpha actin gene which are homologous to similarly located sequences in the chick cardiac alpha gene. Sequential loss of

Figure 7. Nucleotide sequence of exon 1, part of exon 2 and their upstream flanking regions.

The flanking sequences are designated with lower case letters and exons are capitalized and underlined. The CCArGG regions are enclosed in boxes and the associated inverted repeats are designated with arrows. The binding site upstream of exon 1 which was identified by Topol and Parker (personal communication) is marked off with large brackets. The endpoints of 5CCAT deletion 5 and 5CX1CAT deletion 5 are given with small brackets. The TATA box upstream of exon 1 is underlined as is a stretch of As and Ts in the analogous position upstream of exon 2.


-260 ....tggaagt acactcttca
-200 tggcgatata caagacacac acaagcacga acacccagtt gcggaggaaa ttctccgtaa
Δ5
-140 atgaaaaccc aatcggcgaa caattcatac ccatatatgg taaaagtttt gaacgcgact
-80 tgagagcgga gagcattgcg gctgataagg ttttagcgct aagcgggctt tataaaacgg
-20 gctgcgggac cagttttcat ATCACTACCG TTTGAGTTCT TGTGCTGTGT GGATACTCCT
41 CCCGACACAA AGCCGCTCCA TCAGCCAGCA GTCGTCTAAT CCAGAGACAC CAAACCGAAA
101 GACTTAATTT ATATTTATTT AATTAATTTT AATAAAACAC ACCAAATgta agtagctttc
161 cccttcccaa caacaaaaca ccatcgaacc actcccacca agaaaaagca ata.....
591 ..... aatgtacata catacagtat atgcatatta taatctgtaa aactagatca
651 ggttcttgaa aatagtgacg taggcagccg ttttggctga agcagaaatt tttgccggtt
711 tttcaaagtt gtagttgcaa aaatggagaa aaccttcgag cattcgttca tatacacaca
771 ctcacgcgca aaataacgag agagagtgta tgtgtgtgtg agagagcgaa agccagacga
Δ5
831 cggtttgctt ttcgcctcga aacatgacca tatatggtca caaaacttgg ccgccgcaat
891 tcaacacacc agcgctctcc ttcgcaccca tagcgaccat gcggcggagc gagcgagatg
951 gcgagagcga gcgacgccta tggcgacgtc gacgcaggca gcgattgaaa aacgcagtta
1011 actggcattc aacattcacc agccactttc AGTCGGTTTA TTCCAGTCAT TCCTTTCAAA
1071 CCGTGCGGTC GCTTAGCTCA GCCTCGCCAC TTGCGTTTAC AGTAGTTTTC ACGCCTTGAA

these sequences is associated with a progressive loss in promoter activity suggesting that these sequences may be a series of small promoter elements.

Bergsma and coworkers (4) have used CAT fusions to study expression of the chick alpha actin in primary myoblasts. They find that all of the sequences required for optimal expression lie within 200 base pairs of the cap site. Deletion of sequences between -200 and -144 causes a 55% drop in expression. Deletion to -133 causes a reduction to 7% of the starting activity. A sequence CGCCTTCTTTGGG lies between -127 and -139. Partial loss of this sequence is associated with a large drop in CAT activity suggesting that it may play an important role in actin gene regulation.

Parker and coworkers studied the transcription of exon 1 in *in vitro* extracts of Kc cells. They observed that a deletion from -210 to -85 with respect to the exon 1 cap site causes a threefold drop in the level of transcription. This correlates with the loss of the CCArGG sequence. Interestingly, their footprint analyses of this region show a binding site for a factor in the Kc cell extracts which binds between -88 and -117 which is right around the conserved sequence element (Korber, Topol, Price, and Parker, personal communication).

Many *cis* acting sequences that have been found to be important for transcription or its regulation have been shown to be specifically bound by proteins in the appropriate cell extracts. Some examples are a conserved sequence, GGGCGG, found in many genes and shown to be bound by Sp1 (14) glucocorticoid response elements which have been shown to be regions of steroid hormone binding (9) and the TATA box which is specifically bound by B factor in *Drosophila* extracts (37). The fact that the actin CCArGG is a binding site implies some function for this sequence in the transcription of the act5C gene.

The *in vivo* and *in vitro* results are not in complete agreement. Our transient assay results indicate that sequences upstream of -210 are necessary for efficient transcription while these are not required *in vitro*. This discrepancy is not surprising. While important information about regulatory sequences may be gained by *in vitro* assays (14), in some cases there are less strict sequence requirements for *in vitro* than *in vivo* transcription. (23).

Upstream sequences required for exon 2 transcription.

We find that sequences necessary for efficient transcription from exon 2 lie within 450 bases of the exon 2 cap site. Deletion from the 5' end to -450 causes little loss in CAT activity. Deletion to -170 is associated with a fivefold loss in activity.

The exon 2 promoter is an example of the infrequent class of eucaryotic promoters that have neither a TATA box nor a CAAT box. The only recognizable potential promoter element in this upstream region is a 10 base sequence identical to the CCArGG sequence upstream of exon 1. This actually comprises the middle section of a 16 base inverted repeat and is located between -171 and -186 with respect to exon 2. Unfortunately, our deletions did not remove this sequence. If deletion of the CCArGG site causes another twofold drop in activity relative to the deletion up to -193, this would be analogous to the stepwise loss of activity seen for exon 1 with loss of sequences near the cap site. Further experiments are needed to test this.

An interesting phenomenon was observed with respect to the amount of CAT activity obtained from the constructions made with 5CX2CAT. The junction site between actin and the CAT gene in this construction is in the middle of exon 2. The consequence of this fusion is that while transcription can presumably proceed normally from exon 2, there is an inhibition of expression of exon 1 containing transcripts. In this construct exon 1 has a splice donor but

no splice acceptor 5' to the CAT gene. There are several possibilities for what may happen to exon 1 containing transcripts: they are unspliced, they are cleaved at the splice donor but the splicing reaction is never completed due to the lack of a splice acceptor, they are spliced to some cryptic splice acceptor site within the messenger RNA, or they are spliced to the SV40 splice acceptor site at the 3' end of the gene. The latter two possibilities may result in stable RNAs that lack CAT activity.

The curious results are that 5CX2CAT and 5CX2CAT.Sph give relatively low CAT activity at about 20% of the level of 5CCAT. A further deletion which completely removes exon 1 results in a fourfold increase in CAT activity. This result suggests that in the presence of exon 1, exon 2 transcription is partially repressed. One possibility is that when exon 1 is present, this transcription start is favored but most of the exon 1-containing RNAs that are made lack a complete CAT gene. The CAT activity of these constructs is therefore relatively low. When exon 1 is absent, transcription from exon 2 is stronger and CAT activity is at the normal level. In order to test this idea, it would be necessary to look at relative amounts of RNA initiating at exon 1 and exon 2 in the constructs which contain both leader exons.

In summary, we have shown that both exons 1 and 2 are preceded by separate functional promoters. Our data suggest that abortive transcription initiating at exon 1 inhibits initiation of transcription at exon 2. By deletion analyses, we have in part identified sequences that contribute to expression from each of the initiation sites. More experiments are necessary for mapping these promoter regions in detail. Some approaches that may be informative are further mutation or deletion of sequences suspected of being important regulatory elements and an examination of the relative proportion of exon 1 and exon 2 containing transcripts in these constructs.

Acknowledgements

We thank Marie Krempin for expert technical assistance, Elizabeth Keller for helpful discussions and for the valuable observation of the inverted repeat sequence upstream of exon 2, Nevis Fregien for the gift of the puc^{SVO}CAT vector, William Mattox and Mickey Chien-Tsung Hu for technical advice, and Joanne Topol, Bette Korber, and Carl Parker for communicating results prior to publication.

This work was supported by a Public Health Service research grant from the National Institutes of Health.

REFERENCES

1. Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**:729-740.

2. Benyajati, C. and J.F. Dray. 1984. Cloned *Drosophila* alcohol dehydrogenase genes are correctly expressed after transfection into *Drosophila* cells in culture. *Proc. Natl. Acad. Sci. U.S.A.* **81**:1701-1705.

3. Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner. 1983. The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5' end in different developmental stages. *Cell* **33**:125-133.

4. Bergsma, D.J., J.M. Grichnik, L.M. Gossett, and R.J. Schwartz. 1986. Delimitation and characterization of *cis*-acting DNA sequences required for the regulated expression and transcriptional control of the chicken skeletal α-actin gene. *Mol. Cell. Biol.* **6**:2462-2475.

5. Bond, B.J. and N. Davidson. 1986. The *Drosophila melanogaster* actin 5C gene uses two transcription initiation sites and three polyadenylation sites to express multiple mRNA species. *Mol. Cell. Biol.* **6**:2080-2088.

6. Bourouis, M. and G. Richards. 1985. Remote regulatory sequences of the *Drosophila* glue gene sgs-3 as revealed by P-element transformation. *Cell* **40**:349-357.

7. Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72:**248-254.

8. Breathnach, R. and P. Chambon. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50:**349-383.

9. Chandler, V.L., B.A. Maler, and K. R. Yamamoto. 1983. DNA sequences bound specifically by glucocorticoid receptor *in vivo* render a heterologous promoter hormone responsive in vivo. *Cell* **33:**489-499.

10. Charnay, P., R. Treisman, P. Mellon, M. Chao., R. Axel, and T. Maniatis. 1984. Differences in Human alpha- and beta-globin gene expression in mouse erythroleukemia cells: The role of intragenic sequences. *Cell* **38:**251-263.

11. Cox, K.H., L.A. Angerer, J.J. Lee, E.H. Davidson, and R.C. Angerer. 1986. Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. *J. Mol. Biol.* **188:**159-172.

12. Crosby, M.A. and E.M. Meyerowitz. 1986. Regulation of the *Drosophila* glue gene sgs-3: Sequences required for puffing and transcription. *Dev. Biol.* In press.

13. Di Nocera, P.P. and I.B. Dawid. 1983. Transient expression of genes introduced into cultured cells of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **80:**7095-7098.

14. Dynan, W.S. and R. Tjian. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* **316**:774-778.

15. Echalier, G. and A. Ohanessian. 1970. In vitro culture of *Drosophila melanogaster* embryonic cells. *In Vitro* **6**:162-171.

16. Fyrberg, E.A., B.J. Bond, N.D. Hershey, K.S. Mixter, and N. Davidson. 1981. The actin genes of *Drosophila*: Protein coding regions are highly conserved but intron positions are not. *Cell* **24**:107-116.

17. Fyrberg, E.A., K.L. Kindle, N. Davidson, and A. Sodja. 1980. The actin genes of *Drosophila*: A dispersed multigene family. *Cell* **19**:365-378.

18. Fyrberg, E.A., J.W. Mahaffey, B.J. Bond, and N. Davidson. 1983. Transcripts of the six *Drosophila* actin genes accumulate in a stage and tissue specific manner. *Cell* **33**:115-123.

19. Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**:717-728.

20. Gilmour, D.S. and J.T. Lis. 1985. *In Vivo* interactions of RNA polymerase II with genes of *Drosophila melanogaster*. *Mol. Cell. Biol.* **5**:2009-2018.

21. Gorman, C.M., L.F. Moffat, and B.H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.

22. Graham, F. and A. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* **52:**456-457.

23. Grosveld, G.C., C.K. Shewmaker, P. Jat, and R.A. Flavell. 1981. Localization of DNA sequences necessary for transcription of the rabbit beta-globin gene *in vitro*. *Cell* **25:**215-226.

24. Hagenbuchle, O., U. Schibler, S. Petrucco, G.C. Van Tuyle, and P.K. Wellauer. 1985. Expression of mouse *Amy-2*^{a} alpha-amylase genes is regulated by strong pancreas specific promoters. *J. Mol. Biol.* **185:**285-293.

25. Hagenbuchle, O., M. Tosi, U. Schibler, R. Bovey, P.K. Wellauer, and R.A. Young. 1981. Mouse liver and salivary gland alpha-amylase mRNAs differ only in 5' nontranslated sequences. *Nature* **289:**643-646.

26. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28:**351-359.

27. Hunkapiller, M., S. Kent, M. Caruthers, W. Dreyer, J. Firca, C. Griffin, S. Horvath, T. Hunkapiller, P. Tempst, and L. Hood. 1984. A microchemical facility for the analysis and synthesis of genes and proteins. *Nature (London)* **310:**105-111.

28. Laimins, L.A., G. Khoury, C. Gorman, B. Howard, and P. Gruss. 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. *Proc. Natl. Acac. Sci. USA* **79:**6453-6457.

29. Lawson, R., R. Mestril, P. Schiller, and R. Voellmy. 1984. Expression of heat shock-beta-galactosidase hybrid genes in cultured *Drosophila* cells. *Mol. Gen. Genet.* **198:**116-124.

30. Maxam, A. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65:**499-560.

31. McKeown, M. and R.A. Firtel. 1981. Differential expression and 5' end mapping of actin genes in *Dictyostelium*. *Cell* **24:**799-807.

32. Minty, A. and L. Kedes. 1986. Upstream regions of the human cardiac actin gene that modulate its transcription in muscle cells: presence of an evolutionarily conserved repeated motif. *Mol. Cell. Biol.* **6:**2125-2136.

33. Morganelli, C.M. and E.M. Berger. 1984. Transient expression of homologous genes in *Drosophila* cells. *Science* **224:**1004-1007.

34. Morganelli, C.M., E.M. Berger, and H.R.B. Pelham. 1985. Transcription of *Drosophila* small *hsp-tk* hybrid genes is induced by heat shock and by ecdysterone in transfected *Drosophila* cells. *Proc. Natl. Acad. Sci. USA* **82:**5865-5869.

35. Nabeshima, Y., Y. Fujii-Kuriyama, M. Muramatsu, and K. Ogata. 1984. Alternative transcription and two modes of splicing result in two myosin light chains from one gene. *Nature* **308:**333-338.

36. Nasmyth, K. 1985. At least 1400 base pairs of 5'-flanking DNA is required for the correct expression of the *HO* gene in yeast. *Cell* **42**:213-223.

37. Parker, C.S. and J. Topol. 1984. A *Drosophila* RNA Polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. *Cell* **36**:357-369.

38. Periasamy, M., E.E. Strehler, L.I. Garfinkel, R.M. Gubits, N. Ruiz-Opazo, and B. Nadal-Ginard. 1984. Fast skeletal muscle myosin light chains 1 and 3 are produced from a single gene by a combined process of differential RNA transcription and splicing. *J. Biol. Chem.* **259**:13595-13604.

39. Posakony, J.W., J.A. Fischer, and T. Maniatis. 1985. Identification of DNA Sequences Required for the Regulation of *Drosophila* Alcohol Dehydrogenase Gene Expression. *Cold Spring Harbor Symp. Quant. Biol.* **50**:515-520.

40. Robert, B., P. Daubas, M.-A. Akimenko, A. Cohen, I. Garner, J.-L. Guenet, and M. Buckingham. 1984. A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a related pseudogene. *Cell* **39**:129-140.

41. Sanchez, F., S.L. Tobin, U. Rdest, E. Zulauf, and B.J. McCarthy. 1983. Two *Drosophila* actin genes in detail. Gene structure, protein structure, and transcription during development. *J. Mol. Biol.* **163**:533-551.

42. Schibler, U., O. Hagenbuchle, P.K. Wellauer, and A.C. Pittet. 1983. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene *Amy-1*^{a} in the parotid gland and the liver. *Cell* **33**:501-508.

43. Shaw, P., B. Sordat, and U. Schibler. 1985. The two promoters of the mouse alpha amylase gene Amy-1^a are differentially activated during parotid gland differentiation. *Cell* **40:**907-912.

44. Strehler, E.E., M. Periasamy, M. Strehler-Page, and B. Nadal-Ginard. 1985. Myosing Light-Chain 1 and 3 gene has two structurally distinct and differentially regulated promoters evolving at different rates. *Mol. Cell. Biol.* **5:**3168-3182.

45. Tobin, S.L., E. Zulauf, F. Sanchez, E.A. Craig, and B.J. McCarthy. 1980. Multiple actin related sequences in the *Drosophila melanogaster* genome. *Cell* **19:**121-131.

46. Vitek, M.P., C.M. Morganelli, and E.M. Berger. 1984. Stimulation of cytoplasmic actin gene transcription and translation in cultured *Drosophila* cells by ecdysterone. *J. Biol. Chem.* **259:**1738-1743.

47. Young, R.A., O. Hagenbuchle, and U. Schibler. 1981. A single mouse alpha amylase gene specifies two different tissue-specific mRNAs. *Cell* **23:**51-458.

APPENDIX

tage-specific regulation of actin genes in

Drosophila wing cells

(published in Developmental Genetics)

Developmental Genetics 5:219–225 (1985)

Stage-Specific Regulation of Actin Genes in *Drosophila* Wing Cells

Nancy S. Petersen, Beverly J. Bond, Herschel K. Mitchell, and Norman Davidson

The Divisions of Biology (N.S.P., B.J.B., H.K.M.) and Chemistry and Chemical Engineering (N.D.), California Institute of Technology, Pasadena

Extreme and rapid changes in the synthesis of messenger RNAs and proteins accompany differentiation in wing tissues of *Drosophila*. Of the six actin genes, at least three are expressed in wing cells, some during the most extreme changes in cell shape. However, different messages of the set appear, decay, and reappear on a regulated temporal program. These results show that actin expression is stage-specific in a single cell type.

Key words: actin, gene regulation, development, *Drosophila*

INTRODUCTION

Wing tissue of *Drosophila* pupae provides an excellent model system for studies of differentiation. Each wing contains about 32,000 cells, 99% of which are of the same type and in a single layer [1–3]. The wing cells differentiate synchronously in the absence of cell division. One hair per cell is constructed in a multistep process that includes dramatic changes in cell shape [4,5]. While this morphogenesis is in progress rapid changes occur in the rates of synthesis of all the tissue proteins that can be seen on one- and two-dimensional acrylamide gels [1,3,5]. Furthermore, most of the protein changes appear to reflect directly concentrations of the corresponding messenger RNAs [2,6]. Among the most abundant proteins in the differentiating wing tissue are the actins. In this report we describe two peaks of actin synthesis and show that they are due to synthesis, decay, and resynthesis of specific actin messages. Using cloned probes [7–9], we have identified the actin genes expressed in each peak and have shown that the first peak coincides with a developmental time period when there is an extreme change in cell shape [4].

Received for publication February 19, 1985; accepted March 15, 1985.

Address reprint requests to Nancy S. Petersen at the Department of Biochemistry, University of Wyoming, Laramie, WY 82071.

MATERIALS AND METHODS

Fly Stocks

An Oregon R wild-type stock obtained from the Caltech stock center was used in all the experiments described. Flies were kept in cages, and pupae were collected and staged as described previously [10].

Protein Synthesis and Protein Gels

Protein synthesis in pupal wings was followed by removing wings from three pupae of the appropriate age and labeling them in ^{35}S-methionine 10 μCi/μl in MOPS buffer for 25 min at 25°C as described previously [1]. Wing proteins were solubilized by boiling in SDS sample buffer and then run on 7–20% linear gradient gels stabilized with sucrose as described previously [1].

Wing RNA Isolation

Wings were dissected from 200 pupae of the appropriate age and frozen in liquid nitrogen in a small grinder. Fifty microliters each of extraction buffer (0.2 M Tris HCl, pH 7.5, 0.2 M NaCl, 0.04 M EDTA, 1% SDS) and phenol were added to the frozen wings and the mixture was thawed and ground simultaneously. The wing extract was transferred to an Eppendorf centrifuge tube, and 50 μl of chloroform was added. The extract was vortexed for 10 min and centrifuged, and the aqueous phase was reextracted twice with 100 μl of chloroform. The RNA (aqueous phase) was then ethanol-precipitated twice and dissolved in distilled water. The aqueous solution was centrifuged for 5 min in an Eppendorf microfuge and an O.D. 240–300 spectrum was determined to indicate amount of RNA and purity (lack of protein and phenol). When necessary, the RNA was reprecipitated with ethanol once more to remove phenol.

mRNA Translation

RNA isolated as described above was translated in a ^{35}S-methionine translation kit purchased from NEN (Boston, MA) [6,11]. RNA concentrations of 0.15 μg/μl of translation mixture have been previously determined to be in the concentration range in which translation product is proportional to mRNA added [6]. The translation products were run on two-dimensional gels [12,13].

RNA Formaldehyde Gels and RNA Blots

Total wing RNA was run on agarose formaldehyde gels at 5–10 μg/lane as described previously [9]. The gels were blotted to nitrocellulose and hybridized to nick translated actin probes [9]. Hybridization was done at 42°C for at least 24 hr, and washes were done at 65°C in 0.1X SSC to ensure specificity of the hybridization to the gene-specific probes.

Actin Gene Probes

The gene-specific probes used here for each of the actin genes have been described previously [9]. Each corresponds to a 3′ nontranslated region of actin mRNA specific for each actin transcript [9]. Further evidence for the specificity of these probes is that they each show a unique hybridization pattern in these experiments.

RESULTS

Protein synthesis during wing development has been followed in *Drosophila* pupae by dissecting wings at different stages of development and labeling with ^{35}S-methionine for 20–30 min in vitro followed by separation of the proteins on acrylamide gels [1]. Figure 1 shows a typical one-dimensional pattern of labeled wing proteins and the analysis by densitometry. Actin synthesis peaks between 40 and 48 hr after puparium formation and then decreases sharply. At later times, (70–80 hr) actin synthesis again increases and remains high until eclosion. The sharp peak of actin synthesis at 44 hr occurs at the time when wing cells are changing in shape from columnar to flat with very extensive surface convolution [4]. The coincidence in the timing of the actin synthesis and the shape changes suggests that newly synthesized actin is involved in this process.

To determine whether or not changes in overall rates of actin synthesis are due to changes in messenger RNA concentration we have isolated RNA from wings of different ages and translated it in a rabbit reticulocyte system. Figure 2 shows in vitro translation products from 44–48-hr wing mRNA, 52–56-hr wing mRNA, and 78–82-hr wing mRNA. There are multiple spots in the position where actin should be in translation products from 44–48-hr and 78–82-hr message. However, there is no actin detected in the translation products from 55-hr wing message. The major translation product in this region of the gel at 55 hr is the 38 K protein also seen in Figure 1 at the time when actin synthesis decreases. These results indicate that the changes in the rate of synthesis of actin protein in Figure 1 are mostly due to changes in the concentration of actin mRNA.

Further characterization of the changes in actin mRNA concentration during wing development has been done by determining which of the six *Drosophila* actin genes are expressed at 46, 55, and 80 hr of pupal development by using gene-specific probes that are complementary to the unique 3′ untranslated regions of each actin mRNA [9]. Total RNA from each stage of development was subjected to gel blot analysis with these probes. Figure 3 shows the results of hybridization to probes specific for the genes from 5C and 42A, and 79B, in one of these experiments. The hybridizations have been done using each of the probes described by Fyrberg et al [9] (except Act-88F probe B) two or three times and hybridizing to four different mRNA preparations for each time period. The actin 5C transcript is always the most abundantly expressed in wings. Actin 42A is always seen in both peaks, although its relative abundance in the 80-hr peak varied somewhat from one RNA preparation to the next. Hybridization to the actin 79B gene probe was seen in three separate preparations. Its expression is maximum in the 80-hr RNA sample. Transcripts complementary to the 57A, 88F, and 87E (A + B) gene probes were not detected under the same labeling and hybridization conditions used to detect the 5C, 42A, and 79B transcripts.

DISCUSSION

We have shown that at least three of the six *Drosophila* actin genes are expressed during wing development in pupae. Because we have looked only at selected time points, we cannot exclude the possibility that the other actin genes are expressed at

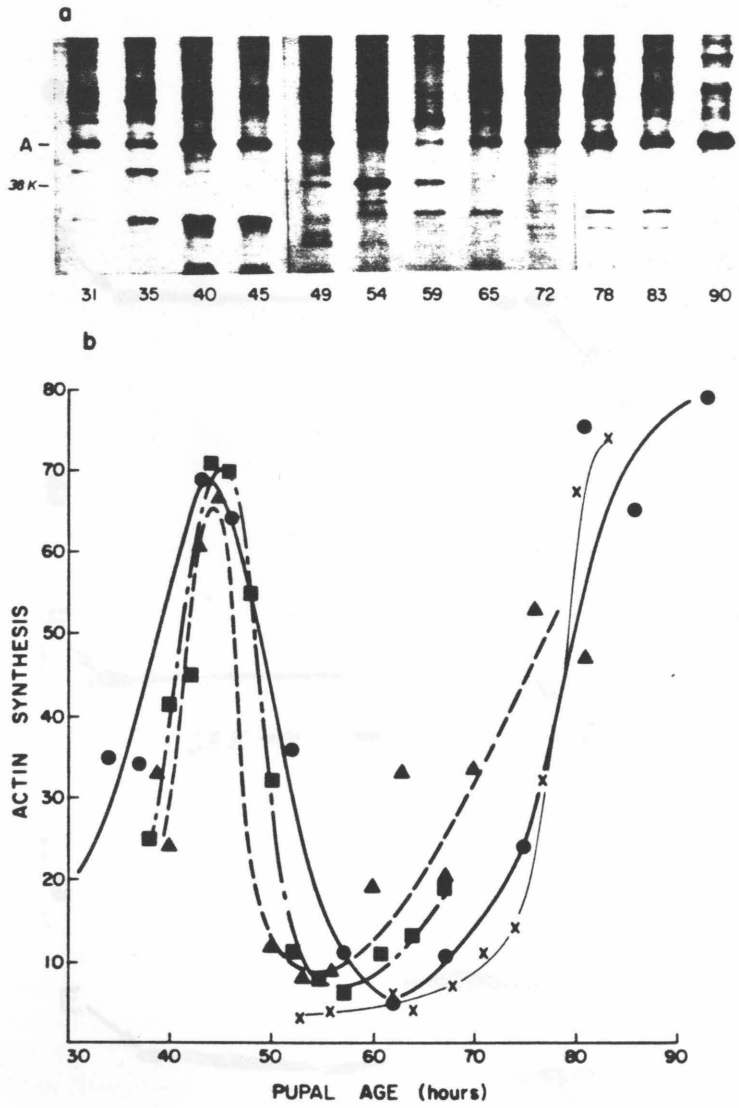


Fig. 1. Actin synthesis in developing wings. a, Autoradiograms of an SDS acrylamide gel showing ^{35}S-labeled proteins synthesized by wings at different times in pupal development. For each sample, wings from three pupae were removed and labeled in ^{35}S-methionine for 25 min at 25°C as described in Materials and Methods. The labeled proteins were dissolved in SDS buffer at 100°C for 2 min and run on 7–20% linear gradient SDS acrylamide gel [1]. The ages of the wing samples are indicated in hr from pupariation. The white prepupa stage is considered to be 0 hr. The position of the actin band is labeled A. Actin is identified by its molecular weight and isoelectric point on two-dimensional gels. b, Relative rates of actin synthesis during pupal wing development. A Joyce-Loebel densitometer was used to measure the density of the actin bands in autoradiograms from five separate experiments similar to the one described in a. The amount of actin synthesized per wing is a function of the age of the wing tissue. Specifically, peaks of actin synthesis are observed at approximately 44 hr after puparium formation and again between 70 hr and eclosion.

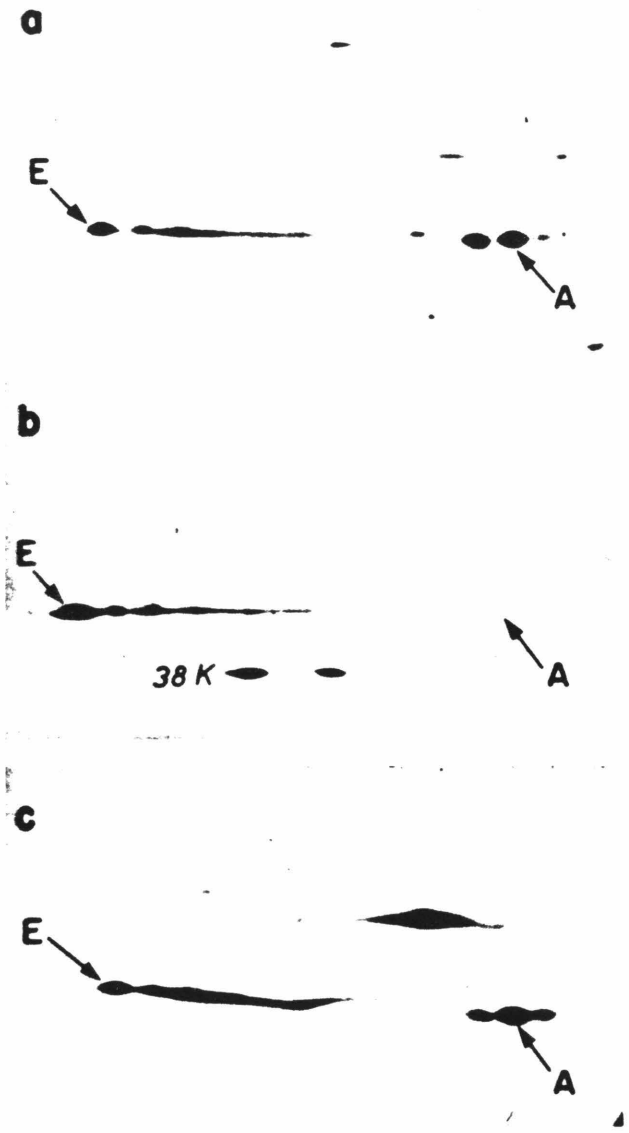


Fig. 2. Actin synthesized from wing mRNA translated in vitro. Total RNA was isolated from wings dissected from pupae at different stages of development and translated in a reticulocyte lysate system using ^{35}S-methionine as described in Materials and Methods. The translation products were solubilized and run on two-dimensional acrylamide gels. The autoradiogram of a portion of the gel shows actin (A) and an endogenously labeled spot (E) for reference. Wing mRNA from three stages has been translated; a, 44–48-hr wing RNA; b, 52–56-hr wing RNA; and c, 78–82-hr wing RNA.

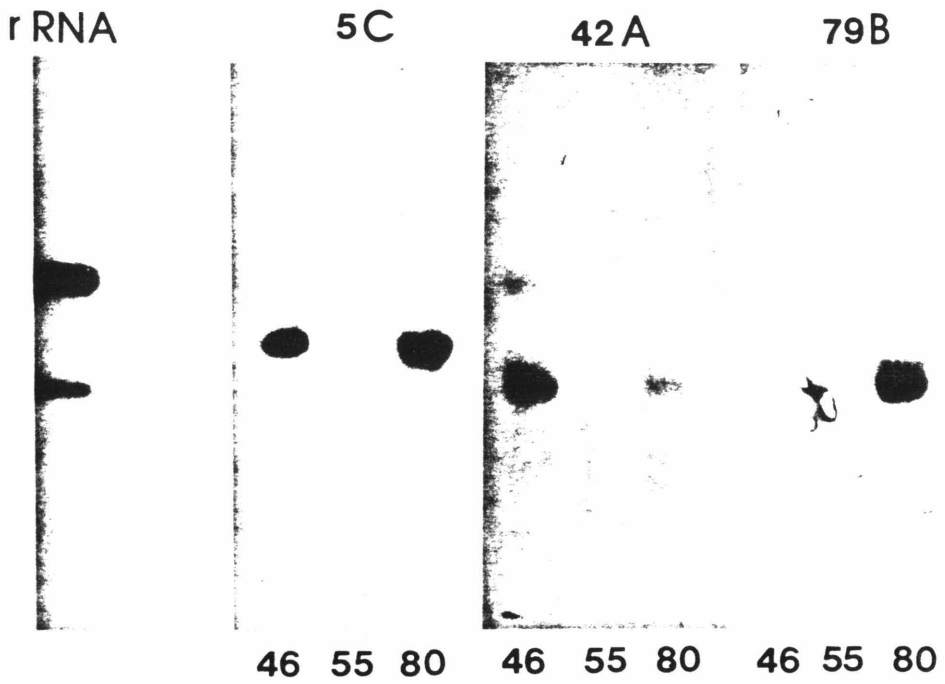


Fig. 3. Identification of actin mRNAs present in developing wings of three different ages. Total RNA was isolated from wings at three developmental stages, 44–48-hr, 52–56-hr, and 78–82-hr as described in Materials and Methods. The RNA was separated on formaldehyde-agarose gels, blotted onto nitrocellulose, hybridized to ^{32}P-labeled gene-specific probes, and washed as described in Materials and Methods. The average age in hr of the pupae from which the RNA was extracted is indicated beneath each lane. The size of the transcripts was determined from E coli ribosomal RNA standards. The Act-5C transcript is 1.9 kb, and the Act-42A and Act-79B transcripts are each 1.6 kb.

other times. Each of the actin genes is expressed with a different time program, suggesting that the synthesis and/or decay of these messages is regulated independently. Wing tissue consists mainly of one type of epithelial cell, each of which produces a single hair. The expression of the major "cytoplasmic" actin genes (5C and 42A) is perhaps not unexpected in such cells. The coincidence of their expression with a major change in cell shape at 44 hr after pupariation [4] suggests that they play a role in this process.

Actins 57A and 87E are genes that code for actins in larval muscle. Actin 88F is expressed abundantly in the indirect flight muscle [9]. Thus their expression is not expected in wing cells.

The actin 79B gene is expressed prominently in developing adult muscle tissues [9], and its expression in wings was unexpected. Expression of this gene in the differentiating wing tissue, however, can hardly be due to this source; even whole pupae in the 24–68-hr range contain very limited amounts of muscle of any kind. Furthermore, the wing tissues used for mRNA preparation for the present experiments were dissected from only the distal portions of the differentiating wings, and muscle tissue is not present at any time in these regions. It is possible that these low-level messages are derived from nerve or bristle cells, which are present in small numbers ($< 2\%$).

CONCLUSIONS

We conclude that three of the six actin genes are expressed in wings, each with its own temporal program of expression. Furthermore, the coincidence in time of the peak of actin synthesis and extreme changes in cell shape suggests that actin gene expression at 44 hr plays a major role in this process.

ACKNOWLEDGMENTS

This work was supported in part by grants HD 19160-02 and GM 25966B from the U.S. Public Health Service.

REFERENCES

1. Mitchell HK, Petersen NS: Rapid changes in gene expression in differentiating tissues of Drosophila. Dev Biol 85:233–242, 1981.
2. Petersen NS, Mitchell HK: Effects of heat shock on gene expression during development: Induction and prevention of the multihair phenocopy in Drosophila. In Schlesinger, Ashburner, Tissieres (eds): "Heat Shock from Bacteria to Man." Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1982, pp 345–352.
3. Mitchell HK, Petersen NS: Gradients of differentiation in wild-type and bithorax mutants of Drosophila. Dev Biol 94:459–467, 1983.
4. Mitchell HK, Roach J, Petersen NS: Morphogenesis of cell hairs in Drosophila. Dev Biol 95:387–398, 1983.
5. Mitchell HK, Petersen NS: Heat shock induction of abnormal morphogenesis in Drosophila. In Schlesinger, Ashburner, Tissieres (eds): "Heat Shock from Bacteria to Man." Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1982, pp 337–344.
6. Petersen NS, Mitchell HK: Recovery of protein synthesis following heat shock: Prior heat treatment affects the ability of cells to translate mRNA. Proc Natl Acad Sci USA 78:1708–1711, 1981.
7. Tobin SL, Zulauf E, Sanchez F, Craig EA, McCarthy BJ: Multiple actin-related sequences in the Drosophila melanogaster genome. Cell 19:121–131, 1980.
8. Fyrberg EA, Kindle KL, Davidson N, Sodja A: The actin genes of Drosophila, a dispersed multigene family. Cell 19:365–378, 1980.
9. Fyrberg EA, Mahaffey JW, Bond BJ, Davidson N: Transcripts of the six Drosophila actin genes accumulate in a stage- and tissue-specific manner. Cell 33:115–124, 1983.
10. Mitchell HK, Petersen NS: Developmental abnormalities in Drosophila induced by heat shock. Dev Genet 3:91–102, 1982.
11. Pelham HRB, Jackson RJ: An efficient mRNA dependent translation system from reticulocyte lysates. Eur J Biochem 67:247–256, 1976.
12. O'Farrell PH: High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021, 1975.
13. Buzin CH, Petersen NS: A comparison of the multiple Drosophila heat shock proteins in cell lines and larval salivary glands by two-dimensional gel electrophoresis. J Mol Biol 148:181–201, 1982.