The Organization and Expression of the light Gene, a Heterochromatic Gene of Drosophila melanogaster

Robert H. Devlin, Brendan Bingham and Barbara T. Wakimoto

Department of Zoology NJ-15, University of Washington, Seattle, Washington 98195

Manuscript received August 30, 1989
Accepted for publication January 25, 1990

ABSTRACT

The light (lt) gene is located in the centromeric heterochromatin of chromosome 2 of Drosophila melanogaster. This gene is necessary for normal levels of pigmentation in a number of adult and larval tissues and is required for viability. Hybrid dysgenic and X-ray induced mutations have been used to identify the gene and compare its organization to that of euchromatic genes. Molecular mapping of lt mutations and its major transcripts has shown that the lt gene is at least 17 kb. By injecting cosmid clones that include this region into lt mutant embryos, we have defined a 30-kb region that can transiently rescue the pigmentation defect in the Malpighian tubules. The major transcription unit of this gene is comprised of exons that are single copy. It is unusual in its organization in having a heterogeneous array of middle repetitive DNA sequences within its intronic and flanking regions.

The molecular isolation of chromosomal sequences and their localization by in situ hybridization has provided much information on the large scale organization of the eukaryotic chromosome. This descriptive approach has been widely used (reviewed in Lima-de-Faria 1983; John and Miklos 1988), but its potential for addressing functional aspects has been best realized in studies of the Drosophila chromosome. With Drosophila, the molecular properties of a region can be correlated with its cytogenetic and genetic properties. Molecular, cytogenetic and genetic strategies have been important in the analysis of constitutive heterochromatin. Heterochromatin is a nearly ubiquitous component of the chromosomes of higher organisms. In contrast to euchromatin, which condenses and decondenses in concert with the cell cycle, heterochromatin appears cytologically to remain condensed (HeitZ 1928). Heterochromatic regions replicate late or remain underreplicated in certain cell types and contain a large proportion of repetitive DNA, including virtually all of the highly repeated sequences (reviewed by Hilliker, Appels and SchaléT 1980; John and Miklos 1988). While constitutive heterochromatin was originally considered genetically inert, studies of Drosophila heterochromatin have shown that it contains several different types of genes. One class includes genes such as the 18 and 28S ribosomal genes which are members of a repetitive gene family. Studies of the ABO (Sullivan and Pimpinelli 1986; Pimpinelli, et al. 1985) and the Responder loci (Wu et al. 1988; Lyttle 1989; Pimpinelli and Dimitri 1989) suggest that these heterochromatic loci may also be comprised of repeated DNA.

A second class of genes has been identified in heterochromatin. There are at least 21 genes, within or very near the centromeric heterochromatin of chromosomes 2 and 3, that are mutable by ethylmethane sulfonate (Hilliker 1976; Marchant and Holm 1988; Schupbach and Wieschaus 1989). Some of these genes are also mutable by the insertion of the transposable P element (data of Holm, reported in Kidwell 1987; this report, Kim and Holm 1988). Since ethylmethane sulfonate and P elements generally induce small lesions, it has been suggested that these autosomal loci are functionally unique and may be structurally similar to conventional euchromatic genes. However, two lines of genetic evidence suggest that the properties of this second class of heterochromatic genes may differ from euchromatic genes (Hilliker 1976). First, genes located within or very near centromeric heterochromatin are subject to position effects if displaced from proximal chromosomal regions by chromosome rearrangements (Schultz and Dobzhansky 1934; Khvostova 1939; Baker 1953; Hessler 1958). Six heterochromatin genes, including the light (lt) gene on chromosome 2 of Drosophila melanogaster show this sensitivity to chromosomal position (Wakimoto and Hearn 1990). This displacement results in a variegated pattern of gene expression. Second, variegating alleles of some heterochromatic genes interact with trans-acting modifiers of position effect variegation in a manner opposite the interactions of these same modifiers with variegating alleles of euchromatic genes. The addition of the Y chromosome (Schultz 1956; Baker and Rein 1962) or mutations that act as genic suppressors of position effect variegation of euchromatic genes (A. Hedrick, M. Hearn, T. Grigliatti and B. Wakimoto, unpublished) enhance the variegation of lt. A model which

Genetics 125: 129-140 (May, 1990)
explains these genetic interactions postulates that the modifiers encode or interact with components involved in chromatin assembly (REUTER and WOLFF 1981; SINCLAIR, MOTTUS and GRIGLIATI 1983; LOCKE, KOTARSKI and TARTOF 1988). This model would then suggest that heterochromatin genes require proteins or a particular chromatin conformation that inhibits the expression of variegating euchromatic genes.

A molecular analysis of heterochromatic genes is important for determining how the regulatory requirements of heterochromatic genes compare to those typical of euchromatic genes and how these properties relate to what is known about the properties of heterochromatin. This report describes molecular studies of the lt gene. Mutations that eliminate lt activity are lethal (HILLIKER 1976). Less extreme reductions in activity result in a cell autonomous decrease in the pigmentation of several adult and larval tissues, including the ommatidia and the Malpighian tubules. The ease of scoring this mutant phenotype has facilitated the recovery of X-ray induced and P element insertion mutations in our laboratory (WAKIMOTO and HEARN 1990 and this report) and in the laboratory of DAVID HOLM (reported in KIDWELL 1987). Here we describe the molecular analysis of the mutations generated in our laboratory and the effects of these mutations on lt transcription. The molecular organization of over 80 kb of genomic DNA that includes the lt gene has been examined. Using a transient expression assay, we have identified a genomic region that permits the expression of lt gene in the Malpighian tubules. The results of these experiments and transcript mapping show that the lt gene is at least 17 kb. Its major transcription unit is comprised of exonic regions that are single copy while several intronic regions and its flanking regions are comprised of a heterogeneous array of middle repetitive DNA sequences.

MATERIALS AND METHODS

Drosophila strains: Cultures were raised on standard cornmeal-molasses-brewer’s yeast media at 25°. The lt* mutations are described in WAKIMOTO and HEARN (1990). A chromosome 2 balancer with a lt mutation, SM1, Cy lt1 was kindly provided by DAVID HOLM. All other mutations are described by LINDSLEY and GREGG (1968). Strains of D. melanogaster surveyed for polymorphisms in the lt region include an isogenic strain y; en bw sp constructed by J. A. KENNISON, and the wild-type strains Oregon R, Sevelen and Canton-S. Other Drosophila species used were D. pseudoobscura, D. viridis and D. simulans.

Isolation of hybrid dysgenic induced lt mutations: For P element insertion mutagenesis, Canton-S females were mated to males from the 2 P strain provided by W. ENGELS. The cultures were raised at 21° and sons from this cross were mated to females from a lt ster strain which does not contain P elements. The progeny of this cross was scored for mutant eye color. Two lt mutations were recovered among 54,045 progeny screened (frequency of 3.7 x 10^-5).

The mutations were balanced over SM1, Cy lt1. The lt469 mutation is a stable, lethal allele of lt for which we were unable to detect a molecular lesion. The lt469 mutation which is unstable under dystrophic conditions, was used in this laboratory for molecular cloning of lt sequences.

Southern blotting and construction of genomic libraries: Small scale preparations of fly DNA for Southern blotting were carried out as described by BENDER, SPIERER and HOGNESS (1983). Restriction enzyme digestion, gel electrophoresis, transfers to nitrocellulose filters and hybridization of DNA blots were performed essentially as described by MANIATIS, FRITSCH and SAMBROOK (1982). Probes were radiolabeled with 32P by nick translation. Southern blots were washed with high stringency using four 15-min washes in 0.1 x SSC, 0.1% SDS at 68°. The filters were exposed to XAR-5 X-ray film for autoradiography. The plasmids containing lt sequences are described in the text. The pr25.1 plasmid which contains a 2.9-kb P element is described by O’HARE and RUBIN (1983). The plasmids containing the middle repetitive elements 2078 and 2228 were isolated by YOUNG (1979) and RUBIN et al. (1981) respectively and were kindly provided by Richard Garber.

To construct genomic libraries from mutants, DNA from adult flies was isolated then partially restricted with the enzyme Sau3A (for the Ith465 and Ith469 libraries) or cut to completion with EcoRI (for the lt469 library) for directed cloning. The DNA was size fractionated on sucrose gradients as described by MANIATIS, FRITSCH and SAMBROOK (1982) then ligated to EMBL4 plasmid arms. The phage were packaged using extracts purchased from Stratagene (La Jolla, CA). The lt469 library was screened for P element sequences using 32P-labeled nick translated probes made from pr25.1. The Ith465 and Ith469 libraries were screened using cloned genomic sequences that flank the P element insertion in the lt469 chromosome.

Two cosmid libraries were screened using the methods described by MANIATIS, FRITSCH and SAMBROOK (1982) with the single-copy probes extending from coordinate –2.4 to 0.0 kb and +13.5 to +15.4 kb (see Figures 2 and 3). The cosRK library, constructed by R. Kelley contains genomic DNA in the vector cosPneo of STELLER and PIRROTTA (1985). The NotBamNot-CoSpeR (COST) library was constructed by J. TAMKUN using DNA from an isogenic y; en bw sp strain and a modified form of the CoSpeR vector (PIRROTTA 1988). The Drosophila sequences cloned in the phage and cosmid vectors were characterized by restriction digestions and blot hybridizations. Fragments were isolated after restriction digests and subcloned into the Bluescript plasmid vectors (Stratagene) for finer level restriction mapping.

Microinjection and transient rescue experiments: Preblastoderm lt ster embryos were injected with cosmIDNA at a concentration of 1 mg/ml using the protocols of SPRADELING and RUBIN (1982). When these individuals reached the third instar larval stage, the Malpighian tubules were dissected in 0.7% NaCl, mounted under a coverslip then scanned for pigment containing cells using a 10X objective lens and the UV epifluorescence optics of a Nikon Microphot-FX. The levels of pigment accumulation in individual cells can be easily monitored after excitation of the pigment by irradiation in the ultraviolet range.

Characterization of RNA transcripts and cDNAs: RNA was isolated by homogenizing tissues in equal volumes of a solution of 100 mM Tris (pH 7.5), 10 mM EDTA, 0.2 M NaCl with phenol:chloroform:isoamyl alcohol (50:49:1, by volume). After centrifugation and repeated extraction of the aqueous phase with phenol:chloroform:isoamyl alcohol, the RNA was precipitated with 0.3 M NaOAc and 2.5 volumes of ethanol. After pelleting and desiccation, the
RNA was dissolved in DEPC-treated water and poly(A\(^+\)) RNA was isolated by chromatography using oligo-dT cellulose (Collaborative Research, Bedford, Massachusetts) as described by MANIATIS, FRITSCH and SAMBROOK (1982). RNA was quantitated by spectrophotometry, fractionated on 2.2 M formaldehyde gels in MOPS buffer (RAVE, CRKVJENJAKOV and BOEDTKE 1979) and blotted to Gene Screen Plus nylon membranes (New England Nuclear). Single-stranded RNA probes were prepared from the T3 or T7 promoters of Bluescript vectors using \(^32\)P-labeled CTP according to the manufacturer’s specifications. Northern blots were hybridized using the conditions described by STROEHER, JORGENSEN and GARBER (1986) and washed as described for Southern blots but at 71°C. RNAs were sized using RNA standards purchased from Bethesda Research Labs (Gaithersburg, MD).

For the analysis of \(lt\) transcripts from mutants RNA was extracted from adults. A Northern blot with 5 µg of poly(A\(^+\)) RNA from each strain was probed with single stranded RNA probes. A probe containing the \(Hinc\)1 fragment extending from coordinates -1.7 to -0.5 kb (Figure 5) was used to detect \(lt\) transcripts. The 1.7-kb actin RNA encoded by the actin 42A gene was detected using single stranded probes that contained the 640 nucleotide sequence complementary to the coding strand found in the 3\(^{']}\) end of the message (FYRBERG et al. 1983). The relative levels of the most abundant \(lt\) transcripts was determined by densitometry of autoradiographs, using the act42A transcripts a standard to control for differences in the amount of RNA loaded in each lane.

The cDNA libraries constructed by L. KAUVAR (POOLE et al. 1985) were screened with a probe extending from coordinate -2.4 to 0.0 kb. Two libraries were used; one constructed from poly(A\(^+\)) RNA from pupae and a second using RNA from 3–12-hr embryos. The cDNAs that hybridized to the probe were subcloned as EcoRI fragments into Bluescript vectors and restriction mapped with \(Bam\) HI, \(Hinc\)1, and HindIII. The ends of these cDNAs were sequenced using dideoxynucleotides and the methods described by SANGER, NICKLEN and GOULSON (1977) and SANGER et al. (1980) using Sequenase (US Biochemicals, Cleveland, Ohio).

Primer extension analysis was performed using a synthetic oligonucleotide of 20 bp (5’T-GAAAAATGCAACAGCT-GA1-3’) that is complementary to the coding strand found in the 3\(^{’}\) region of several of the cDNAs. This sequence is located 3.5 bp from the end of the longest cDNAs, c35. The primer was end labeled with \(^32\)P using T4 polynucleotide kinase (Bethesda Research Laboratories), annealed to 10 µg of poly(A\(^+\)) RNA of adults and extended with AMV reverse transcriptase (Bethesda Research Laboratories), using the protocol described by WILLIAMS and MASON (1983). Annealings of the primer and RNA at temperatures from 45\(^\circ\) to 65\(^\circ\) gave identical qualitative results, although quantitative differences in the amount of extension product were observed. Reverse transcriptase products were sized on a 6% acrylamide gel, using a sequence reaction as a size standard.

**RESULTS**

**Molecular mapping of \(lt\) mutations:** In a screen for \(P\) element induced mutations, we recovered one unstable \(lt\) allele, \(lt^{kds}\). Flies heterozygous for this mutation and the \(It^1\) allele have an orange eye color. Twenty sublines were established, each from a male heterozygous for \(lt^{kds}\) and a \(lt^1\) \(stu^3\) marked chromosome. These lines were maintained by repeated back-crossing to a \(lt^1\) \(stu^3\) M strain and were continually monitored for stability of the mutant phenotype. From these lines, we isolated derivatives of the \(lt^{kds}\) allele that resulted in wild type, intermediate or stable orange eye color phenotypes. DNA was extracted from each of the sublines, digested with EcoRI and the \(P\) element containing fragments were identified by Southern blot analysis. The comparison among the lines showed that an 8.2-kb EcoRI, \(P\) element-containing fragment was present in mutable \(lt^{kds}\) lines (e.g., the \(lt^{kds1}\) and \(lt^{kds1B}\) lines) but absent in the lines that had reverted to give a wild type level of eye pigmentation (Figure 1A). This 8.2-kb EcoRI fragment was cloned from the \(lt^{kds}\) strain, isolated by hybridization to a \(P\) element containing probe (O’HARE and RUBIN 1983) and identified by comparing its restriction map to the map deduced from the genomic Southern blots. The results of these experiments showed that the \(lt^{kds}\) chromosome contained a 1.5-kb \(P\) element inserted into an 8.0-kb EcoRI genomic fragment (Figure 2).

The insertion site is in a 2.4-kb BamHI-PstI fragment, which was previously shown to contain the sites of insertions of two of the mutations isolated by HOLM (DEVLIN et al. 1990).

The analysis of the revertant lines showed that changes in the \(lt\) phenotype were correlated with changes in the cloned region. To examine the nature of the changes seen in the other derivative lines, sequences that flank the \(P\) element insert were used to isolate the corresponding fragments from recombinant phage libraries made from two derivative lines, \(lt^{kds1}\) and \(lt^{kds1B}\). One subline \(lt^{kds1B}\), had very low levels of eye pigmentation. Unlike the ethylmethane sulfonate and X-ray induced hypomorphic mutations in which pigment levels are correlated with viability, \(lt^{kds1}\) shows only a moderate effect on viability; flies heterozygous for \(lt^{kds1}\) and EMS 40-12, a lethal allele of \(lt\), survive with 60% of the viability of their wild-type sibs. This line differs from the original line because it contains an additional 500 bp of sequence to the left of the \(P\) element (Figures 1 and 2). We have determined by restriction mapping and sequence analysis that this difference is due to a tandem duplication of genomic sequences (data to be presented elsewhere). Phenotypically wild type flies have been recovered from this derivative. We examined five independent revertant lines by Southern blotting and found that in each case, the precise or nearly precise excision of the \(P\) element occurred and the duplication was retained. This result shows that the duplication is compatible with high levels of \(lt\) function. Imprecise excision of the \(P\) element was also observed and in the case of \(lt^{kds1B}\), a small region of less than 50 bp containing \(P\) element sequences was retained. This derivative yields a darker red but not fully wild-type level of eye pigmentation. The viability of \(lt^{kds1B}\) is increased relative to its parent \(lt^{kds1}\) chromosome, to 80% of the wild-type levels.
The importance of the cloned region for \textit{lt} function was further suggested by the analysis of additional \textit{lt} mutations. We surveyed the 14-kb region shown in Figure 2 (from coordinates $-7.6$ to $+6.0$ kb) for molecular lesions using sequences from $-2.4$ to $0.0$ kb. The \textit{lt} mutant isolated by Bridges in 1924 (LINDSLEY and GRELL 1968) is deleted for a 200 $\pm$ 50-bp region (around coordinate $+2.0$ kb) compared to the wild type \textit{Oregon-R}, \textit{Canton-S} and \textit{Sevelen} chromosomes. The parent chromosome of \textit{lt} is not available for comparison so we cannot be certain that this difference is responsible for the mutant phenotype.

Comparisons to the parental \textit{Canton-S} chromosome were made for 56 X-ray induced \textit{lt} mutations (WAKIMOTO and HEARN 1990). Thirty-seven mutant alleles, including 12 which variegate for the \textit{lt} gene appeared identical to the parental \textit{Canton-S} chromosome in the Southern blot analyses. Seventeen alleles are deleted for the 14-kb region. These are all lethal alleles of the \textit{lt} locus; some were shown by complementation analyses to be also deleted for adjacent genes (WAKIMOTO and HEARN 1990). Two mutations showed detectable restriction site differences. The \textit{lt}^\textit{84} mutation is a lethal allele associated with a 30-bp deletion within the $-0.3$ to $0.0$ interval between the \textit{Hin}cII and \textit{Bam}HI sites (Figure 2). The \textit{lt}^\textit{90} allele is hypomorphic and associated with a lesion to the left of the \textit{Eco}RI site at coordinate $+0.5$. This defect produces a \textit{PstI} fragment that was 4.8 kb smaller than wild type. In summary, the molecular analyses show that mutations affecting \textit{lt} function map throughout a region of at least 3 kb, from coordinates $-0.5$ to $+3.5$.

**Analysis of \textit{lt} expression by transient rescue:** Overlapping cosmid clones have been isolated that together cover 80 kb, including the area to which the \textit{lt} mutations were mapped (Figure 3). Each cosmid clone contains from 25 to 40 kb of genomic DNA from \textit{lt} Drosophila strains. To assay for the expression of the wild type gene from the cloned sequences, we microinjected cosmid DNA into preblastoderm embryos homozygous for \textit{lt}, and then monitored the injected individuals for pigmentation of the larval Malpighian tubule cells.
The Malpighian tubule cells of the lt' host strain have a low level of pigmentation with a uniform distribution of small pigment granules (550 larvae examined). The results of transient expression assays after injection of the DNA of five different cosmids are compared in Figure 3. With the cosmids cosRK5, cosT6 and cosRK7, 2-4% of the injected individuals showed some heavily pigmented cells. In these individuals, the number of cells rescued varied from only a single cell to 16 of the approximately 480 cells present in the Malpighian tubules of the third instar larva (JANNING 1986). The left and right endpoints of cosRK7 and cosT6, respectively, define a region approximately 30 kb in length (from coordinates -11 to +19 kb) that promoted the highest levels of rescue. CosRK11 whose right endpoint maps at -3.8 kb showed a significantly lower success rate; only a single cell in the 668 individuals scored was rescued. CosRK12, whose left end maps at coordinate +1.0 failed to rescue the mutant phenotype.

**Analysis of transcripts in the lt region:** To identify the lt transcription unit, RNA was extracted from various tissues and stages of development and hybridized to a series of subclones from coordinate -5 to +15 kb. Seven species of poly(A+) RNAs, ranging in size from 13 kb to 1.0 kb were detected in adults (Figure 4A). The two most prevalent species were 3.2 kb (transcript A) and 3.6 kb (transcript B) in length. Our survey suggests that A and B are widely expressed. We have detected both transcripts in embryos, larvae, pupae, and adults (Figure 4B) and in isolated tissues such as the ovaries of 5-day adult females, in brain, imaginal discs, fat body, gut, Malpighian tubules, salivary glands of the larvae (data not shown). The five minor transcripts, transcripts C through G are observed in most developmental stages but were most easily detected in whole adult and ovary RNA. Four of these transcripts hybridize with the single stranded probe from -2.4 kb to 0.0 that detects A and B. The single stranded probe of this same region but of opposite orientation detects only transcript C. The A and B transcripts are strongly implicated as important for lt' function since they differ qualitatively in two of the mutant strains. The lt' mutation, which is associated with a 200-bp deletion, produces A and B transcripts that are approximately 500 bp shorter than the wild-type messages. The size of the
C transcript is also reduced in the \( lt' \) mutant (data not shown). The importance of transcripts D through G for \( lt \) function is not known. We have not detected an effect of mutations on the lower molecular weight transcripts. Because of the large size and variable ability to detect the D transcript in wild type strains, we have been unable to determine if the mutations alter this transcript.

The DNA sequences coding for the transcripts A and B were examined in more detail by probing Northern blots of 5 \( \mu \)g of poly(A\(^+\)) RNA from adults with a variety of single stranded probes from the region. The results of these studies are summarized by Figure 5. Both transcripts span a region of at least 17 kb since they are detected with probes containing sequences located from -1.7 to 0.0 kb and with probes extending from +12.5 to +15.4 kb. This 17 kb is within the region important for \( lt \) function as defined by the cosmid rescue experiments. The transcript mapping studies also revealed differences between A and B. Two probes that contain sequences from coordinates 0.0 to +1.0 kb detect only B. This suggests the possibility that A and B are alternatively processed messages or that B represents a precursor to A.

The transcript mapping was extended by the analysis of cDNAs. These cDNAs were isolated from the early embryonic and pupal cDNA libraries constructed by L. Kauvar (Poole et al. 1985) by probing with sequences from -2.4 to 0.0 kb. Ten of the largest cDNAs recovered from the embryo library and two recovered from the pupal library were analyzed. These ranged in size from 1.1 kb to 2.9 kb and shared a consistent restriction map for all but the 3' most regions. Heterogeneity at the 3' end of the cDNAs was due to two factors. Sequence analysis of the ends of the cDNAs showed that ten of the cDNA clones terminated with poly A tails; five of these had a poly A tail that began 133 bp downstream of the site of
The Drosophila light Gene

A

D

C

B

> 0

< 4.40

2.37

1.35

ACTIN

FIGURE 4.—Northern blot analysis of the RNAs from the lt gene. In panel (A), the left lane shows an autoradiographic exposure of a Northern blot containing 10 µg of poly(A+) adult RNA and probed with a radiolabeled double stranded Pst-BamHI fragment from -2.4 to 0.0 kb. This autoradiograph was overexposed to show the five minor transcripts A through G, in addition to the most abundant transcripts A and B. The shorter exposure of the same autoradiograph (right lane) shows that A and B are two distinct species of RNA of 3.2-kb and 3.6-kb, respectively. Panel (B) shows an autoradiographic exposure of a Northern blot containing 5 µg of poly(A+) RNA isolated from ovaries or from whole animals of the Canton-S strain. A single stranded probe containing sequences from the HincII-HincII fragment from coordinate -1.7 kb to -0.5 kb was used as the probe. Transcripts A and B are detected throughout development. Panel (C) shows an autoradiographic exposure of a Northern blot containing 3 µg of poly(A+) RNA from Canton-S, Itb+/+ deficiencies heterozygotes and It+/ homogygotes. The blot was probed with the same single stranded probe as in panel B to detect the It transcripts and with a single stranded probe to detect the act42A actin mRNA. The actin transcript provides a size standard to compare the sizes of transcripts A and B in the mutant and wild type strains.

poly A addition in the other five. In addition, restriction mapping of the cDNAs provided evidence for differential splicing near the 3' end of the transcription unit; two of the clones possess a HincII fragment 60 bp smaller than the other cDNAs (Figure 5).

Most or all of the cDNAs we analyzed are likely to be cDNAs of transcript A. This conclusion is based on three lines of evidence. First, the abundance of A would predict that its cDNA would be most readily recovered. Second, none of the cDNAs contain the B specific regions. Third, the length of the longest cDNA, c35 is 2.9 kb. To estimate how far RNAs from the lt region might extend past c35, a primer extension assay was performed. Adult poly(A+) RNA was annealed to an oligonucleotide complementary to the coding strand and located 35 bp from the 5' end of c35. This oligonucleotide primer was extended by reverse transcriptase and yielded product of 189 bp (Figure 6). This evidence suggests that c35 is a nearly full length cDNA. Its length is consistent with the size estimate of the 3.2 kb transcript A, with the addition of the uncloned 5' extension (134 bp) and a poly(A+) tail.

The c35 cDNA was used as a probe to Southern blots containing restriction digests of the cosmid clones to map the major exonic and intronic regions. This experiment was more sensitive than the experiments using the single stranded probes to map the A and B transcripts on Northern blots and allowed us to identify a third transcribed region. This region that hybridized to c35 was within the fragment extending from +1.2 to +2.2 kb (Figure 5). These results suggest that the major lt transcript contains at least two introns and that sizes of these introns are no larger than 1.5 kb and 8.1 kb. Both of the approaches we used to identify transcribed regions are limited to the detection of hybridizing fragments with lengths greater than 50 bp. Hence, additional small exons may exist throughout the two intronic regions we defined and these would not have been detected in our experiments.

Two of the regions defined as exonic regions (from coordinates -1.7 to 0.0 and from +1.2 to +5.6) correspond to the fragments to which the lt mutations that affect transcript size were mapped. In particular, the P element in the ItbDNA is inserted at -0.5 kb with the orientation of the P element transcription unit opposite to that of the major lt transcripts. The 200
bp deletion in \(lt^1\) maps to coordinate +2.0 kb. We have not detected molecular lesions in any of the \(lt^e\) mutations in the region from +12.5 to +15.4 kb.

**Characterization of the copy number of sequences in the \(lt\) region:** Previous studies by one of us (Devlin et al., 1990) showed the single copy fragment that contained sites of the \(P\) element insertions (coordinates -2.4 to 0.0 kb) was flanked by repetitive DNA. In order to further examine the copy number of sequences within the \(lt\) region and its relationship to the transcription unit, we hybridized restriction fragments across the cloned region to EcoRI restricted genomic DNA. The 80-kb region cloned in cosmids was divided into 40 fragments ranging in size from 0.2 kb to 4.1 kb (average length of 2 kb). To assess the variability in organization of these sequences among strains, we included in our survey three different laboratory strains of *D. melanogaster* (Canton-S, Oregon R and \(yw\) bw sp) and *D. simulans, D. pseudoobscura* and *D. viridis.*

The results showed that single copy regions comprise only a small proportion of this region in *D. melanogaster.* Each single copy fragment is less than 3 kb in length and together, they represent approximately 10% (8.2 kb) of the 80 kb surveyed (Figure 7). The single copy regions coincide closely with the transcribed regions. The two largest single copy fragments (from -2.4 to 0.0 kb and from +12.5 to +15.4 kb) correspond to the exons defined by the hybridization experiments with c35. The third (from 0.0 to +0.5) resides in the region containing sequences specific to transcript B. When c35 is used as a hybridization probe of DNA extracted from flies, only fragments predicted from the restriction maps of the cosmids are detected. Hence, the exonic region of c35 contained within the region from coordinates +1.2 to +2.2 is also single copy.

The bulk of the 80-kb region is comprised of repeated sequences. Interestingly, two fragments hybridize to just 3 and 6 EcoRI bands in genomic DNA and these are located just upstream of the transcribed sequences (from -2.4 to -5.2 kb). The remaining regions hybridize to over 20 EcoRI fragments. The middle repetitive sequences are heterogeneous; we observe at least twelve different repeat patterns. Experiments to measure cross hybridization between fragments within the cosmid walk have not been performed. However, comparisons of the EcoRI patterns show that at least two of these repeats are present within the large intron of the major transcript (coordinate +8.0 to +12.0 kb) and also downstream of the gene (+25.4 to +32.0 kb). These same regions also contain sequences that hybridize to the previously characterized middle repetitive elements 2078 (Young 1979, Young and Schwartz 1981) and 2228 (Rubin et al., 1981). Our blot hybridizations show that
The Drosophila light Gene

FIGLTRE (137) (Estimation of the 5' extent of light transcripts by primer extension analysis. A 32P-labeled oligonucleotide complementary to a 20 nucleotide sequence at the 5' end of the c35 cDNA was hybridized to 10 µg of poly(A)+ RNA from adults, extended with reverse transcriptase and analyzed on a sequencing gel. The autoradiograph shows that ½ and ½ of the reaction (middle and right lanes, respectively) yielded an extension product sized at 189 nucleotides (NT) using a sequencing reaction as a size standard (left four lanes).

Figure 6.—Estimation of the 5' extent of light transcripts by primer extension analysis. A 32P-labeled oligonucleotide complementary to a 20 nucleotide sequence at the 5' end of the c35 cDNA was hybridized to 10 µg of poly(A)+ RNA from adults, extended with reverse transcriptase and analyzed on a sequencing gel. The autoradiograph shows that ½ and ½ of the reaction (middle and right lanes, respectively) yielded an extension product sized at 189 nucleotides (NT) using a sequencing reaction as a size standard (left four lanes).

the other middle repetitive elements in the it region are similar to the 2078 and 2228 elements in the following properties. Each element is present at approximately the same copy number in different laboratory strains of D. melanogaster but there is variability among strains in the sizes of the hybridizing bands.

Using the single copy fragments from D. melanogaster or the c35 cDNA as probes, we detected one or two fragments in the DNA of D. simulans, D. pseudoobscura and D. virilis under high stringency hybridization conditions. The repetitive DNA fragments hybridize to as many or nearly as many bands in the DNA of D. simulans as in D. melanogaster, albeit with less intensity. None of the repeated elements were detected in the more distantly related species D. pseudoobscura and D. virilis (data not shown).

DISCUSSION

We have used a hybrid-dysgenic-induced mutation to recover a region of heterochromatin that contains the it gene. The molecular studies of the properties of wild type and mutant alleles establishes that the region we isolated indeed contains the it gene. We have shown that the it gene encodes multiple transcripts; the two most abundant transcripts are altered in mutants. The transcripts are present in ovaries and throughout development in a wide variety of cell types. This wide distribution of transcripts suggests that the it gene may encode a general cellular function. Expression in ovaries is consistent with the strong maternal effect on pigmentation of the Malpighian tubules. A larva homozygous for a hypomorphic allele shows a wild type level of pigmentation if its mother had carried a normal allele (NICKLA 1972). The mapping of the most abundant transcripts, transcripts A and B, and the analysis of cDNAs provides evidence for alternative splicing near the 5' and 3' ends of the transcription unit, and for the use of alternative poly(A)+ addition sites for transcript A.

The A and B transcripts map within a 17-kb region that is included in the region that supports the transient expression of the it+ structural sequences but it is limited in defining its regulatory requirements since the mechanism of rescue after cosmid injection is unknown. Although it is likely that rescue is due to the transcription of the gene directly from the cosm id molecules, expression of chromosomally integrated sequences is a formal possibility. We have not attempted to isolate DNA from larvae that develop from the injected embryos. We note that the frequency of rescue we achieved is relatively low compared to the success of transient expression after injection of plasmids containing the ry+ (RUBIN and SPRADLING 1982) or Adh genes (MARTIN et al. 1986). This may be due to the instability of the larger cosmid molecules after injection and/or the low probability of obtaining within a cell a concentration of molecules high enough to permit rescue. Alternatively, higher levels of it+ expression may depend on sequences not included in the cosmids we injected.

In situ hybridization of cloned sequences to polytene chromosomes show that the it gene lies proximal to the well banded region on the salivary gland chromosomes and within the region referred to as beta-heterochromatin (DEVLIN et al. 1990; WAKIMOTO and HEARN 1990). β-Heterochromatin was described by HEITZ (1934) as the diffusely staining material adjac-
cent to the more compact α-heterochromatin in the 
chromocenter of salivary gland nuclei. β-Heterochro-
matin is replicated in salivary gland chromosomes 
(Gall, Cohen and Polan 1971; Lakhotia 1974), 
transcriptionally active (Biessmann et al. 1981; Mik-
los et al. 1984) and enriched in middle repetitive 
DNA (reviewed in Spradling and Rubin 1981). Stud-
ies by Miklos et al. (1984; 1988) have shown that 
the β-heterochromatin of the X chromosome and the en-
tire chromosome 4 contain middle repetitive elements 
terspersed with the short single copy regions. Al-
though the X chromosome sequences cloned by Mi-
klos et al. (1984) and Healy, Russell and Miklos 
(1988) were shown to be near the uncoordinated locus, 
the relationship of the repeated DNA to transcription 
units of known genetic function has not been re-
ported. The 80-kb region we characterized has molec-
ular properties similar to the β-heterochromatin of the 
X and chromosome 4 regions cloned by Miklos et 
al. (1988) but contains a larger proportion of repeti-
tive DNA. DNA fragments that make up about 90% 
of the region we cloned contain middle repetitive 
DNA; the remaining 10% lack middle repetitive DNA 
and these single copy fragments correspond to the 
exonic regions of the lt transcripts.

The lt region contains a heterogeneous array of 
middle repetitive elements. These elements include 
the previously cloned sequences, 2078 (Young 1979) 
and 2228 (Rubin et al. 1981) which are present within 
the intron and downstream of the gene. The 2078 
sequence is a dispersed, moderately repetitive sequence present in 25 to 60 copies in euchromatin and heterochromatin (Young 1979; Young and Schwartz 1981). We have not performed an exten-
sive characterization of the other repeats from the lt 
region. However the evidence from genomic South-
ern blots shows that these repeats are also polymorphic 
in different laboratory strains of D. melanogaster. The 
lt gene encodes a function that is essential in D. 
melanogaster and sequences that hybridize to the lt 
gene are detected in the DNA of the other Drosophila 
species. The lack of conservation of the repetitive 
DNA that surrounds the gene in D. melanogaster 
suggests that the organization of the lt gene may be quite 
different in the more distantly related species, D. 
pseudoobscura and D. virilis. Alternately, a similar in-
terspersion of single copy and repeated sequences may 
be present in these species, but the repeats may not 
be homologous to those in D. melanogaster. A compara-
tive study of the lt gene in these different species 
should be informative to determine if the homologous 
gene is heterochromatic and if its chromosomal position 
is correlated with the presence of middle repetitive 
elements within and surrounding the gene.

Extended chromosome walks covering 100- to 250-
kilobase regions within euchromatin have shown there are 
no transposable elements found near the wild type 
alleles of the Bithorax Complex (Bender, Spierer 
and Hoggness 1983), the Antennapedia Complex 
(Garber, Kuriowa and Gehring 1983; Scott et al. 
1983), the achaete-scute Complex (Campuzano et al. 
1985) and the Notch gene (Artavanis-Tsakonas, 
Muskavitch and Yedvobnick 1983; Kidd, Lockett 
and Young 1983) of D. melanogaster. Surveys of the 
variation among alleles present in natural populations 
of D. melanogaster have shown that insertions within 
the introns and immediately adjacent to transcribed 
sequences are observed at a low frequency (e.g., see 
Langley et al. 1988). Hence, compared to the func-
tionally unique euchromatic genes that have been 
analyzed thus far, the high density of middle repetitive 
elements within and immediately surrounding the lt 
gene is unusual. However, this type of organization is 
clearly compatible with lt function and it may be
The abundance of middle repetitive elements may be higher in proximal chromosomal regions where recombination is reduced in frequency (Charlesworth, Langley and Stephan 1986) and may not be of fundamental importance for the expression of heterochromatic genes. However, we view the presence of repeated DNA within and immediately surrounding the $lt$ gene as potentially interesting because of the genetic properties of rearrangements that remove $lt$ from its normal chromosomal position and cause variegated expression. The sensitivity of the $lt$ gene to its displacement from proximal chromosomal regions may be due to its movement away from factors concentrated in centromeric heterochromatin (Wakimoto and Hearn 1990). The molecular cloning of the $lt$ gene and definition of its transcripts will permit us to use germline transformation to examine the regulatory requirements of this gene. This approach should allow us to examine if the expression of $lt$, its interactions with modifiers of position effect variegation and its sensitivity to chromosomal position are dependent upon proximity to the repetitive DNA sequences we have isolated.

We thank R. Kelley and A. Spradling for the gift of the cosRK library, L. Tamkun and M. Scott for providing the NotBamNotcosPnR library and Mark Hamilton for help with the figures. We are grateful to M. Hearn, S. Henikoff, R. Levie, S. Parks and J. Tomkel for suggestions on the manuscript. This work was supported by National Institutes of Health grant GM34977 and the Chicago Community Trust Searles Scholars Program.

**LITERATURE CITED**


Kim, J. K., and D. G. Holm, 1988 Hybrid dysgenic mutagenesis of genes located in the centric heterochromatic region of chromosome 3 in Drosophila melanogaster (abstracts of the XVth International Congress of Genetics). Genome 30 (Suppl. 1): 71.


Maniatis, T., E. F. Fritsch and J. Sambrook, 1982 Molecular
R. H. Devlin, B. Bingham and B. T. Wakimoto  


Communicating editor: M. T. FULLER