Studies of gene structure:

I. Expression of human α-globin genes in COS cells

I. Isolation and characterization of the myosin light chain genes from *Drosophila melanogaster*

Thesis by

Vann Phillips Parker

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1985

(Submitted January 9, 1985)

Acknowledgments

I would like to thank everyone I knew while at CalTech as they have contributed, for better or worse, to my graduate life. Alas, that's a feat even Meryl Streep couldn't pull off. Instead I've made the following list of ten acknowledgments:

1. My parents, Helen and Walt, who had enough confidence in me to insure I would make it through.
2. The various members of Prufrock house and eating groups, who will undoubtably be remembered as part of the good old days.
3. Richard Parker, who tried to teach me about swimming with sharks (and Mark Silver who taught me how to swim around them.)
4. Cathy O'Connell, who introduced me to a world outside of the lab.
5. Tom Maniatis, from whom I learned oh so many things and Pamela Mellon with whom I collaborated in Tom's lab.
6. Scott Falkenthal and Bill Mattox, each of whom possessed the rare quality of being both close collaborators and friends.
7. Mark Garfinkel, whose quest for knowledge and evening visits I already miss (no matter what I may have said at the time.)

8. Tim Hunkapiller, whose help with analysis of my data cannot be overlooked.

9. Norman Davidson, whose love of science is truly infectious and who gave me enough rope to do what I wanted--but pulled it back in before I hung myself.

10. Carol, to whom I dedicate this thesis, that we may now continue our lives together.

I would also like to acknowledge the National Science Foundation which supported me through the first three years of my graduate career and the National Institutes of Health which provided support for the remainder of my tenure.

Abstract

The first part of this thesis describes the establishment of COS cells as an effective host for the replication and expression of genes cloned into plasmids containing the SV40 viral replication origin. The human α-globin gene was cloned into a pBR322 derivative which contained a 311 base pair SV40 fragment. These plasmids were replicated to high copy number. The α-globin gene was transcribed to produce high levels of RNA which was indistinguishable from human α-globin mRNA. Deletion mutants were generated which defined the region from 55 to 87 base pairs upstream from the mRNA cap site as necessary for high levels of transcription of this gene. This region includes the conserved sequence CCAAT found in a similar position upstream from many eukaryotic genes.

The second part of this thesis describes the cloning of muscle specific genes from Drosophila melanogaster and a detailed analysis of the two myosin light chain genes. A genomic library was screened with RNA isolated from the late pupal stage of development. Recombinant DNA clones which mapped to 24 independent locations were identified. The clones containing the myosin light chain genes were identified from within this group. A cDNA library was generated from late pupal RNA. Clones homologous to each of the myosin light chains were identified and their cloned

inserts were sequenced. Their classification as *Drosophila* myosin light chain genes was established by comparing the derived amino acid sequence to the protein sequence of the chicken myosin light chains. Each myosin light chain is single copy. The alkali myosin light chain maps to the chromosomal locus 98B. It hybrid selects RNA which may be translated *in vitro* to yeild several polypeptides of molecular weight 18,000 to 19,000 daltons. The gene for myosin light chain-2 maps to the chromosomal locus 99E and is shown to encode two proteins of apparent molecular weights 26,000 and 17,000 daltons. Each protein has a putative divalent cation binding domain.

Table of Contents

Acknowledgments ii

Abstract iv

Introduction 1

Chapter 1. Identification of sequences required for transcription of the human α-globin gene in a new SV40 host-vector system. 20

Chapter 2. *Drosophila melanogaster* has only one myosin alkali light-chain which encodes a protein with considerable amino acid homology to chicken myosin alkali light chain. 31

Chapter 3 The unique myosin light chain-2 gene of *Drosophila melanogaster* encodes two polypeptides which are abundantly expressed in muscle. 42

Introduction

This thesis, as my graduate career, is divided into two parts. The first part describes the culmination of my activities in Tom Maniatis' laboratory. This work was carried out in close collaboration with Pamela Mellon. Our interest was in studying the regions flanking genes which are necessary for the control of that gene's expression.

We were utilizing the technique of DNA mediated co-transformation (23,34) to study globin gene expression. We were frustrated by the length of time needed to assay stably transformed lines. The process required clonally selecting transformants, growing cells until the desired co-transformed element could be verified and then assaying for the desired trait--a process which required a month or more for each construction tested. Transient assay systems were being developed which allowed the analysis of transfected DNAs as early as two days after applying DNA to mammalian cells. Mulligan and Berg had developed SV40 based vectors which had genes cloned into the early or late regions of SV40 and could be propagated with complementing helper virus (21). Major drawbacks to this system were that the cloned DNA was constrained to a size of less than 2.5 kb and signals for transcription and RNA processing were from the vector. A second generation of vectors which they developed could be propagated as plasmids in E.coli, thus bypassing the size limitations, but these vectors still relied on SV40 signals for initiation of transcription, RNA splicing and poly adenylation. Similar

systems were being exploited by Schaffner's group (3).

We helped to develop a new host-vector system for the transient expression of introduced DNAs. Gluzman, working at Cold Spring Harbor, generated cell lines transformed by SV40 DNA which had deletions in the region responsible for viral DNA replication (16). He showed that several of these cell lines were capable of complementing virus which could not produce SV40 T antigen. The cell lines were derived from the monkey cell line CV-1 and were designated COS for CV-1, origin defective SV40 transformed cells (15). Myers and Tjian working at Berkeley were interested in replication of SV40. They generated a plasmid containing the SV40 origin of replication by inserting a 311 bp fragment from SV40 into pBR322. They showed that this plasmid, pSV01, was replicated to high copy number in the COS cells (22). We obtained the cell lines and plasmid to test the efficacy of the system for the expression of foreign DNA cloned into pSV01.

Our initial experiments involved cloning of the human β-globin gene into pSV01. We established that the DNA could be cloned into either side of the replication origin without apparent effect on the levels of DNA replication. We also saw that fragments as large as 7.6 kb did not seem to affect the level of DNA replication. Much to our dismay, however, we had extreme difficulty detecting β-globin mRNA from these cells. We constructed new vectors and tried recombinants with the human α-globin gene, both

of which were more amenable to this assay system.

Lusky and Botchan (20) had defined a region in pBR322 which inhibited the replication of SV40 molecules in cis. They transfected COS cells with a plasmid containing both pBR322 and SV40 DNA sequences. After 48 hours in culture many of the plasmids which had replicated contained deletions of pBR322 sequences. They selected one such deletion molecule which proved to be stable upon introduction into either E.coli or COS cells. Furthermore this plasmid was replicated to much higher levels in COS cells than the original plasmid. Using plasmids derived from their mutant we enhanced the level of DNA replication of our plasmids between five- and tenfold in COS cells. We also found that the α-globin gene gave approximately 100-fold more mRNA than similar β-globin constructs. We now know that the β-globin gene requires an enhancer element to attain high levels of expression upon re-introduction into mammalian cells (first described in 3).

Chapter 1 of this thesis is a report on the use of the COS cell system to dissect the regions 5' to the α-globin gene necessary for high levels of expression.

The second part of this thesis describes experiments on Drosophila melanogaster muscle genes I undertook in the laboratory of Norman Davidson. Previous work in this laboratory involved the isolation of the six actin genes (14) and the unique myosin heavy chain gene (26) from Drosophila melanogaster. We extended the studies of the

genes of the major myofilbrillar components by isolating and characterizing the myosin light chain genes from *Drosophila*. The major proteins of muscle are myosin and actin. These are ubiquitous proteins which are probably present in all eukaryotic cells. Their interaction has been implicated in cellular processes such as mitosis, cytokinesis and motility as well as the contraction of muscle. In muscle cells, actin is polymerized into thin filaments and myosin into thick filaments, which then assemble into an alternating array. Muscle contraction occurs when these two filaments are powered to slide past one another.

Myosin is a hexameric protein complex consisting of two heavy chains and two pairs of light chains, which we refer to as the alkali light chain (ALC or MLC-ALK) and myosin light chain-2 (MLC-2). Each myosin molecule has two globular head regions attached to a long α-helical tail. The rod-like tail is responsible for assembly into thick filaments. The head region is the site of ATP hydrolysis, the energy transduced to power muscle contraction. The head region also forms the cross-bridges to the actin filament. The two pairs of light chains are located on the head (see below).

Upon entering this laboratory I began working with Scott Falkenthal and William Mattox. They had screened a genomic library of *Drosophila* DNA to isolate genes for muscle specific transcripts. During the late pupal stage

of development synthesis of adult indirect flight muscle proteins occurs. RNA isolated from flies synchronized at this stage was use to generate a probe for muscle genes. By counterscreening with a probe from a non-myogenic stage of development 24 independent groups of muscle specific clones were identified. Preliminary characterizations of the clones involved hybrid selection and *in vitro* translation of RNAs abundantly transcribed in late pupae. Further characterizations were done by localizing the clones by *in situ* hybridization to *Drosophila* polytene chromosomes. Scott was initially interested in isolating the myosin light chain-2 gene. He chose as a candidate λdmpT102, a clone which hybrid selected an RNA which could be translated to give a protein which comigrated on a 2-dimensional gel with what we believed to be a myosin light chain polypeptide. I became interested in a set of clones which included eleven of the 73 clones we initially examined and was shown to encode two proteins. The clone λdmpT57 was a member of this group. A comparison was made of the amino acid sequences derived from translation of the open reading frames in λdmpT102 and λdmpT57 to the amino acid sequences of known myofibrillar components. We discovered that λdmpT102 encoded the alkali light chain and λdmpT57 encoded the myosin light chain-2. These are described in chapters 2 and 3 of this thesis.

Evolution of the myosin light chains.

The two myosin light chains have evolved from a common ancestor and belong to a family of related proteins which also includes calmodulin, troponin C and parvalbumin (4). These proteins arose from a primordial one-domain Ca^{++} binding protein which was duplicated to form a two-domain protein and then duplicated again to form a four domain protein. These proteins evolved rapidly from one another, eventually fixing mutations which are presumed advantageous. Goodman points out that in MLC-2 the Ca^{++} binding region of domain 2 had a deletion causing loss of divalent metal ion binding (17). During evolution this mutation was followed by a period of rapid amino acid substitution which was followed by a greatly decreased rate of change. He hypothesizes that this reflects an important structural function which has been well conserved in MLC-2. It should be noted that the two different classes of myosin light chains now share less than 25% sequence homology (8) and (presumably) serve different functions.

Analysis of myosin light chain function.

The myosin light chains have been implicated in the control of muscle contraction. Under normal conditions actin and myosin are inhibited from sliding past one another by one or more of three different mechanisms. Each of these modes of regulation are affected by the intracellular Ca^{++} concentration. The concentration of

Ca^{++} in relaxed muscle cells is less than 10^{-7}M. When this level becomes 10^{-6} to 10^{-5}M contraction is stimulated. In vertebrate striated muscle regulation is imparted by the thin filament. Ca^{++} acts by binding to troponin C, a protein of the thin filament, to relax the inhibition of the actin-myosin interaction. The other regulatory mechanisms involve the myosin light chain-2.

In some muscle types Ca^{++} binds directly to MLC-2, activating the myosin ATPase. This type of regulation has been shown in scallop muscle (30). Addition of Ca^{++} to scallop myosin allows ATP hydrolysis in the absence of actin. Upon treatment with EDTA one of the MLC-2 subunits is dissociated from the myosin complex. This myosin is desensitized as its ATPase activity is no longer regulated by calcium. Kendrick-Jones et al. showed that readdition of the removed scallop MLC-2 restored full activity (18). In addition they showed that MLC-2 from many different species could associate with the desensitized scallop myosin restoring ATPase activity. The alkali light chains did not complement the deficiency. They defined two classes of MLC-2 based on the ATPase activity of the reconstituted myosin. The MLC-2 of molluscs and vertebrate smooth muscle restored Ca^{++} sensitivity to free myosin as well as actomyosin. MLC-2 from vertebrate striated muscle was incapable of restoring Ca^{++} sensitivity to free myosin. The muscles from which the latter group were isolated also failed to show myosin linked regulation.

The third type of Ca^{++} control in muscle contraction is phosphorylation. Phosphorylation of MLC-2 occurs in vertebrates when the intracellular levels of Ca^{++} rise to 10^{-5} M. At this concentration calmodulin binds four Ca^{++} ions and is then capable of activating myosin light chain kinase (MLCK). This complex phosphroylates a serine residue near the amino terminus of MLC-2. When the calcium level drops, Ca^{++} is released from calmodulin inactivating MLCK. A Ca^{++} insensitive phosphatase then removes the phosphate from MLC-2. Adelstein and Conti reported phosphorylation of human platelet myosin resulted in an increased ATPase activity (1). Others have reported an absolute requirement for phosphorylation (28, reviewed in 2).

Phosphorylation has been reported to modulate contraction by different modes. Craig et al. reported that phosphorylation of chicken thymus (non-muscle) or gizzard (smooth muscle) MLC-2 induces long range structural effects. These lead to conformational changes in the myosin heavy chain rod causing it to unfold (9). The extended myosin molecules are then capable of assembling into filaments. Phosphorylation of the MLC-2 provides a second level of control in mouse skeletal muscle. The primary regulation in these cells is troponin C inhibition. Phosphorylation was shown to decrease the amount of ATP needed for contraction (10). There is increasing evidence for multiple levels of control in

muscle contraction. Calcium ion concentration has been shown to be important for phosphorylated smooth muscle myosin control (6). All three levels of control have been suggested to exist in rabbit skeletal muscle. Pulliam et al. have shown that at physiological ionic strengths rabbit skeletal muscle is calcium sensitive and may be further modulated by the extent of phosphorylation (24).

The alkali light chains were believed to be essential for myosin linked ATPase until recently. Myosin which has its MLC-2 removed by treatment with EDTA (see above) or DTNB (5,5'-dithiobis-(2-nitrobenzoic acid)) still possesses ATPase activity. Treatment at pH 11.1 to remove the alkali light chains results in loss of the ATPase. The alkali light chains were thus classified as "essential" light chains as it was believed they were essential for ATPase activity (33). This classification has recently been questioned by the work of Wagner and Giniger who were able to remove the ALC from myosin heads with mild detergent and antibody columns. They showed that complete removal of light chains did not result in complete loss of ATPase (31). Thus the role of the alkali light chains is unknown.

Location of myosin light chains.

The location and proximity of the two light chains has been studied in a series of recent experiments. Wallimann and coworkers have used antibodies to decorate the myosin light chains. They can visualize the region of antibody

binding in the electron microscope thereby inferring the positions of the light chains (13). From these experiments they concluded that the scallop myosin light chains were both located near the junction of the myosin heavy chain head and tail. Previous biochemical data had shown that the myosin light chains are extended molecules (29). Crosslinking of the light chains was achieved using bifunctional reagents. It was found that the two types of light chains were apposed to one another over a major portion of their length (32). The functional significance of their location and movement of the light chains during contraction remain as prospects for further investigation.

Future study of the Drosophila melanogaster myosin light chains

In vertebrates there are several distinct isoforms of the myosin light chains which are encoded by different genes. There are different alleles for smooth, skeletal and cardiac muscle as well as for embryonic and adult stages. We have shown that both the alkali light chain and the myosin light chain-2 genes in *Drosophila* are single copy (11, and chapter 3 of this thesis). The myosin heavy chain gene in *Drosophila* is also single copy (5,26). There is evidence, however, that all three of these genes encode multiple RNA transcripts which are translated to yeild more than one polypeptide during different stages of development. For the alkali light chain differential RNA

splicing leads to proteins which have different carboxy termini (12). In the myosin heavy chain transcripts there are different length transcripts generated through development (26). The myosin light chain-2 has two major transcripts which produce at least two polypeptides in an as yet unidentified manner.

The single copy nature of these genes should facilitate their study by the powerful technique of P-element mediated gene transfer. It has been shown that genes may be stably introduced into the germline of *Drosophila* by cloning them adjacent to the DNA segment responsible for transposition of DNA (the P element) in hybrid dysgenic crosses of *Drosophila*, and microinjecting the DNA into suitable embryos (27). Lethal deficiencies exist in the regions of both the alkali light chain (df 98B-98F) and myosin light chain-2 (df 99E-99F) loci. If either of these deficiencies could be complemented by re-introduction of the myosin light chain gene then many experiments could be undertaken to discern essential features of the myosin light chain polypeptide and transcription unit. (Since each of these deficiencies encompasses a large amount of DNA it is likely that genes other than the myosin light chains lead to the lethal phenotype. Many of the same experiments could still be done, but their interpretation would be somewhat more difficult.)

There are several structural homologies between the

Drosophila myosin light chains and their vertebrate counterparts which have been implicated in function (see above). Using the techniques of site directed _in vitro_ mutagenesis it would be possible to alter or delete specific codons from the myosin light chain coding region (25). These altered genes could then be re-introduced into _Drosophila_ cells to assay their effect. The myosin light chain-2 gene has several sites which could be tested. There is a conserved serine phosphorylation site at amino acid position 67 ("*" in figure 1). This could be changed to a cysteine, an amino acid of similar physical characteristics under the reducing conditions of intracellular space (M.White, personal communication) which cannot be phosphorylated, by changing the serine TCG codon to TGC, which encodes cysteine. The putative Ca^{++} binding domain of this protein extends from amino acid 80 to 107.

```
DMLC2   1 MADEKKKVKKKKTKEEGGTSETASEAASEAATPAPAATPA 40

                                     *
DMLC2  41 PAASATGSKRASGGSRGSRKSKRAGSS-VFSVFSQKQIAE 79
CDTNB   1               PKKAKRRAAEGSSNVFSMFDQTQIQE 26

           n  nn  na a ag  a  an  nn  n
DMLC2  80 FKEAFQLMDADKDGIIGKNDLRAAFDSVGK-IANDKELDA 118
CDTNB  27 FKEAFTVIDQNRDGIIDKDDLRETFAANGRLNVKNEELDA 66

DMLC2 119 MLGEASGPINFTQLLTLGANRMATSGANDEDEVVIAAFKT 158
CDTNB  67 MIKEASGPINFTVFLT-MFGEKLK-GADPED-VIMGAFKV 104

DMLC2 159 FDND--GLIDGDKFREMLMNFGDKFTMKEVDDAYDQMVID 196
CDTNB 105 LDPDGKGSIKKSFLEELLTTQCDRFTPEEIKNMWAAFPPD 144

DMLC2 197 DKNQID-TAALIEMLTGKGEEEEEEAA              222
CDTNB 145 VAGNVDYKNICY-VITH-GEDKEGE                167
```

Figure 1. Comparison of _Drosophila_ and Chicken MLC-2

Amino acids involved in the chelation of Ca^{++} include the acidic residues at positions 88, 90, 92 and 99 (denoted by "a" in figure 1); and the glycine residue ("g") at position 93. Flanking these residues are hydrophobic amino acids thought to form α-helices important for proper orientation of the Ca^{++} binding domain ("n" in figure 1). The requirement for divalent metal ion binding to the light chain could be tested by disruption of this region. The structural requirements predicted for Ca^{++} binding (19) could be tested by altering the codons for individual amino acids. A third region of interest is the highly conserved region from amino acid 115 to 134. There are only four amino acid differences in this region between *Drosophila* and chicken myosin light chains. Two of these changes are between non-conservative amino acids: lysine for glycine and valine for glutamine. It has been proposed that this is the domain responsible for interaction with the myosin heavy chain (17). It would be interesting to see if the differences in the myosin light chains have evolved in concert with compensating changes in the myosin heavy chain. By changing the *Drosophila* codons to match the chicken amino acid replacements perhaps this question could be addressed. There is a 54 residue amino terminal extension of the *Drosophila* polypeptide relative to the chicken sequence. To discern whether this sequence is necessary a construct could be made which deletes this

upstream leader segment. This construct could then be re-introduced into *Drosophila* embryos.

Without trying to address any ethical considerations here, I would like to propose some unique experiments which could be attempted using genes from other species. The myosin light chain genes have the unusual structure of the ATG codon for the initiation of translation forming the 3' boundary of an exon. It has been shown that hybrid genes whose intron boundaries are from different sources are properly excised (7). The possibility thus exists of constructing a gene which has a first exon from *Drosophila* and a hybrid intron followed by the protein coding sequence for a vertebrate myosin light chain. Such a gene would presumably contain all of the signals necessary for efficient transcription and translation in *Drosophila* and the sequence for a vertebrate protein. The assembly of the vertebrate myosin light chain into *Drosophila* musculature could be followed through the use of specific antibodies.

The alkali light chain gene of *Drosophila* has two different patterns of RNA splicing which result in polypeptides which differ in their carboxy terminal 14 amino acids (12). These gene segments could be altered by changing the sequences at the splice junction sites or within the protein coding segments so that only one of the two carboxy terminal amino acid sequences is used. Any effects could be monitored by re-introduction of the gene into *Drosophila*. There is also a putative Ca^{++} binding

domain in the alkali light chain. It could be tested as described above for the myosin light chain-2.

In vertebrates control of muscle contraction by myosin light chain varies with muscle type. In *Drosophila* there are six actin genes. It is conceivable that regulation of the interaction of *Drosophila* myosin with the different actins varies. Since the myosin genes in *Drosophila* are single copy, if different control mechanisms are present it would imply one or more of the following: there are additional factors which modulate control of muscle contraction present in different cells, the myosin light chains are processed to yeild different polypeptides in different cells, and the potential for control by each mechanism exists in all muscle types. Thus, specific alterations introduced into the myosin light chains may only have an effect at a particular time of development and in specific cells.

I hope that the data presented in the final two chapters of this thesis provide the foundation and inspiration for these further studies.

References

1. Adelstein,R.S., and M.A.Conti. 1975. Phosphorylation of platelet myosin increases actin-activated myosin ATPase activity. Nature (London) 256: 597-598.

2. Adelstein,R. and A.Eisenberg. 1980. Regulation and kinetics of the actin-myosin-ATP interaction. Ann.Rev. Biochem. 49: 921-956.

3. Banerji,J.,S.Rusconi, and W.Schaffner. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell 27: 299-308.

4. Barker,W.C.,L.K.Ketcham, and M.O.Dayhoff. 1978. The troponin C superfamily, p. 273-284. In M.O.Dayhoff (ed.), Atlas of protein sequence and structure, vol.5. National Biomedical Research Foundation, Silver Springs, Md.

5. Bernstein,S.I.,K.Mogami,J.J.Donady,and C.P.Emerson. 1983. Drosophila muscle myosin heavy chain is encoded by a single gene located in a cluster of muscle mutations. Nature (London) 302: 23-34.

6. Chacko,S.,M.A.Conti, and R.S.Adelstein. 1977. Effect of phosphorylation of smooth muscle myosin on actin activation and Ca^{2+} regulation. Proc.Natl.Acad.Sci.U.S.A. 74: 129-133.

7. Chu,G., and P.A.Sharp. 1981. A gene chimaera of SV40 and mouse β-globin is transcribed and properly spliced. Nature 289: 378-382.

8. Collins,J.H. 1976. Homology of myosin DTNB light chain with alkali light chains, troponin C and parvalbumin. Nature (London) 259: 699-700.

9. Craig,R.,R.Smith, and J.Kendrick-Jones. 1983. Light-chain phosphorylation controls the conformation of vertebrate non-muscle and smooth muscle myosin molecules. Nature (London) 302: 436-439.

10 Crow,M.T., and M.J.Kushmerick. 1982. Myosin light chain phosphorylation is associated with a decrease in the energy cost for contraction in fast twitch mouse muscle. J.Biol.Chem. 257: 2121-2124.

11. Falkenthal,S.,V.P.Parker,W.W.Mattox,and N.Davidson. 1984. Drosophila melanogaster has only one myosin alkali light-chain gene which encodes a protein with considerable amino acid sequence homology to chicken myosin alkali light chains. Mol. Cell. Biol. 4: 956-965.

12. Falkenthal,S.,V.Parker, and N.Davidson. 1985. Developmental variations in the splicing pattern of transcripts from the Drosophila myosin alkali light chain gene result in different carboxy-terminal amino acid sequences. Proc.Natl.Acad.Sci., in press.

13. Flicker,P.F.,T.Wallimann, and P.Vibert. 1983. Electron microscopy of scallop myosin: location of regulatory light chains. J.Mol.Biol. 169: 723-741.

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. Gluzman,Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23: 175-182.

16. Gluzman,Y.,R.Frisque, and J.Sambrook. 1980. Origin-defective mutants of SV40. Cold Spring Harbor Symp. Quant.Biol. 44: 293-300.

17. Goodman,M. 1980. Molecular evolution of the calmodulin fanily. pp.347-354 in Siegel,F.L., E.Carafoli, R.H. Kretsinger,D.H.Maclennan, and R.W.Wasserman, (eds.). Calcium-binding proteins: structure and function. Elsevier, New York.

18. Kendrick-Jones,J.,E.M.Szentkiralyi, and A.G.Szent-Gyorgyi 1976. Regulatory light chains in myosins. J.Mol.Biol. 104: 747-775.

19. Kretsinger,R.H. 1980. Structure and evolution of calcium-modulated proteins. Crit.Rev.Biochem. 8: 119-174.

20. Lusky,M., and M.Botchan. 1981. Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature (London) 293: 79-81.

21. Mulligan,R.C., and P.Berg. 1980. Expression of a bacterial gene in mammalian cells. Science 209: 1422-1427.

22. Myers,R.M., and R.Tjian. 1980. Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. Proc.Natl.Acad. Sci.U.S.A. 77:6491-6495.

23. Pellicer,A.,D.Robins,B.Wold,R.Sweet,J.Jackson,I.Lowy, J.M.Roberts,G.-K.Sim,S.Silverstein, and R.Axel. 1980. Altering genotype and phenotype by DNA-mediated gene transfer. Science 209: 1414-1421.

24. Pulliam,D.L.,V.Sawyna, and R.J.C.Levine. 1983. Calcium sensitivity of vertebrate skeletal muscle myosin. Biochem. 22: 2324-2331.

25. Razin,A.,T.Hirose,K.Itakura, and A.D.Riggs. 1978. Efficient correction of a mutation by use of chemically synthesized DNA. Proc.Natl.Acad.Sci. 75: 4268-4270.

26. Rozek,C.E., and N.Davidson. 1983. Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell 32: 23-34.

27. Rubin,G.M., and A.C.Spradling. 1982. Genetic transformation of Drosophila with transposible element vectors. Science 218: 348-353.

28. Sobieszek,A., and J.V.Small. Regulation of the actin-myosin interaction in vertebrate smooth muscle: activation via a myosin light-chain kinase and the effect of tropomyosin. J.Mol.Biol. 112: 559-576.

29. Stafford,W.F., and A.G.Szent-Gyorgyi. 1978. Physical characterization of myosin light chains. Biochemistry 17: 607-614.

30. Szent-Gyorgyi,A. 1975. Calcium regulation of muscle contraction. Biophys.J. 15: 707-723.

31. Wagner,P.D., and E.Giniger. 1981. Hydrolysis of ATP and reversible binding to F-actin by myosin heavy chain free of all light chains. Nature (London) 292: 560-562.

32. Wallimann,T.,P.M.D.Hardwicke, and A.G.Szent-Gyorgyi. 1982. Regulatory and essential light-chain interactions in scallop myosin: II. Photochemical cross-linking of regulatory and essential light-chains by heterobifuncional reagents. J.Mol.Biol. 156: 153-173

33. Weeds,A.G. 1969. Light chains of myosin. Nature (London) 223: 1362-1364.

34. Wigler,M.,R.Sweet,G.-K.Sim,B.Wold,A.Pellicer,E.Lacy, T.Maniatis,S.Silverstein, and R.Axel. 1979. Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell 16: 777-785.

Chapter 1

Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system.

(Reprinted from Cell 27: 279-288 (1981).)

Identification of DNA Sequences Required for Transcription of the Human α1-Globin Gene in a New SV40 Host-Vector System

Pamela Mellon,*† Vann Parker,† Yakov Gluzman‡ and Tom Maniatis*†
* Harvard University
Department of Biochemistry and Molecular Biology
7 Divinity Avenue
Cambridge, Massachusetts 02138
† California Institute of Technology
Division of Biology
Pasadena, California 91125
‡ Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

Summary

We have developed a rapid and simple method for studying the transcription of cloned eucaryotic genes, which involves transfecting SV40-transformed monkey cell lines (COS cells) with derivatives of the plasmid pBR322 that contain the SV40 viral replication origin but lack regions necessary for viral transcription (SV-ORI vectors). Because COS cells produce SV40 T antigen and are permissive for SV40 viral replication, transfected SV-ORI plasmids replicate to a high copy number. SV-ORI plasmids carrying a human α-globin gene are also replicated in COS cells. Moreover, the α-globin gene is faithfully transcribed to produce high levels of RNA, which is accurately processed to produce authentic α-globin mRNA. We have used this transcription system to demonstrate that a sequence located between 55 and 87 base pairs upstream from the mRNA capping site is required for efficient transcription of the α-globin gene in COS cells.

Introduction

The availability of DNA sequence information and the recent development of assays for transcription of cloned eucaryotic genes have made it possible to study the structure and function of eucaryotic promoters. A number of different transcription assays have been described: in vitro transcription in cell free extracts (Weil et al., 1979; Manley et al., 1980); microinjection of cloned eucaryotic genes into Xenopus oocytes (Brown and Gurdon, 1977; Kressman et al., 1977; Grosschedl and Birnstiel, 1980a; McKnight et al., 1981); transcription of genes stably introduced into mammalian cells in culture (Mantei et al., 1979; Wigler et al., 1979; Wold et al., 1979; for review see Pellicer et al., 1980); and transcription of cellular genes carried on SV40 viral DNA vectors (Hamer and Leder, 1979; Mulligan et al., 1979; Hamer et al., 1980; Mulligan and Berg, 1980; Gruss et al., 1981).

We describe the use of a new eucaryotic host-vector system to study the sequences necessary for transcription of a cloned human α-globin gene (Lauer et al., 1980). This cellular transcription system is based on the development of CV-1 monkey cell lines transformed with a replication-origin-defective SV40 DNA molecule. These cell lines, designated COS (CV-1, origin, SV40), do not contain a complete copy of the SV40 genome, but produce high levels of large T antigen and are permissive for SV40 DNA replication. They also efficiently support replication of defective viruses with deletions in the early region (Gluzman, 1981) and of bacterial plasmids containing only a 228 base pair (bp) SV40 origin fragment (Myers and Tjian, 1980). Deletion analyses have shown that this origin fragment lacks sequences essential for efficient early gene transcription in vivo (Benoist and Chambon, 1981; Gruss and Khoury, 1981; M. Fromm and P. Berg, personal communication). In addition, the sites of initiation of the major late mRNAs are 43 to 147 bp outside this origin fragment (Ghosh et al., 1978). Thus this system provides a means of amplifying cloned eucaryotic genes in a mammalian cell and studying their promoters free from the interference of highly active viral promoters and splicing sites.

We demonstrate that derivatives of the plasmid pBR322 that contain the SV40 viral replication origin (SV-ORI vectors) can be used efficiently to replicate cloned human α-globin genes in COS cells and that the amplified gene copies are accurately transcribed from their own promoters to give high levels of α-globin mRNA. Moreover, the α-globin transcripts are spliced to produce mature mRNA that is identical to authentic human α-globin mRNA. We have used this system to identify 5′ flanking sequences that are necessary for α-globin transcription in cells in culture. These studies demonstrate that a maximum of 87 bp of DNA upstream from the mRNA capping site is required for accurate and efficient transcription of the human α-globin gene in this system.

Results

Replication of SV-ORI Plasmids in COS Cells

A variety of plasmids containing the SV40 replication origin was constructed and assayed for DNA replication in COS cells (a detailed description of these plasmids is presented in Figure 1). The extent of replication was measured by a restriction endonuclease cleavage assay described by Peden et al., 1980. The assay is based on the fact that the restriction enzyme Dpn I will cut DNA that has been methylated at the sequence $G^{me}ATC$ by propagation in *dam*$^{+}$ Escherichia coli, but will not cut the unmethylated version of the same sequence. Since mammalian cells do not contain an analogous methylase, plasmid DNA that has replicated in cells in culture will not be cleaved by the Dpn I enzyme. SV-ORI plasmid DNAs were introduced into COS cells by a modified calcium phosphate precipitation protocol (Experimental Procedures). Low molecular weight DNA was isolated 5 and 48 hr after transfection by the procedure of Hirt (1967)

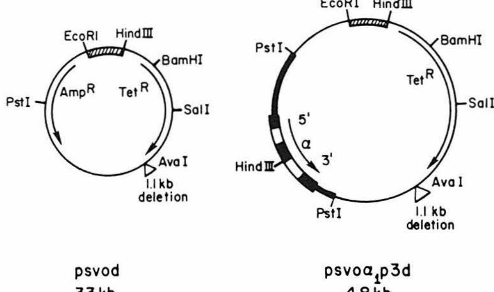


Figure 1. Restriction Maps of an SV-ORI Vector and an SV-ORI/α-Globin Gene Recombinant

The vector pSVOd was constructed as follows. The original SV40 origin plasmid, pSVO1 was constructed by inserting the 311 bp Eco RII G fragment of SV40 into the Eco RI site of pBR322 with Eco RI linkers (Myers and Tjian, 1980). A vector that does not contain sequences inhibitory to replication of pBR322 DNA in mammalian cells was selected from pSVO1 by Lusky and Botchan (1981). This vector, pML2-RIIG, contains a deletion from 1120 to 2490 on the pBR322 map, which removes the inhibitory sequence and a portion of the tetracycline resistance gene. The vector pSVOd was derived from pML2-RIIG by reinserting sequences from 1120 to the Ava I site at 1424 to reconstruct the tetracycline resistance gene. In addition, the sequences between the Hind III site in pBR322 and the Hind III site in the SV40 insert were deleted to remove one each of the two Eco RI and Hind III sites. Plasmid pSVOd does not contain the pBR322 sequences inhibitory to replication in mammalian cells, carries both ampicillin and tetracycline resistance markers and has unique Pst I, Eco RI, Hind III, Bam HI and Sal I sites.

Human α-globin genes were inserted into both pSVO1 and pSVOd. A 1.5 kb Pst I fragment containing the α1-globin gene was inserted at the Pst I site in either orientation. The notation pSVOα1p3d signifies: plasmid (p) with an SV40 origin (SVO) containing α1-globin (α_1) in the Pst site (P), oriented from the 3' to 5' direction clockwise (3), and containing a deletion of the replication inhibitory sequences in pBR322 (d). Thin line: pBR322 sequences. Striped box: SV40 sequences. Thick line: human α-globin flanking sequences. Black boxes: α-globin mRNA coding sequences. Open boxes: α-globin intervening sequences. Arrows: transcriptional orientation of tetracycline and ampicillin resistance genes and the human α1-globin gene. Figures are approximately to scale, and some restriction sites are shown for vector sequences only.

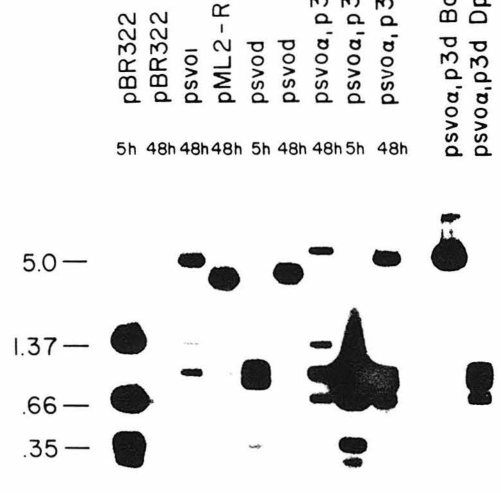


Figure 2. Replication of SV-ORI Plasmid DNAs in COS cells

COS cells were transfected with plasmid DNAs as described in Experimental Procedures. At 5 or 48 hr after transfection, low molecular weight DNA was prepared as described by Hirt (1967). DNA samples were treated with DNAase-free RNAase followed by digestion with the restriction enzymes Dpn I and Bam HI. DNA purified from approximately 10^5 cells was fractionated by electrophoresis on a 1.5% agarose gel, transferred to nitrocellulose (Southern, 1975) and hybridized with ^{32}P-labeled pML2-RIIG DNA. The filter was exposed to Kodax XR-5 film for 12 hr without an intensifying screen. The plasmids and transfection times are indicated above each lane. The two lanes on the far right of the figure contain 100 ng of pSVOα1p3d DNA digested with Bam HI or Dpn I.

and was digested with Dpn I and Bam HI. The recombinant plasmids contain one Bam HI and multiple Dpn I sites. Plasmid DNA that has replicated in mammalian cells therefore should be cut once by Bam HI and not at all by Dpn I to produce a full-length linear plasmid DNA. Figure 2 shows that pBR322 DNA (not containing the SV40 replication origin) is not replicated in COS cells, as evidenced by the fact that the DNA is digested with Dpn I 48 hr after transfection. In contrast, plasmid DNA containing the SV40 replication origin is extensively replicated after 48 hr (Figure 2). The deletion of a sequence in the plasmid DNA that inhibits DNA replication in mammalian cells (Lusky and Botchan, 1981; see also Myers and Tjian, 1980) significantly increases the amount of replicated DNA (Figure 2). Furthermore, the insertion of fragments of DNA as large as 8 kb into the SV-ORI plasmid does not substantially inhibit DNA replication (Figure 2 and our unpublished data). The expected requirement of T antigen for the replication of SV-ORI plasmids is confirmed by the finding that only low levels of Dpn I-resistant plasmid DNA are observed in normal monkey cells (CV-1 cells) after transfection with the α-globin SV-ORI plasmids (data not shown).

The number of copies of replicated SV-ORI plasmid DNA in COS cells 48 hr after transfection was estimated by DNA blotting experiments with known amounts of plasmid DNA as standards (data not shown). We estimate that there are approximately 100 ng of SV-ORI DNA per 5×10^5 cells or 40,000 copies per cell. If 10% to 20% of the cells are assumed to be transfected (Chu and Sharp, 1981), each cell with replicated SV-ORI DNA contains between 200,000 and 400,000 copies of DNA after 48 hr. SV40-infected monkey cells contain approximately 250,000 copies of SV40 DNA late in infection (Tooze, 1980). Thus the efficiency of replication of SV-ORI recombinants in COS cells is approximately the same as that of wild-type SV40 in normal monkey cells. We have tested two independently derived COS cell lines, COS-

7 and COS-1 (Gluzman, 1981), and no difference in the extent of replication of SV-ORI plasmids was observed.

Transcription of the Human α-Globin Gene

The feasibility of using the SV-ORI/COS cell system to study eucaryotic gene transcription was demonstrated by use of SV-ORI/α-globin gene recombinant plasmids. A 1.5 kb Pst I fragment, containing the intact α1-globin gene, 525 bp of 5′ flanking DNA sequences and 83 bp of 3′ flanking DNA sequences, was inserted into pSV01 and pSVOd (Figure 1). These plasmids were used to transfect COS cells, and cytoplasmic RNA was harvested 48 hr later. A variety of procedures was used to analyze the α-globin mRNA produced. Figure 3 shows the results of an RNA blotting experiment comparing the poly(A) RNA isolated from transfected COS cells with that from human blood obtained from the umbilical cord (cord blood). Using a ^{32}P-labeled α-globin cDNA plasmid (Wilson et al., 1978) as a hybridization probe, we detect a 9S

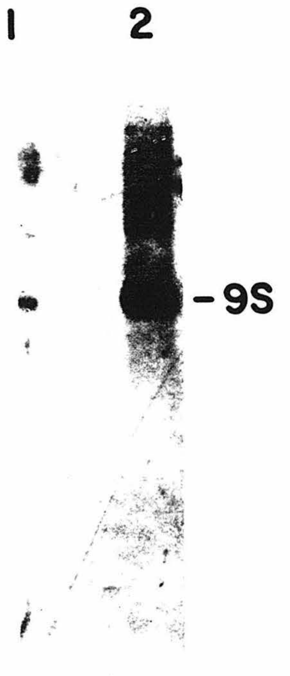


Figure 3. Transcription of the Human α1-Globin Gene in COS Cells

Poly(A)-containing cytoplasmic RNA was prepared from COS cells 48 hr after transfection. RNA blot hybridization analysis was performed as described in Experimental Procedures with a 1% agarose-formaldehyde gel and blotting to nitrocellulose. The ^{32}P-labeled probe was an α-globin cDNA plasmid (pJW101, Wilson et al., 1978). Lane 1, 2 ng of poly(A)-selected human cord blood RNA with 2 μg sea urchin RNA carrier; lane 2, 2 μg of poly(A) RNA from COS cells transfected with pSVOα_1p3. Exposure time: 4 days with an intensifying screen.

RNA both in transfected COS cell RNA and in human cord blood RNA. The size of the α-globin mRNA in transfected COS cells suggests that the cloned α1-globin gene is accurately transcribed in COS cells and that its transcript is faithfully spliced and polyadenylated. We did not detect high levels of α-globin mRNA in CV-1 cells transfected with an SV-ORI plasmid or in COS cells transfected with α-globin plasmids that do not carry an SV40 replication origin (data not shown). Since we have shown that transfected DNA does not replicate significantly under either of these conditions, we conclude that high levels of α-globin transcription of COS cells requires a high DNA copy number. This conclusion is supported by the observation that α-globin recombinants containing the replication-inhibitory sequence of pBR322 (Lusky and Botchan, 1981) produce less RNA, by a factor of 5–10, than α-globin/pSVOd recombinants in COS cells.

The size of the α-globin mRNA produced in COS cells is approximately the same as that of α-globin mRNA in human cord blood (Figure 3, see also Figure 6B). To establish that these two mRNAs are identical, we carried out more extensive structural studies. First, we demonstrated that the two mRNAs have identical 5′ ends. A 52 bp DNA fragment from the first exon of the α-globin was end-labeled, strand-separated and used to prime reverse transcription of transfected COS cell α-globin mRNA and human α-globin mRNA (Proudfoot et al., 1980). The 3′ end of the primer is 23 nucleotides from the cap of α-globin mRNA (Baralle and Brownlee, 1978; Liebhaber et al., 1980). The products of primer extension are 75 nucleotides in length for both mRNAs, as predicted from the published mRNA sequence data (Figure 4A). Thus the 5′ end of the α-globin mRNA produced in transfected COS cells is coterminal with that of human α-globin mRNA.

Second, we demonstrated that the two mRNAs have identical 3′ ends. A restriction fragment that terminates within exon III of the α-globin gene was 3′-end-labeled with ^{32}P and used to map the 3′ end of COS cell α-globin mRNA and cord blood α-globin mRNA as described in the legend to Figure 4B and in Experimental Procedures. The fragments protected from S1 nuclease digestion by both types of RNAs are the size predicted from the sequence of α-globin mRNA (Proudfoot and Longley, 1976, see Figure 4B).

Third, we demonstrated that the α-globin mRNA from transfected COS cells is accurately spliced, using an S1 nuclease hybridization assay that detects all three exons. Plasmid DNA containing the α-globin gene was uniformly labeled with ^{32}P by growth in bacteria. A Pst I restriction fragment was isolated by gel electrophoresis, denatured and annealed to either the transfected COS cell RNA or human cord blood RNA. The hybrid was digested with S1 nuclease, and the products were fractionated by polyacrylamide gel electrophoresis. Protected DNA fragments of 133,

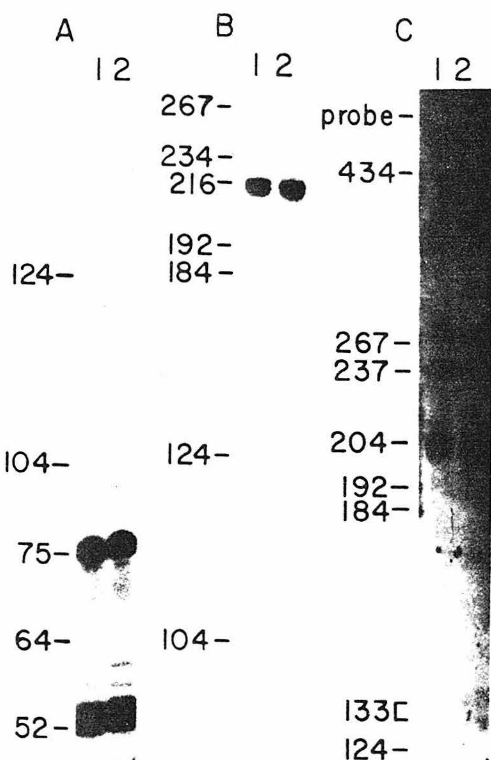


Figure 4. Structural Analysis of Human α1-Globin RNA from COS Cells

Cytoplasmic RNA was prepared from COS cells 48 hr after transfection with an SV-ORI vector containing the α1-globin gene. (A) Primer extension analysis of the 5′ end of α1-globin RNA (Proudfoot et al., 1980) The primer was a 5′-end-labeled 52 bp Hinf I–Hae III fragment from the first exon of the α1-globin gene (+23 to +75 bp in the sequence; Liebhaber et al., 1980), which was extended 23 bp with reverse transcriptase to generate a 75 bp DNA fragment. Lane 1, 2 μg of poly(A)-selected cytoplasmic RNA from COS cells transfected with pSVOα_1p3; lane 2, 10 ng of poly(A)-selected RNA from human cord blood with 2 μg of sea urchin RNA carrier.

(B) S1 nuclease analysis of the 3′ end of the α-globin RNA was performed according to the procedure of Weaver and Weissman (1979). A 3′-end-labeled 600 bp Bst EII fragment that extends from a site within exon III of the α1-globin gene (see Figure 5) to a site approximately 400 bp downstream from the poly(A) addition site was used as a probe. The S1 nuclease-resistant fragments were fractionated in a 4% acrylamide–7 M urea gel (Maniatis et al., 1975b). Lane 1, 2 μg of poly(A)-selected RNA isolated from COS cells transfected with pSVOα_1p3; lane 2, 10 ng of poly(A)-selected human cord blood RNA with 2 μg sea urchin carrier RNA.

(C) S1 nuclease analysis of α1-globin RNA with a uniformly labeled duplex DNA as probe. Samples were subjected to electrophoresis on an 8% acrylamide–7 M urea gel. Lane 1, 20 μg of cytoplasmic RNA from COS cells 48 hr after transfection with pSVOα_1p3d; lane 2, 50 ng of poly(A)-selected human cord blood RNA with 20 μg of sea urchin carrier RNA.

204 and 237 nucleotides in length, corresponding to exons 1, 2 and 3, are predicted from the DNA sequence (Leibhaber et al., 1980). As shown in Figure 4C, fragments of this size are observed with both the COS cell and the erythroid cell RNAs. The small 5′ exon fragments appear as diffuse bands with both kinds of RNA, but the fact that the 5′ primer extension analysis shows discrete bands (Figure 4A) indicates that the heterogeneity is an artifact of S1 nuclease digestion. We conclude from these analyses that the COS cell α-globin mRNA is structurally indistinguishable from normal human α-globin mRNA.

By reconstruction experiments in which known amounts of human cord blood mRNA are used, we estimate that 1% to 4% of the cytoplasmic poly(A)-containing RNA in COS cells is α-globin mRNA. If we assume that a typical fibroblast has 0.3 pg of mRNA (Lewin, 1980), there are approximately 10,000 to 40,000 copies of α-globin mRNA per COS cell after 48 hr. If 20% of the cells are transfected (see above), there must be 200,000 copies of globin mRNA per cell or approximately 1 mRNA molecule for each DNA copy in the cell. This estimate is consistent with numbers that Hamer and Leder (1979) have reported for mouse α-globin RNA production using a lytic SV40 virus vector. (A typical erythroblast with four α-globin genes contains on the order of 50,000 to 75,000 molecules of α-globin mRNA; Paul et al., 1977.)

We have shown that the high level of transcription of the α-globin gene in COS cells does not depend on the orientation of the α-globin-gene insert with respect to the SV40 replication origin. The 1.5 kb Pst I α1-globin gene fragment was inserted into the Pst I site of pSVOd in both orientations, and a 4.2 kb α2-globin gene fragment was inserted into one of the two Eco RI sites of pSVO1. The level of mRNA produced in each case did not depend on the orientation of the insert. This observation and the fact that the 5′ ends of the α-globin mRNA from COS cells map to the mRNA capping site of the gene indicate that the α-globin genes must be transcribed from their own promoters.

Identification of Sequences Required for Transcription of the α-Globin Gene

To identify the sequences required for transcription of the human α1-globin gene in COS cells, we constructed a set of cloned α1-globin genes with deletions of different amounts of 5′ flanking sequences. The deletions were generated by digesting the α1-globin/SV-ORI plasmid with restriction enzymes that recognize sites located at known distances from the mRNA capping site of the gene. The digested DNA was rendered blunt-ended by T4 DNA polymerase, then closed by ligation. As shown in Figure 5, deletions were constructed at −525, −220, −87, −55 and +157 bp from the mRNA capping site. Each of the deleted DNAs was fused with the same site in pBR322, thus they all have common 5′ flanking plasmid sequences. The transcriptional activity of these α-globin plasmids was tested by transfection into COS cells. An equal amount of a SV-ORI/β-globin gene plasmid was included in the transfection mix as an internal

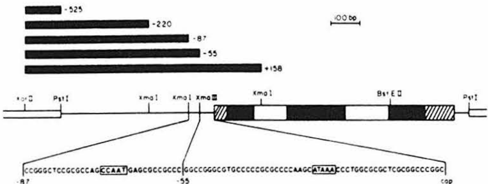


Figure 5. Deletions of the 5′ Flanking Region of the Human α1-Globin Gene

The plasmid pSVOα_1p3 was digested to completion with Xor II and partially digested with Pst I, Xma I or Xma III, and the staggered ends were made flush with T4 DNA polymerase. The plasmid fragments were isolated from 1% agarose gels, recircularized by ligation, then transfected into E. coli. Positions of the −525, −220 and +157 deletions were determined by restriction mapping and therefore are approximate. Positions of the −88 and −55 deletions were determined by direct DNA sequence analysis (Maxam and Gilbert, 1977). Double line: pBR322 sequences. Single line: α1-globin flanking sequences. Black boxes: α1-globin coding sequences. Open boxes: α1-globin intervening sequences. Striped boxes: α1-globin untranslated mRNA sequences. The sequence of the 88 bases immediately 5′ of the mRNA start (Leibhaber et al., 1980) is shown with the highly conserved CCAAT and ATAAA sequences boxed.

control. Results of the experiment are presented in Figure 6. Using the 3′ end assay, we find that the amount of α1-globin RNA made from the −220 plasmid is reproducibly two times greater (by scintilliation counting of excised bands) than that produced from the −525 and −87 plasmids (though the plasmids replicate to the same level; data not shown). We do not understand the reason for this difference. Deletion of the additional 32 bp region between −87 and −55 bp from the mRNA capping site decreases the amount of α-globin RNA transcript by a factor of 10–20. The faint band observed in the −55 lane may be the result of transcriptional readthrough from weak promoter sites elsewhere in the vector, since a faint band also is observed with the +157 plasmid, which is missing all of the 5′ flanking sequence and the entire first exon of the gene. The level of the 3′ end β-globin mRNA is approximately the same in each RNA sample (Figure 6A), indicating that the decrease in α-globin mRNA in some of the deletions is not due to differential efficiency of transfection or degradation of RNA during analysis. Similar results were obtained when the RNA blotting assay was used to study the α-globin gene deletions (Figure 6B). As in the case of the 3′ end assay, the amount of α-globin mRNA observed with the −220 plasmid is slightly increased compared to that observed with the −525 and −87 plasmids. We have not, however, detected any RNA bands with the −55 and +157 deletions, even with 2-week exposures (a decrease in the signal by a factor of 20 would have been detected). The presence of discrete 3′ mRNA ends when no RNA bands are present in the blotting analysis supports the suggestion that the 3′ ends result from transcriptional readthrough from various weak promoter sites upstream with polyadenylation at the 3′ end of the α-globin gene. Thus the number of accurate mRNA transcripts between the −87 and −55 deletions must decrease by a factor of at least 20. We conclude that more than 55 but fewer than 87 base pairs of 5′ flanking sequence are required for accurate transcription of α-globin gene in COS cells.

Discussion

In this paper, we describe the use of a rapid and convenient system for studying the transcription of cloned eucaryotic genes. This SV40-derived system has several useful features. First, SV-ORI plasmids can be readily manipulated in vitro and propagated in bacterial cells and then efficiently replicated and transcribed after transfection into COS cells. Second, COS cells provide the T antigen required for SV40 replication. Thus no helper virus is necessary. Third, there are no known size constraints on the recombinant plasmids. Fourth, their replication does not require the presence of actively transcribed viral genes whose promoters, splicing junctions and termination sites may interfere with the analysis of cloned gene transcripts. Finally, in contrast to the stable cotransformation procedure (Wigler et al., 1979), in which each assay requires several weeks and in which variable but low levels of mRNA are produced, the SV-ORI/COS cell transcription assay requires only 48 hr and the cloned gene is transcribed to give reproducibly high levels of mRNA. Thus it should be possible to clone and amplify large eucaryotic genes and to study their transcription easily regardless of their orientation with respect to the SV40 origin fragment. Alternative transcription systems that involve the transient introduction of cloned DNA into cells in culture have been developed independently, in which large segments of SV40 DNA (Mulligan and Berg, 1980; Banerji et al., 1981) or polyoma DNA (Dierks et al., 1981a and 1981b; J. de Villiers and W. Schaffner, personal communication) are used. These vectors contain active early and late viral promoters and RNA splicing and termination sites.

In principle, the SV-ORI/COS cell system could be used to study the transcription of any mammalian gene. In fact, the 4.2 kb rat α2U-globulin gene, which contains six intervening sequences (Kurtz, 1981), is transcribed and accurately processed in COS cells to produce a faithful 1000-nucleotide α2U-globulin mRNA at levels greater than 10,000 copies per cell (D. Kurtz, personal communication). In addition, the 1.3 kb herpes virus thymidine kinase gene, which contains no intervening sequences (McKnight, 1980) is efficiently transcribed in COS cells (S. Silverstein, personal communication). In contrast, the amount of mRNA produced from a cloned human β-globin gene is less by a factor of 100 than that made from the human α-globin gene, regardless of the location or orientation of the β-globin gene in pSVOd or pSVO1

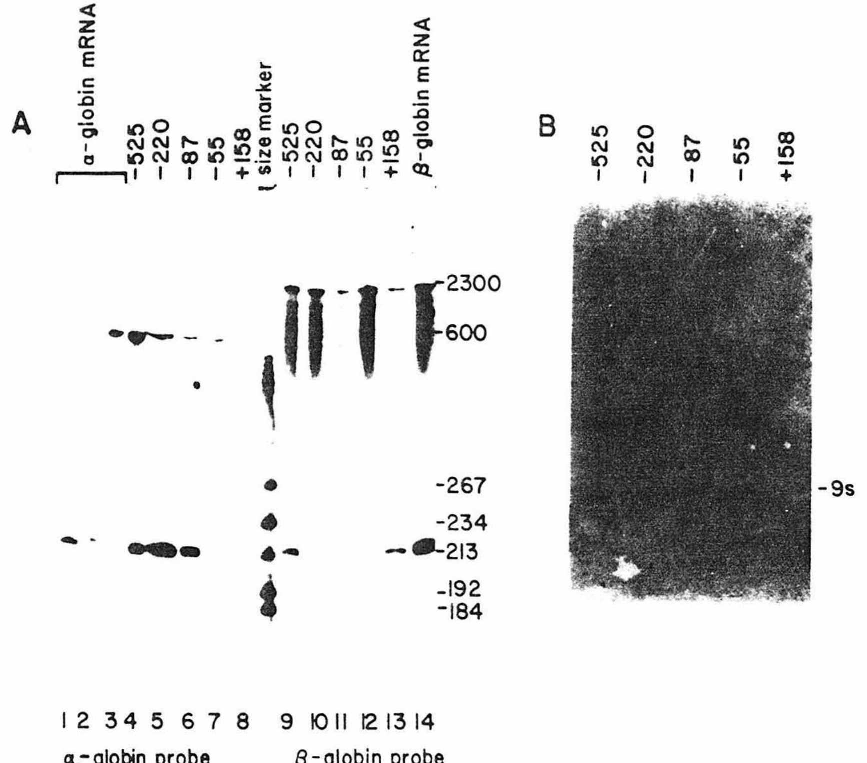


Figure 6. Analysis of the α-Globin Gene Deletion Mutants

(A) S1 nuclease analysis. Each of the five deleted α-globin plasmids described in Figure 5 was transfected into COS cells along with an equal amount of the pSVO1 plasmid containing a 5.2 kb Bgl II fragment that includes the entire human β-globin gene and flanking sequences. We analyzed 2 μg of poly(A)-selected cytoplasmic RNA for 3′ α1-globin sequences as described in Figure 3B. A second 2 μg aliquot of each RNA sample was also analyzed for the presence of human β-globin RNA sequences. By visual comparison of the band intensities produced from 3 ng of cord blood RNA, using the α-globin and β-globin probes, we estimate that the amount of β-globin mRNA produced in COS cells is less, by a factor of 100, than the amount of α-globin mRNA. The 3′-end-labeled probe was a 2.3 kb DNA fragment extending from the Eco RI site in the third exon of the human β-globin gene to the Bgl II site located 2.1 kb from the poly(A) addition site. The bands at 0.6 and 2.3 kb result from reannealing of the intact double-stranded DNA probe and were also detected with samples containing only sea urchin RNA. Lane 1, 10 ng; lane 2, 3 ng; lane 3, 1 ng of poly(A)-selected human cord blood RNA, with 2 μg of sea urchin carrier RNA. Lanes 4 through 8: RNA from COS cells transfected with deletion plasmids and analyzed with an α-globin probe. Lanes 9 through 13: RNA from COS cells transfected with deletion plasmids analyzed with a β-globin probe. Lane 14: 3 ng poly(A)-selected human cord blood RNA with 2 μg sea urchin RNA.

(B) RNA blot hybridization analysis. The α-globin gene deletions originally cloned in pSVO1 were transferred into pSVOd (removing the pBR322 replication inhibitory sequence) and transfected into COS cells. We subjected 10 μg of total cytoplasmic RNA to electrophoresis on a 2% agarose-formaldehyde gel and transferred it to nitrocellulose as described (Experimental Procedures). Though the α-globin mRNA produced in COS cells comigrated with human cord blood RNA in a 1% agarose-formaldehyde gel, it was occasionally slightly larger in a 2% agarose-formaldehyde gel (data not shown). We attribute this difference to heterogeneity in the length of the 3′ poly(A) tail since the 5′ and 3′ ends and the sizes of the three exons of the COS cell α-globin RNAs are identical to the human cord blood RNA (Figure 4).

(Figure 6 and R. Treisman and N. Proudfoot, personal communication). J. Sklar and P. Berg (personal communication) have observed a similar difference in α- and β-globin gene expression using a different SV40-based vector system (Mulligan and Berg, 1980). The reasons for this difference are under investigation.

The SV-ORI/COS cell system offers many advantages over existing eucaryotic transcription systems; however, it also has limitations. First, because the gene is amplified to several thousand copies per cell, its transcriptional properties may not accurately reflect those of normal cells. Second, the use of the vector is limited to COS cells, so gene expression cannot be studied in specific differentiated cell types. Third, preliminary studies suggest that α-globin mRNA (our unpublished observations), α2U-globulin mRNA (D. Kurtz, personal communication) and herpes virus thymidine kinase mRNA (S. Silverstein, personal communication) are not efficiently translated in COS cells. Thus it may not be possible to use the SV-ORI/COS

cell system to produce large amounts of protein without the use of an SV40 promoter.

Globin Gene Promoters

Comparison of 5′ flanking sequences of several mammalian α-globin and β-globin genes revealed two highly conserved sequences: ATAAA, located at 24 to 32 bp, and CCAAT, located at 67 to 77 bp, on the 5′ side of the mRNA capping site (Efstratiadis et al., 1980; Liebhaber et al., 1980). Both these sequences are also found in the corresponding locations of many other eucaryotic genes (Goldberg, 1979; Benoist et al., 1980). A region that includes the ATAAA sequence but not the CCAAT sequence is required for faithful in vitro transcription of the chicken conalbumin gene (Corden et al., 1980), the major adenovirus late genes (Hu and Manley, 1981) and the rabbit β-globin gene (Dierks et al., 1981a; Grosveld et al., 1981). The ATAAA sequence was directly implicated in transcription in vitro by the observation that a single base change (ATAAA to AGAAA) reduces the level of in vitro transcription from the conalbumin gene by a factor of 10 (Wasylyk et al., 1980). In contrast, the deletion of a similar AT-rich sequence 5′ to the SV40 early genes has little or no effect on the amount of early transcript made in vitro or in vivo, but does affect the site of mRNA initiation (Gluzman et al, 1980a, 1980b; Rio et al., 1980; Benoist and Chambon, 1981; Ghosh et al, 1981; Mathis and Chambon, 1981). As in the SV40 early genes, the −30 bp region of the rabbit β-globin promoter region is involved in positioning the mRNA start site both in vitro and after transfection into cells in culture (Grosveld et al., 1981; G. Grosveld and R. Flavell, personal communication; Dierks et al., 1981a). In contrast to SV40, however, deletion of this sequence in the rabbit β-globin promoter region decreases transcription in cells in culture by a factor of 6 (G. Grosveld and R. Flavell, personal communication) to 30 (Dierks et al., 1981a).

Our deletion analysis localizes an additional sequence required for transcription of the α-globin gene in cells in culture by demonstrating that deletion of 5′ flanking sequences to the position −87 does not significantly affect transcription, but the deletion of an additional 32 bp to the position −55 bp from the mRNA capping site drastically reduces the level of transcription. A similar analysis of deletion mutants in the rabbit β-globin promoter region in mouse cells (Dierks et al., 1981a and 1981b) and in HeLa cells (G. Grosveld and R. Flavell, personal communication) indicates that the region between −76 to −66 bp in mouse cells and −100 to −54 bp in HeLa cells is required for efficient transcription after transfection. The deletion of these sequences decreases the amount of accurately initiated transcript by a factor of 5-20. Similarly, the deletion of a sequence from −109 to −52 bp upstream from the mRNA capping site of the herpes thymidine kinase gene decreases its transcription by a factor of 50 in Xenopus oocytes (McKnight et al., 1981). In the case of the α-globin gene, we have not observed any 9S α-globin mRNA from the −55 deletion plasmid under conditions in which the amount of RNA detected would have decreased by a factor of 20 (Figure 6B and data not shown).

In contrast to the examples cited above, deletion of the corresponding region from the sea urchin H2A histone gene has little effect on its transcription in oocytes (Grosschedl and Birnstiel, 1980a). This observation and the absence of discernible sequence homology in the −70 to −80 bp region of some eucaryotic genes suggests diversity in the mechanism of eucaryotic RNA polymerase–promoter interactions. However, the conservation of the CCAAT sequence in the −70 to −80 bp region among mammalian globin genes over 500 million years is remarkable (Efstratiadis et al., 1980). More thorough molecular genetic analysis will be required to determine whether the CCAAT sequence is directly involved in globin transcription.

The fact that the ATA region is required for both in vitro and in vivo transcription of the globin gene and the CCAAT region is necessary only in vivo suggests that the two regions have separate functions. Considering the low efficiency of in vitro transcription (fewer than 1 in 100 DNA templates are transcribed in vitro; Weil et al., 1979; Manley et al., 1980), it is possible that factors which ordinarily interact with the −70 to −80 bp region to promote transcription are missing from the in vitro extracts. Alternatively, it is possible that globin genes after transfection into cells in culture are assembled into histone-DNA complexes that promote efficient RNA polymerase–DNA interactions.

Our analysis demonstrates that sequences farther than 87 bp from the 5′ end of the α-globin gene are not required for transcription in COS cells. In contrast, sequences farther upstream are required for transcription of the sea urchin histone H2A gene in Xenopus oocytes (Grosschedl and Birnstiel, 1980b) and the SV40 early genes (Benoist and Chambon, 1981; Gruss and Khoury, 1981) and polyoma early genes (C. Tyndall, U. Novak and R. Kamen, personal communication) in cells in culture. Although these sequences have no homology to each other, they have similar effects on transcription. In addition, these SV40 and polyoma sequences can enhance the transcription of a cloned rabbit β-globin gene transfected into HeLa cells (Banerji et al., 1981) or mouse 3T6 cells (J. de Villiers and W. Schaffner, personal communication; Dierks et al., 1981a). The enhancing sequences act only in cis, and their activity is not influenced by their orientation or, within limits, by their location (Banerji et al., 1981; M. Fromm and P. Berg, personal communication). It is worth noting that the SV40 sequence that enhances β-globin gene expression in HeLa cells is not present in SV-ORI vectors.

In conclusion, analysis of a number of different

genes in a variety of transcriptional assay systems has revealed similarities as well as differences in the sequence requirements for eucaryotic gene transcription. Whether the differences are indicative of heterogeneity in promoter function that operates in the natural state or are related to the assay system has not been established.

Experimental Procedures

Enzymes and Reagents

Restriction enzymes were obtained from New England BioLabs or Bethesda Research Laboratories or were prepared in the laboratory by M. Alonso. The Klenow fragment of E. coli DNA polymerase I was purchased from Boehringer Mannheim, and T4 polynucleotide ligase and E. coli DNA polymerase I holoenzyme were provided by M. Alonso. T4 DNA polymerase was purchased from Bethesda Research Laboratories. Reverse transcriptase was provided by J. Beard, Life Sciences Inc. S1 nuclease, prepared by the procedure of Weigand et al. (1975), was provided by E. Fritsch. RNAase-free DNAase was purchased from Worthington Biochemicals and repurified on agarose 5′ paraamino-phenylphosphoryl-uridine 2′ (3′)-phosphate (Miles) as described by Maxwell et al. (1977). We purchased α-^{32}P nucleotide triphosphates (2000–3000 Ci/mmole) from Amersham. Ribonucleoside-vanadyl complexes were prepared as described by Berger and Birkenmeier (1979). Oligo(dT) cellulose was purchased from Collaborative Research.

Purification and Labeling of DNA

The plasmid pSVO1 was provided by R. Myers and R. Tjian and pML2-RIIG by M. Lusky and M. Botchan. We prepared pBR322 subclones of the human α1-globin gene and α2-globin gene from the bacteriophage λ clone λHαG2 (Lauer et al., 1980). The pSVO1/human β-globin gene recombinant was prepared from the bacteriophage λ clone λHβG3 (Lawn et al., 1978). Plasmids were propagated in *dam*$^+$ E. coli K12 strains. DNA fragments were labeled by nick translation under the conditions of Maniatis et al. (1975a). We 3′-end-labeled restriction fragments using the Klenow fragment of E. coli polymerase I according to the conditions described by Lawn et al. (1980). Plasmid DNA was uniformly labeled with ^{32}P by growth in 40 ml of dephosphorylated L broth containing 4 mCi of ^{32}P.

COS Cell Transfection

COS and CV-1 cells were propagated as described by Gluzman (1981). Cells were subcultured at a dilution of 1:4, 1 day before transfection. We prepared calcium phosphate-DNA coprecipitates as described by Chu and Sharp (1981) using 10–20 μg of plasmid DNA per milliliter. We added 1 ml of precipitate to the 10 ml of media present in each 100 mm plate of cells, and allowed it to remain for 5 to 7 hr. The cells were shocked with 1 ml of 25% glycerol in medium (Frost and Williams, 1978) for 1 min at room temperature and were rinsed with 37°C medium, then fresh medium was added to the plate. This glycerol shock procedure increased α-globin RNA levels at least tenfold (data not shown). Cells were harvested 40 to 68 hr after transfection with essentially equivalent results. Although COS-7 cells were used routinely, COS-1 cells were equally effective for DNA replication. A derivative of COS-1 was equally effective for α-globin RNA production. Transfection precipitates formed with 5–25 μg of plasmid DNA per milliliter gave similar results; however, plasmid DNA concentrations below 1 μg/ml resulted in lowered amounts of α-globin RNA production.

RNA Purification and Analysis

Cytoplasmic RNA was prepared from COS cells as described by Favaloro et al. (1980). Poly(A) selection was carried out according to the procedure of Aviv and Leder (1972). Primer extension on α-globin mRNA templates was carried out as described by Proudfoot et al. (1980). S1 nuclease analysis of the 3′ end of α-globin and β-globin RNA was carried out according to the technique of Weaver and Weissman (1979). Approximately 100 ng of agarose-gel-purified, end-labeled DNA probe was annealed to the RNA sample at 51°C (β-globin probe) or 55°C (α-globin probe). S1 nuclease analysis with a uniformly labeled DNA probe was carried out at 51°C according to the procedure of Berk and Sharp (1977).

RNA blotting onto nitrocellulose was carried out with agarose-formaldehyde gels (B. Seed and D. Goldberg, personal communication; see for example Goldberg, 1980). The RNA used in blotting experiments was treated with RNAase-free DNAase in 100 μl of 10 mM Tris (pH 7.5), 10 mM $MgCl_2$ and 2 mm ribonucleotide-vanydyl complex, at 37°C for 30 min, extracted with phenol and ethanol-precipitated.

Acknowledgments

We are grateful to R. Myers and R. Tjian and to M. Lusky and M. Botchan for discussions and for providing plasmids prior to their publication. We thank W. Schaffner, R. Flavell, C. Weissman, P. Berg, S. Silverstein and D. Kurtz and their coworkers for communicating results prior to publication. We are indebted to R. Treisman and N. Proudfoot for discussions regarding their work on transcription of the β-globin gene in COS cells and to E. Leahy and S. Collins for excellent technical assistance. This work was funded by the National Institutes of Health. P. M. was supported by a National Institutes of Health postdoctoral fellowship, and V. P. was supported by a National Institutes of Health training grant.

We appreciate Monica Shander's contributions to the early stages of this work and express our sorrow for her untimely death.

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, 1981; revised September 24, 1981

References

Aviv, H. and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidilic acid-cellulose. Proc. Nat. Acad. Sci. USA *69*, 1408–1412.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Baralle, F. E. and Brownlee, G. G. (1978). AUG is the only recognizable signal sequence in the 5′ non-coding regions of eukaryotic mRNA. Nature *274*, 84–87.

Benoist, C. and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature *290*, 304–310.

Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980). The ovalbumin gene-sequence of putative control regions. Nucl. Acids Res. *8*, 127–142.

Berger, S. L. and Birkenmeier, C. S. (1979). Inhibition of intractable nucleases with ribonucleoside-vanadyl complexes: isolation of messenger ribonucleic acid from resting lymphocytes. Biochemistry *18*, 5143–5149.

Berk, A. J. and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell *12*, 721–732.

Brown, D. P. and Gurdon, J. B. (1977). High-fidelity transcription of 5S DNA injected into Xenopus oocytes. Proc. Nat. Acad. Sci. USA *74*, 2064–2068.

Chu, G. and Sharp, P. A. (1981). SV40 DNA transfection of cells in suspension: analysis of the efficiency of transcription and translation of T antigen. Gene *13*, 197–202.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P. and Uedinger, L. (1980). Promoter sequences of eukaryotic protein-coding genes. Science *209*, 1406–1414.

Dierks, P., Wieringa, B., Marti, D., Reiser, J., van Ooyen, A., Meyer,

F., Weber, H. and Weissman, C. (1981a). Expression of β-globin genes modified by restructuring and site-directed mutagenesis. ICN-UCLA Symposium, in press.

Dierks, P., van Ooyen, A., Mantei, N. and Weissman, C. (1981b). DNA sequences preceding the rabbit β-globin gene are required for formation in mouse L cells of β-globin RNA with the correct 5' terminus. Proc. Nat. Acad. Sci. USA *78*, 1411–1415.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. and Proudfoot, N. J. (1980). The structure and evolution of the human β-globin gene family. Cell *21*, 653–668.

Favaloro, J., Treisman, R. and Kamen, R. (1980). Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Meth. Enzymol. *65*, 718–749.

Frost, E. and Williams, J. (1978). Mapping temperature-sensitive and host-range mutations of adenovirus type 5 by marker rescue. Virology *91*, 39–50.

Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. and Weissman, S. M. (1978). Heterogeneity and 5' terminal structures of the late RNAs of simian virus 40. J. Mol. Biol. *126*, 813–846.

Ghosh, P. K., Lebowitz, P., Frisque, F. J. and Gluzman, Y. (1981). Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Nat. Acad. Sci. USA *78*, 100–104.

Goldberg, D. (1980). Isolation and partial characterization of the Drosophila alcohol dehydrogenase gene. Proc. Nat. Acad. Sci. USA *77*, 5794–5798.

Goldberg, M., (1979). Sequence Analysis of Drosophila Histone Genes. Ph.D. thesis, Stanford University, Stanford, California.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell *23*, 175–182.

Gluzman, Y., Frisque, R. and Sambrook, J. (1980a). Origin-defective mutants of SV40. Cold Spring Harbor Symp. Quant. Biol. *44*, 293–300.

Gluzman, Y., Sambrook, J. and Frisque, R. (1980b). Expression of early genes of origin-defective mutants of simian virus 40. Proc. Nat. Acad. Sci. USA *77*, 3898–3902.

Grosschedl, R. and Birnstiel, M. L. (1980a). Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Nat. Acad. Sci. *77*, 1432–1436.

Grosschedl, R. and Birnstiel, M. L. (1980b). Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequences are essential for promotion of H2A histone gene transcription in vivo. Proc. Nat. Acad. Sci. USA *77*, 7102–7106.

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

Gruss, P. and Khoury, G. (1981). Expression of simian virus 40-rat preproinsulin recombinants in monkey kidney cells: use of preproinsulin RNA processing signals. Proc. Nat. Acad. Sci. USA *78*, 133–137.

Gruss, P., Dhar, R., Khoury, G. (1981). Simian virus 40 tandem repeated sequences as an element of the early promoter. Proc. Nat. Acad. Sci. USA *78*, 943–947.

Hamer, D. and Leder, P. (1979). Expression of the chromosomal mouse β-maj-globin gene cloned in SV40. Nature *281*, 35–40.

Hamer, D. H., Kaehler, M. and Leder, P. (1980). A mouse globin gene promoter is functional in SV40. Cell *21*, 697–708.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. *26*, 365–369.

Hu, S-L. and Manley, J. (1981). DNA sequence required for initiation of transcription in vitro from the major late promoter of adenovirus 2. Proc. Nat. Acad. Sci. USA *78*, 820–824.

Kressman, A., Clarkson, S., Telford, J. and Birnsteil, M. (1977). Transcription of Xenopus $tDNA_1^{met}$ and sea urchin histone DNA injected into the Xenopus oocyte nucleus. Cold Spring Harbor Symp. Quant. Biol. *42*, 1077–1082.

Kurtz, D. T. (1981). Rat α_{2u} globulin is encoded by a multigene family. J. Mol. Appl. Genet. *1*, 29–38.

Lauer, J., Shen, C.-K. J. and Maniatis, T. (1980). The chromosomal arrangement of human α-like globin genes: sequence homology and α-globin gene deletions. Cell *20*, 119–130.

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. and Maniatis, T. (1978). The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. Cell *15*, 1157–1174.

Lawn, R. M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980). The nucleotide sequence of the human β-globin gene. Cell *21*, 647–651.

Lewin, B. (1980). Gene Expression, *2* (New York: John Wiley and Sons).

Liebhaber, S., Goossens, M., Kan, Y. W. (1980). Cloning and complete nucleotide sequence of human 5'-α-globin gene. Proc. Nat. Acad. Sci. USA *77*, 7054–7058.

Lusky, M. and Botchan, M. (1981). Inhibition of SV40 replication in simian cells by specific pBR322 DNA sequences. Nature *293*, 79–81.

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

Maniatis, T., Jeffrey, A. and van Sande, H. (1975b). Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. Biochemistry *14*, 3787–3794.

Manley, J. L., Fire, A., Lang, A., Sharp, P. A. and Gefter, M. L. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Nat. Acad. Sci. USA *77*, 3855–3859.

Mantei, N., Boll, W. and Weissman, C. (1979). Rabbit β-globin mRNA production in mouse L cells transformed with cloned rabbit β-globin chromosomal DNA. Nature *281*, 40–46.

Mathis, D. and Chambon, P. (1981). The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature *290*, 310–315.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

Maxwell, I. H., Maxwell, F. and Hahn, W. E. (1977). Removal of RNase activity from DNase by affinity chromatography on agarose-coupled aminophenyl phosphoryl-uridine-2' (3')-phosphate. Nucl. Acids Res. *4*, 241–246.

McKnight, S. L. (1980). The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucl. Acids Res. *8*, 5949–5964.

McKnight, S. L., Gavis, E. R., Kingsbury, R. and Axel, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell *25*, 385–398.

Mulligan, R. C. and Berg, P. (1980). Expression of a bacterial gene in mammalian cells. Science *209*, 1422–1427.

Mulligan, R. C., Howard, B. H. and Berg, P. (1979). Synthesis of rabbit β-globin in cultured monkey kidney cells following infection with a SV40 β-globin recombinant genome. Nature *277*, 108–114.

Myers, R. M., and Tjian, R. (1980). Construction and analysis of simian virus 40 origins defective in tumor antigen binding and DNA replication. Proc. Nat. Acad. Sci. USA *77*, 6491–6495.

Myers, R. M., Rio, D. C., Robbins, A. K. and Tjian, R. (1981). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. Cell *25*, 373–384.

Paul, J., Affara, N., Conkie, P., Gilmour, R., Harrison, P., MacGilliuray, A. Malcom, S., Windass, J. (1977) Transcriptional control in higher eukaryotes. In Molecular Biology of the Mammalian Genetic Apparatus, 2. P. Ts'o, ed. (Amsterdam: North Holland Publishing Co.), pp. 135–148.

Peden, K. W. C., Pipas, J. M., Pearson-White, S. and Nathans, D. (1980). Isolation of mutants of an animal virus in bacteria. Science *209*, 1392–1396.

Pellicer, A., Robins, D., Wold, B., Sweet, R., Jackson, J., Lowy, I., Roberts, J. M., Sim, G-K., Silverstein, S. and Axel, R. (1980). Altering genotype and phenotype by DNA-mediated gene transfer. Science *209*, 1414–1421.

Proudfoot, N. J. and Longley, J. I. (1976). The 3′ terminal sequences of human α and β globin messenger RNAs: comparison with rabbit globin messenger RNA. Cell *9*, 733–746.

Proudfoot, N. J., Shander, M. H. M., Manley, J. L., Gefter, M. L. and Maniatis, T. (1980). Structure and in vitro transcription of human globin genes. Science *209*, 1329–1335.

Rio, D., Robbin, A., Myers, R. and Tjian, R. (1980). Regulation of simian virus 40 early transcription in vitro by a purified tumor antigen. Proc. Nat. Acad. Sci. USA *77*, 5706–5710.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Tooze, J., ed. (1980) Molecular Biology of Tumor Viruses, Part II: DNA Tumor Viruses. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teole, R. and Chambon, P. (1980). Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in T-A-T-A box homology sequence. Proc. Nat. Acad. Sci. USA *77*, 7024–7028.

Weaver, R. F. and Weissman, C. (1979). Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucl. Acids Res. *6*, 1175–1193.

Weil, P. A., Luse, D. S., Segall, J. and Roeder, R. G. (1979). Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell *18*, 469–484.

Wiegand, R. C., Godson, G. N. and Radding, C. M. (1975). Specificity of the S1 nuclease from aspergillus oryzae. J. Biol. Chem. *250*, 8848–8855.

Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. and Axel, R. (1979). Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell *16*, 777–785.

Wilson, J. T., Wilson, L. B., deRiel, J. K., Villa-Komaroff, L., Efstratiadis, A., Forget, B. G. and Weissman, S. M. (1978). Insertion of synthetic copies of human globin genes into bacterial plasmids. Nucl. Acids. Res. *5*, 563–581.

Wold, B., Wigler, M., Lacy, L., Maniatis, T., Silverstein, S. and Axel, R. (1979). Introduction and expression of a rabbit β-globin gene in mouse fibroblasts. Proc. Nat. Acad. Sci. USA *76*, 5688–5694.

Chapter 2

Drosophila melanogaster has only one myosin alkali light-chain gene which encodes a protein with considerable amino acid sequence homology to chicken myosin alkali light chain.

(Reprinted from Molecular and Cellular Biology 4: 956-965.)

(1984)

Drosophila melanogaster Has Only One Myosin Alkali Light-Chain Gene Which Encodes a Protein with Considerable Amino Acid Sequence Homology to Chicken Myosin Alkali Light Chains

SCOTT FALKENTHAL,[1]†* VANN P. PARKER,[2]‡ WILLIAM W. MATTOX,[2] AND NORMAN DAVIDSON[1]

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

Received 9 January 1984/Accepted 27 February 1984

A chimeric lambda DNA molecule containing the myosin alkali light-chain gene of *Drosophila melanogaster* was isolated. The encoded amino acid sequence was determined from the nucleic acid sequence of a cDNA homologous to the genomic clone. The identity of the encoded protein was established by two criteria: (i) sequence homology with the chicken alkali light-chain proteins and (ii) comparison of the two-dimensional gel electrophoretic pattern of the peptides synthesized by in vitro translation of hybrid-selected RNA to that of myosin alkali light-chain peptides extracted from *Drosophila* myofibrils. There is only one myosin alkali light-chain gene in *D. melanogaster*; its chromosomal location is region 98B. This gene is abundantly expressed during the development of larval as well as adult muscles. The *Drosophila* protein appears to contain one putative divalent cation-binding domain (an EF hand) as compared with the three EF hands present in chicken alkali light chains.

Myosin light chains are proteins which occur abundantly and in a defined stoichiometry in myofibrils. They are members of an evolutionarily related group of calcium-binding proteins known as the troponin C superfamily, which includes calmodulin, troponin C, and the myosin alkali and 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB) light chains. The primary amino acid sequence has been determined for at least one vertebrate example of each of these polypeptides (2). The principal sequence homology between these proteins resides in the putative Ca^{2+}-binding domains, which are known as EF hands (14). The roles of all of these proteins, except for the myosin alkali light chain, in muscle function have been determined (10, 13, 29; R. A. Murphy, M. O. Askoy, P. F. Dillon, W. T. Gerthoffer, and K. E. Kanim, Fed. Proc. **42:**51–57, 1983).

The skeletal muscle myosin alkali light chains are so named because of the high pH required to dissociate them from the myosin heavy chain (39). For vertebrate muscles, they are sometimes called MLC-1 and MLC-3. The two skeletal muscle alkali light chains of mammals and chickens, which have molecular weights of about 21,000 (MLC-1) and 17,000 (MLC-3), are virtually identical in sequence over their C-terminal 141 residues, but diverge in sequence at the amino terminus. MLC-1, depending upon the tissue from which it is isolated, has an additional alanine-proline- or alanine-lysine-rich sequence of 40 amino acids at its amino terminus. There is evidence that in rats the two proteins are encoded by a single gene (L. Garfinkel, R. Gubits, B. Nadal-Ginard, and N. Davidson, manuscript in preparation). At one time, these peptides were thought to be essential for the actin-activated adenosine triphosphatase activity of myosin (16, 32, 38), but recent in vitro studies have suggested that such is not the case (36). Thus, the function of the alkali light chains is unknown.

Drosophila melanogaster is an advantageous organism for further studies of structure and function of gene products such as the myosin light chains because one may use both molecular and genetic approaches. Here, we describe our initial steps in isolating and identifying the myosin alkali light-chain gene of *D. melanogaster* and in determining from the nucleic acid sequence the primary structure of the protein. We also report on the reiteration frequency and developmental expression of this gene.

MATERIALS AND METHODS

Isolation of RNA. Total cellular RNA was prepared by homogenizing developmental-stage whole animals in 4 M guanidine thiocyanate–1 M 2-mercaptoethanol–0.05 M sodium acetate–0.001 M EDTA (pH 6.0) and banding in cesium chloride (11, 35). Polyadenylate-containing [poly(A)$^+$] RNA was selected by oligodeoxythymidylate-cellulose (type T3; Collaborative Research) chromatography as described by Anderson and Lengyel (1).

Isolation and purification of DNA. Charon phage DNA was isolated as described by Yen and Davidson (41) with the modifications described by Snyder et al. (33). Plasmid DNA was isolated as described by Fyrberg et al. (11). High-molecular-weight pupal DNA was prepared from 40- to 60-h pupae as follows. Two grams of pupae frozen in liquid nitrogen were ground to a fine powder in a mortar at −70°C. The frozen powder was homogenized by 10 to 15 strokes with a B pestle Dounce homogenizer in 30 ml of homogenization buffer (50 mM Tris, pH 7.2, 25 mM KCl, 5 mM $MgCl_2$, 350 mM sucrose, 0.15 mM spermine, 0.15 mM spermidine) and then filtered through Nitex cloth to remove cuticular debris. The nuclei were collected by centrifugation in a Sorvall HB-4 rotor at 4°C for 15 min at 3,000 rpm. The pelleted nuclei were washed with 30 ml of homogenization buffer and centrifuged to reduce the mitochondrial contamination of the nuclear pellet. The washed and pelleted nuclei were then suspended in 1.0 ml of nuclear suspension buffer (60 mM NaCl, 10 mM Tris, pH 7.2, 10 mM EDTA, 0.15 mM spermine, 0.15 mM spermidine). The nuclei were then lysed by the addition of 2.0 ml of lysis buffer (200 mM Tris, pH 8.5, 30 mM EDTA, 2% (wt/vol) Sarkosyl). Proteinase K was added to a concentration of 50 μg/ml, and the solution was gently mixed. After 2 to 4 h at 42°C, CsCl and ethidium

* Corresponding author.

† Present address: Department of Genetics, 927 Biological Sciences Building, The Ohio State University, Columbus, OH 43210.

‡ Present address: Amgen, Newberry Park, CA 91320.

bromide were added to concentrations of 0.925 g/ml and 500 μg/ml, respectively. The solution was spun in a table-top clinical centrifuge at top speed for 20 min to float the denatured protein. The cleared solution was centrifuged for 18 h at 53,000 rpm (20°C) in a Beckman VTi65 rotor. The banded genomic DNA was visualized by UV irradiation, and the band was collected by side puncture. The ethidium bromide was removed by butanol extraction. The sample was diluted to 0.5 ml and dialyzed twice against 5,000 volumes of 10 mM Tris (pH 8.0)–1 mM EDTA. The salt concentration was adjusted to 0.15 M with sodium acetate, and the DNA was precipitated by the addition of 2.0 volumes of absolute ethanol. After 1 h at −20°C, the DNA was collected by centrifugation, rinsed with 70% ethanol, and air dried at 4°C for 24 h. The DNA was resuspended by overlaying the pellet with 10 mM Tris (pH 8.0)–1 mM EDTA and allowing it to hydrate for 48 to 72 h at 4°C.

Screening libraries. A *Drosophila* genomic DNA library prepared by J. Lauer was used. It consisted of randomly sheared Canton S embryonic DNA of 12 to 20 kilobases (kb) inserted into Charon 4 via synthetic *Eco*RI linkers (17). The library was screened by the high-density plaque hybridization technique of Benton and Davis (4).

Subcloning fragments of genomic clones. DNA fragments of chimeric lambda clone λdmpT102 were obtained by digestion with *Hin*dIII and *Eco*RI. These fragments were ligated into plasmid pBR322 as described previously (41).

DNA labeling and hybridizations. ^{32}P-labeled pupal cDNA probes were prepared from pupal poly(A)$^+$ RNA by using oligo(dT) primers (P-L Biochemicals) by the method of Mullins et al. (24). Preparation of nick-translated probes and hybridization of ^{32}P-labeled DNA probes to filter-bound DNA were performed essentially as described by Mullins et al. (24). Nick-translated probes were hybridized at a concentration of 10^5 dpm/ml (10^8 dpm/μg), whereas the labeled cDNA probes were hybridized at a concentration of 1×10^6 to 2×10^6 dpm/ml (10^7 dpm/μg).

Modified Okayama and Berg technique for cDNA library preparation. The goal of this approach was to make a cDNA library with efficient utilization of poly(A)$^+$ RNA and with long inserts which contained the entire 3′ untranslated region, the protein-coding region, and most of the 5′ untranslated region, following the general spirit of the method of Okayama and Berg (28). Their procedure can be readily adapted and indeed simplified by using a modern cloning vector with a suitable polylinker. The first version of our procedure is depicted in Fig. 1. In this proposed procedure, intermediate 4 is treated with dGTP and terminal transferase to give intermediate 5 and to provide a G tail for oligodeoxycytidylate priming of second-strand synthesis from intermediate 6. In this procedure, we used Mn^{2+} ion as a catalyst based on the recommendation of Deng and Wu (9). If successful, this procedure would have regenerated the *Eco*RI site upon final ligation. In fact, none of the resulting clones examined at the sequence level had a GC oligonucleotide segment or the *Eco*RI site. However, this procedure has worked well in the hands of other investigators (F. K. Lin, personal communication). We presume that our lack of success in this step was due to a failure of the terminal transferase tailing reaction. We believe that an alternative and equally effective procedure would be to carry out the replacement synthesis directly on intermediate 4. This procedure, which is analogous to that used by Charles Rice (personal communication), relies on RNase H action to generate the primers for second-strand synthesis. Several of our clones for other genes have been shown to include up to ca. 10 nucleotides from the 5′ end of the mRNA. Thus a 5′-terminal primer is hardly necessary. This revised procedure would result in converting the *Bam*HI site into a blunt end. The final ligation would thus preserve the *Eco*RI site.

It should be noted that with the pUC8 vector there is an upstream *lac* promoter and translation start system. Thus, the resulting cDNA clones are useful for expression of the encoded proteins in *E. coli*.

pUC8 DNA (10 μg) was digested with 50 U of *Pst*I in 40 μl containing 100 mM Tris-hydrochloride (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, and 100 μg of bovine serum albumin (Pentex) per ml. After 60 min the solution was heated to 65°C for 15 min and then quickly cooled. Poly $(T)_{50-80}$ tails were added to the *Pst*I-digested DNA with terminal transferase. The reaction solution (200 μl) contained the following: 10 μg of *Pst*I-cut pUC8 DNA, 0.25 mM unlabeled TTP, 50 μCi of [^{3}H]TTP (77.8 Ci/mmol), 0.2 M cacodylate, 0.05 M Tris (pH 7.0), 1 mM $CoCl_2$. This was incubated for 10 min at 37°C, after which dithiothreitol and terminal transferase were added to concentrations of 0.1 mM and 110 U/ml, respectively. After 20 min the nucleic acid was precipitated by the addition of 2 volumes of absolute ethanol. The precipitated nucleic acid was collected by centrifugation, washed once with 90% ethanol, and dried under vacuum.

At this point, the vector has polythymidylate tails at each end of the molecule. To act as a good primer there must be only one tail per molecule. One tail is eliminated by digesting the molecule with *Bam*HI, whose recognition site is 12 nucleotides 5′ to the T-tailed *Pst*I site. The small resulting oligonucleotide was separated from the tailed vector by gel filtration on Sephadex G-150.

The T-tailed DNA was resuspended in 200 μl of 100 mM Tris (pH 7.5)–50 mM KCl–10 mM $MgCl_2$–100 μg of BSA per ml. *Bam*HI (10 U) was added, and the solution was incubated at 37°C. After 1 h, EDTA was added to a concentration of 50 mM, and the solution was layered onto a 5.0-ml Sephadex G-150 column (0.5 by 28 cm) equilibrated with 50 mM NaCl–10 mM Tris (pH 8.0)–10 mM EDTA. Fractions containing the excluded peak were pooled, and the nucleic acid was recovered by ethanol precipitation and centrifugation.

First-strand cDNA synthesis was performed in a volume of 100 μl containing 2 μg of poly(T)-tailed pUC8 DNA, 10 μg of pupal poly(A)$^+$ RNA, 50 mM Tris (pH 8.3), 50 mM KCl, 8 mM $MgCl_2$, 4 mM sodium PP_i, 2 mM DTT, 2 mM TTP, 2 mM dCTP, 2 mM dATP, 2 mM dGTP, and 5 μCi of [^{32}P]dCTP (410 Ci/mmol). Reverse transcriptase (54 U) was added, and the solution was incubated at 46°C for 45 min. After phenol extraction, the aqueous phase was layered onto a 5.0-ml Sephadex G-150 column (0.5 by 28 cm) equilibrated in 50 mM NaCl–10 mM Tris (pH 8.0)–10 mM EDTA. The excluded peak fractions were pooled, and the nucleic acid was recovered by ethanol precipitation and centrifugation. The second-strand replacement synthesis was done with RNase H and DNA polymerase 1 as described by Okayama and Berg (28). The resulting double-stranded DNA molecules were circularized by blunt end ligation as follows. The DNA was ethanol precipitated after second-strand synthesis, collected by centrifugation, and resuspended in 0.5 ml of a solution containing 50 mM Tris (pH 7.8), 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP, and 50 μg of BSA per ml. Incubation with 2,000 U of T4 DNA ligase proceeded for 12 h at 12°C. Bacterial cells (*Escherichia coli* MC1061) were transformed, and 5% of the transformation mix was plated onto L plates containing 50 μg of ampicillin per ml, whereas the remainder of the library was amplified for 3 h in liquid culture and then frozen in 50% glycerol at −40°C. Our library contained

200,000 independent recombinants from 2 μg of starting vector DNA.

Purification of *Drosophila* myosin. *Drosophila* indirect flight muscle myofibrils were purified from 24- to 48-h adults essentially as described by Mogami et al. (23). Myosin was prepared by high-salt extraction of isolated myofibrils as described by Whalen et al. (40). EDTA was present at a concentration of 10 mM during the isolation of myofibrils and extraction of actomyosin, which probably resulted in the dissociation of DTNB light-chain peptides from the myosin (V. P. Parker, S. Falkenthal, and N. Davidson, manuscript in preparation).

Electrophoresis of RNA in formaldehyde gels. Electrophoresis of RNA in formaldehyde gels and transfer to nitrocellulose paper was performed as described by Rozek and Davidson (31). Prehybridization, hybridization with 10% dextran sulfate, and the washing of filters after hybridization were done as described by Mullins et al. (25).

Positive selection and translation of RNA. RNA was selected by a procedure similar to that described by Ricciardi et al. (30). Bacteriophage DNA (10 μg) was denatured by heating and applied to nitrocellulose filters (0.3 by 0.3 cm) in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 sodium citrate). Up to 15 filters containing different cloned *Drosophila* DNA sequences were prehybridized in 200 μl of 70% formamide–0.4 M NaCl–0.1 M PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)] (pH 6.5) for 2 h at 50°C. Hybridizations were done for 4 h at 50°C in 100 μl of the same buffer containing 100 to 150 μg of pupal poly(A)$^+$ RNA (70 to 75 h postpupariation). After hybridization, the filters were washed en masse 10 times in 5 ml of 1× SSC–0.5% sodium dodecyl sulfate at 65°C, then three times in 5 ml of 0.01 M Tris (pH 7.8)–1 mM EDTA at room temperature. The filters were then placed in individual vials, and the hybridized RNA was eluted in boiling distilled water and recovered by ethanol precipitation. Translation was in a commercial (Bethesda

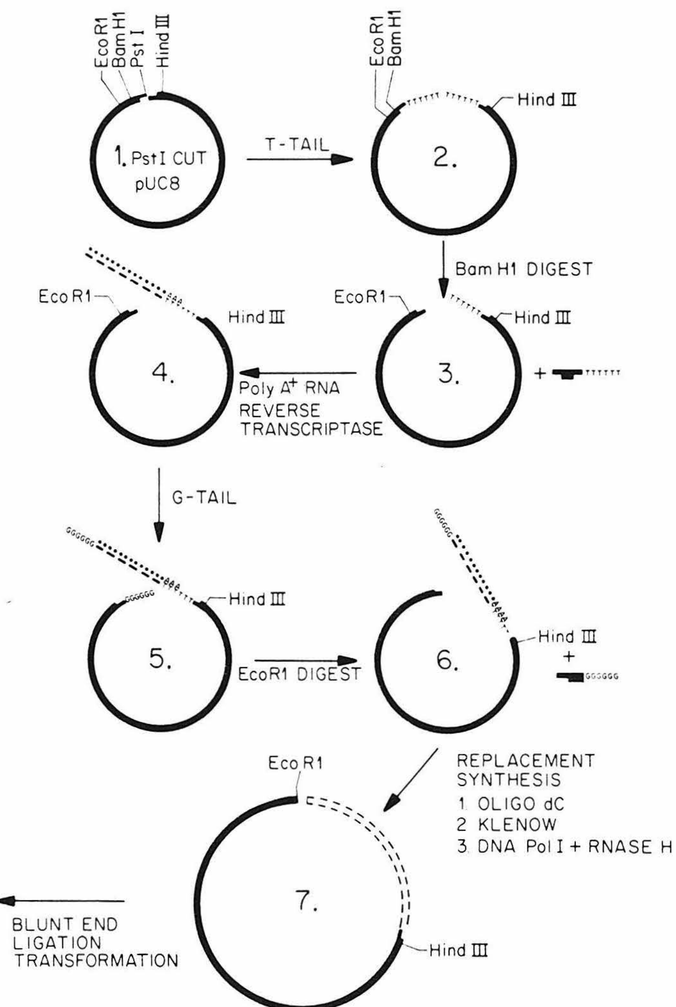


FIG. 1. Diagrammatic representation of the cDNA cloning strategy.

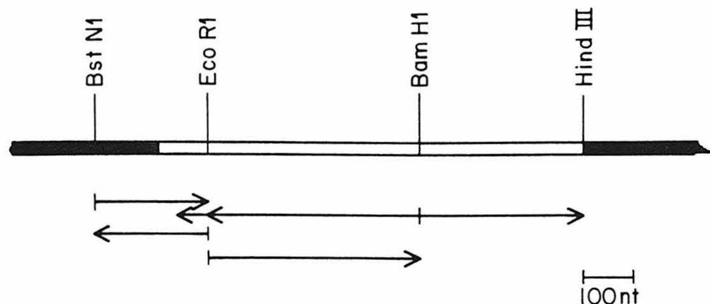


FIG. 2. Sequencing strategy of cDNA clone pcMLC-1.10. The DNA sequencing strategy is such that a vertical line (|) denotes ^{32}P labeling at the 3′ terminus. The arrows indicate the direction and extent of sequencing from the labeled terminus.

Research Laboratories) micrococcal nuclease-treated rabbit reticulocyte translation system. For each 30-μl translation, 50 μCi of 800-Ci/mmol [^{35}S]methionine (New England Nuclear) was used. Translation was terminated after 60 min by the addition of 50 ng each of RNase A and DNase I and subsequent incubation at 4°C for 30 min.

One-third of each translation assay was analyzed on a two-dimensional polyacrylamide gel as described by O'Farrell (27). The lysis buffer was modified to contain 0.1% sodium dodecyl sulfate (6). Molecular weights were determined by electrophoresing ^{14}C-labeled protein standards (Bethesda Research Laboratories) in an adjacent slot. Radioactively labeled proteins were detected by fluorography by the method of Laskey and Mills (15). Dried gels were exposed to preflashed Kodak XAR X-ray film.

DNA sequencing. The DNA sequence of a cDNA clone was determined by the method of Maxam and Gilbert (20) with modifications as described by Snyder et al. (34). The sequencing strategy used is shown in Fig. 2. Regions which were sequenced for one strand only were confirmed by a comparison with the DNA sequence of the chromosomal gene (S. Falkenthal, V. P. Parker, and N. Davidson, manuscript in preparation).

Amino acid homology comparisons. One homology comparison was carried out with a standard dot matrix program that asks for matches of five out of eight contiguous amino acids, but accepts conservative amino acid replacements (Val, Ileu, and Leu; Arg and Lys; Gln and Asn; Glu and Asp) as a match. In practice, this provides a rather stringent comparison of amino acid sequence homology (T. Hunkapiller, personal communication). The second two-dimensional matrix analysis, denoted the "best-fit" program, is useful for comparisons of sequences which are more distantly related. A match of the central amino acid is given a score of one, and matches at increasing distances from the center are given a reduced added score [M. Hunkapiller, S. Kent, M. Carruthers, W. Dreyer, J. Firca, C. Griffin, S. Horvath, T. Hunkapiller, P. Tempst, and L. Hood, Nature (London), in press].

RESULTS

Isolation of genes abundantly expressed in the indirect flight muscle. Our approach for isolating the myosin light-chain genes was based on the assumption that the concentration of their mRNAs would be high during the developmental stage when maximal synthesis of adult musculature occurs. Protein labeling studies showed that this occurred 70 to 75 h after puparium formation (data not shown). In vitro translation of RNA isolated from dissected thoraces at this developmental stage revealed that greater than 50% of the incorporation of [^{35}S]methionine was into myofibrillar proteins (data not shown). Accordingly, cDNA synthesized from pupal thoracic poly(A)$^+$ RNA was used to screen a Canton S random shear library of *Drosophila* genomic DNA in the vector Charon 4 (17). The 73 positive phages so selected were all plaque purified. These isolates were counterscreened with [^{32}P]cDNA homologous to early embryo RNA, a developmental time in which muscle-specific genes are not expressed, and with actin and myosin heavy-chain probes. The isolates which screened negatively with the above hybridization probes were rescreened with [^{32}P]cDNA synthesized from RNA extracted from the dorsal-lateral indirect flight muscle of late pupae (70 to 75 h). By restriction mapping, the resulting positive clones fell into 24 groups of nonoverlapping DNA inserts.

Initial characterization of these clones involved cytological localization of the DNA inserts by in situ hybridization to salivary gland chromosomes (Table 1). It may be seen that these inserts, representing genes abundantly expressed in the indirect flight muscle, do not show a general pattern of clustering on the *Drosophila* chromosomes. Tight clustering, as has been demonstrated for the cuticle genes on the second

TABLE 1. In situ localization of λdmpT recombinant clones

Clone	Band
31[a]	100B
49	28C
50[a]	97A
57[a]	99E
61[a]	17A
63	36B
73	88F
75	30B
85[a]	64C
101	102EF
102	98B
104	30EF
106	Repeated
115	72DE
116	53F
120	64F
121[a]	66F
123	94E

[a] Clone blots and in vitro translation of hybrid-selected RNA indicated that more than one gene is contained within the *Drosophila* insert (data not shown).

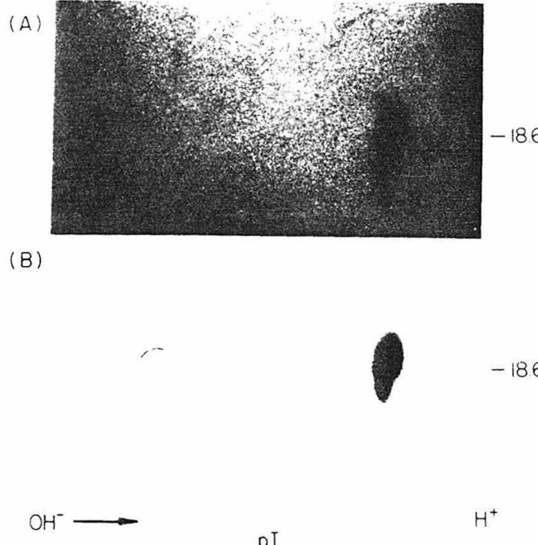


FIG. 3. Translation of myosin light-chain mRNA in a rabbit reticulocyte lysate. Myosin light-chain mRNA was selected by hybridization to filter bound λdmpT102 DNA as described in the text. This RNA was translated in an mRNA-dependent rabbit reticulocyte lysate containing [^{35}S]methionine, and the translation products were coelectrophoresed with 5 μg of *Drosophila* myosin purified from adult myofibrils on a 16% polyacrylamide gel. (A) Coomassie brilliant blue-stained gel. (B) Autoradiographic exposure (36 h) of the gel in A.

chromosome (33, 34) and the glue protein genes on the third chromosome (21), is not excluded by this analysis. Clone blots with labeled pupal cDNA as a hybridization probe and in vitro translation of hybrid-selected RNA revealed that 11 of the inserts contained only one gene which was expressed during pupal myogenesis, whereas the other inserts contained two or possibly three separate but closely linked transcription units (data not shown). Therefore, there is some tight clustering of genes abundantly expressed in the indirect flight muscle. Only one clone, λdmpT73, which probably contains a tropomyosin gene (3) hybridized in situ to the same region, 88F, where an actin gene expressed in the indirect flight muscle maps (12). Mutations which result in dominant flightless behavior have been found at regions 36B, where the myosin heavy-chain gene maps, and 88F (5, 23). None of our other clones mapped in these two regions.

Identification of the clone which contained the myosin light-chain sequence. The DNA insert which most probably encodes a myosin light-chain gene was identified by hybrid selection of RNA and in vitro translation (see above). The resulting ^{35}S-labeled polypeptides were compared to purified *Drosophila* myosin light-chain protein by two-dimensional gel electrophoretic analysis. The in vitro translation products of two of these clones (λdmpT75 and λdmpT102) had molecular weights of 17,000 to 20,000, the molecular weight range expected for the myosin light-chain protein. Whereas the polypeptide encoded by λdmpT75 had a more basic isoelectric point (data not shown), that encoded by λdmpT102 had the identical electrophoretic mobility of the extracted myosin light-chain protein (Fig. 3). The molecular weight of the latter protein(s) ranged from 18,000 to 19,000. Note that there is heterogeneity in both the molecular weight and the isoelectric point of the myosin light-chain isolated from adult muscle as well as for the in vitro-synthesized translation products.

Isolation and sequencing of a myosin light-chain cDNA clone. To achieve a more positive identification of the protein encoded by the insert of λdmpT102, a cDNA clone for which it codes was isolated and sequenced. Initially, the coding region of λdmpT102 was localized onto 3.2 kb of DNA by clone blots with [^{32}P]cDNA from total pupal poly(A)$^{+}$ RNA as a hybridization probe. This region contains two contiguous restriction fragments, a 1.8-kb *Hin*dIII-*Eco*RI fragment and a 1.35-kb *Eco*RI fragment. These two fragments were subcloned (see Fig. 4 for relevant restriction maps) and used to screen a cDNA library.

The cDNA library was prepared from late pupal poly(A)$^{+}$ RNA by a simplified procedure modeled on that of Okayama and Berg (28) to increase the probability of obtaining long inserts. Of 10,000 recombinants screened (5% of the total library), 10 clones containing sequences homologous to the insert of p102.6 were obtained. The clone, pcMLC-1.10, which contained the longest insert (880 nucleotides), was

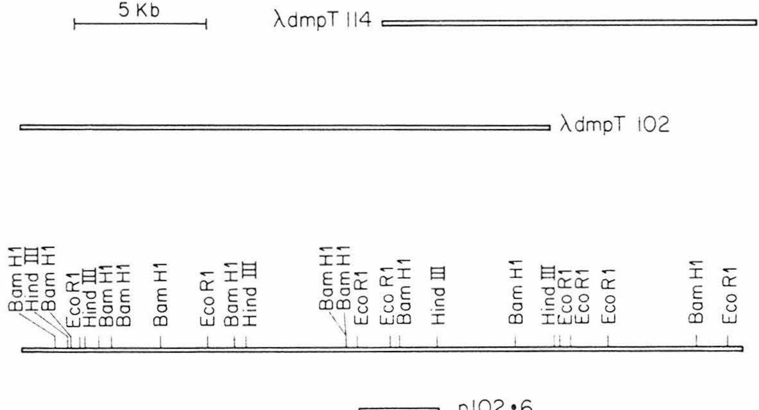


FIG. 4. Restriction endonuclease map of λdmpT102 and λdmpT114. A composite map of the two bacteriophage lambda clone inserts is shown. The subcloned coding region is shown below the map.

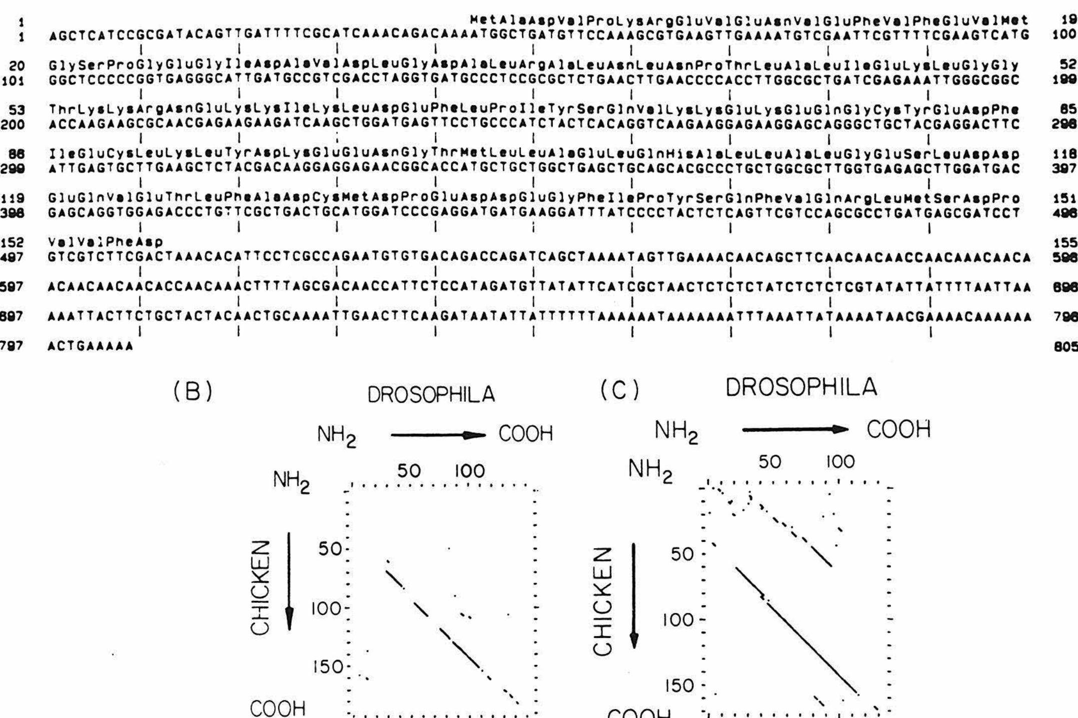


FIG. 5. Predicted protein coding sequence of the insert of pcMLC-1.10. (A) The amino acid sequence of the putative myosin light chain is depicted over the encoding DNA sequence of the 880-nucleotide insert of pcMLC-1.10. The DNA sequence was determined as described in the text. (B) Homology search between the predicted amino acid sequence of the *Drosophila* protein and the amino acid sequence of chicken skeletal muscle myosin light chain-1 protein. The comparison is such that every eight amino acids of the *Drosophila* protein are compared with the chicken protein (a positive score is marked if five out of eight amino acids match). A diagonal line indicates homology. (C) A best-fit comparison between the predicted amino acid sequence of the *Drosophila* protein and the chicken skeletal muscle myosin light chain-1 protein calculated by the method of Hunkapiller et al. (in press). The comparison is such that 10 amino acids of the Drosophila protein are compared with the chicken protein, allowing conservative amino acid replacements. The confidence limit for this fit was set at 99%.

sequenced (Fig. 5A). An AUG codon at nucleotide 44 begins a reading frame 456 nucleotides in length, which terminates with a TAA at nucleotide 509. Downstream from this termination codon are two additional in-frame termination signals (TAA and TGA at nucleotides 552 and 561, respectively).

The amino acid sequence of this reading frame was deduced and compared with the amino acid sequence of bovine brain calmodulin (37), rabbit troponin C (7), chicken skeletal muscle myosin light chain-1 (MLC-1) (19), and chicken skeletal muscle myosin light chain-2 (MLC-2) (18). These evolutionarily related proteins all possess the divalent metal binding (Ca^{2+}, Mg^{2+}) structure referred to as the EF hand (14).

By the stringent dot matrix amino acid homology comparison described above, only the chicken skeletal muscle myosin alkali light chains (MLC-1 and MLC-3) gave a significant positive score with the *Drosophila* protein (Fig. 5B). This fact and the in vitro translation study described above are the primary bases on which we identified the gene as that for the *D. melanogaster* alkali light chain.

In all regions where the dot matrix method showed homology between the *Drosophila* protein and the chicken MLC-1 protein, the latter is identical in sequence with the chicken MLC-3 protein (see Fig. 8 and below). The *Drosophila* protein has approximately the same molecular length as does the chicken MLC-3 protein, without the amino-terminal tail of the MLC-1 protein. A more detailed comparison of these sequences is presented later. We note here simply that overall the region of homology extends from amino acids 32 to 144 of the *Drosophila* protein with amino acids 72 to 167 and 31 to 143 of chicken MLC-1 and MLC-3, respectively (see Fig. 8). On this basis, we identified this gene as the *Drosophila* myosin alkali light chain (MLC-ALK).

The *Drosophila* MLC-ALK and chicken MLC-1 sequences were compared by the best-fit program which is designed to identify more distant sequence relationships (Fig. 5C). This comparison confirmed the major segment of homology noted above. In addition, it revealed that the sequence from amino acids 90 to 105 of *D. melanogaster* is related to the amino terminus proximal sequence from amino acids 56 to 71 of chicken MLC-1.

Developmental expression of the MLC-ALK gene. The MLC-ALK gene was selected by using a probe isolated from the pupal stage of development. The question arose as to whether the MLC-ALK gene was transcribed at other times in development, particularly at other times of muscle biosynthesis. To answer this question, total poly(A)$^+$ RNA was

isolated from animals at various developmental stages, electrophoresed on denaturing formaldehyde agarose gels, transferred to nitrocellulose paper, and hybridized with nick-translated MLC-ALK-coding DNA (pl02.6). The results (Fig. 6) demonstrated that the MLC-ALK gene is transcribed during late embryogenesis (the time of formation of the larval musculature), during the larval instars (a time of rapid tissue, and therefore muscle, growth), and during the late stages of pupariation (the time of rapid synthesis of the indirect flight muscle and other adult muscles). There was no accumulation of MLC-ALK transcripts during early pupal development, the time of histolysis of larval musculature (8). The most abundant RNA showed a broad distribution with a mean molecular length of around 0.95 kb. There is a less intense band at 1.3 kb, which was most evident in Fig. 6 in the late pupal indirect flight muscle RNA lane, because of the greater fraction of muscle-specific RNA in this preparation. A 3.2-kb band which we believe is an unspliced nuclear transcript (unpublished data) is faintly visible in several lanes.

Reiteration frequency of the MLC-ALK gene. There are indications of multiple polypeptides in the in vitro translation products of RNA selected by the MLC-ALK gene we isolated (Fig. 3). Additionally, cross-hybridizing RNAs are expressed in several stages of development (Fig. 6). Similar results were observed for the actin genes of *D. melanogaster*, which comprise a small multigene family (11), and for the myosin heavy-chain gene which is single copy (5, 31). If the MLC-ALK gene is reiterated, all of the copies must reside at the chromosomal region 98B (Table 1), because this is the only site labeled by in situ hybridization.

We carried out genome blot studies to resolve this question. The gel blots were hybridized to the MLC-ALK probes

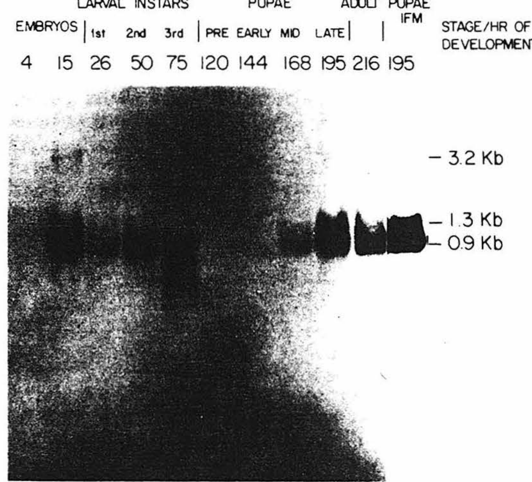


FIG. 6. Developmental expression of myosin alkali light chain mRNAs. Poly(A)$^+$ RNA (1 μg) from different developmental stages of synchronized populations of *D. melanogaster* was subjected to electrophoresis on 1.5% agarose gels containing 2.2 M formaldehyde. The RNA was transferred to nitrocellulose and hybridized with a nick-translated probe synthesized from p102.6 DNA. Time (in hours) was calculated from the time of egg deposition. Pupae were resynchronized by floatation on water at 5 h post-pupariation (22). *E. coli* 16S and 23S rRNAs as well as HeLa 28S and 18S rRNAs were used as length standards.

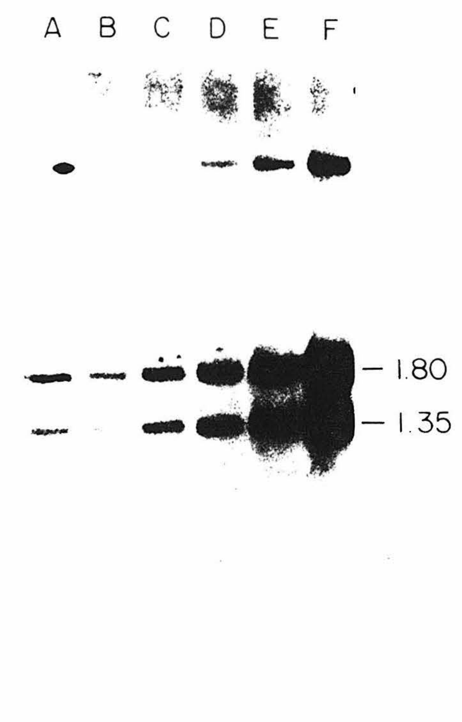


FIG. 7. Genomic representation of the coding region of λdmpT102. (A) Genomic *Drosophila* Canton S DNA (4 μg) was digested with *Eco*RI and *Hin*dIII, and the resulting fragments were separated on 1% agarose gels. In adjacent lanes of the same gels, amounts of *Eco*RI-*Hin*dIII-restricted λdmpT102 DNA equivalent to (B) 0.5, (C) 1.0, (D) 3.0, (E) 5.0, and (F) 25.0 copies per genome (a haploid genome size of 1.6×10^8 base pairs was assumed) plus 4 μg of *Eco*RI-*Hin*dIII-restricted *E. coli* DNA were subjected to electrophoresis. After electrophoresis, the fragments were transferred to nitrocellulose and hybridized with ^{32}P-labeled nick-translated p102.6 DNA.

at a moderate stringency. Quantitative comparisons of intensity were made by genome reconstruction experiments. All these data (Fig. 7) show that at the moderate stringency of the hybridization experiments there is only one MLC-ALK gene per haploid genome.

DISCUSSION

We cloned a *D. melanogaster* gene coding for a myosin alkali light-chain protein. The gene is expressed to provide a relatively high concentration of RNA during those specific times in development when extensive myofibrillar assembly occurs, and it is not expressed before larval myogenesis or during early pupariation when larval muscles are undergoing histolysis (8). We showed by in vitro translation of hybrid selected RNA that the gene encodes a polypeptide which comigrates in two-dimensional gels with myosin alkali light-chain protein extracted from adult indirect flight muscle. The strongest evidence that the gene encodes the MLC-ALK protein was obtained from the nucleotide sequence of a cDNA clone derived from the mRNA transcript of the gene.

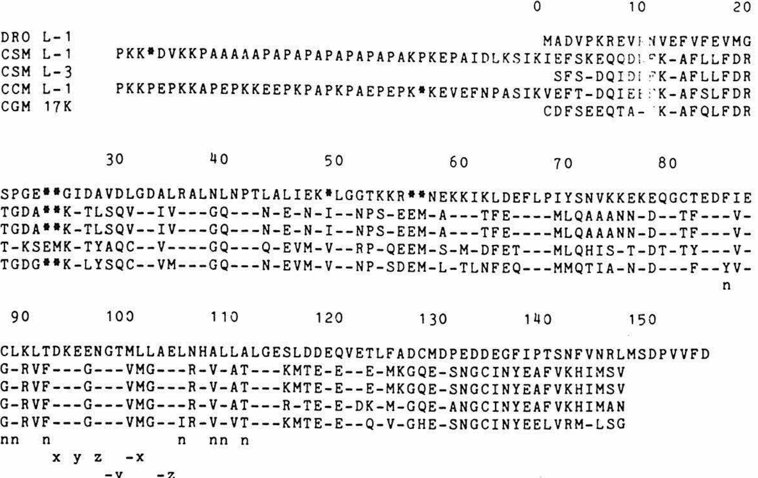


FIG. 8. Comparison of the amino acid sequence of *Drosophila* myosin light chain with the sequence of chicken myosin light chains (7, 19). Abbreviations: DRO L-1, *Drosophila* myosin alkali light chain; CSM L-1, chicken skeletal muscle myosin light chain-1; CSM L-3, chicken skeletal muscle myosin light chain-3; CCM L-1, chicken cardiac muscle myosin light chain-1; CGM 17K, chicken gizzard muscle 17,000-dalton myosin light chain. *Drosophila* amino acids identical to those of the chicken myosin light chains are indicated by a hyphen. The absence of a particular amino acid in the sequence is denoted by an asterisk; n indicates those residues which form the core of the E and F α helices. X, Y, Z, −X, −Y, and −Z are residues which might be involved in the binding of divalent metal ions in EF hand domain III.

The corresponding amino acid sequence showed a high degree of sequence homology with that of the chicken alkali light chains and much less homology with any other of the evolutionarily related calcium ion-binding proteins.

The consensus EF hand structural domain of the calcium-binding proteins consists of an α helix of 9 amino acids (the E helix), a loop of 12 amino acids (6 of which have side chains containing oxygen atoms which are capable of octahedral coordination to a Ca^{2+} ion), and an α helix of eight amino acids (the F helix) (14). This structure is denoted n(2), n(5), n(6), n(9), X(10), Y(12), Z(14), −Y(16), −X(18), −Z(21), n(22), n(25), n(26), n(28), n(29). The numbers denote the relative positions of residues, where n represents the hydrophobic or inner core of the α helix and X, Y, Z, −X, −Y, and −Z represent the calcium ligands. The amino acids at these positions are variable from protein to protein. To bind Ca^{2+}, there must be at least four aspartic and glutamic acid residues in the calcium loop, but others may be serine, threonine, asparagine, and glutamine. In some calcium-binding proteins, one of the six positions is occupied by glycine which is capable of coordinating Ca^{2+} through the oxygen of a hydrogen-bonded water molecule (14). The vertebrate myosin alkali light chains do not, in fact, bind Ca^{2+} in vitro (39). However, because of the very similar amino acid composition of their EF hand domains with those of other calcium-binding proteins, they are included in this group.

A comparison of the amino acid sequence of the *Drosophila* MLC-ALK gene with those of the highly homologous chicken skeletal, cardiac, and gizzard alkali light chains is shown in Fig. 8. The region of greatest homology between the chicken proteins and the *Drosophila* protein extends from amino acids 31 to 124 of the latter. There is an EF hand (denoted EF hand III) of the chicken proteins in a region of very high sequence homology with the *Drosophila* protein, amino acids 85 to 112. The chicken MLC-ALK proteins have four domains which are identified as resembling the EF hand domains of other calcium-binding proteins. However, due to an insertion of two amino acids in the Ca^{2+} loop of EF hand domain II, it is believed that this structure has been disrupted. Weeds et al. (39) note that only the EF hand in domain III has four acidic residues among its six ligating groups.

Sequence comparisons of aligned amino acids for chicken MLC-3 and the *Drosophila* protein are shown in Table 2. The highest degree of homology is in domain III, with a moderate degree of homology in domain II and very low homology in domains I and IV. Analysis of domain II did not show a good EF hand structure. Therefore, we believe that the only potential calcium-binding site of the *Drosophila* protein is that found in the region homologous to domain III of the chicken proteins.

The best-fit matrix analysis showed a definite but more distant relationship between domain III of the *Drosophila* protein and domain I of the chicken protein (Fig. 5C), calmodulin, and troponin C (data not shown). This is reasonable because the four Ca^{2+}-binding domains of calcium-binding proteins are thought to have arisen from an ancestral Ca^{2+} binding domain that underwent two gene duplication events followed by sequence divergence (14). Similar comparisons show sequence homology between this domain of the *Drosophila* protein with the EF hand III domains of bovine brain calmodulin and of rabbit skeletal muscle troponin C (data not shown). This EF hand domain contains the high-affinity Ca^{2+}–Mg^{2+}-binding site.

We conclude that the *Drosophila* alkali light chain contains one EF hand domain which is related in sequence to those of the family of calcium-binding proteins.

The *Drosophila* MLC-ALK gene is present at only one copy per haploid genome and is transcribed during larval and pupal myogenesis. In this respect, it is analogous to the myosin heavy-chain gene (5, 31). However, the latter shows developmental differences in transcript lengths and splicing patterns (31) which lead to developmental differences in amino acid sequences (C. Rozek, personal communication).

TABLE 2. Homology by domain of the chicken and *Drosophila* proteins

Domain (amino acids)	% Homology		% Protein homology allowing conservative amino acid changes	% DNA homology discounting third-base changes	% Nonhomology due only to third-base changes	Protein homology/DNA homology
	Protein[a]	DNA[b]				
I (1–32)	15	34	21	39	8	0.44
II (33–76)	46	52	50	59	15	0.88
III (77–115)	58	57	70	68	33	1.02
IV (116–157)	15	27	18	30	4	0.55

[a] Homology between the chicken skeletal muscle myosin light chain-3 and *Drosophila* myosin alkali light chain (Fig. 4).

[b] Homology between the *Drosophila* myosin light-chain sequence and that of chicken skeletal muscle myosin light chain-3 cDNA clone (26).

There are indications of heterogeneity in molecular weight and isoelectric point for the *Drosophila* MLC-ALK in vitro translation product and for this protein extracted from adult indirect flight muscle. This heterogeneity could be due to post-translational modification, such as phosphorylation. Alternatively, or in addition, there could be subtle differences in transcripts, thereby encoding proteins which differ slightly in amino acid composition, which are not revealed by the RNA gel blots. We have evidence that some of this heterogeneity is due to differential splicing generating transcripts encoding two proteins which differ at their carboxy-terminal ends. Studies of these and related questions dealing with the fine structure of the gene will be reported in a later communication (Falkenthal et al., in preparation).

ACKNOWLEDGMENTS

We thank E. Fyrberg and C. Rozek for their generous gifts of actin and myosin heavy-chain-coding DNAs and T. Hunkapiller for help in the execution of the protein homology comparisons. We also thank H. Mitchell and N. Peterson for guidance and direction in pupal dissection and sharing their results of pupal indirect flight muscle development.

S.F. was supported by a Muscular Dystrophy Postdoctoral Fellowship and NRSA Fellowship 1 F32 GM08895-01. The research was supported by Public Health Service grants from the National Institutes of Health.

LITERATURE CITED

1. **Anderson, K. V., and J. A. Lengyel.** 1979. Rates of synthesis of major classes of RNA in Drosophila embryos. Dev. Biol. **70**:217–231.
2. **Barker, W. C., L. K. Ketcham, and M. O. Dayhoff.** 1978. The troponin C superfamily, p. 273–284. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5. National Biomedical Research Foundation, Silver Spring, Md.
3. **Bautch, V. L., R. V. Storti, D. Mischke, and M. L. Pardue.** 1982. Organization and expression of Drosophila tropomyosin genes. J. Mol. Biol. **162**:231–250.
4. **Benton, W. D., and R. W. Davis.** 1978. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
5. **Bernstein, S. I., K. Mogami, J. J. Donady, and C. P. Emerson.** 1983. Drosophila muscle myosin heavy chain is encoded by a single gene located in a cluster of muscle mutations. Nature (London) **302**:393–397.
6. **Brandhorst, B. P.** 1976. Two-dimensional gel patterns of protein synthesis before and after fertilization of sea urchin eggs. Dev. Biol. **52**:310–317.
7. **Collins, J. H., J. D. Potter, M. J. Horn, G. Wilshire, and N. Jackman.** 1973. The amino acid sequence of rabbit skeletal muscle troponin C: gene replication and homology with calcium binding proteins from carp and hake muscle. FEBS Lett. **36**:268–272.
8. **Crossley, A. C.** 1978. The morphology and development of the Drosophila muscular system, p. 499–560. *In* M. Ashburner and T. R. F. Wright (ed.), The genetics and biology of Drosophila, vol. 2b. Academic Press, Inc., New York.
9. **Deng, G., and R. Wu.** 1981. An improved procedure for utilizing terminal transferase to add homopolymers to 3′ termini of DNA. Nucleic Acids Res. **9**:4173–4188.
10. **Ebashi, S.** 1963. Third component participating in the superprecipitation of "natural actomyosin." Nature (London) **200**:1010–1012.
11. **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.
12. **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.
13. **Klee, C. B., T. H. Crouch, and P. G. Richman.** 1980. Calmodulin. Annu. Rev. Biochem. **49**:489–515.
14. **Kretsinger, R. H.** 1980. Structure and evolution of calcium modulated proteins. Crit. Rev. Biochem. **8**:119–174.
15. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative fluorography of acrylamide gels. Eur. J. Biochem. **56**:335–341.
16. **Lowey, S., and D. Risby.** 1971. Light chain from fast and slow muscle myosins. Nature (London) **234**:81–85.
17. **Maniatis, T., R. C. Hardison, E. Lacy, Jr., J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**:687–702.
18. **Matsuda, G., T. Maita, Y. Kato, J. Chen, and T. Umegane.** 1981. Amino acid sequences of the cardiac L-2A, L-2B, and gizzard 17,000 M light chains of chicken muscle myosin. FEBS Lett. **135**:232–236.
19. **Matsuda, G., T. Maita, and T. Umegane.** 1981. The primary structure of L-1 light chain of chicken fast skeletal muscle myosin and its genetic implication. FEBS Lett. **126**:111–113.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **Meyerowitz, E. M., and D. S. Hogness.** 1982. Molecular organization of a Drosophila puff site that responds to ecdysone. Cell **28**:165–176.
22. **Mitchell, M. K., and A. Mitchell.** 1964. Mass culture and age selection in Drosophila. Dros. Inf. Ser. **39**:135–137.
23. **Mogami, K., S. C. Fujita, and Y. Hotta.** 1982. Identification of Drosophila indirect flight muscle myofibrillar proteins by means of two-dimensional electrophoresis. J. Biochem. **91**:643–650.
24. **Mullins, J. I., J. W. Casey, M. O. Nicolson, K. B. Burck, and N. Davidson.** 1981. Sequence arrangement and biological activity of cloned feline leukemia virus proviruses from a virus-productive human cell line. J. Virol. **38**:688–703.
25. **Mullins, J. I., J. Casey, M. Nicolson, and N. Davidson.** 1980. Sequence organization of feline leukemia virus DNA in infected cells. Nucleic Acids Res. **8**:3287–3305.
26. **Nabeshima, Y., Y. Fuju-Kuriyama, M. Muramatsu, and K. Ogata.** 1982. Molecular cloning and nucleotide sequences of the complementary DNAs to chicken skeletal muscle myosin two alkali light chain mRNAs. Nucleic Acids Res. **10**:6099–6110.
27. **O'Farrell, P. H.** 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**:4007–4021.
28. **Okayama, H., and P. Berg.** 1982. High efficiency cloning of full length cDNAs. Mol. Cell. Biol. **2**:161–170.
29. **Pemrick, S. M.** 1980. The phosphorylated L2 light chain of

skeletal myosin is a modifier of the actomyosin ATPase. J. Biol. Chem. **255**:8836–8841.

30. **Ricciardi, R. P., J. S. Miller, and B. E. Roberts.** 1979. Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation. Proc. Natl. Acad. Sci. U.S.A. **76**:4927–4931.
31. **Rozek, C. E., and N. Davidson.** 1983. Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell **32**:23–34.
32. **Sarkar, S., F. A. Streter, and J. Gergely.** 1971. Light chains of myosins from white, red, and cardiac muscles. Proc. Natl. Acad. Sci. U.S.A. **68**:946–950.
33. **Snyder, M., J. Hirsh, and N. Davidson.** 1981. The cuticle genes of Drosophila: a developmentally regulated gene cluster. Cell **25**:165–177.
34. **Snyder, M., M. Hunkapiller, D. Yuen, D. Silvert, J. Fristrom, and N. Davidson.** 1982. Cuticle protein genes of Drosophila: structure, organization, and evolution of four clustered genes. Cell **29**:1027–1040.
35. **Ullrich, A., J. Shine, J. Chirguin, R. Pictet, E. Tischer, W. J. Rutter, and H. M. Goodman.** 1977. Rat insulin genes: construction of plasmids containing the coding sequences. Science **196**:1313–1319.
36. **Wagner, P. D., and E. Giniger.** 1981. Hydrolysis of ATP and reversible binding to F-actin by myosin heavy chains free of all light chains. Nature (London) **292**:560–562.
37. **Watterson, D. M., F. Sharief, and T. Vanaman.** 1980. The complete amino acid sequence of the calcium dependent modulator protein (calmodulin) of bovine brain. J. Biol. Chem. **255**:962–975.
38. **Weeds, A. G., and S. Lowey.** 1971. Substructure of the myosin molecule II. The light chains of myosin. J. Mol. Biol. **61**:701–725.
39. **Weeds, A., P. Wagner, R. Jakes, and J. Kendrick-Jones.** 1977. Structure and function of myosin light chains, p. 222. *In* R. H. Wasserman et al. (ed.), Calcium binding proteins and calcium function. Elsevier/North-Holland Publishing Co., Amsterdam.
40. **Whalen, R. G., G. S. Butler-Browne, and F. Gros.** 1978. Identification of a novel form of myosin light chain present in embryonic muscle tissue and cultured muscle cells. J. Mol. Biol. **126**:415–431.
41. **Yen, P. H., and N. Davidson.** 1980. The gross anatomy of a tRNA gene cluster at region 42A of the Drosophila melanogaster chromosome. Cell **22**:137–148.

Chapter 3

The unique myosin light chain-2 gene of *Drosophila melanogaster* encodes two polypeptides which are abundantly expressed in muscle

ABSTRACT

Recombinant DNA clones encoding the *Drosophila melanogaster* homolog of the vertebrate myosin light chain-2 (MLC-2) have been isolated. This single copy gene maps to the chromosomal locus 99E. The nucleotide sequence was determined for a 3.4 kb genomic fragment containing the gene and for two cDNA clones generated from late pupal mRNA. Comparison of these sequences shows that the gene contains two introns. Extension of a primer homologous to the mRNA reveals two start sites for transcription 12 nucleotides apart. The sequence TATA is not present prior to the mRNA cap site. There are two major sites of poly A addition separated by 356 nucleotides. The protein sequence derived from translation of the cDNA sequence shows a high degree of homology with that for the DTNB myosin light chain (MLC-2) of chicken. A lower degree of sequence homology was seen in comparisons with other evolutionarily related calcium binding proteins. RNA blots show high levels of expression of multiple transcripts present during the developmental time stages when muscle is being produced. *In vitro* translation of hybrid selected RNA produces two polypeptides which comigrate on 2-dimensional gels with proteins from *Drosophila* actomyosin.

INTRODUCTION

Myosin is a hexameric protein which forms the core of the thick filaments of muscle. All myosin filaments share the same architecture of two heavy chains and two pairs of light chains. In vertebrates, there are different non-allelic forms of these proteins which are utilized in different tissues and at different stages of development. The evolution of these proteins can be studied by comparing their sequences both within and between species.

In scallop muscle contraction is stimulated by the binding of Ca^{++} to the regulatory myosin light chains (RLCs) (29). In vertebrate smooth muscle regulation is imparted by the binding of calcium to the homologous light chain (11), but myosin from striated muscle has lost this function (26). In addition to RLC these myosin light chains have been refered to as the DTNB light chains, the phosphorylatable light chain (PLC), and myosin light chain-2. In this paper we refer to the *Drosophila* protein as myosin light chain-2 (MLC-2). Another level of control in some types of muscle is phosphorylation of myosin (for review see (1)).

In myogenic cells terminal differentiation is marked by several phenomena. In skeletal muscle, formation of myotubes from the fusion of mononucleated myoblasts is preceded by drastic changes in gene expression. Transcription and translation of genes for muscle specific

isoforms of proteins used in the contractile apparatus are induced approximately 1000 times their previous rates while some non-muscle proteins are repressed by a similar level (19). After fusion has begun the components of the thin filament: mainly α-actin, tropomyosin and the troponins; and those of the thick filament: myosin heavy chain and the two myosin light chains, are assembled into an interdigitating lattice. The regulation of this complex set of events is poorly understood.

Several laboratories have isolated recombinant DNA clones for the major myofibrillar components of *Drosophila melanogaster* in order to gain a better understanding of muscle development and control, the evolution of muscle proteins and to study the genetics of these genes. Previous reports have presented detailed characterizations of the *D.melanogaster* actin (17,51), tropomyosin (3,27), myosin heavy chain (6,50) and myosin alkali light chain genes (15). In this report we describe the characterization of the myosin light chain-2 gene of *D.melanogaster*. The DNA sequence is presented from both the genomic region and its transcribed RNAs. The protein sequence translated from this region is compared to other MLC-2 protein sequences. We describe the developmental expression of this myosin light chain gene and show that it is present in a single copy on the *Drosophila* genome.

MATERIALS AND METHODS

Enzymes and chemicals

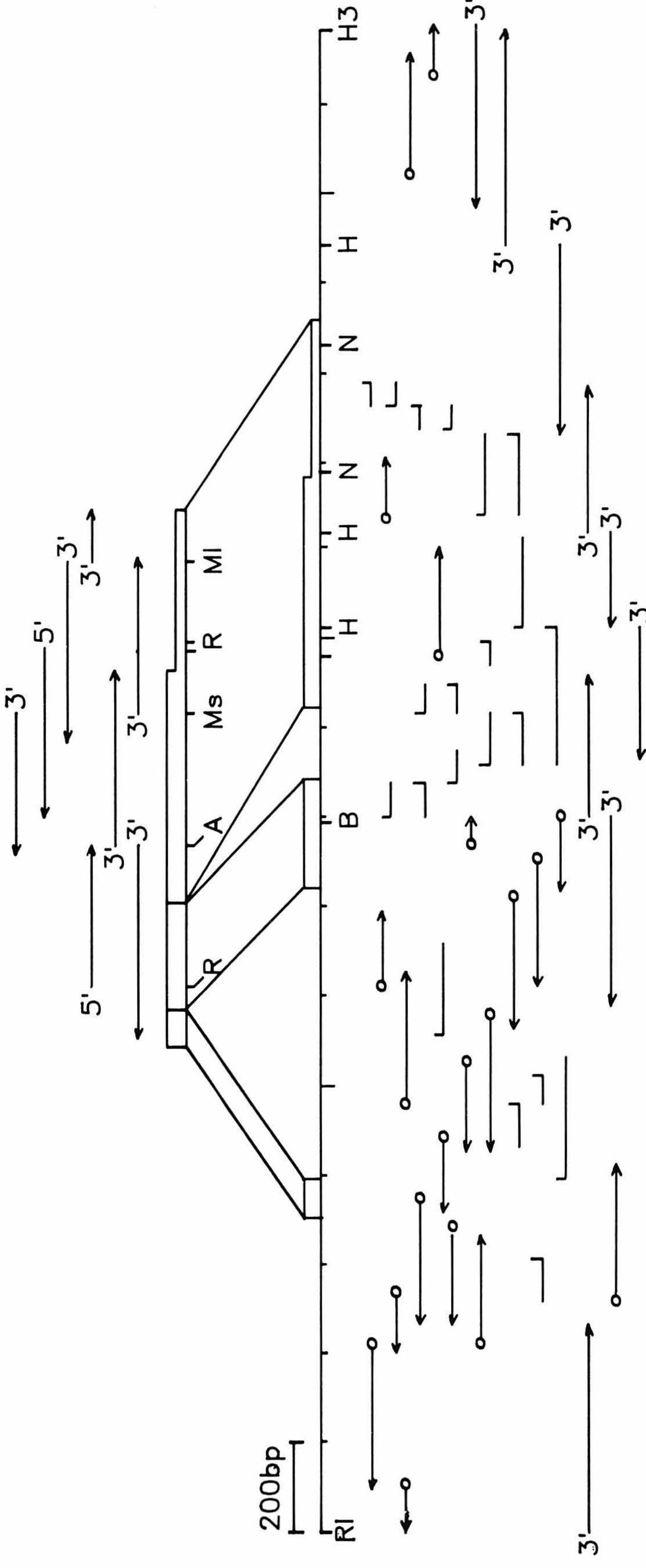
Restriction enzymes and enzymes used for labeling DNA were purchased from New England Biolabs, Bethesda Research Labs or Boehringer-Mannheim. Nuclease S1 was from Sigma or Miles. All reactions were performed as recommended by the manufacturer or as described by Maniatis et al. (34) unless otherwise stated.

Generation of plasmid subclones

DNA was subcloned from the lambda clone λdmpT57 (see below) by digestion with EcoRI and HindIII and ligation of the resulting fragments into pBR322 as described (7,34). A clone containing a 3.4 kb EcoRI-HindIII fragment which includes the coding region from λdmpT57 was identified. This subclone, pLP5734, was used for subsequent cloning and analyses. Plasmid subclones used for DNA sequence analysis were generated by digesting the 1.8 kb BamHI-HindIII fragment of pLP5734 with HpaII and ligation into the AccI site of pUC8 (56). An NdeI subclone was generated by digesting the HpaII subclone containing the region 3' to the open reading frame of pLP5734 (see below) with NdeI, ligation into the calf intestinal alkaline phosphatase treated NdeI site of pBR322 and selection by colony hybridization(23).

Figure 1. Sequencing strategy

The upper part of the figure is the sequencing strategy for the cDNA clones. The rectangular region (▭) denotes the transcribed DNA present in mature mRNA, (▭) sequences specific to the longer transcript (see text). The lower part of the figure shows the organization of the exons transcribed from the genomic DNA and the sequencing strategy of the 3.4 kb EcoRI-HindIII fragment. The direction and extent of sequencing are shown. Symbols used for sequences read from left to right on the map are (└──): M13 clones, (o⟶): Bal 31 generated M13 clones (see materials and methods), (3'⟶): 3' end labeled restriction fragments, and (5'⟶): 5' end labeled fragments. Restriction enzyme sites are: A: AvaI, B: BamHI, H: HpaII, H3: HindIII, Ml: MluI, Ms: MstII, N: NdeI, R: RsaI, and RI: EcoRI.


200bp
RI
B
H
N
N
H
H3
R
A
Ms
R
MI
5'
3'

Generation of M13 subclones

M13 subclones were initially generated by digesting pLP5734 with Sau3A or HpaII and ligating the resulting fragments into M13mp8 (37) which had been digested with BamHI or AccI, respectively. The 3.4 kb Drosophila DNA insert in pLP5734 was isolated by agarose gel electrophoresis and nick translated to probe for clones with desired inserts by the method of Benton and Davis (4). The original vector, pBR322, was used to counterscreen plaques.

The EcoRI-BamHI region was underrepresented by this shotgun approach (see figure 1). To generate molecules which would allow us to rapidly sequence the entire insert region we used the double strand exonuclease Bal31 (32). This enabled us to isolate clones with different amounts of DNA deleted from one of their endpoints. The technique is shown diagrammatically in figue 2. Ten micrograms of pLP5734 was linearized with HindIII, phenol extracted, ethanol precipitated and resuspended in 100 µl of Bal 31 buffer. An amount of Bal 31 titrated to digest approximately 40bp/end/min was added. Aliquots were removed at five minute intervals and the reactions stopped either by addition of EGTA to 25 mM or by phenol extraction and ethanol precipitation. Reactions were diluted and digested with EcoRI. Molecules generated in this manner can be ligated into a vector with one blunt end and one EcoRI end. We experienced difficulty using HincII to

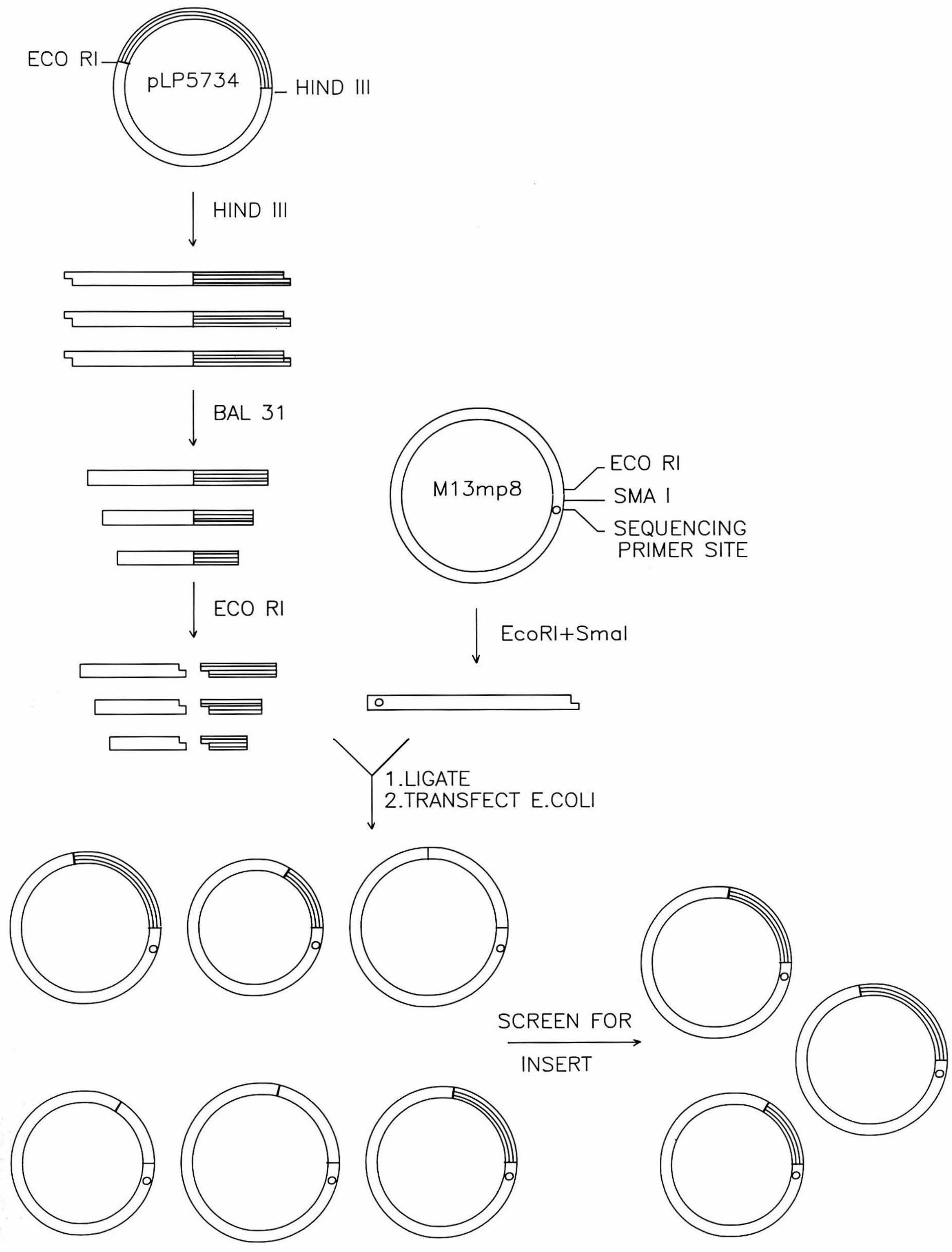
produce a blunt end for ligation. As an alternative we digested 5 μg of M13mp8 DNA with HindIII, added each of the four dNTPs to 20 μM, 1 μl of the Klenow fragment of DNA polymerase I (2.5 units, Boehringer-Mannheim) and incubated at room temperature for 20 minutes. Reactions were terminated by phenol extraction and ethanol precipitation. These linearized, blunt-ended molecules were then recut with EcoRI to generate the desired vector. Ligations were incubated at 14°C for 18 hours using approximately 50 ng of vector and 100 ng of insert in 10 μl. Plaques were screened as described above for the selection of M13 recombinants. An analogous scheme was used to generate deletions from the BamHI and EcoRI sites of pLP5734. Recombinant DNA molecules used for DNA sequence analysis are shown in figure 1. A similar method used to generate nested deletions for sequencing has been described by Poncz et al. (44).

All experiments involving organisms harboring recombinant DNA molecules were carried out in EK1 vectors under P1 or P2 levels of containment.

Labeling of DNA fragments with ^{32}P

Fragments were labeled on their 5' termini as follows. Five micrograms of restriction enzyme cleaved DNA was incubated with calf alkaline intestinal phosphatase for 30 minutes at 37°. The DNA was phenol extracted, ethanol precipitated and resuspended in 30 μl kinase buffer (50 mM

Figure 2. Diagrammatic representation of Bal 31 exonuclease technique for generating subclones (see text for details).


ECO RI
pLP5734
HIND III
HIND III
BAL 31
M13mp8
ECO RI
SMA I
SEQUENCING
PRIMER SITE
ECO RI
EcoRI+SmaI
1.LIGATE
2.TRANSFECT E.COLI
SCREEN FOR
INSERT

Tris pH 9, 10 mM $MgCl_2$, 1 mM EDTA, 1 mM spermidine, 10 mM dithiothreitol). One microliter γ-^{32}P-ATP (7000 Ci/mmol, ICN; usually 100-250 µCi) and 1 µl T4 polynucleotide kinase were added and the reaction was incubated at 37° for 1 hour. Labeled fragments were separated from unincorporated ^{32}P either by passage over a G-50 column or fragment isolation by gel electrophoresis (see below). If a second restriction enzyme digestion was necessary the kinase was inactivated with phenol and the DNA ethanol precipitated before proceeding.

DNA fragments were labeled on their 3' termini using α-^{32}P-dNTPs (3000 Ci/mmol; Amersham, NEN, or ICN) to fill in the recessed ends produced by restriction enzyme cleavage. To insure homogeneously labeled ends only the last nucleotide to be replaced was usually labeled. For 2 pmol of BamHI generated ends (GATC replacement), 3 pmol of α-^{32}P-dCTP would be used with unlabeled dATP, dGTP, and TTP at a final concentration of 20 µM each in 20 µl HaeIII buffer (6 mM NaCl, 6 mM TrisHCl pH7.5, 6 mM MgCl, 100 µg/ml BSA, 6 mM β-mercaptoethanol). DNA polymerase I Klenow fragment (2-5 units) was added and the reaction was allowed to proceed 30-60 minutes at room temperature. The reaction was terminated by incubation at 68° for 10 minutes. The salt concentration was adjusted as necessary and a second restriction enzyme was added. Specific activities of 10^6 cpm per end were obtained by either 3'or 5' labeling methods.

DNAs were nick translated as previously described (48).

Gel electrophoresis and isolation of fragments

Labeled DNA fragments used for sequencing by the hemical modification method (36) were isolated on 4.75% polyacrylamide: 0.25% bis-acrylylcystamine (Biorad) gels in 100 mM TBE (100 mM Tris base, 100 mM boric acid, 2 mM EDTA, pH 8.3) as described by Garfinkel et al. (18).

RNA was fractionated on a 1% formaldehyde-agarose gel and transferred to nitrocellulose filters as described (34).

Two-dimensional gel electrophoresis of proteins followed the technique of O'Farrell (42). Ampholines of pH range 3.5 to 10 were used for the first dimension, 12% polyacrylamide gels were used for the second dimension. ^{14}C protein standards were from BRL. Drosophila proteins were extracted from adult animals as described (15).

Sequencing gels were 5%, 8%, and 20% polyacrylamide (20:1::acrylamide:bisacrylamide), 50% urea in 100 mM TBE. Gels were 0.4 mm thick and 40 or 80 cm long.

DNA sequence analysis

DNA sequencing was done by both the chemical modification technique (36) and the dideoxy chain termination method (52) using M13 phage vectors (37). The sequencing strategies for pLP5734 and the cDNA clones

pCDM20 and pCDM38 are shown in figure 1.

S1 Nuclease mapping of transcripts

Transcripts were mapped by S1 nuclease digestion of DNA-mRNA hybrids using the Weaver and Weissman modifications of the technique of Berk and Sharp (5,16,57). Restriction enzyme generated fragments of DNA were labeled as described above. In a typical reaction approximately $6x10^4$ cpm of end labeled probe was hybridized to 7.5 μg of late pupal poly A^+ mRNA at 56° in S1 hybridization buffer (80% formamide, 40 mM PIPES, 400 mM NaCl, 1 mM EDTA, pH 6.8) for 5 hours. Hybridizations were diluted with 300 μl of S1 prep buffer (280 mM NaCl, 30 mM NaOAc pH 4.4, 4.5 mM $Zn(OAc)_2$, 20 μg/ml denatured salmon sperm DNA) at 0° and placed immediately on ice. S1 nuclease was added and the reactions were allowed to incubate for 30 minutes at the indicated temperature. S1 digestions were terminated by addition of 75 μl of S1 stop buffer (2.5 mM NH_4OAc, 50 mM EDTA, 400 μg/ml yeast tRNA) and precipitated using 400 μl of isopropanol. Pelleted samples were resuspended in sequencing gel loading dye (36), boiled and applied to an appropriate gel.

RESULTS

Isolation of an abundantly expressed myogenic clone

We have previously described the isolation of recombinant DNA clones encoding Drosophila genes abundantly expressed during muscle formation (15). We used RNA from animals 70-75 hours after puparium formation to screen a genomic library in the lambda phage vector charon 4. After exclusion of non-myogenic clones and those encoding actin or myosin heavy chain gene sequences we examined 59 clones in further detail. Preliminary characterizations of these clones by restriction enzyme analysis and blotting with a cDNA probe revealed one group of clones which included eleven independent phage isolates. A restriction endonuclease map of these clones is shown in figure 3.

The region hybridizing to late pupal cDNA was localized to a 3.4 kb EcoRI-HindIII fragment. This region was subcloned from the lambda phage λdmpT57 into the plasmid vector pBR322 and is designated pLP5734.

Multiple RNA transcripts from one region

Other Drosophila muscle specific genes have been shown to display a specific pattern of expression during different developmental stages (15,50,51). We wished to determine at what stages of development the gene in pLP5734 was expressed and the sizes of any transcripts which were present. RNA blots of poly A^+ RNA isolated from different

Figure 3. Restriction endonuclease map of overlapping λ clones and the subclone pLP5734.

The EcoRI-HindIII insert of pLP5734 is 3.4 kb. Abbreviations for restriction enzymes used are B: BamHI, R: EcoRI and H: HindIII.

H B B B H B R B R B R R R

pLP5734
dmpT3
dmpT4
dmpT5
dmpT6
dmpT9
dmpT56
dmpT20
dmpT55
dmpT2
dmpT84
dmpT57

developmental stages were probed with pLP5734 (figure 4). There are two major transcripts which are approximately 1.4 and 1.1 kb in length and a minor transcript of 2.7 kb. These transcripts are abundant in late embryonic, late pupal, second instar larval and adult RNAs. They are detected at less intensity in first and third instar larvae and are undetected in early embryonic and early pupal RNAs. This is the same pattern of expression seen with the myosin heavy chain gene (50) and the myosin alkali light chain gene (15) when the same protocols for selecting RNA from different stages of development were used.

Copy number and cytogenetic location

To determine the copy number of the sequences within λdmpT57 we prepared a blot of BamHI-EcoRI digested *Drosophila* DNA which had known equivalents of digested λdmpT57 in adjacent lanes (figure 5). The blot was probed with λdmpT57 DNA at moderate stringency. By comparison of radiographic intensities it was determined that the cloned DNA was single copy within the *Drosophila* genome.

The complexity of the *Drosophila* genome is low enough that it is possible to probe a blot of cloned DNA with nick translated genomic DNA. Since the DNA on the filter is in excess of the homologous sequence in the probe, the intensity of signal is proportional to the length of the fragment and the copy number of the sequence. Standards of actin (six copies in the genome),

Figure 4. Developmental expression of mRNAs from pLP5734.

Poly A^+ RNA (1 μg) isolated from different developmental stages of synchronized populations of *Drosophila melanogaster* was separated by electrophoresis on a 1.5% agarose, 2.2 M formaldehyde gel. After transfer to nitrocellulose the RNA was hybridized to a nick translated probe of pLP5734 DNA. Time in hours is from egg deposition. Pupae were resynchronized at 5 hour post-pupariation by floatation (38) and assumed to be 125 hours old. RNA size standards are *E. coli* 16S (1547 nt) and 23S (2904 nt) ribosomal RNA, and HeLa cell 18S (1760 nt) and 28S (5090) ribosomal RNAs (9,10).

LARVAL INSTAR PUPAE

EMBRYO 1st 2nd 3rd PRE EARLY MID LATE ADULT

2 15 26 50 75 120 144 168 195 216

–28S

–23S

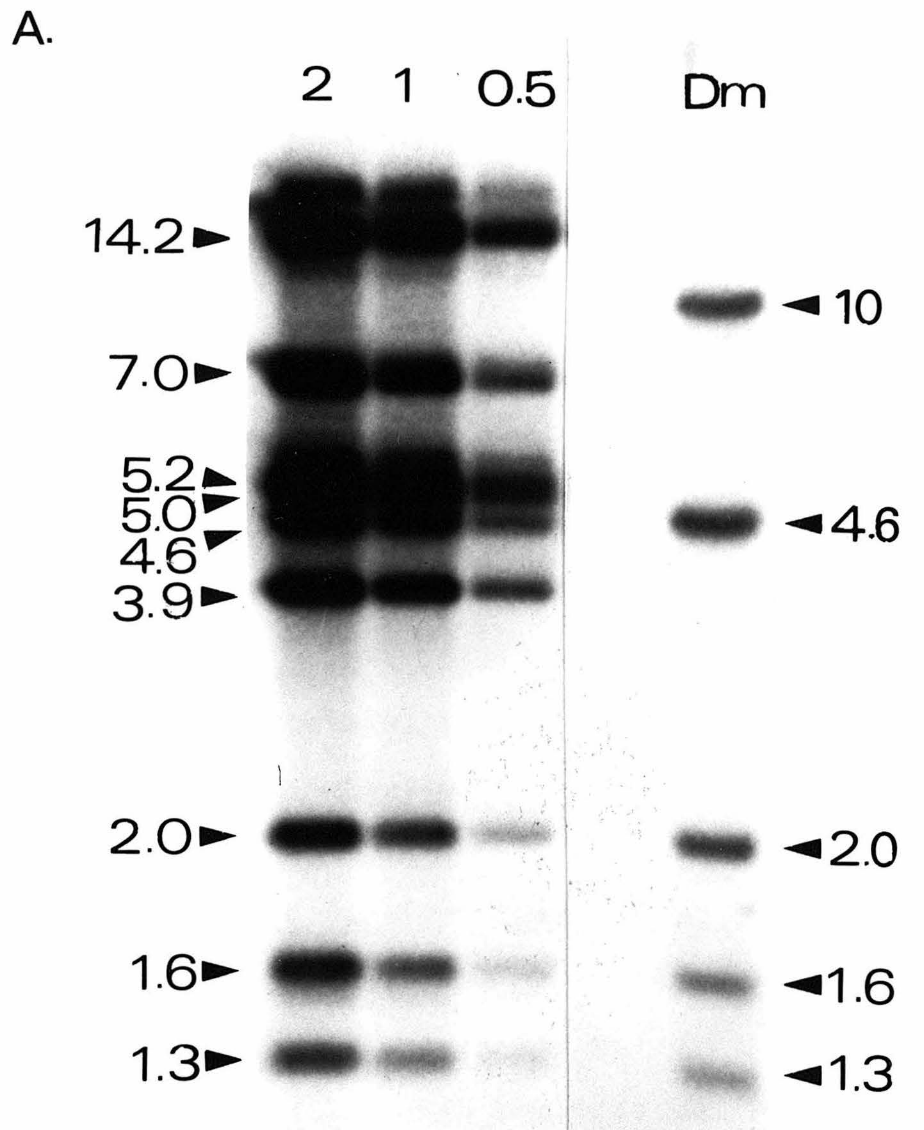
–18S

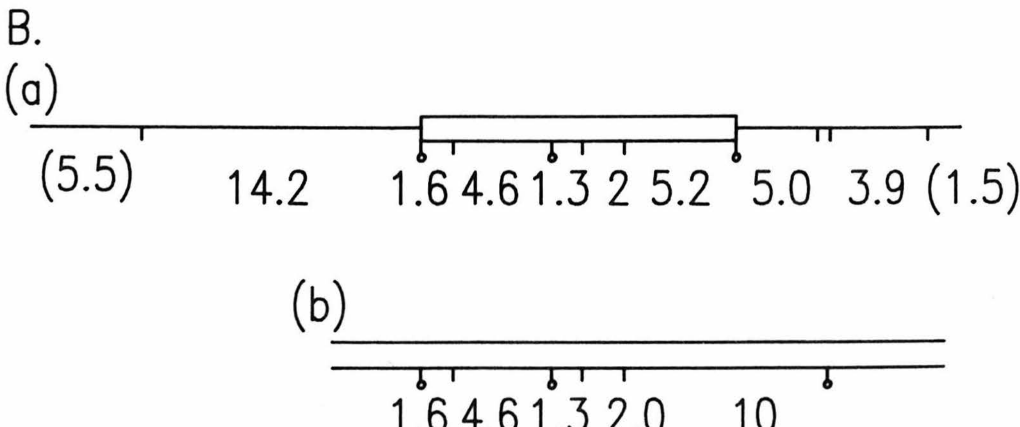
–16S

Figure 5. Reiteration frequency of λdmpT57 in *Drosophila* DNA.

A. DNAs were digested with EcoRI and BamHI. E.coli DNA was added to the λdmpT57 digests as carrier. Fragments were separated by electrophoresis on a 0.7% agarose gel and transferred to nitrocellulose. The nitrocellulose filter was probed with nick translated λdmpT57 DNA. Gel lanes had 2, 1 or 0.5 genomic equivalents of λdmpT57 DNA or 1 μg of *Drosophila* genomic DNA (Dm) applied as indicated. Sizes are in kilobases.

B. Map of expected restriction endonuclease fragment sizes: (a) λdmpT57 and (b) *Drosophila* genomic DNA. Symbols used |: BamHI and ⫯: EcoRI ; ———— λ DNA, ▭ *Drosophila* DNA.


A.
2
1
0.5
Dm
14.2
7.0
5.2
5.0
4.6
3.9
2.0
1.6
1.3
10
4.6
2.0
1.6
1.3


B.
(a)
(5.5)
14.2
1.6 4.6 1.3 2 5.2
5.0
3.9 (1.5)
(b)
1.6 4.6 1.3 2.0
10

copia clone IIB11 (approximately 40 copies per genome), λcDM2100 (highly repetitive) and λcDM2301 (single copy) were included for comparison. This experiment also indicated that λdmpT57 was uniquely represented in the genome (data not shown).

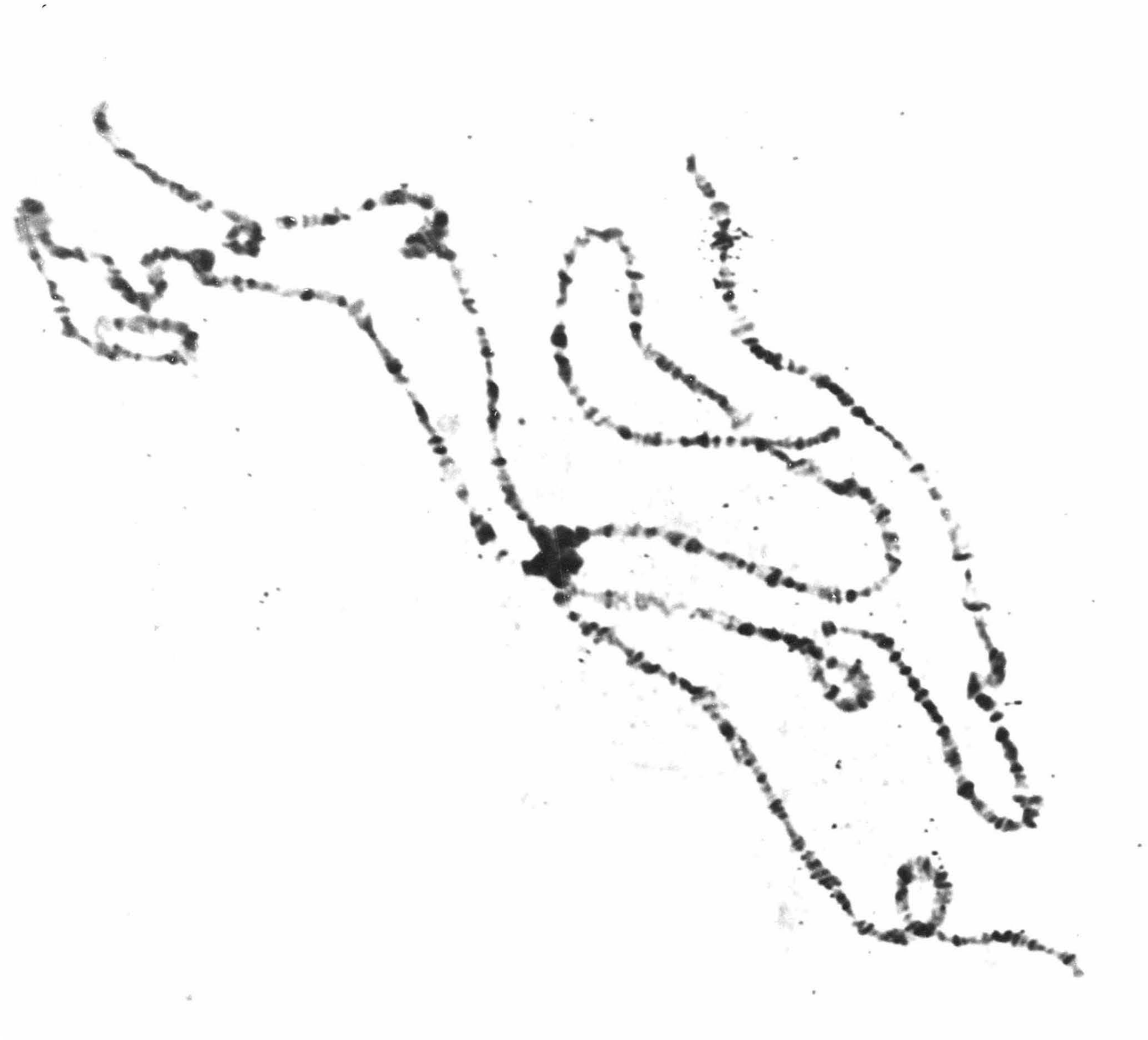
The cytogenetic locus of λdmpT57 was determined using both ^{3}H-cRNA and ^{35}S labeled nick translated DNA as position 99E. ^{35}S labelling has the distinct advantage of shortening the exposure time necessary for identification by twenty-fold, but suffers somewhat from a lack of resolution. An *in situ* hybridization using ^{35}S is shown in figure 6. There is a single site of hybridization at 99E, even upon prolonged exposure.

Hybrid selection and in vitro translation

To help identify the gene product(s) encoded by the cloned region we characterized the polypeptides produced by hybrid selection and in vitro translation (47). DNA from the clone λdmpT55 was spotted onto a nitrocellulose filter. PolyA^{+} RNA from late pupae was hybridized to select the mRNAs transcribed from this cloned region during muscle formation. Following elution from the filter the mRNA was translated in an mRNA dependent rabbit reticulocyte system using ^{35}S-methionine. Products were analyzed by one and two-dimensional gel electrophoresis. Figure 7 shows a two-dimensional gel of the two *Drosophila* proteins produced. The apparent molecular weights (M_r) of

Figure 6. *In situ* hybridization of pLP5734 DNA to *Drosophila* polytene chromosomes.

The 3.4 kb subclone pLP5734 was nick translated using ^{35}S-dATP to a specific activity of 10^6 cpm/μg. 2×10^5 cpm were used to hybridize to *Drosophila* polytene chromosomes. After 24 hour exposure the slide was developed and the chromosomes stained.



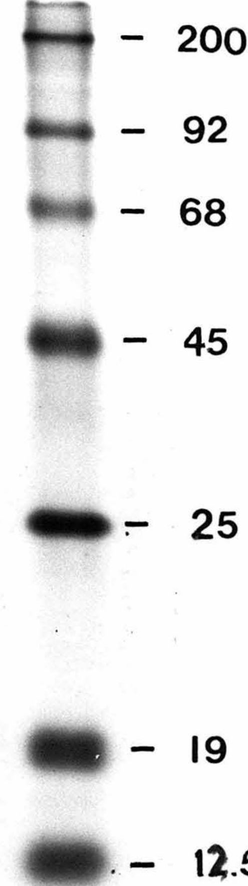
the protein products are 26 kd and 17 kd. Each has a pI of 5. We occasionally saw an additional spot migrating slighty slower than the 17 kd band in the SDS-polyacrylamide dimension. Hybrid selection using the 1.8 kb BamHI-HindIII fragment of pLP5734 gave the same pattern.

DNA sequence analysis of pLP5734

The sequence of the cloned insert is presented in figure 8. All regions were sequenced on both strands except bases 2723-2885 and 2886-3041 which were sequenced in the same direction from two different clones; and nucleotides 1-392 which were sequenced from several clones in both directions. We occasionally experienced difficulty reading nucleotides adjacent to inverted repeats using M13 based dideoxy sequencing. An example of this is the sequence from 1224 to 1241. The last base of the sequence AATCCCTATTAAGGGATA follows an inverted repeat (underlined) of six nucleotides and gave a signal in all four lanes. Raising the temperature of polymerization to 30° did not alleviate this problem entirely. Nucleotides which gave an ambiguous signal were determined by a combination of the following: sequencing from the opposite direction, using the chemical modification technique, or using an M13 clone generated using Bal31 which interrupted the inverted repeat.

Figure 7. *In vitro* translation of hybrid selected RNA.

RNAs transcribed from the DNA in λdmpT55 during late pupal development were selected by hybridization to filter bound λdmpT55 DNA as described in materials and methods. RNA was released from the filter by boiling and translated in a rabbit reticulocyte extract. Products of *in vitro* translation were coelectrophoresed with proteins isolated by extraction of adult *Drosophila*.


200
92
68
45
25
19
12.5
OH⁻
H⁺

Isolation and characterization of cDNA clones

To map the transcribed regions of pLP5734 and to determine the nature of the heterogeneity in the different length transcripts, cDNA clones homologous to this region were isolated. We have previously described the construction of a cDNA library in the vector pUC8 using polyA+ RNA isolated from Drosophila 70-75 hours after puparium formation (15). The 1.8 kb HindIII-BamHI fragment of pLP5734 was isolated, nick translated using α-^{32}P-dCTP and used to probe 25 plates containing 700-1000 colonies each from this library. Thirty positively hybridizing colonies were selected and streaked out to obtain isolated colonies. Two cDNA clones with relatively large inserts were chosen for further study. The clones pCDM20 and pCDM38 were sequenced using the chemical modification methods as shown in figure 1.

As shown in figure 8, the cDNA sequence was in complete agreement with the genomic sequence and was used to identify the exons in the latter. The sequence of pCDM20 localized two intervening sequences in the genomic sequence. The junctions of the intervening sequences follow the consensus rules previously described for eukaryotic genes (8) and for those of Drosophila (54). The poly A tail from this clone was approximately 200 nucleotides in length and began at nucleotide 2361 of our sequence, 21 nucleotides 3' to the first A in the recognition sequence AATAAA for poly A addition (45). The

Figure 8. DNA sequence of pLP5734.

The DNA sequence was determined as described in the text and shown in figure 1. The sites mapped for initiation and termination of transcription are denoted by an arrow (▲). Intron sequences at the intron-exon junctions are underlined. The predicted amino acid sequence is depicted over the DNA sequence. C20 is the 5' end of the cDNA clone pCDM20.

```
   1  GAATTCTCGGCGCTCTTCATATTGCACAATGTGCGGGCATTTCGAAAACAGAAAAAGGAACGTAGGCGAGCAAAACATCGAGTATATAGGCGTGAATAGAATACATACATATGCTATATG   120
 121  TTGGCATATATGTATGCTGTCCGCCGCCAAGGAAAATGTAGGTGATTTCAGCGCAAGAGGACACCTTTCGGAGGTGGCCTCTGGCAGAGAGACGCACCGAGTAGTGAAGCAGAGCAAAAT   240
 241  AAAAGAAAACAAAACAGAATCCGAGGATGGGGAAATGTATCCACAACAGAGAGCCGCTATATGAACCAATCCGAATCGCCCCGAAATAAATTCCGAACATATGTATAGGCTTCTGCTGCC   360
 361  CAAAAGCCAAAGCCAAAGCAATTTCTCTGCCTTTAAAATTAGATTTTTGATTAGAGCAGCGCCCGATGGATGAGTAAACACCTCGGATACAGCTACACCATAGGCTATAGCTATAGCCAT   480
 481  GGCTATATCTATATAGTTGTAGCTGCCCGCCCCCGCCGATCTCCTGGGCGGACCTCGCCTTGCAGATACTAAAAAAGGCAGCACGAATATTGCAGATACGGATACAGCGAATGGCGCCCG   600
                                                                                                                         c20
 601  AAGATCCCCGAAGGTGGCCGAACCTCCACGAGCGGCCCCAATGCTTGGAAGAAGAAGAGCTCCTCCGAGCTCTTAAAACATCATTGTCTTGGCGATAAAATCAGCCAGAGATTCAGTGTG   720

   1                                          Met                                                                                1
 721  CTCTGACGGCTCGACTCGATTAGACAACACATCAAAGTTGAGCACCATGGTAAGGACTCGTCCCGGAATATCACTACAATCCAACTGAACCGAACTGAACCAGAACCGAATCGAACTGGA   840
 841  CCCCAGTGCCTGGATATATGCAGCAAGGATAATGCCGATCAGTGCAGTGAATCGTTTGCATTCGAGAATGACCCCGAGGATATGAATATGAAGTGAACCATAACTCTAAATCATGGTGAT   960
 961  CCCGATTCACCTTCTTCACATACGCAGTGAGAATAGTGCAGAAAGTTCTATAAATGGATCGATCCGATCCGAGTGTTACGCGTCAGTAATGTAATCTGATATCCGTGCCGGGGAGCGAGC  1080
1081  GAGACAGCGAGTGGCCACGGCGGAGAGCTTACGTAATATGCGTCCGGTATATTTTTAGAATGAGTGGAAAACTTTCGCCTGCAGATACATATTTTCTGGCCTAGTCGGGTAGGCCTTTTG  1200
1201  AAAAGCATACCTTACATCCCTTAAATCCCTATTAAGGGATAACTTTCTACCCATATAATTCCAATTGAAGATCCTTCAACGCGGATAACTTACTAAGATCCCTAAAACGAATAGTATAAC  1320

1321  TTCCGGAATCGGAACTTTCAAGTGACAGCTAAAGACGGACTATAACGACTAATGACCGAGTAGATACCCCCTTTCTGCCACGAGTCCCTACTAATAAATCCCTCTTTTCTCTCCTCAG     1438

   2  AlaAspGluLysLysLysValLysLysLysLysThrLysGluGluGlyGlyThrSerGluThrAlaSerGluAlaAlaSerGluAlaAlaThrProAlaProAlaAlaThrProAlaPro    41
1439  GCCGATGAGAAGAAGAAGGTTAAGAAGAAGAAGACCAAGGAAGAGGGTGGTACTTCCGAAACCGCTTCTGAGGCCGCATCCGAGGCAGCAACCCCGGCACCAGCTGCCACTCCTGCCCCG  1558

  42  AlaAlaSerAlaThrGlySerLysArgAlaSerGlyGlySerArgGlySerArgLysSerLysArgAlaGlySerSerValPheSerValPheSerGlnLysGlnIleAlaGluPheLys    81
1559  GCCGCATCTGCCACTGGTTCGAAGAGAGCGTCGGGCGGATCCCGTGGCTCCAGGAAGTCGAAGCGCGCTGGCTCCTCGGTCTTCTCTGTGTTCTCCCAGAAGCAGATCGCCGAGTTCAAG  1678

  82  Glu                                                                                                                       82
1679  GAGGTGAGTTTTGGGCGTTACTCATATATCGCATCATATGTCGGATCACTCGAGACCCATCCCATCATCCTGTCATCCTACCATCTATCGCACTTGTCAAGGACCAAGTCACCAACACAC  1798

  83                                      AlaPheGlnLeuMetAspAlaAspLysAspGlyIleIleGlyLysAsnAspLeuArgAlaAlaPheAspSerValGly   108
1799  TCTTTCTTTCTATCTCTCTCTGTCTCTGTGCGGGATCGATAGGCCTTCCAACTCATGGATGCCGACAAGGACGGTATTATTGGCAAGAACGATCTGCGCGCTGCCTTCGACTCCGTCGGC  1918

 109  LysIleAlaAsnAspLysGluLeuAspAlaMetLeuGlyGluAlaSerGlyProIleAsnPheThrGlnLeuLeuThrLeuPheAlaAsnArgMetAlaThrSerGlyAlaAsnAspGlu   148
1919  AAGATCGCCAACGACAAGGAGTTGGACGCCATGCTGGGCGAGGCCTCGGGTCCGATCAACTTCACCCAGTTGCTGACCCTGTTCGCCAACCGCATGGCCACCTCCGGTGCCAACGATGAA  2038

 149  AspGluValValIleAlaAlaPheLysThrPheAspAsnAspGlyLeuIleAspGlyAspLysPheArgGluMetLeuMetAsnPheGlyAspLysPheThrMetLysGluValAspAsp   188
2039  GACGAAGTTGTTATTGCTGCCTTCAAAACATTCGATAACGATGGTCTCATCGACGGTGACAAATTCCGCGAAATGCTCATGAACTTCGGTGACAAGTTCACCATGAAGGAGGTTGATGAT  2158

 189  AlaTyrAspGlnMetValIleAspAspLysAsnGlnIleAspThrAlaAlaLeuIleGluMetLeuThrGlyLysGlyGluGluGluGluGluGluAlaAla                    222
2159  GCCTACGATCAGATGGTGATCGACGACAAGAACCAGATCGATACCGCCGCCCTGATCGAGATGCTCACCGGCAAGGGTGAAGAGGAGGAGGAGGAGGCCGCCTAAGGCACATCCGATCAC  2278
2279  ATCCTAACTTCTAACTCAACATCCATTTTACAGCCAGAATCAGAACGTTAAATCAATCAGTCAAATCAGTAATAATAAACGTAATTAAAAGAAACGCTCGTTTACATATGTAGATTCTCA  2398
2399  GTATCTTGTAATCTCTTTTGTTAACCATCACCAAGTACCTTGTACCAAGAACAACACCACCAACCAGATCAGCAACATCCGATCAAACTGAACCTCAAAAGAACTTCTTCAGCAACATTT  2518
2519  GGTAGCGATCAGAACCATTTACCATGTATCTGTCCATCATAGTTAGGGATACTTACGATCACGTTCACTGTCACGGCAAGAACTGACGCGTAGTTTTCTAGTTGATTTACTGTTAAATTT  2638
2639  TACACGTAAAATTTATAAAACATATGCTATCTATTTTATTTAAAACACCTATAAAATCAACAAAAAAAACTTAAAAACAACAAACAAGCCGTGTCTATTTTTATATCGCTGTCCGTTGAA  2758
2759  GAACTAAAGAGAAGAACCCCCAGAGAGAGATGATTACCAGATCGAGATCGAGATCGGAATCGAAATGAGGATGTATGATGTATGATTTATGATGTAGGGAGGACGAAGACGATGTTGATG  2878
2879  GGTAGCCCGGCTGTATTCGTTAGCCCTGCACCCGCGAGTTCAATTTGATTATGAATATGAATTTTAGCCAGTTTTGTGAGTTACAATGTTTTTCGTTCAAGAAATCTCGCAGGACCATGA  2998
2999  GAGAGAGAGGTAAACCCATAACAAACTATTGTTAGGGATCATTTAACTGAGAGCCAAATATATATCCTACTATCACAAAATATTTATCGTTATACATATCTATCGTTCGTTACTTACTTA  3118
3119  TTTATCAGTTATGAGTTATCAATATTATTGCAAGTTATCGAGAGTTAGGCTAATAACAGAGAAACATTAAGTTCATTCACTCAAATTGTTTAATGCGAATTTATATAACTTAACCATCGC  3238
3239  CACCACCACAACTTTCTGCTTATGAACTATCAGCGATATAAACCACCCAAACGAACAAACAACCCAAAATGATATCTGAAATATGAAGAAACATTCAAATAACAAACCCAGTGTTTCTAG  3358
3359  TTAATGCGTTTTGAAGCTT                                                                                                       3377
```

insert of pCDM20 is 918 nucleotides long (excluding the polyA sequence).

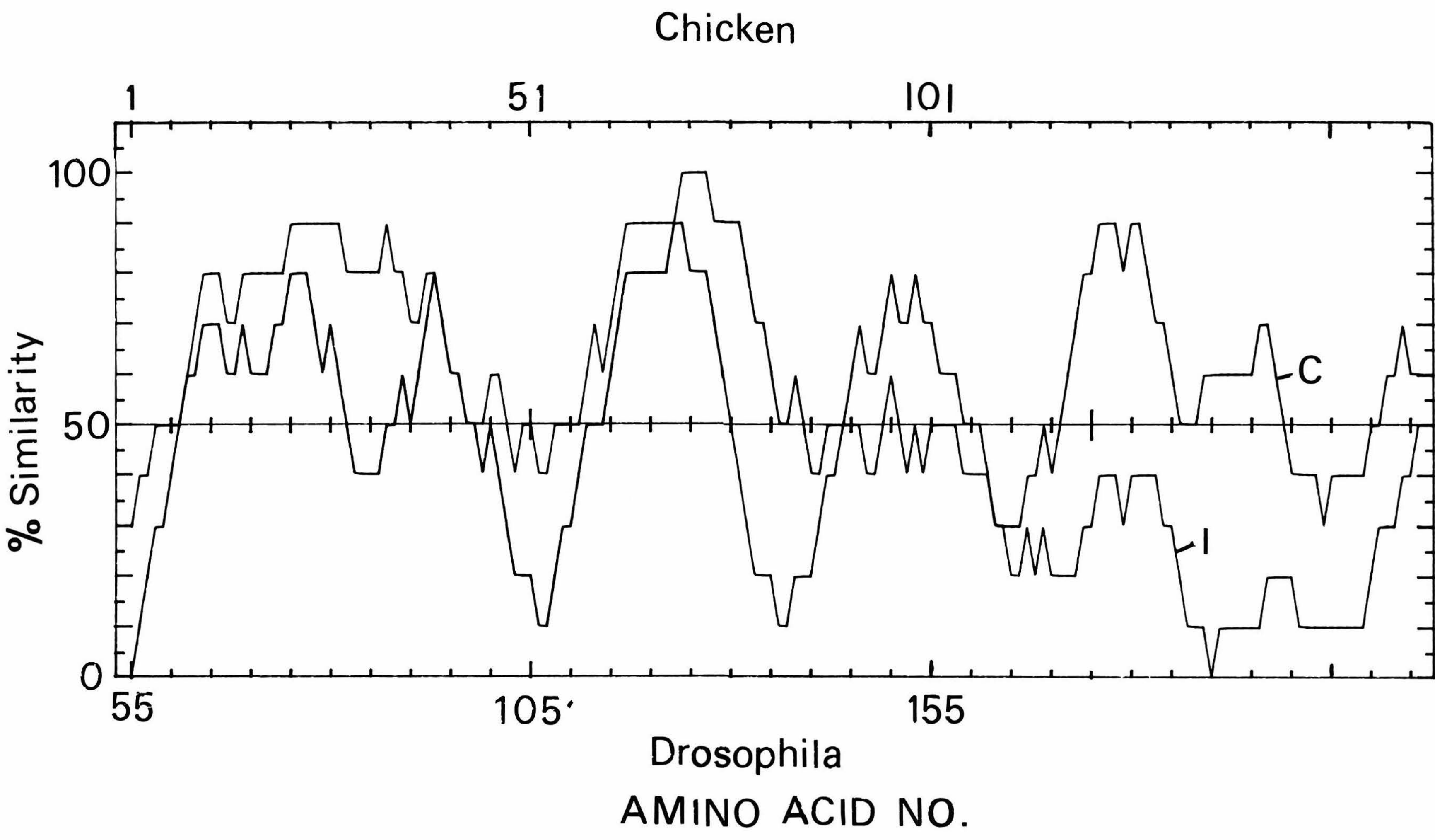
The insert of pCDM38 contains 1111 bp of genomic sequence and a polyA segment of greater than 150 nucleotides. It is colinear with pCDM20 starting at nucleotide 1443 of figure 8. It has the same intron spliced out at position 1619-1840 but extends in the 3' direction for an additional 356 nucleotides beyond the site of poly A addition of pCDM20. This clone does not include the 5' exon of pCDM20 but appears to have terminated reverse transcription prematurely or to have initiated second strand synthesis internally.

Homology to chicken DTNB myosin light chain

Examination of the cDNA sequence revealed an open reading frame of length 666 nucleotides following an ATG codon. Dot matrix sequence comparisons (25) were used to compare the amino acid sequence derived from the *Drosophila* DNA sequence with the protein sequence of the chicken myosin light chains and other myofibrillar components (data not shown). Homology was evident between the *Drosophila* sequence and the members of the troponin C superfamily (2). The *Drosophila* sequence showed a significant match only with the myosin light chain-2 sequence. A histogram showing the homology of the *Drosophila* sequence with the chicken DTNB myosin light chain is shown in figure 9. In this comparison stretches of 10 amino acids are compared.

Figure 9. Similarity comparison of the predicted *Drosophila* protein sequence with the chicken DTNB myosin light chain.

The protein sequences for the *Drosophila* polypeptide and the chicken DTNB light chain were aligned by dot matrix comparisons (data not shown). These histograms plot the percent homology of the two sequences over stretches of ten amino acids. The comparisons were done requiring an exact match of the two sequences (trace I) and allowing for conservative amino acid replacements (trace C). Conservative replacements are any amino acid within the same functional group. Since all matches in I would also be scored in C, the percent similarity in C will always be equal to or greater than I.



Chicken
1
51
101
100
50
0
% Similarity
C
I
55
105′
155
Drosophila
AMINO ACID NO.

A match in all 10 positions would be 100%, in 5 would be 50%, etc. Comparisons were made either requiring identical amino acids for a match (I) or allowing for conservative amino acid replacements (C). Conservative amino acid replacements were considered any amino acid with the same functional side chain (basic, acidic, polar, or non-polar). In the calcium binding domain from amino acid 65 to 101 of *Drosophila* there is 65% (25 of 38) identity with the chicken DTNB myosin light chain (55). The region from amino acid 115 to 134 has an 80% identity (16 of 20) between these sequences. The carboxy proximal regions of these two proteins are less homologous (approximately 25% identity). We have thus designated the gene the *Drosophila* myosin light chain-2. The size of the protein predicted from the DNA sequence is 24 kd, corresponding to the larger product of *in vitro* translation of this gene.

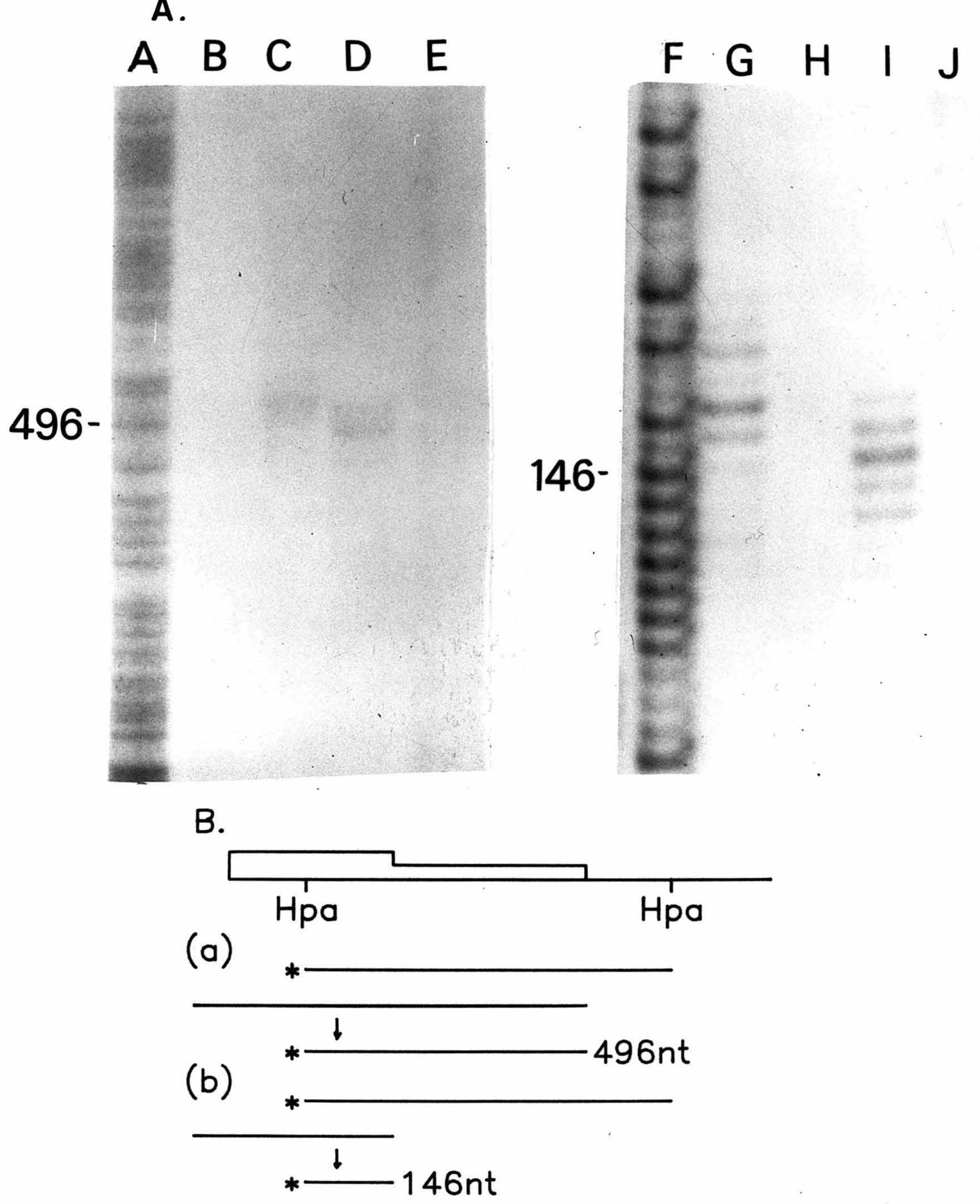
S1 nuclease analysis

To confirm the 3' endpoints of transcripts within pLP5734 an S1 nuclease experiment was performed. Hybrids formed between late pupal RNA and end labeled DNA were digested with S1 and the products separated by electrophoresis on a denaturing acrylamide gel (figure 10). Using a probe which had been labeled at a HpaII site (position 2227), protected fragments of 146 and 496 nucleotides were obtained. These correspond to the position of the polyA tails in the cDNA clones pCDM20 and

Figure 10. S1 nuclease mapping of the 3' ends of myosin light chain mRNAs.

A. Late pupal RNA was hybridized to a probe labeled at the HpaII site at nucleotide 2227. Following S1 nuclease digestion the products were run on a 5% polyacrylamide, 50% urea gel. To size the products accurately two loadings (A-E and F-J) were applied. (A) and (F) purine specific sequence ladder (gift of L.Yu); (B) and (H) tRNA + Miles S1; (C) and (G) 1 μg late pupal RNA + Miles S1; (D) and (I) 1 μg late pupal RNA + Sigma S1; (E) and (J) no S1.

B. Diagrammatic representation of expected fragment size predicted by cDNA clones (a) pCDM38 and (b) pCDM20.


A.
A
B
C
D
E
F
G
H
I
J
496-
146-
B.
Hpa
Hpa
(a)
496nt
(b)
146nt

pCDM38, respectively. We observed heterogeneity of 3 to 5 nucleotides in the length of the fragments protected by nuclease S1. Similar results have been reported previously. In addition, enzymes which were purchased from different commercial sources digested the hybrids to end products which differed from each other by 2 nucleotides. We found that varying the temperature used in the S1 nuclease digestion from 0° to 20° did not alleviate the observed heterogeneity. We believe the different digestion products result from digestion of A-T base pairs formed at the end of the duplex.

S1 nuclease mapping was also carried out using the cDNA clone pCDM20. Alternative patterns of RNA splicing would be revealed by shorter S1 protected fragments than those predicted from the pCDM20 sequence. Probes were generated by 3' end labeling the NcoI site which is at the initiation of translation (CCATGG) in exon 1, and by 5' end labeling the MstII site at the termination of translation (CCTAAGG) within exon 3. No evidence for alternative splicing patterns downstream from exon 1 to the poly A site of pCDM20 and upstream from exon 3 to the 5' end of pCDM20 was observed in these experiments (figure 11).

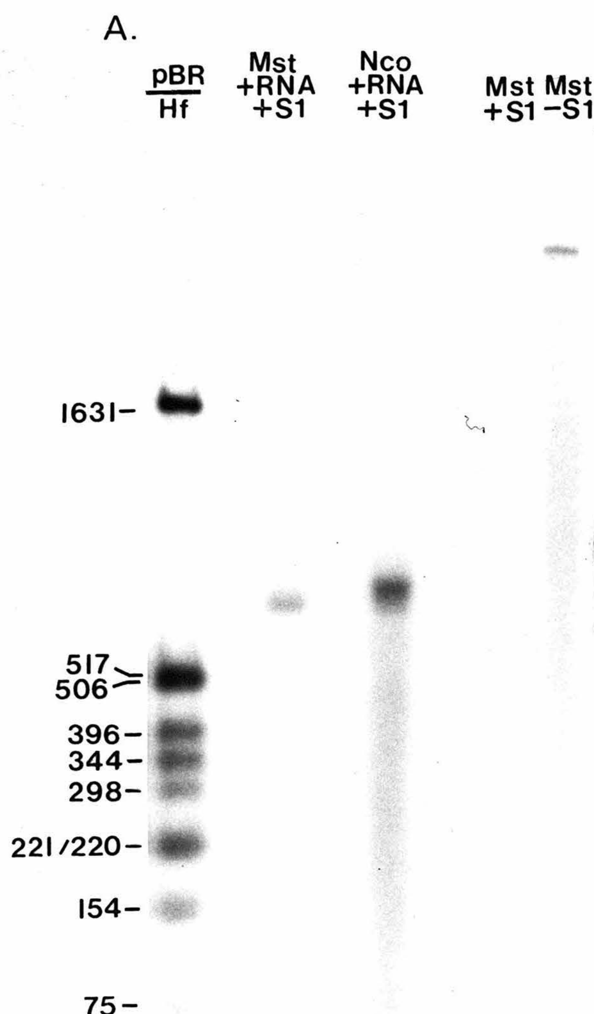
Analysis of transcription initiation by primer extension

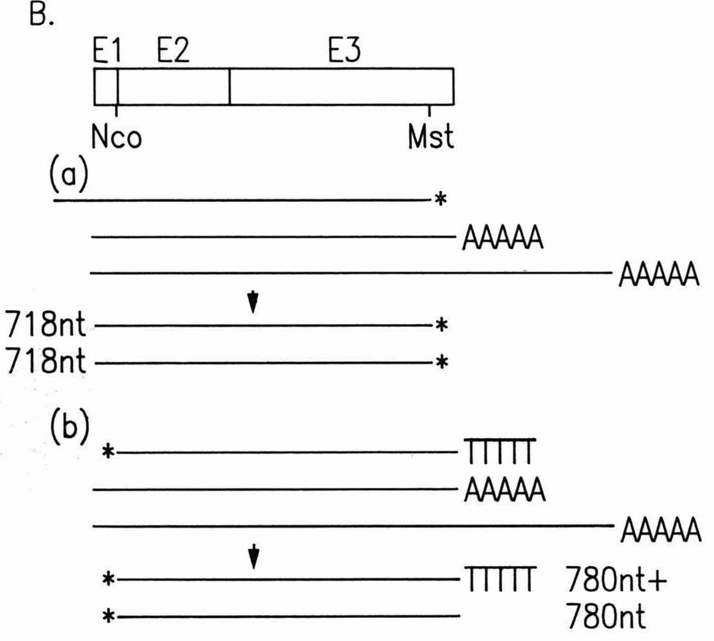
As a further test of the possibility that there was a different first exon spliced onto the two 3' exons, we used a 51 nucleotide fragment from the middle exon (bases 1439-

Figure 11. S1 nuclease mapping of myosin light chain mRNAs with the cDNA clone pCDM20.

A. Late pupal RNA was hybridized to probes generated by 3' end labeling the NcoI site or 5' end labeling the MstII site of pCDM20. Fragments protected from S1 nuclease digestion were analyzed on a 2% alkaline agarose gel. (m) pBR/Hf: pBR322 DNA digested with HinfI, end labeled with ^{32}P; (a) Mst + RNA + S1: 5' end labeled pCDM20 digested with MstII + late pupal RNA + S1; (b) Nco + RNA + S1: 3' end labeled pCDM20 digested with NcoI + late pupal RNA + S1; Mst + S1: the probe in (a) with tRNA and S1; and Mst - S1: the probe in (a) with tRNA but no S1.

B. Diagrammatic representation of expected fragment sizes predicted by cDNA clones. E1, E2 and E3 refer to exons 1,2 and 3. The products shown are those expected for RNAs homologous to pCDM20 and pCDM38 for the (a) MstII end labeled probe and (b) NcoI end labeled probe.


A.
pBR
Hf
Mst
+RNA
+S1
Nco
+RNA
+S1
Mst
+S1
Mst
-S1
1631-
517
506
396-
344-
298-
221/220-
154-
75-

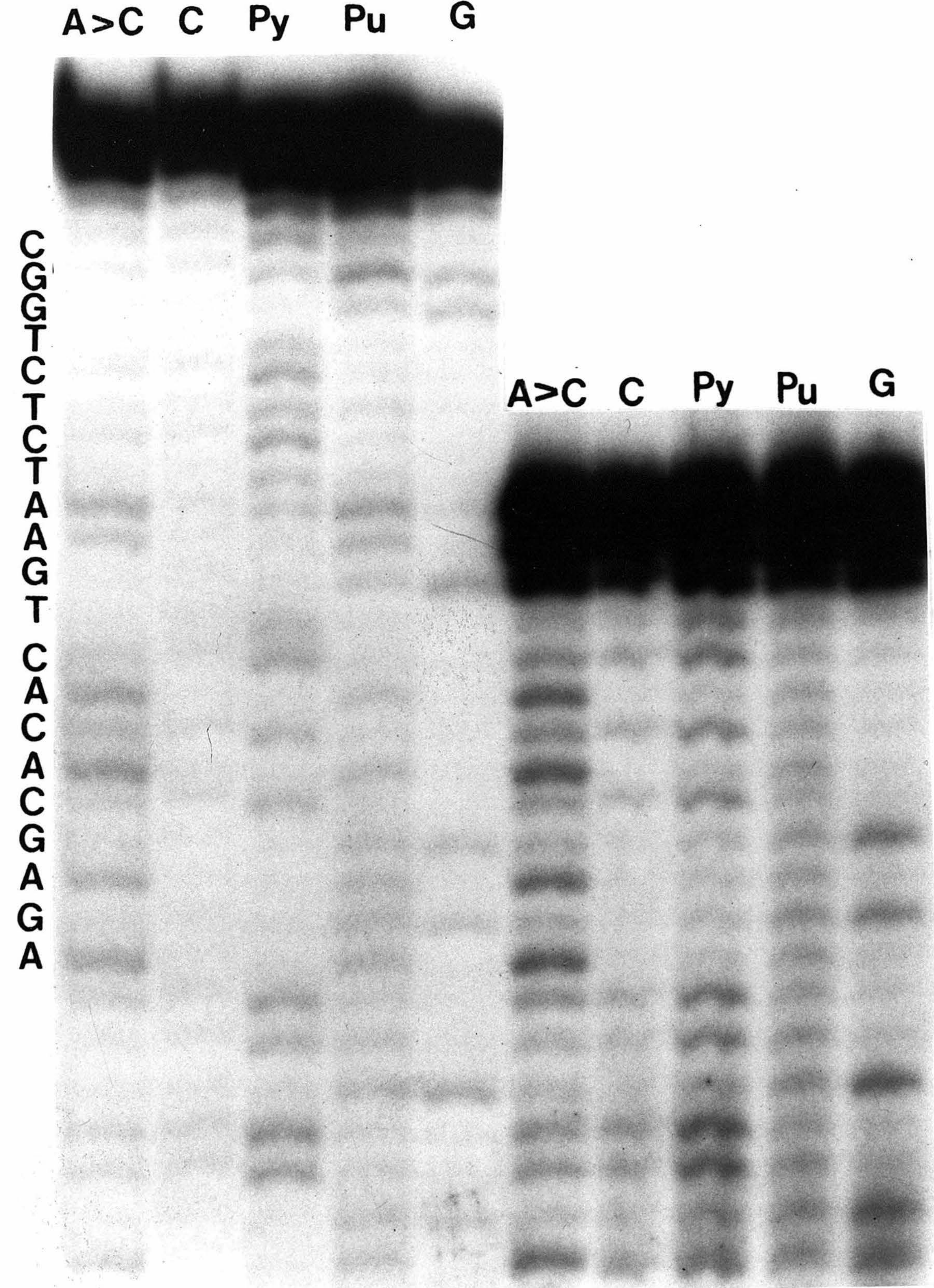

B.
E1
E2
E3
Nco
Mst
(a)
AAAAA
AAAAA
718nt
718nt
(b)
TTTTT
AAAAA
AAAAA
TTTTT
780nt+
780nt

1489) to map the 5' end(s) of the RNA transcripts by primer extension. Five micrograms of the clone pCDM20 was digested with RsaI, labeled on its 5' termini by polynucleotide kinase using γ -^{32}P-ATP and redigested with NcoI. Isolation of a single stranded primer was facilitated because RsaI leaves a blunt end while NcoI has a 4 nucleotide, 5' overhang causing the two strands to separate well on a denaturing gel. The isolated labeled fragment was then used to extend a late pupal RNA template (20). Final products of length 109 and 121 nucleotides were observed. To determine the exact nature of these products the reaction was run preparatively and the two extension products were eluted from a gel. We recovered approximately 10^3 cpm of ^{32}P labeled fragments from each band. These were divided into five reactions each for DNA sequence analysis (36). Chemical modification reactions were allowed to proceed 40% longer than usual to aid in the visualization of shorter fragments. The sample was loaded onto one gel which was allowed to expose to X-ray film for 20 days (figure 12). The result showed that the two sequences have mRNA 5' cap sites separated by 12 nucleotides. To rule out the possibility that this result was an artifact of reverse transcription we performed several experiments. S1 nuclease analysis using a probe labeled at the NcoI site of pLP5734 showed two products as expected (data not shown). Primer extension experiments using different primers also confirmed the positions of the

two start sites.

Figure 12. DNA sequence analysis of primer extension products.

A 51 nucleotide primer from exon 2 was extended to products of 109 and 121 nucleotides in length. These were isolated from a gel and sequenced by the chemical modification technique (36). Base specific cleavage reactions were adenine and cytosine (A>C), cytosine specific (C), pyrimidine specific (Py), purine specific (Pu), and guanine specific (G). The sequence read alongside the larger extension product is the complement of nucleotides 704 to 724 of figure 8.


A>C
C
Py
Pu
G
C
G
G
T
C
T
C
T
A
A
G
T
C
A
C
A
C
G
A
G
A
A>C
C
Py
Pu
G

DISCUSSION

We have cloned a gene from *Drosophila melanogaster* which is abundantly expressed during muscle development. Analysis of the cDNA shows that the gene encodes a myosin light chain-2. The gene is single copy within the genome and located at position 99E on *Drosophila* polytene chromosomes.

There are two structural features of the myosin light chain-2 which have been implicated in function. They are phosphorylation and the binding of calcium ions. Myosin light chain-2 is part of a family of calcium binding proteins including parvalbumin, troponin C, alkali light chains, and calmodulin. These proteins are thought to have evolved by domain duplication followed by gene duplication from an ancestral calcium binding protein capable of binding one calcium ion to a family of proteins capable of binding four calcium ions (2). Through later evolution the ability of some of the domains in these proteins to bind Ca^{++} has been lost. Thus, mammalian myosin light chain-2 has lost the ability to bind Ca^{++} at three binding sites, parvalbumin appears to have lost two, whereas troponin C has retained all four sites (31). Figure 12 is an alignment of the DTNB myosin light chains of rabbit skeletal muscle (12), chicken skeletal muscle (55) and chicken cardiac muscle (35), the myosin light chain 2 of chicken smooth muscle (33) and the regulatory light chain

of scallop (29) with the D. melanogaster myosin light chain sequence. The calcium binding domain of region 1 (amino acids 70 to 101) is designated following the rules of Kretsinger (31). The Drosophila sequence has acidic residues in four of the six chelating vertices (X,Y,Z, and -Z in figure 12) as required. There is also a glycine residue at position -Y which may augment Ca^{++} binding through a hydrogen bond with water (31). The hydrophobic alpha helix residues are denoted by an n. We would thus predict that the Drosophila myosin light chain-2 is capable of binding calcium with high affinity.

The site of phosphorylation has been identified in the rabbit DTNB light chain as serine residue 15 and in chicken smooth muscle as serine residue 19. It has been proposed that phosphorylation is a recent evolutionary event as other members of the troponin C family are not phosphorylated at this site and the scallop regulatory light chain is not phosphorylated. The Drosophila protein has a serine which aligns with the rabbit serine-15 and shows conservation of the preceding two amino acids as well. It has been reported that the D. melanogaster protein is phosphorylated, and that this phosphorylation is necessary for assembly of the myosin filament (24).

A third highly conserved region is in the ancestral calcium binding domain 2 of the aligned myosin light chains, where Drosophila, rabbit and chicken sequences have 14 of 18 identical amino acids. Goodman et al. (21) have

suggested that this striking conservation in sequence for rabbit, chicken and scallop reflects the conserved function

Figure 13. Comparison of the *Drosophila* protein sequence with myosin light chain sequences from other organisms.

The *Drosophila* sequence (DMLC2) is aligned with the chicken skeletal muscle DTNB myosin light chain (CDTNB) (55) the chicken cardiac DTNB MLC (CCARD) (35), the rabbit skeletal muscle DTNB MLC (RDTNB) (12), the chicken smooth muscle MLC (CSMLC) (33), and the scallop EDTA regulatory light chain (SEDTA) (28). Identity to the *Drosophila* sequence is denoted by a "." and a "-" is used to denote the absence of a residue at that position. Above amino acids 81 to 108 of the *Drosophila* sequence the letters X, Y, Z, -Z, -X and -Y denote the position of amino acids which may be involved in calcium binding, n refers to the position of hydrophobic residues which form the adjacent α helices and G and I refer to glycine and isoleucine residues often found in Ca^{++} binding domains (31).

Figure 13

```
DMLC2      1  MADEKKKVKKKKTKEEGGTSETASEAASEAATPAPAATPA   40

DMLC2     41  PAASATGSKRASGGSRGSRKSKRAGSS-VFSVFSQKQIAE   79
CDTNB      1                PKKAKRRAAE...N...M.D.T..Q.   26
CCARD      1                 PKKA.K.VE.G.N...M.E.T..Q.   25
RDTNB      1              PKKAKR.AAAEG...N...M.D.T..Q.   28
CSMLC      1          SSKRKRPQRAKAKTTKAT.N..AM.D.S..Q.   32
SEDTA      1                       ADKAA.G.LTKLP....Q.   19

                                 -Y -X
               n nn   nX Y ZG I  -Zn  nn  n
DMLC2     80  FKEAFQLMDADKDGIIGKNDLRAAFDSVGK-IANDKELDA  118
CDTNB     27  .....TVI.QNR....D.D...ET.AAN.RLNVKNE....   66
CCARD     26  .....TI..QNR..F.D.A...DT.AAL.RLNVKNE..ED   65
RDTNB     29  .....TVI.QNRN...D.E...DT.AAM.RLNVKEED...   68
CSMLC     33  .....NMI.QNR..F.D.E..HDMLA.M..-NPT.EY.EG   71
SEDTA     20  M....SMI.V.R..FVS.D.IK.ISEQL.R-TPD....T.   58

DMLC2    119  MLGEASGPINFTQLLTLGANRMATSGANDEDEVVIAAFKT  158
CDTNB     67  .IK.........VF..-MFGEKLK-..DP..-.IMG...V  103
CCARD     66  .VK..P......VF..-MFGEKLK-.TDP.E-TILN...I  102
RDTNB     69  .-K.........VF..MMFGEKLKK...P..-.ITG...V  106
CSMLC     72  .MS..P......MF..-MFGEKLN-.TDP..-.IRN..AC  108
SEDTA     59  ..K..P..L...MF.S----DKLS-.TDS.E-TIRN..AM   92

DMLC2    159  FDND--GLIDGDKFREMLMNFGDKFTMKEVDDAYDQMVID  196
CDTNB    104  L.P.GK.S.KKSFLE.L.TTQC.R..PE.IKNMWAAFPP.  143
CCARD    103  ..PEGK.H.KA.YIK....TQEGR.SQE.INQMFAAFPP.  142
RDTNB    107  L.PEGK.T.KKQFLE.L.TTQC.R.SQE.IKNMWAAFPP.  146
CSMLC    109  ..EEAS.F.HE.HL..L.TTM..R..DE...EM.REAP..  148
SEDTA     93  ..EQENKKLNIEYIKDL.EDM.NN.NKD.MRMTFKEAPV-  131

DMLC2    197  DKNQID-TAALIEMLTGKGEEEEEEAA               222
CDTNB    144  VAGNV.YKNICY-VI.H-..DK.G.                 166
CCARD    143  VSGNL.YKNLCY-VI.H-...KD                   163
RDTNB    147  VGGNV.YKNICY-VI.H-.DAKD.Q                 169
CSMLC    149  K.GNFNYVEFTR-I.KH-.AKDKDD                 171
SEDTA    132  EGGKF.YVKFTAMIK-.S..D.A                   153
```

of myosin heavy chain binding.

We wish to also note similarities of amino acid sequence of the *Drosophila* protein with vertebrate alkali myosin light chains. The alkali myosin light chains of vertebrate striated muscle are made up of two classes of proteins, LC1 and LC3. In chicken these proteins are identical over their carboxy terminal 141 amino acids. As a result of differential RNA splicing LC1 has 42 amino acids and LC3 has 8 amino acids on their amino termini prior to the shared region (41). Amino acids 11 to 23 of the LC1 protein are AAAPAPAPAPAPA. The *Drosophila* MLC-2 sequence has some striking similarities to this pattern. The amino terminus of the chicken DTNB light chain aligns with amino acid 54 of the *Drosophila* sequence. Within this 53 amino acid extension is the sequence AATPAPAATPAPA (residues 30 to 42) which is identical in 10 of 13 residues (underlined) to the chicken alkali light chain. The function, if any, of this region is unclear, but we find it noteworthy that a region in the alkali light chain of vertebrates is found in the MLC-2 of *Drosophila*. We find no other significant region of homology between these proteins.

The 5' flanking regions of most eukaryotic genes have been shown to possess two signals for transcription. Near nucleotide position -65 from the transcription initiation site is the sequence CCAAT. It has been shown that this sequence is important for high levels of expression of

genes in *in vitro* assays (13). At position -20 to -30 is the conserved sequence TATA which was first described by Goldberg. The *Drosophila* sequence contains a CCAAT sequence 62 and 74 nucleotides upstream of the position of the two mRNA start sites (figure 8). There is no TATA sequence preceding the mapped transcription start. The two start sites of transcription observed with this gene are consistent with the hypothesis that the TATA is important in determining the positioning of the mRNA start site (22). Some other genes in Drosophila (53), yeast (40), and humans (14) have been shown to lack the TATA sequence.

There are two major transcripts from the genomic region of the *Drosophila* MLC-2, with estimated lengths of 1.4 and 1.1 kb (see figure 4) DNA sequencing of a cDNA clone and S1 nuclease analysis of RNA-DNA hybrids show that the longer transcript includes an additional 350 nucleotides of sequence downstream from the termination site of the shorter transcript. The sequence AATAAA is thought to be involved in polyadenylation of mRNA (45). This sequence occurs 21 nucleotides before the poly A sequence of the shorter transcript. The longer transcript terminates 21 nucleotides after the sequence AAAAAAAA.

A question left unresolved in the present study is the relation of the DNA sequence to the two polypeptides encoded by this gene. Hybrid selection and *in vitro* translation produces two protein products with apparent molecular weights of 17 and 26 kd, which comigrate on a

two-dimensional gel with abundant proteins extracted from *Drosophila* indirect flight muscle actomyosin. The open reading frame deduced from the cDNA sequence predicts a protein of length 223 amino acids which would have a molecular weight of 24 kd, apparently corresponding to the larger of the polypeptides observed. Since this is a single copy gene and we find no evidence for alternative RNA splicing patterns, we see no obvious explanation for the two translation products. Apparently identical protein spots were observed by Mogami et al. (39) in gels from adult *Drosophila* myofibrils (their spots 181 and 184), but no functional identifications were carried out. There are several mechanisms not addressed in the current study by which the shorter polypetide could be generated. These include: protease action or other specific modification which is present in both the *in vitro* translation system and *in vivo* in *Drosophila*; initiation of translation at the internal methionine at amino acid position 87 of the present sequence (see figure 8); or initiation elsewhere within the gene. We note, however that the ATG codon at position 767 (amino acid 1 in our protein sequence) has a good consensus sequence for initiation of translation as defined by Kozak (30). The resolution of this problem remains for future work.

Finally we note one additional analogy between the *Drosophila* MLC-2 gene structure reported here, the *Drosophila* alkali myosin light chain gene (S. Falkenthal,

V. Parker and N. Davidson, in press), and the myosin light chain-3 genes of chicken (41), rat (43) and mouse (49). In all cases the initiation ATG codon occurs precisely at the 3' end of the first exon. Thus the intriguing possibility exists that this structural characteristic may apply to myosin light chain genes for higher eukaryotes over a large evolutionary distance.

REFERENCES

1. Adelstein,R. and A.Eisenberg. 1980. Regulation and kinetics of the actin-myosin-ATP interaction. Ann.Rev.Biochem. 49: 921-956.

2. Barker,W.C.,L.K.Ketcham, and M.O.Dayhoff. 1978. The troponin C superfamily, p. 273-284. In M.O.Dayhoff (ed.), Atlas of protein sequence and structure, vol.5. National Biomedical Research Foundation, Silver Springs, Md.

3. Bautch,V.L.,R.V.Storti,D.Mischke, and M.L.Pardue. 1982. Organization and expression of *Drosophila* tropomyosin genes. J.Mol.Biol. 162: 231-250.

4. Benton,W.D., and R.W.Davis. 1978. Screening of λgt recombinant clones by hybridization to single plaques *in situ*. Science 196: 180-182.

5. 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.

6. Bernstein,S.I.,K.Mogami,J.J.Donady,and C.P.Emerson. 1983. *Drosophila* muscle myosin heavy chain is encoded by a single gene located in a cluster of muscle mutations. Nature (London) 302: 23-34.

7. Bolivar,F.,R.L.Rodriguez,P.J.Greene,M.C.Betlach,H.L. Heyneker,H.W.Boyer,J.H.Crosa, and S.Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2: 95-113.

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

9. Brosius,J.,T.J.Dull, and H.F.Noller. 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. Proc.Natl.Acad.Sci.U.S.A. 77: 201-204.

10. Brosius,J.,M.L.Palmer,P.J.Kennedy, and H.F.Noller. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc.Natl.Acad.Sci. U.S.A. 75: 4801-4805.

11. Chacko,S.,M.A.Conti, and R.S.Adelstein. 1977. Effect of phosphorylation of smooth muscle myosin on actin activation and Ca^{2+} regulation. Proc.Natl.Acad.Sci. U.S.A. 74: 129-133.

12. Collins,J.H. 1976. Homology of myosin DTNB light chain with alkali light chains, troponin C and parvalbumin. Nature (London) 259: 699-700.

13. Dierks,P.,A.van Ooyen,N.D.Cochran,C.Dobkin,J.Reiser, and C.Weissman. 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. Cell 32: 695-706.

14. Efstratiadis,A.,J.W.Posakony,T.Maniatis,R.M.Lawn,C. O'Connell, R.A.Spritz,J.K.DeRiel,B.G.Forget, S.M. Weissman, J.L.Slightom,A.E.Blechl,O.Smithies,F.E. Baralle,C.C.Shoulders, and N.J. Proudfoot. The structure and evolution of the human β-globin gene family. Cell 21: 653-668.

15. Falkenthal,S.,V.P.Parker,W.W.Mattox,and N.Davidson. 1984. Drosophila melanogaster has only one myosin alkali light-chain gene which encodes a protein with considerable amino acid sequence homology to chicken myosin alkali light chains. Mol. Cell. Biol. 4: 956-965.

16. 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.

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. Garfinkel,M.D.,R.E.Pruitt, and E.M.Meyerowitz. 1983. DNA sequences, gene regulation and modular protein evolution in the Drosophila 68C glue gene cluster. J.Mol.Biol. 168: 765-789.

19. Garrels,J.I. 1979. Changes in protein synthesis during myogenesis in a clonal cell line. Dev.Biol. 73: 134-152.

20. Ghosh,P.K.,V.B.Reddy,M.Paitak,P.Lebowitz, and S.M. Weissman. 1980. Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. Methods Enzymol. 65: 580-595.

21. Goodman,M.,J.-F.Pechere,J.Haiech,J.G.Demaille. 1979. Evolutionary diversification of structure and function in the family of intracellular calcium-binding proteins. J.Mol.Evol. 13: 331-352.

22. Grosschedl,R.,and M.Birnstiel. 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl.Acad.Sci.U.S.A. 77: 1432-1436.

23. Grunstein,M., and D.Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc.Natl.Acad.Sci. U.S.A. 72: 3961-3965.

24. Hayashi,Y.,and Y.Hotta. 1982. Post-eclosional phosphorylation of Drosophila indirect flight muscle (IFM) myofibrillar proteins. Dev.Growth Differ. 24: 417.

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

26. Jakes,R.,F.Northrup, and J.Kendrick-Jones. 1976. Calcium binding regions of myosin "regulatory" light chains. FEBS Lett. 70: 229-234.

27. Karlik,C.C.,J.W.Mahaffey,M.D.Coutu, and E.A.Fyrberg. 1984. Organization of contractile protein genes within the 88F subdivision of the Drosophila melanogaster third chromosome. Cell 37: 469-481.

28. Kendrick-Jones,J., and R.Jakes. 1977. pp. 28-40 In Riecker,G.,A.Weber, and J.Goodwin (eds.). International Symposium on Myocardial Failure. Springer-Verlag, New York.

29. Kendrick-Jones,J.,E.M.Szentkiralyi, and A.G.Szent-Gyorgi 1976. Regulatory light chains in myosins. J.Mol.Biol. 104: 747-775.

30. Kozak,M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucleic Acids Res. 9: 5233-5252.

31. Kretsinger,R.H. 1980. Structure and evolution of calcium-modulated proteins. Crit.Rev.Biochem. 8: 119-174.

32. Legerski,R.J.,J.L.Hodnett, and H.B.Gray,Jr. 1978. Extracellular nucleases of pseudomonas Bal31 III. Use of the double-strand deoxyribonuclease activity as the basis of a convenient method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucleic Acids Res. 5: 1445-1464.

33. Maita,T.,J.Chen, and G.Matsuda. 1981. Amino-acid sequence of the 20,000-molecular-weight light chain of chicken gizzard-muscle myosin. Eur.J.Biochem. 117: 417-424.

34. Maniatis,T.,E.F.Fritsch, and J.Sambrook. 1980. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press: Cold Spring Harbor,NY.

35. Matsuda,G.,T.Maita,Y.Kato,J.Chen, and T.Umegane. 1981. Amino acid sequences of the cardiac L-2A, L-2B, and gizzard 17,000 M light chains of chicken muscle myosin. FEBS Lett. 135: 232-236.

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

37. Messing,J.,and J.Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19: 269-276.

38. Mitchell,M.K., and A.Mitchell. 1964. Mass culture and age selection in Drosophila. Dros.Inf.Ser. 39: 135-137.

39. Mogami,K.,S.C.Fujita, and Y.Hotta. 1982. Identification of Drosophila inidrect flight muscle myofibrillar proteins by means of two-dimensional electrophoresis. J.Biochem. 91: 643-650.

40. Montgomery,D.L.,D.W.Leung,M.Smith,P.Shalit,G.Faye, and B.D.Hall. 1980. Isolation and sequence of the gene for iso-2-cytochrome c in Saccharomyces cerevisiae. Proc. Natl.Acad.Sci.U.S.A. 77: 541-545.

41. Nabeshima,Y.,Y.Fujii-Kwriyama,M.Murimatsu, 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.

42. O'Farrell,P.H. 1975. High resolution two-dimensional electrophoresis of proteins. J.Biol.Chem. 250: 4007-4021.

43. 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. 21: 13595-13604.

44. Poncz,M.,D.Solowiejczyk,M.Ballantine,E.Schwartz, and S.Surrey. 1982. "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc.Natl.Acad.Sci. U.S.A. 79:4298-4302.

45. Proudfoot,N.J.,and G.G.Brownlee. 1976. 3' non-coding region sequences in eukaryotic messenger RNA. Nature (London) 263: 211-214.

46. Pulliam,D.L.,V.Sawyna, and R.J.C.Levine. 1983. Calcium sensitivity of vertebrate skeletal muscle myosin. Biochem. 22: 2324-2331.

47. Ricciardi,R.P.,J.S.Miller, and B.E.Roberts. 1979. Purification and mapping of specific RNAs by hybridization-selection and cell-free translation. Proc.Natl.Acad.Sci.U.S.A. 76: 4927-4931.

48. Rigby,P.W.J.,M.Dieckmann,C.Rhodes, and P.Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J.Mol.Biol 113: 237-251.

49. 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 chain 1 and 3, a second locus corresponds to a related pseudogene. Cell 39: 129-140.

50. Rozek,C.E., and N.Davidson. 1983. Drosophila has one myosin heavy chain gene with three developmentally regulated transcripts. Cell 32: 23-34.

51. Sanchez,F.,S.L.Towbin,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.

52. Sanger,F.,S.Nicklen, and A.R.Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc.Natl.Acad. Sci.U.S.A. 74: 5463-5467.

53. Snyder,M., and N. Davidson. 1983. Two gene families in a small region of the Drosophila genome. J.Mol.Biol. 166: 101-118.

54. Snyder,M.,M.Hunkapiller,D.Yuen,D.Silvert,J.Fristrom, and N.Davidson. 1982. Cuticle protein genes of Drosophila: structure, organization, and evolution of four clustered genes. Cell 29: 1027-1040.

55. Suzuyama,Y.,T.Umegane,T.Maita, and G.Matsuda. 1980. The amino acid sequence of the L-2 light chain of chicken skeletal muscle myosin. Hoppe-Seyler's Z. Physiol. Chem. 361: 119-127.

56. Vieira,J., and J.Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19: 259-268.

57. Weaver,R.F., and C.Weissmann. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. 7: 1175-1193.