Structure and expression of a muscle specific gene which is adjacent to the *Drosophila* myosin heavy-chain gene and can encode a cytochrome *b* related protein

Robin J. Levin, Patricia L. Boychuk, Colleen M. Croniger, Jeffrey A. Kazzaz⁺ and Charles E. Rozek

Department of Biology, Case Western Reserve University, Cleveland, OH 44106, USA

Received February 13, 1989; Revised and Accepted June 23, 1989

EMBL accession no. X15008

ABSTRACT

We report the characterization of a transcription unit at the second chromosome locus 36B, designated TU-36B which is adjacent to the 3' end of the *Drosophila* myosin heavy-chain (*Mhc*) gene. We have isolated and sequenced a complementary DNA clone and the region of genomic DNA which represents this gene. The sequencing studies reveal that this gene contains one intron, the mRNA is 1480 nucleotides in length, the TU-36B mRNA is transcribed in an orientation opposite to the *Mhc* mRNAs, and the poly(A) addition site of this gene is 99 nucleotides downstream of poly(A) addition site A-2 of the *Drosophila* *Mhc* gene. The mRNA contains a continuous open reading frame which can encode a protein product of 47,000 daltons in molecular weight. The proposed protein shares homology with cytochrome *b* proteins. Comparison of *in situ* hybridization of *Mhc* specific and TU-36B specific probes to tissue sections demonstrates that both mRNAs are predominantly transcribed in the same muscle tissues of the developing fly.

INTRODUCTION

Numerous genetic loci have been identified in *Drosophila* and an increasing number of these loci have been isolated and characterized on a molecular level. In many cases it has been demonstrated that these genetic loci are complex and often contain one or more transcription units which encode various gene products (1, 2, 3, 4, 5, 6, 7). Within some genetic loci, the transcription units are functionally related and coordinately transcribed (2, 4, 7) while in other genetic loci, the transcription units are not functionally related nor are they coordinately transcribed (8, 5).

One example of a gene locus in *Drosophila* where related genes are clustered and coordinately transcribed is the 88F locus on the third chromosome which encodes actin, tropomyosin, and tropomyosin-related genes (1, 9, 10). A number of flightless mutations which genetically map at the 88F locus have been isolated (11). The DNA alterations responsible for several of these flightless mutations have been characterized and are insertions, deletions, or single base pair changes. Several of these mutations have been identified within the actin gene (10, 12, 13) while others map within a tropomyosin gene (3).

Mogami and Hotta (11) also identified flightless mutations which map at or near the

second chromosome locus 36B, which has been shown to contain the myosin heavy chain gene (14, 15). The DNA lesions associated with four of these flightless mutations have been characterized and are deletions or insertions of DNA in the 5' end of the *Mhc* gene (16). Other flightless mutations which map at or near the *Mhc* locus remain uncharacterized. These uncharacterized mutations may result from subtle molecular changes in the *Mhc* gene which have not yet been detected, or the mutations may reside in transcription units which are near the *Mhc* gene and encode muscle related polypeptides.

To further characterize chromosomal locus 36B and flightless mutations whose primary alterations are unknown, we have initiated a search to identify and characterize transcription units which are near the *Mhc* gene. In the present communication we detail the characterization of a transcribed region adjacent to the 3' end of the *Mhc* gene. The mRNA transcribed from this gene is expressed in the same muscle tissues as the *Mhc* gene and can encode a protein with a molecular weight of 47,000 daltons, which shares homology with cytochrome b proteins.

METHODS AND MATERIALS

Materials

Restriction enzymes, the Klenow fragment of DNA polymerase I, alkaline phosphatase and polynucleotide kinase were obtained from Boehringer Mannheim or Bethesda Research Laboratories. Avian myeloblastosis virus (AMV) reverse transcriptase was obtained from Seikagaku America, Inc. RNA size standards were from Bethesda Research Laboratories. Sequenase was obtained from United States Biochemicals. ³²P - nucleotide - 5'-triphosphates (3000 Ci/mmol) were obtained from ICN. ³⁵S - deoxyadenosine 5' - [alpha-Thio] Triphosphate (1200 Ci/mmol) and [gamma - ³²P] - adenosine 5' - Triphosphate (6000 Ci/mmol) were obtained from NEN.

Screening the Drosophila cDNA Library

Two *Drosophila melanogaster* cDNA libraries cloned into lambda gt10 were screened (17). These libraries were constructed from 12-24 hr. embryonic RNA and 7-9 day late pupal RNA. Libraries were screened by differential plaque hybridization using the methods of Benton and Davis (18) and St. John and Davis (19).

RNA Blotting

Total RNA were isolated as previously described (20) and was electrophoresed in agarose gels containing 2.2M formaldehyde, transferred to nitrocellulose, and hybridized with ³²P -labeled probes as described (15).

Isolation of DNA, Restriction Endonuclease Mapping and Gel Electrophoresis of DNA

Bacteriophage DNA was isolated as described by Snyder et al. (7). Plasmid DNA was isolated as described by Fyrberg et al. (20). Restriction endonuclease maps of cloned DNA were generated by single and double digests with various restriction endonucleases and agarose gel electrophoresis.

DNA Sequencing and Analysis

Genomic DNA and cDNA restriction fragments were subcloned into M13 vectors. single-stranded template DNA was isolated, and both strands of DNA were sequenced according to standard methods (21, 22). Areas with GC compressions were resolved using 7-deaza-dGTP or dITP (23).

The sequence was assembled and analyzed by computer using the Intelligenetics PC/GENE, the University of Wisconsin Genetic Corporation Group (UWGCG), and Beckman's Microgenie software packages. The Genbank, EMBL, National Biomedical Research Foundation Protein Identification Resource (NBRF/PIR) and Swiss-Prot databases were searched for homology.

Isolation of Restriction Fragments for Cloning and Labeling of Fragments for use as Hybridization Probes

DNA was digested with the appropriate restriction enzymes and run on agarose gels and recovered by electroelution (24). The DNA fragments were uniformly labeled using random oligonucleotide primers and ^{32}P -labeled nucleotide triphosphates (25).

Construction of Transcription Vectors and Synthesis of Complementary RNA Probes

Restriction fragments were cloned into PT-7 transcription vectors (United States Biochemical) in an orientation which would result in the synthesis of complimentary RNA (antisense) to *Mhc* and TU-36B mRNAs respectively. A negative hybridization control (sense RNA for *Mhc*) was synthesized from a transcription vector which contained a *Mhc*-specific restriction fragment cloned in the opposite orientation.

Linearized transcription vectors (80 ug/ml) were incubated in 40mM Tris-HCL, pH 8.0, 15mM MgCl₂, 5mM dithiothreitol, 10uM ribonucleotide triphosphates, 0.5 mg/ml bovine serum albumin, 20 units of T7 RNA polymerase (donated by David Samols), and 100uCi of 2 different 3H-ribonucleotide triphosphates (30uCi/mmol, ICN) at 37°C for 30 mins. The reaction was then treated with DNase I to remove the DNA template and chromatographed over a Sephadex G-50 column developed in 50mM Tris-HCL, pH 7.6, 250mM NaCl, 0.2% SDS, 2mM EDTA.

Figure 1. Restriction maps for cDNA and genomic DNA clones representing the TU-36B transcript and hybridization of RNA gel blots with sequences homologous to *Mhc* and TU-36B

- A. Shown in the diagram are the restriction maps of TU-36B cDNA and genomic DNA. The restriction maps are aligned to indicate the common restriction sites. The position of exon A of the *Mhc* gene is indicated below the DNA map. Also indicated is the 1.55 kb EcoRI genomic DNA fragment used for hybridization in B.
- B. Total RNA (10 μ g) isolated from 48 hour larvae (1) and newly enclosed adults (2) were electrophoresed in a 1% agarose gel containing 2.2 M formaldehyde. The RNA was transferred to nitrocellulose and hybridized with a probe representing the 1.55 kb EcoRI fragment shown in A.

In situ Hybridization to Tissue Sections

Sixteen hour old embryos were dechorionated with 4% hypochlorite for 1-2 mins. and then washed thoroughly with 130 mM NaCl, 7 mM Na₂HPO₄, 3 mM

Na₂H₂PO₄-H₂O (PBS) and blotted dry. They were then incubated in 450 μ l PBS, 60 μ l 37% formaldehyde, and 4.5 ml heptane for 10 mins. with moderate shaking. The aqueous phase was removed, replaced with 5 ml of methanol and agitated vigorously for 5 mins. The devitellinized embryos were then washed twice in methanol and placed in 70% ethanol for storage at -20°C. Seventy-five hour old pupae were prepared in the following manner. The pupal cases were removed by dissection, and the pupae were fixed in 4% paraformaldehyde/PBS for 30 mins. Embryos and pupae were prepared as follows for embedding in paraffin. The samples were dehydrated consecutively through 30%, 50%, 70%, 90%, 95% and 100% ethanols, cleared in xylenes and embedded in paraffin as described (25).

Paraffin sections of 15 Hr. embryos (7 μ m) and 75 hr. pupae (10 μ m) were mounted on poly-lysine coated slides, dewaxed in xylenes and rehydrated through graded ethanols (100%, 95%, 80%, 60%, 30%). Sections were then incubated at room temperature in 0.2M HCl (20 mins.), double distilled H₂O (5 mins.) and PBS (1 min). The sections were postfixed in 4% paraformaldehyde which was dissolved in PBS (20 mins.), rinsed in PBS (1 min), acetylated with 0.25% acetic anhydride, 0.1 M triethanolamine, pH 8.0 (10 mins.), rinsed in PBS (1 min.) and dehydrated through ethanols as above (26).

Sections were hybridized with 3H-labeled cRNA probes at a concentration of 15,000 dpm/ μ l. Hybridization conditions and posthybridization treatment were as previously described (26) with the exception that the hybridization reactions were done at 50°C for 48 hrs.

The slides were coated with Ilford K5.D emulsion diluted 1:1 with H₂O and exposed for 3 weeks at 4°C. After the proper exposure time the slides were developed for 2-3 mins. in Kodak D19:H₂O 1:1, 30 seconds in Kodak Stop, and 2 mins. in Kodak Fixer. After thorough rinsing with water the slides were lightly stained with 0.5% Methylene Blue, dehydrated through 30%, 50%, 70%, 90% and 100% ethanols, cleared in xylene and mounted in CoverBond.

RESULTS

Identification of a Transcription Unit Adjacent to the *Drosophila* Mhc gene.

We have initiated a search for transcribed regions which are in close proximity to the *Mhc* gene of *Drosophila*. Our strategy has been to use labeled genomic DNA restriction fragments which are contiguous with the *Mhc* gene as hybridization probes to RNA gel blot filters. We began this search downstream of the 3' end of the *Mhc* transcription unit since we have previously deduced the structure of exons at the 3' end of the *Mhc* gene (27). We uniformly labeled a 1.55 kb EcoRI genomic DNA fragment (Fig. 1A) which we have previously demonstrated is single-copy in the *Drosophila* genome

5'		
← Hind III ~ 50nt		
gtggctctcactcttatacaggaaaaagtaacagtagtgaataagagaatacaaacgatacttacattgaagggcgccacgattgcgac		80
tagagtatgtgtgcatgtcatttattttccataagtcagaactttgtttaaaagaaaagtataagataaatacttacacttatatt		178
acttacttatacttatactatactttagcttcataaagtaattgtgtacaagattacattatgtctaccaagcaagaaaaatgaact		267
caacgtgtagccatgtcaataaacccatagtaaatcgtgaattttatccctacaaattattttattttatggagctcttagagataaga		356
ctacccccacaactctcaaaagctgattgctttcggttagaccacgcgacgcaagcgagcctcaaaagtcgcgcataccataacatt		445
gttatggccaccgcactgacctattgtcgccgaagctgctcgaaccgccatgtggtgcttagacatttcaccaaatgaattagcgat		534
tagaacatacctgtctaaaaatagtgtagaattaaatagacgtaatagtaaaagcggcagctgcgtctctctgataacacggcg		623
cttatcaaaatggaaacgttttggtaaactctgattttcggtttataataatcaaaagagatttgcgctacacataaaataacggga	Pvu II	712
ccactgcgggtgctcttAGTTAGTCGTTGGAATCCACGCGAAAGGTGCTCGGATCGGTAAAGTTTGTCACAGACGCTCAAGGATCA	Aat II	801
MetValIleGluGluTrpLysLysSerGlyIleAlaThrLysPheProThrTyrArgAsnSerAlaLeuIleThr		891
AAAGGACTCGGCAATATGGTAATCGAGGAGTGAAAAAGAGTGGCATTGCCACCAAGTTTCCACATACCGAAATTCGGCCCTAATCACC		981
ThrHisSerTrpGlnLysGlyLysArgGlnAspAspGlyAlaGluGlyLeuTrpArgIleAsnAspGlyIleTyrAspPheThrSerPhe		1071
ACACACAGCTGGCAAAAGGGTAACCGGACGACGATGGAGCCGAAGGACTGTGGCGCATCAACGATGGCATCTATGACTTCACCTCGTTC	Pvu II	1161
IleAspLysHisProGlyGlyProPheTrpIleArgGluThrLysGlyThrAspIleThrGluAlaPheGluAlaHisHisLeuThrThr		1251
ATTGACAAGCATCTGGCGGTCCTTCTGGATTCTGAGACCAAGGGCACCACATCACCAGAGCCTTTGAGGCACATCATCTGACCACG		1341
AlaProGluLysMetIleAlaLysTyrLysValArgAspAlaAlaGluProArgIleTyrThrLeuThrLeuGluGluGlyGlyPheTyr		1431
GCTCCCGAAAGATGATCGCCAAATACAGGTGAGGGATGCGGCTGAACCGAATCTACACACTGACCTTGAGGAGGGCGGATCTAC		1521
LysThrLeuLysGluArgValArgGluGlnLeuLysThrIleAspLysArgProLysLysSerAsp		1612
AAGACGCTAAAGGAGCGAGTACGCGAGCACTAAAGACCATAGACAGCGGCCGAAAGAGAGTGATgtgagtgttaattcccaaaaa	Xba III	1702
agaataccttattgaaatggcggtgcgttagatttcatctcagtttaagtaattcttatcagttacttattgataggagtccat		1792
attaagttgaactgccggctcataataaattttactaactaaaacatctaattctacatttcacgtgtcgcagcctaataattatgc		1882
ccccactgtattcggcgccgtgtgtgattatatttaattatttgcacgctgaaccctaattaatgataacttttaaaatttatc		1972
LeuIleHisLeuGlyLeuValValSerLeuTyrLeuLeuGlyIleAlaSerAla		2062
atttatattataatataataatatacatttcttttagCTCATTACCTGGGCTCGTTGTCGCTCTACCTGTTGGGCTACCGCAGTGCC		2152
LysTyrAsnSerLeuLeuAlaLeuValAlaSerValAlaLeuCysTrpThrValIleValSerHisAsnTyrPheHisArgArgAsp		2242
AAATACACAGCCTTTTGGCTTTGGCTTGGCGAGTGTGGCCCTCTGCTGGACGGTGATCGTGTCCACACACTACTTTTCATCGTGGGAC		2332
AsnTrpGlnMetTyrAlaPheAsnLeuGlyMetMetAsnPheAlaProTrpArgValSerHisAlaLeuSerHisHisThrTyrProAsn		2421
AACTGGCAGATGTACGCTTTAATCTGGGCATGATGAACCTTGGCGCTTGGCGTGTCCTGCTGTCGCCACACACCTATCCGAAT	Eco	2511
SerTyrLeuAsnLeuGluLeuSerMetPheGluProLeuLeuCysTrpValProAsnProHisIleLysSerLysLeuMetArgTyrVal		2601
TCGTATCTCAATCGGAACCTCGCGATGTTGCAACCTTGTGCTGTGGTGCCCAATCCGACATCAAGGCAAGTTGATCGCTTATGTG	R1	2671
SerTrpValThrGluProValAlaTyrAlaLeuAlaPhePheIleGlnMetGlyThrArgIlePheTyrSerLeuArgHisThrAsnIle		
TCCTGGGTACGGAGCGGTGGCCTATGCCCTCGCTTCTCATACAAATGGGCACTCGTATTTCATTTCGCTGCGCCACACCAACATC		
LeuTyrTrpHisAspLeuLeuProLeuThrIleProIleAlaIleTyrLeuGlyThrGlyGlySerLeuGlyIleTrpIleCysValArg		
CTCTACTGGCAGCAGCTACTGCCACTGACTATACCATTTGCCATATACCTGGGAACCGGGGATCTTTAGGCATTGGATCTGCGTGC		
GlnTrpLeuAlaMetThrSerIleAlaSerPheSerPheCysLeuIleGlyLeuAsnAlaAlaHisHisAspProGluIleTyrHisGlu		
CAGTGGCTCGCGATGACGTCGATAGCCAGCTTATGCTTCTGCTTGTGGTCTGAATGCGGGCGCATCAGATCCGGAATCTATCAGCAG	Aat II	
GlyAspAlaAsnArgGluAspGlyAspTrpGlySerSerArgTrpThrArgLeuSerThrAlaValIleSerSerGlyArgSerSerTrp		
GGTGATGCCAATCGTGAAGATGGTACTGGGGCTTCCAGGTGGACAGATTATCGACCGGGTGATCTCAAGTGGTCGAGTCTCTGG		
CysSerProThrLeuValThrMetCysSerThrIleSerSerArgProTrpIleMetAlaSerCysProProSerIleArgCysPheThr		
GTCTACCCACTTGTGGTACCATGTGCTCCACCATCTCTCCGACCTGGATCAGCTCTGCCCCCTCTATCCGGTGCTTTACC	Bst E II	
ArgLeuSerMetSerSerArgValIleSerAlaSerAlaThrThrSerSerThr***		
AGACTCTCGATGAGTTCAGGGTCATCTCCGAGTGCAACCACATCGAGCATGATTGGCCAGCACAAGCAACTATTGGCGATCGAG	Fsp I	
CCCAATCCAGGGCTCTGGGGCGGGAAGTAGGACTTAATTGTATGGGGCTTAAGTAAAGTAAATAATATATCCACACATCTTAT		
TAGTTTACaaagacctgtgacaatttcttaacagtagtaccagtaaatatgaatgaatgtccagtcgaatgattgatcaggaa		
aataatacacatttgttaaatgatttcatatttttgtgctttttgtatacttttcatattagtgtca 3'	Poly A	
	Poly A-2 of MHC	
	Exon A MHC gene	

(15), and separately hybridized this probe to blots of total RNA isolated from larval (48 hours after egg deposition) and early adult stages of development. We detect the expected *Mhc* mRNAs (27) and an additional RNA which is approximately 1.6 kb in length with this hybridization probe in both stages of development examined (Figure 1-B). Previously, we had failed to detect *Mhc* mRNA in adults; however, our previous RNA samples were isolated from flies which were greater than five days old. This suggests that *Mhc* mRNA is unstable in late adult stages of development which coincides with the observation that adult flight muscle deteriorates with age (28). The hybridization probe also detects several other size classes of RNA during larval and early adult stages of development. We do not know the identity of these RNAs; however, we note that they are not detected in poly(A)+RNA (data not shown). Since the RNA used for RNA blotting experiments was total RNA, we may be detecting precursors to the respective RNAs or, alternatively, other RNA species which have limited regions of homology with the hybridization probe.

In order to determine the structure of the 1.6 kb RNA we used the 1.55 kb hybridization probe to screen two recombinant cDNA libraries cloned into lambda gt10 (17). One library was constructed from first and second larval instar poly(A)+ RNA, and the second library was constructed from late pupal (7 to 9 day old) poly(A)+ RNA. Approximately 5×10^5 recombinant phage were plated and screened from each library. Duplicate library filters were hybridized with a labeled restriction fragment containing only *Mhc* sequences to identify clones representing only *Mhc* mRNAs. By this criteria, six

Figure 2. The genomic sequence of DNA representing the TU-36B gene and the proposed open reading frame of the deduced amino acid sequence

DNA was sequenced by the method of Sanger et al., (21). The genomic sequence represents DNA beginning near the HindIII restriction endonuclease site and extends into DNA representing exon A of the *Mhc* gene (see Figure 1A). Nucleotides depicted in capital letters represent the TU-36B cDNA sequence. The position of the intron was established by comparing genomic DNA sequence with the cDNA sequence, and the intron as well as intergenic regions are denoted by lower case letters. The poly(A) addition site (following nucleotide 2519), the consensus polyadenylation signal AATAAA (nucleotides 2487-2492) and the start site of transcription (↑, nucleotide 730) are indicated. A potential TATA box (nucleotides 694-708) and two potential CAAT boxes (nucleotides 629-634 and 678-683) are indicated upstream of the start site of transcription. The position of the most 3' polyadenylation site (poly A-2, preceding nucleotide 2619 of the *Mhc* gene is shown - (Note since TU-36B mRNA is transcribed in an opposite orientation to the *Mhc* mRNAs, the *Mhc* sequence presented represents the complement to the *Mhc* mRNA). Consensus splice junction sequences are underlined in the intron as well as a putative consensus branch point sequence (TTAAT). The deduced amino acid sequence begins at the first methionine codon (M) (nucleotides 817-819), and a second potential initiating methionine codon (M) is also indicated (nucleotides 1084-1086). Various restriction endonuclease recognition sites are indicated below their recognition sequences which were used to define the structure of the TU-36B gene.

recombinant DNA clones were isolated from the first and second larval instar cDNA library and four recombinant DNA clones from the late pupal cDNA library which only hybridize with the 1.55 kb EcoRI fragment. DNA was isolated from the individual clones and subjected to restriction enzyme analysis. The cDNA clones contained inserts ranging from 0.7 kb to 1.5 kb and all cDNA clones contained several common restriction fragments (data not shown). We chose to further characterize the cDNA clone containing the largest insert (1.5 kb) which was isolated from the first and second instar cDNA library. Shown in Figure 1A are the restriction maps of the cDNA clone and the genomic DNA homologous to this cDNA. The common restriction enzyme recognition sites present in the two cloned DNAs are indicated. The orientation of the cDNA clone relative to the genomic DNA was deduced from the order of the unique restriction endonuclease sites shown in Figure 1A. We have tentatively named this cDNA and the mRNA it represents "transcription unit at chromosome locus 36B" (TU-36B). To our knowledge this is the first transcription unit identified and characterized in detail which is in close proximity to the 3' end of the *Mhc* gene. Figure 1A also illustrates the close proximity of the final 3' exon of the *Mhc* gene (Exon A) relative to the TU-36B transcription unit.

Structure of the TU-36B Gene

The structure of the TU-36B gene was determined by comparing the sequence of the TU-36B cDNA clone to the sequence of genomic DNA homologous to the cDNA clone. The genomic sequence of the TU-36B gene is shown in Figure 2. The genomic sequence begins approximately fifty nucleotides from the Hind III site (Figure 1A) and extends 2671 base pairs downstream into exon A of the *Mhc* gene (27). The entire sequence of the cDNA clone (1480 nucleotides) was determined and compared to the genomic sequence to deduce the structure of the TU-36B gene. The sequencing studies are summarized below.

3' End of the TU-36B RNA. The structure of the 3' end of TU-36B RNA was determined as follows. An EcoRI restriction fragment of the cDNA clone was subcloned into an M13 vector and sequenced. The sequence of one end of the fragment contained a stretch of polyadenylic acid (poly(A)) that was eighteen nucleotides in length. The poly(A) stretch was contiguous with an EcoRI recognition sequence and 110 nucleotides upstream of the poly(A) stretch was a Fsp I restriction endonuclease recognition sequence. The restriction analysis of the cDNA clone and genomic DNA (Figure 1A) revealed a single unique Fsp I restriction site in both clones. The close proximity of an Fsp I recognition sequence to an EcoRI recognition sequence (within 130 nts) suggests that the EcoRI site identified by sequencing represents one of the synthetic EcoRI linkers added to the double-stranded cDNA during cloning. This cDNA sequence was then compared to the genomic DNA sequence spanning the Fsp I site. The genomic

sequence was identical with the cDNA sequence up to nucleotide 2523 at which point the sequences diverged. The site of divergence is coincident with the poly(A) stretch found in the cDNA clone. Thus, this sequence information defines the right end of the cDNA insert as shown in Figure 1A. We propose that the site of poly(A) addition (Figure 2) is at or near the cytosine residue at position 2319 of the genomic sequence for the following reasons. First, the sequence C(A)_N conforms with the consensus sequence for poly(A) addition (29); second, twenty seven nucleotides upstream from the cytosine residue is a consensus polyadenylation sequence AATAAA (30); and third, nine nucleotides downstream of the cytosine residue is a small G/T cluster (31). None of our data precludes the possibility that up to four of the adenosine residues following the deoxyribosylcytosine in the cDNA clone may be derived from transcription of the genomic DNA rather than from post-transcriptional polyadenylation.

Thus, identification of the structure and location of the 3' end of the TU-36B gene and comparison of this sequence to the previously determined sequence at the 3' end of the *Mhc* gene indicates that the TU-36B poly(A) addition site is only ninety-nine nucleotides downstream of poly(A) addition site A-2 of the *Mhc* gene (27), and that their respective mRNAs are convergently transcribed.

Exon/Intron Structure and Start Site of Transcription of the TU-36B Gene.

Comparison of the cDNA and genomic sequences identifies a single intron within the TU-36B gene. Examination of Figure 2 shows that the splice junction donor and acceptor sequences match the consensus sequences for these RNA processing signals. In addition, at nucleotides 1539-1543 is a putative consensus branch point sequence (TTAAT) preceding the consensus acceptor sequence (32, 33). Based on nuclease S-1 analysis using genomic DNA clones hybridized with RNA isolated from various stages in development and restriction mapping of the remaining TU-36B cDNA clones which contained smaller inserts, we have discovered no indications of alternate RNA processing for the TU-36B mRNA (data not shown).

Primer extension analysis to map the start site of transcriptions revealed two major extension products 59 and 60 nucleotides longer than the single-stranded primer with both larval and late pupal RNA (data not shown). The extended fragments coincide (+/- one nucleotide) with the first nucleotide of the cDNA sequence (nucleotide 730 in Figure 2). Inspection of genomic sequences surrounding this position reveal that the sequence is an almost identical match (TTCAGTT) with the proposed initiation sequence at the 5' end of *Drosophila* mRNAs (ATCAG/TTC/T) (34, 35). Thus, this result suggests that a single initiation site of transcription is at or near nucleotide 730 (Figure 2), and that the cDNA clone sequenced contains a full-length representation of the TU-36B mRNA.

Upstream of the start site of transcription we have indicated putative eukaryotic promoter elements based on their sequence homology and conserved positions relative

[illegible]

to the apparent start site of transcription. Between 21 and 37 nucleotides upstream of the start site of transcription is an AT-rich sequence flanked by GC bases which may represent the TATA box (36) for this gene. Also indicated are the positions of two potential CAAT boxes which are in conserved positions relative to the start site of transcription (37).

We have identified several nucleotide differences by comparing the cDNA and genomic DNA sequences. These discrepancies may be a consequence of *Drosophila* strain differences used in construction of the libraries and not due to cloning artifacts (37). The consequences of these nucleotide changes in terms of protein primary structure are discussed below.

Primary Structure Of The Deduced TU-36B Protein

We have identified a continuous open reading frame in the cDNA sequence which extends for 1242 nucleotides (414 amino acids). The other two possible reading frames were eliminated from consideration since they contain numerous stop codons and the potential coding regions do not conform with codon usage in *Drosophila* (39). The proposed open reading frame is shown in Figure 2. The protein sequence begins at a methionine codon encoded by nucleotides 817-819. Downstream at nucleotides 1084-1086 is a second methionine codon which could also serve as the initiating codon. We have indicated these two alternatives, since the nucleotide sequences preceding both methionine codons conform with the proposed consensus sequence C/AAAA/CAUG flanking the translational start sites of numerous *Drosophila* mRNAs (40).

The open reading frame would terminate at nucleotide 2387-2389 (TGA). Codon preference in this reading frame conforms with codon usage in *Drosophila* (39). Thus, if this mRNA is translated *in vivo*, initiation of protein synthesis at the first methionine residue would result in an unmodified protein of approximately 47,000 daltons in molecular weight while initiation at the second methionine would result in a protein of approximately 38,000 daltons in molecular weight.

Comparison of the cDNA and genomic sequences reveal that there are five nucleotide differences in the protein coding region. These differences may reflect *Drosophila* strain differences used for constructing the two recombinant DNA libraries

Figure 3. Amino acid sequence comparisons of TU-36B and cytochrome b proteins

Shown is a comparison of the sequence homologies between the proposed TU-36B protein, yeast cytochrome b2 (42), and chicken cytochrome b5 (43). Amino acids are designated by the single letter code. Sequence alignments were determined with the aid of a computer. Dashes represent gaps imposed for optimal sequence alignment. Shaded areas represent identical homology. Vertical bars indicate conservative amino acid replacements between TU-36B and yeast cytochrome b2.

(39, 41). A nucleotide change at position 1747 (G-->C) would result in no amino acid substitution. The remaining four nucleotide changes would result in amino acid substitutions (NT 1748, C-->G and NT 1750, T-->C = Pro-->Ala; NT 1782, C-->T = Thr-->Ile; NT 2174, G-->C = Gly-->Arg). Each of these changes would be nonconservative. The consequences of these substitutions, if any, on the biological activity of the proposed protein are unknown. Computer modeling of the proposed protein sequence does not reveal any statistically significant features. Finally, we have searched the nucleic acid and protein databases for homology with the DNA sequence and proposed open reading frame as well as potential amino acid sequences in the other two reading frames. We found no significant DNA sequence homologies; however, the proposed open reading frame shown in Figure 2 is homologous with cytochrome b proteins from various sources. The homology with two cytochrome b proteins is shown in Figure 3. The TU-36B protein has 17.6% identity with yeast cytochrome b-2 (42) and when conservative amino acid replacements are considered, the proteins are 41% homologous. The chicken cytochrome b5 protein (43) has 15.6% homology with the TU-36B protein. Interestingly, when we compare the nucleic acid sequence of TU-36B and yeast cytochrome b2 (42), we cannot identify any significant DNA homology; however, the TU-36B gene is similar to the yeast gene since it is a nuclear gene and single-copy.

Tissue Specific Expression of TU-36B mRNA

We have also examined and compared the expression of TU-36B and *Mhc* mRNAs by *in situ* hybridization to tissue sections. We have used late embryonic (16 hours after egg deposition-late stage 16) and late pupal (75 hours after pupariation) stages of development for *in situ* hybridization to tissue sections, since they can be readily staged and isolated, their muscle structures are well formed, and in control experiments TU-36B mRNA as well as *Mhc* mRNAs are present in these stages in development (data not shown). The hybridization probes used for *in situ* hybridization were *in vitro* synthesized complementary RNA (cRNA).

In situ hybridization of ³H-labeled cRNAs to tissue sections are shown in Figures 4 and 5. Sections of 16 hour old embryos are shown in Figure 4. We detect specific expression of the TU-36B mRNA predominantly in the embryonic body wall (somatic musculature), and the walls of midgut structures (visceral musculature) (Figure 4A). In similar sections of embryos (Figure 4B) we detect expression of *Mhc* mRNA in the same muscle containing tissues: the body wall and midgut. Figure 4C represents an embryo section hybridized with a non-complementary probe as a control in which the low background of exposed silver grains is evident.

In situ hybridization of cRNAs to parasagittal sections of late pupae are shown in Figure 5. We detect specific expression of TU-36B mRNA in numerous muscle tissues

Figure 4. Localization of TU-36B and Mhc transcripts in late embryonic stages of development

Sections of late stage 16 embryos (16 hours after egg deposition) were separately hybridized with ^3H -labeled TU-36B and *Mhc* specific cRNA probes. Each panel shows a photograph of the hybridized sections under phase contrast (-1) and dark-field illumination (-2). The embryo section shown in Panel 4A was hybridized with a TU-36B specific probe. The embryo section shown in Panel 4B was hybridized with a *Mhc* specific probe. Panel 4C shows embryo sections hybridized with a non-specific control probe. Anterior ends (Ant) of the embryos are indicated. Embryonic structures indicated; ph, pharynx; mg, midgut. Autoradiographic exposures were for 21 days. Magnification is 50X.

(Figure 5A). In the thorax, we detect expression in the dorsal median indirect flight muscle and the tergosternal muscles. We also detect expression in leg musculature as well as the ventral muscles of the abdomen. Figure 5B shows a parasagittal section of a pupa hybridized with a *Mhc* specific cRNA probe. Similar to the result shown in Figure 5A we detect specific expression in muscles within the thorax, legs, and abdomen with the

Mhc cRNA probe. In general the hybridization pattern with the TU-36B and *Mhc* specific probes are similar. Figure 5C represents a pupal section hybridized with a non-complementary probe.

In summary, the results of *in situ* hybridization indicate that the TU-36B mRNA is expressed in similar tissues as *Mhc* mRNAs and that the expression of the TU-36B mRNA is predominantly muscle specific.

DISCUSSION

We report the identification and characterization of a gene named TU-36B which is adjacent to the *Drosophila Mhc* gene. We have shown that the gene contains one intron and that a mRNA of at least 1480 nucleotides is transcribed from this gene. The most downstream poly(A) addition site of the *Mhc* gene (A-2) is 99 nucleotides from the single poly(A) addition site of the TU-36B gene, thus the two genes are convergently transcribed. We have identified a continuous open reading frame in the TU-36B mRNA which could encode a primary translation product of 47,000 daltons in molecular weight, which has homology with cytochrome b proteins. Finally, *in situ* hybridization to tissue sections demonstrate that TU-36B mRNA is expressed within the same muscle tissues as *Mhc* mRNAs.

The predominant expression of TU-36B mRNA within muscle containing tissues and the homology of the TU-36B protein product with cytochrome b provide strong evidence that this gene plays a role in muscle cell metabolism. We have used two stages in *Drosophila* development for *in situ* hybridization where the muscular systems are well differentiated (44, 45). We detect similar patterns of *in situ* hybridization using cRNA probes specific for TU-36B mRNA and *Mhc* mRNAs. The *Mhc* cRNA that we have used will detect all four *Mhc* mRNAs (27), thus this probe should identify all muscle mRNA containing tissues in the fly. The detection of a similar pattern of hybridization with a TU-36B probe correlates with the biological role of cytochrome b proteins which are associated with mitochondria.

Cytochrome b2 in yeast is a bifunctional tetrameric protein which carries two

Figure 5. Localization of TU-36B and *Mhc* transcripts in late pupal stages of development

Parasagittal sections of late pupae (75 hours after puparium formation) were separately hybridized with 3H-labeled TU-36B and *Mhc* specific cRNA probes. Each panel shows a photograph of the hybridized sections under phase contrast (-1) and dark-field illumination (-2). The pupal section shown in Panel 5A was hybridized with a TU-36B specific probe. The pupal section shown in Panel 5B was hybridized with a *Mhc* specific probe. Panel 5C shows a pupal section hybridized with a non-specific control probe. Pupal structures indicated; hd, head; cm, cephalic musculature; lg, legs; dmifm, dorsal median indirect flight muscles; tgm, tergosternal muscles; vma, ventral muscles of the abdomen. Autoradiographic exposures were for 21 days. Magnification is 25X.

prosthetic groups; one heme and one flavin molecule per subunit of 55,000 MW (42). Examination of Figure 3 reveals that the TU-36B protein and the cytochrome b proteins share greater homology in the amino half of the respective proteins. This domain in cytochrome b has been shown to bind heme (51). Thus, there appears to be high conservation of amino acids in TU-36B within an important structural domain.

The TU-36B gene is similar to other muscle genes in *Drosophila* since it is single-copy and not a member of a gene family. On the other hand, it differs from most of the single copy muscle genes identified thus far in *Drosophila* since there is no apparent alternate processing of the TU-36B mRNA (*Mhc*, [15]; MLC-alk-3, [46]; MLC-alk-2, [47, 48]; tropomyosin, [1]). We cannot rule out that there may be diversity generated by post-translational processing; however, based on our evidence, it appears that a single protein product could provide all the necessary biological functions throughout development.

The close proximity and co-expression of the TU-36B and *Mhc* genes represents another example in *Drosophila* of the clustering of functionally related genes. Two cytochrome c genes have been characterized at chromosomal locus 36A10-11, which is close to the 36B locus (50). Thus, these four genes define a larger muscle specific gene cluster. In the Introduction we mentioned that many functionally related genes are clustered in the *Drosophila* genome (2, 4, 7). The 88F locus (on the 3rd chromosome) of *Drosophila* is a well characterized example where several muscle genes are spatially close in the genome and functionally related (12). Additionally, several ethyl-methanesulfonate (EMS)-induced flightless mutations at this locus have been characterized in detail and have been shown to result from DNA alterations within an actin gene and a tropomyosin gene (10, 49, 12, 13).

EMS-induced flightless mutations have also been genetically mapped at or near the *Mhc* locus (11). The primary defects of four of these mutations have been shown to result from insertions or deletions within the 5' end of the *Mhc* gene (16). There are additional flightless mutations which map at this locus whose primary defects have not yet been identified. We have isolated genomic DNA from six of these uncharacterized flightless mutations and compared the organization of their DNA to TU-36B genomic DNA by restriction enzyme analysis (Levin and Rozek, unpublished results). At this level of sensitivity we have not detected any alterations in the organization of DNA from the mutants. The inability to detect differences using this methodology is not surprising since several of the EMS-induced mutations within the 88F actin gene have been shown to be single point mutations (12). Attempts to rescue flightless mutations by introduction of copies of the TU-36B "wild type" gene by transformation may provide additional information on the nature of some of these uncharacterized flightless mutations.

Finally, we would like to consider several features of the organization and

expression of the TU-36B and *Mhc* genes and the consequences, if any, on RNA transcription and RNA processing. These features are the unusually close proximity of the two genes, convergent directions of transcription, and apparent co-expression within similar muscle tissues. Our experimental evidence indicates that the poly(A) addition site of the TU-36B gene is separated by 99 nucleotides from the most distal poly(A) addition site of the *Mhc* gene. Birnstiel et al., (29) has summarized consensus sequences implicated in controlling selection of the site for polyadenylation. These are the nearly ubiquitous AATAAA sequence preceding the site of poly(A) addition and the less common but frequent occurrence of TG clusters downstream of the poly(A) addition site. We have examined in detail the sequences surrounding the poly(A) addition site of the TU-36B gene and poly(A) addition site A-2 of the *Mhc* gene (Figure 2). Each of these poly(A) addition sites have the consensus features - the canonical AATAAA sequence upstream and short TG clusters downstream of the sites of poly(A) addition. Besides these sequence requirements, an accumulating body of experimental evidence demonstrates that transcription must proceed a considerable distance downstream for accurate poly(A) addition to occur. The extent of transcription beyond the poly(A) addition site may be as short as 30 nucleotides, however, the preponderance of evidence suggests that hundreds of nucleotides must be transcribed downstream for accurate poly(A) addition to occur. We have preliminary evidence that transcription of the *Mhc* and TU-36B genes do in fact proceed beyond their respective poly(A) addition sites (Croniger and Rozek, unpublished results).

Our observations that the TU-36B and *Mhc* genes are co-expressed in the same muscle tissues and that *Mhc* transcription continues downstream into the TU-36B transcription unit suggests that occasionally two RNA polymerase II transcription complexes may come into contact during transcription. We have no information on rates of transcription for these genes, so we cannot speculate on how often polymerase molecules may come into contact. One possible consequence of contact of converging polymerases is the stalling of transcription. We are currently investigating if such a phenomena is occurring during the convergent transcription of these two genes in the same tissues.

ACKNOWLEDGMENTS

We thank Margaret Chidlaw, Thomas Bizon, and Kari Maier for their generous help with various aspects of this research. We also thank Drs. Susan Celniker and Mark Garfinkel for their helpful comments on the manuscript. This study was supported by National Institutes of Health Grant No. GM33297 to C.E.R.

⁺Present address: Cold Spring Harbor Institute, PO Box 100, Cold Spring Harbor, NY 11724, USA

REFERENCES

1. Bautch, R. V. and Storti, V. L. (1983) *Proc. Natl. Acad. Sci. USA* 80, 7123-7127.
2. Jones, C. W. and Kafatos, F. C. (1980) *Nature* 284, 635-638.
3. Karlik, C. C., Mahaffey, J. W., Coutu, M. D. and Fyrberg, E. A. (1984) *Cell* 37, 469-481.
4. Meyerowitz, E. M. and Hogness, D. S. (1982) *Cell* 28, 165-176.
5. Rawls, J. M., Freund, J. N., Jarry, B. P., Louis, C., Seagraves, W. A. and Schedl, P. (1986) *Mol. Gen. Genet.* 202, 493-499.
6. Snyder, M. and Davidson, N. (1983) *J. Mol. Biol.* 166, 101-118.
7. Snyder, M. P., Hirsh, J. and Davidson, N. (1981) *Cell* 25, 165-177.
8. Henikoff, S., Keene, M. A., Fechtel, K. and Fristrom, J. W. (1986) *Cell* 44, 33-42.
9. Fyrberg, E. A., Bond, B. J., Hershey, N. D., Mixter, K. S. and Davidson, N. (1981) *Cell* 24, 107-116.
10. Karlik, C. C., Coutu, M. D. and Fyrberg, E. A. (1984) *Cell* 38, 711-719.
11. Mogami, K. and Hotta, Y. (1981) *Mol. Gen. Genet.* 183, 409-417.
12. Karlik, C. C., Saville, D. L. and Fyrberg, E. A. (1987) *Mol. and Cell Biol.* 7, 3084-3091.
13. Mahaffey, J. W., Coutu, M. D., Fyrberg, E. A. and Inwood, W. (1985) *Cell* 40, 101-110.
14. Bernstein, S. S., Mogami, K., Donady, J. and Emerson, C. P. (1983) *Nature* 302, 393-397.
15. Rozek, C. E. and Davidson, N. (1983) *Cell* 32, 23-34.
16. Mogami, K., O'Donnell, P. T., Bernstein, S. I., Wright, T. R. F. and Emerson, C. (1986) *Proc. Natl. Acad. Sci. USA* 83, 1393-1397.
17. Poole, S. J., Kauvar, L. M., Drees, B. and Kornberg, T. (1985) *Cell* 40, 37-43.
18. Benton, W. D. and Davis, R. W. (1977) *Science* 196, 180-182.
19. St. John, T. P. and Davis, R. W. (1979) *Cell* 16, 443-452.
20. Fyrberg, E. A., Kindle, K. L., Davidson, N. and Sodja, A. (1980) *Cell* 19, 365-378.
21. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci.* 74, 5463-5467.
22. Maxam, A. M. and Gilbert, W. (1980) *Meth. Enzymol.* 65, 499-560.
23. Mizusawa, S., Nishimura, S. and Seela, (1986) *Nucl. Acids Res.* 14, 1319-1324.
24. Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) *Anal. Biochem.* 112, 295-298.
25. Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* 132, 6-13.
26. Ingham, P. W., Howard, K. R. and Ish-Horowicz, D. (1985) *Nature (London)* 318, 439-445.
27. Rozek, C. E. and Davidson, N. (1986) *Proc. Natl. Acad. Sci.* 83, 2128-2132.
28. Takahashi, A., Philpott, D. E., and Miquel, J. (1970) *J. Geront.* 25, 222-228.
29. Birnstiel, M. L., Busslinger, M. and Strub, K. (1985) *Cell* 41, 349-359.
30. Proudfoot, N. J. and Brownlee, G. G. (1976) *Nature (London)* 263, 211-214.
31. McLauchlin, J., Gaffney, D., Whitton, J. L. and Clements, J. B. (1985) *Nucl. Acids Res.* 13, 1347-1368.
32. Keller, E. B. and Noon, W. A. (1984) *Proc. Natl. Acad. Sci. USA* 81, 71417-7420.
33. Keller, E. B. and Noon, W. A. (1985) *Nucl. Acids Res.* 13, 4971-4981.
34. Hultmark, P., Klemenz, R. and Gehring, W. J. (1986) *Cell* 44, 429-438.
35. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. and Davidson, N. (1982) *Cell* 29, 1027-1040.
36. Goldberg, M. L. (1979) Thesis, Stanford University.
37. Bucher, P. and Trifonov, E. N. (1986) *Nucl. Acids Res.* 14, 10009-100026.
38. Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B., Hogness, D. (1983) *Science* 221, 23-29.
39. O'Connell, P. and Rosbash, M. (1984) *Nucl. Acids Res.* 12, 5495-5513.
40. Cavener, D. R. (1987) *Nucl. Acids Res.* 15, 1353-1361.
41. Bowtell, David D. L., Simon, M. A. and Rubin, G. M. (1988) *Genes and Develop.* 2, 620-634.

42. Guiard, B. (1985) EMBO J. 4, 3265-3272.
43. Nobrega, F. G. and Ozols, J. (1971) J. Biol. Chem. 246, 1706-1717.
44. Campos-Ortega, J. A. and Hartenstein, V. (1985) Springer-Verlag, New York.
45. Crossley, A. C. (1978) (ed. Ashburner and T. R. F. Wright), pp.499-560. Academic Press, New York.
46. Falkenthal, S., Parker, V. P., Mattox, W. W. and Davidson, N. (1984) Mol. Cell Biol. 4, 956-965.
47. Parker, V. P., Falkenthal, S. and Davidson, N. (1985) Mol. Cell Biol. 5, 3058-3068.
48. Toffenetti, J., Mischke, D. and Pardue, M. L. (1987) J. Cell Biol. 104, 19-28.
49. Karlik, C. C. and Fyrberg, E. A. (1985) Cell 41, 57-66.
50. Limbach, K. J. and Wu, R. (1985) Nucl. Acids Res. 13, 631-644.
51. Pompon, D. and Lederer, F. (1976) Eur. J. Biochem. 68, 415-423.