Enhancer of rudimentary^{p1}, *e(r)*^{p1}, a Highly Conserved Enhancer of the *rudimentary* Gene

Edward Wojcik,*[,1] Anne Marie Murphy,*[,2] Hanna Fares,*[,3] Khoi Dang-Vu* and Stuart I. Tsubota[†,4]

**Department of Biology, University of Michigan, Ann Arbor, Michigan 48109, and †Department of Biology, Saint Louis University, St. Louis, Missouri 63103*

Manuscript received June 29, 1994
Accepted for publication August 22, 1994

ABSTRACT

A hybrid dysgenesis-induced mutation, *enhancer of rudimentary*^{p1} (*e(r)*^{p1}), is a recessive enhancer of a weak *rudimentary* mutant phenotype in *Drosophila melanogaster*. The *e(r)* gene was cloned using *P* element tagging and localized to region 8B on the *X* chromosome. It encodes a 1.0-kb and a 1.2-kb transcript. The 1.0-kb transcript is present in both adult males and females, while the 1.2-kb transcript is predominantly found in females. The difference in the lengths of the two *e(r)* transcripts is caused by two different polyadenylation sites spaced 228 bp apart. The amounts of both of these transcripts are drastically reduced in the *e(r)*^{p1} mutant. The *P* element in *e(r)*^{p1} is inserted in the 5′-untranslated leader region near the start of transcription. It may be producing its effect by suppressing transcription and/or by providing transcription termination and polyadenylation signals. The putative *e(r)* protein is 104 amino acids in length and bears no striking resemblance to protein sequences in GenBank or PIR. While its biochemical function is unknown at this time, sequence analysis indicates that the *e(r)* protein is highly conserved and, presumably, functionally very important. The amino acid sequences of the *D. melanogaster* and the *Drosophila virilis* proteins are 95% identical.

THE *rudimentary* gene of *Drosophila melanogaster* encodes the first three enzymatic activities of the pyrimidine biosynthetic pathway (JARRY and FALK 1974; NORBY 1973; RAWLS and FRISTROM 1975). Not surprisingly, *r* mutations are pyrimidine auxotrophs (FALK and NASH 1974; NORBY 1970). Most Drosophila media contain enough pyrimidines to allow survival of *r* mutants to adulthood; however, both sexes exhibit a characteristic truncation of the wings. The phenotypes suggest that, although *r* encodes basic metabolic functions that are necessary at all stages in development, its regulation is temporally and spatially regulated to ensure normal development of the wings and ovaries. Consistent with this idea is that *r* expression is highest in embryos (BROTHERS *et al.* 1978), adult females (RAWLS 1979; TSUBOTA and SCHEDL 1986), and imaginal discs (S. TSUBOTA, unpublished results), and that the expression of *r* in the ovaries is localized to the nurse cells (AMBROSIO and SCHEDL 1984).

The *r* mutant wing phenotype is very sensitive to changes in *r* expression and is a powerful tool in the isolation of *cis*-acting regulatory mutations at the *r* gene and *trans*-acting mutations which affect the wing phenotype. In a screen for mutations which changed the wing phenotype of hypomorphic *r* mutations, 20 mutations were isolated out of 7793 flies (TSUBOTA and SCHEDL 1986). Eighteen of these mutations had changes in the 5′-control region of the *r* gene.

The remaining two mutations mapped to other genes and demonstrate that changes in the wing phenotype can be used to identify *trans*-acting mutations as well as *cis*-acting mutations. One of these mutations called *enhancer of rudimentary*^{p1}, *e(r)*^{p1}, enhances the wing truncation of hypomorphic *r* mutants. This study deals with the initial characterization of *e(r)*^{p1}, its effects on *r* expression, and the cloning and sequencing of the wild-type *e(r)* gene.

MATERIALS AND METHODS

***D. melanogaster* mutations:** The *r* alleles, r^{hd1}, $r^{hd1\text{-}12}$ and r^{hd4}, are hypomorphic mutations and have been previously described (TSUBOTA and SCHEDL 1986). *e(r)*^{p1} was originally incorrectly designated as the *r* allele, $r^{hd1\text{-}16}$ (TSUBOTA and SCHEDL 1986). The mutations *y* (yellow body color) and *w* (white eye color) have been previously described (LINDSLEY and ZIMM 1992). *Df(1)KA14* (7F1-2; 8C6) is a large deficiency which includes the *e(r)* region (LINDSLEY and ZIMM 1992).

Cloning of *e(r)*^{p1}: Given that *e(r)*^{p1} was isolated in a hybrid dysgenesis mutagenesis, it was probably caused by an insertion of a *P* element. An *in situ* with a *P* element probe to the polytene chromosomes indicated that there was a *P* element in 8B on the *X* chromosome, the vicinity of the genetic map position of *e(r)*. DNA from *e(r)*^{p1} was partially digested with *Sau*3A and the DNA in the range of 15–20 kb was ligated into the *Bam*HI site of EMBL3. The resulting library was screened with a *P* element probe and positive clones were isolated. Each clone was hybridized to the polytene chromosomes of a

The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession number L36921.

[1] Present address: Department of Biochemistry, Medical Sciences Institute, The University, Dundee DD1 4HN, Scotland.

[2] Present address: Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218.

[3] Present address: Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599.

[4] To whom correspondence should be addressed.

wild-type strain, which does not contain any *P* elements, to determine if any came from the *e(r)* region. One clone hybridized to 8B and was used to screen a wild-type genomic library in EMBL3, made by cloning fragments generated by a partial *Sau*3A digestion (TSUBOTA *et al.* 1989).

Northern Blot analysis: Poly(A)$^+$ RNA was isolated from either adult females or males from the appropriate strains using established procedures (CHOMCZYNSKI and SACCHI 1987). For each lane 7 μg of RNA were used in a 0.9% agarose-formaldehyde gel (MANIATIS *et al.* 1982). The gel was blotted with a nylon filter (Micron Separations, Inc.) and prehybridized and hybridized in a solution containing 5 × SSPE, 1.0% SDS, 10 × Denhardt's solution, 150 μg/ml denatured salmon sperm DNA, and 50% formamide at 52°. ^{32}P-labeled probes were prepared using random priming (FEINBERG and VOGELSTEIN 1983) and the Boehringer Mannheim kit. For the *r* transcript, a cDNA fragment from pCrud5 (SEGRAVES *et al.* 1984) was labeled, for *e(r)* a 6.2-kb *Sal*I fragment containing the wild-type *e(r)* gene (see Figure 3) was labeled, and for *Ras2*, a *Ras2* genomic fragment (MOZER *et al.* 1985) was labeled.

cDNA Isolation: Two cDNA libraries were screened using the 6.2-kb *Sal*I fragment containing the *e(r)* gene as a probe. The first library was from adults (Stratagene), while the second library was from embryos (POOLE *et al.* 1985).

Primer extension analysis: Primer extension analysis was performed essentially as previously described (KINGSTON 1987). Reverse transcriptase was purchased from Stratagene. The extension reaction was performed utilizing poly(A)$^+$ RNA isolated from both adult male and female wild-type flies. The primer was expected to hybridize within 125 bp of the putative start of transcription as estimated from the length of the longest isolated cDNA. The length of the resulting extension product was determined by comparison against a DNA-sequencing ladder generated with the same primer on a genomic DNA clone. The primer sequence was 5′-CGTTCGAACGCCGGACT-3′.

Site-directed mutagenesis: The *e(r)* open reading frame was chosen on the basis of maximum length, consensus with the Drosophila start of translation (CAVENER 1987), and codon usage (GRIBSKOV *et al.* 1984). Site-directed mutagenesis was performed according to KUNKEL *et al.* (1991). The oligonucleotide used in the mutagenesis had the sequence, 5′-CCATCCTATAGGTATAGCCGGGTG-3′. This oligonucleotide replaces the seventh and ninth codons (see Figure 6A) with the nonsense codon, UAG, to produce a null mutation which drastically truncates the *e(r)* protein. The mutations were made in a 2.0-kb *Xba*I-*Sal*I fragment. After the mutagenesis, this entire fragment was sequenced to ensure that only the two nonsense mutations had been generated.

Transformation with *e(r)*: Two different *e(r)* transformation vectors were constructed, one with the wild-type 6.2-kb *Sal*I fragment and one with a mutated *Sal*I fragment (see the previous section). In the making of the wild-type construct, the 6.2-kb *Sal*I fragment was first cloned into Bluescript KS–. The fragment was then excised with *Xho*I and *Eco*RI and cloned into pCaSpeR 4 (PIRROTTA 1988). The mutant construct was made by replacing the wild-type *Xba*I-*Sal*I fragment contained within the 6.2-kb *Sal*I with the mutated fragment. Each construct along with the helper plasmid 25.7wc (KARESS and RUBIN 1984) was then used to transform a *y w* strain of *D. melanogaster* using published protocols (SPRADLING 1986).

DNA sequencing: All clones were made using the plasmids Bluescript SK– and KS– (Stratagene) and sequenced using the dideoxy procedure (SANGER *et al.* 1977). All genomic and cDNA clones were sequenced on both strands and, where necessary, nested deletions were constructed with a Stratagene Exo III/Mung Bean Deletion Kit.

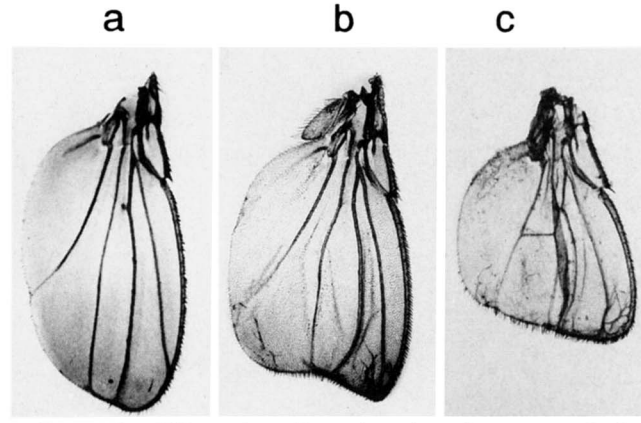

FIGURE 1.—Effect of $e(r)^{p1}$ on the wing phenotype of a hypomorphic *r* allele. Wing phenotypes. (a) *cv* $e(r)^+$ r^{rd1} male, (b) *cv* $e(r)^{p1}$ r^{rd1} male, (c) $e(r)^{p1}$ r^{rd1}/*Df(1)KA14* r^{rd1} female. The wings of the males display the *crossveinless*, *cv*, phenotype. This phenotype does not effect the *rudimentary* phenotype.

Cloning of the *Drosophila virilis* *e(r)* gene: A fragment containing the *D. melanogaster e(r)* open reading frame (see Figure 6A) was amplified using the polymerase chain reaction (SAIKI *et al.* 1988). This fragment was used to probe a *D. virilis* genomic library obtained from WILLIAM GELBART's laboratory at Harvard University. This library was constructed by cloning large *Sau*3A fragments into the *Bam*HI site of EMBL3. One clone was isolated, from which the *e(r)* coding region was subcloned into Bluescript KS– and sequenced by the dideoxy procedure (SANGER *et al.* 1977).

RESULTS

Phenotypes of $e(r)^{p1}$: When first isolated, $e(r)^{p1}$ was presumed to be a *r* allele and given the name $r^{hd1\text{-}16}$ (TSUBOTA and SCHEDL 1986). Further studies showed that it was not allelic to *r* and behaved as a recessive enhancer of hypomorphic *r* alleles such as r^{hd1}, r^{hd4} and $r^{hd1\text{-}12}$. For example, r^{hd1} produces a slightly truncated wing, whereas $e(r)^{p1}$ r^{rd1} results in a much more severely truncated wing (Figure 1). However, $e(r)^{p1}$ is probably not a null mutation, because the heterozygote, $e(r)^{p1}$ r^{rd1}/*Df(1)KA14* r^{rd1}, has more severely truncated wings than either the homozygous female or the hemizygous male, $e(r)^{p1}$ r^{rd1} (Figure 1). *Df(1)KA14* is a deficiency that completely deletes the *e(r)* gene. In a wild-type *r* background, $e(r)^{p1}$ does not display a mutant wing phenotype, so its effect can only be seen in a mutant *r* background.

To determine if $e(r)^{p1}$ could be affecting *r* transcript levels, poly(A)$^+$ RNA from $e(r)^{p1}$ and $e(r)^+$ adult males and females was examined by Northern blotting (Figure 2). The filter was scanned with a laser densitometer to determine relative amounts of *r* transcript standardized to the levels of *Ras2* (MOZER *et al.* 1985). This analysis indicated that $e(r)^{p1}$ males had about 50% the amount of *r* transcript as their $e(r)^+$ counterparts. The shift in the size of the *r* transcript in $e(r)^{p1}$ males is not reproducible and is not a result of the $e(r)^{p1}$ mutation. Little difference was seen between the two classes of females. It can be

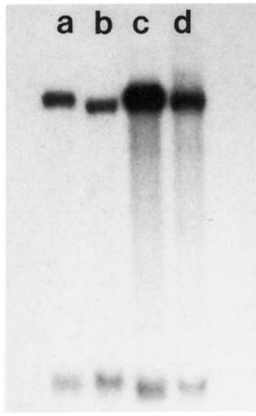

FIGURE 2.—Effect of $e(r)^{p1}$ on the *rudimentary* transcript. Poly(A)$^+$ RNA from adult males or females of the designated stocks was examined with Northern blotting using a probe for the *r* gene. The higher molecular weight bands are the *r* transcript and the lower molecular weight bands are the *Ras2* transcript that was used as an internal control. (a) $e(r)^+$ r^+ males, (b) $e(r)^{p1}$ r^+ males, (c) $e(r)^+$ r^+ females, (d) $e(r)^{p1}$ r^+ females.

concluded that $e(r)^{p1}$ is not drastically affecting *r* transcript levels in adults and it is unlikely that $e(r)^{p1}$ is acting at the level of the *r* transcript.

Cloning of *e(r)*: Because $e(r)^{p1}$ was isolated in a P-M hybrid-dysgenesis screen, it is most likely the result of a *P* element insertion. This contention is supported by the instability of $e(r)^{p1}$ and the presence of a *P* element in chromosomal position 8B, the vicinity of the genetic map position of *e(r)*. To clone $e(r)^{p1}$, a genomic library of $e(r)^{p1}$ was screened for *P* element clones. Each clone was hybridized to Oregon R chromosomes to determine the chromosomal position of the non-*P* element DNA. One clone hybridized to 8B. This putative $e(r)^{p1}$ clone was used to probe a wild-type genomic library. Four clones were isolated which span a region of about 28 kb (Figure 3). The putative $e(r)^{p1}$ *P* element is inserted into a 6.2-kb *Sal*I fragment.

To determine whether the *e(r)* gene had been cloned, revertants of $e(r)^{p1}$ were isolated. Two flies out of 9,201 were obtained in screens for revertants for $e(r)^{p1}$. The *P* element responsible for the $e(r)^{p1}$ mutation should have been mobilized in the production of these revertants. Any changes that might have occurred in the cloned region in these two flies were examined by Southern analysis. In one of the revertants, the *P* element within the 6.2-kb *Sal*I fragment had excised precisely, whereas, in the other, the *P* element had excised imprecisely. Together these revertants argue that the isolated clones were indeed of the *e(r)* gene.

The imprecise excision resulted in an unusual sequence which was revealed by DNA sequencing. The genomic region flanking and including the *P* element insertion site from this line was amplified by the polymerase chain reaction method and cloned into Bluescript for DNA sequencing. Three independent clones were sequenced. In this revertant, the *P* element had precisely excised and left behind an 89-bp insertion at its original site (Figure 4). This insertion is flanked by the characteristic *P* element site duplication of 8 bp. The insertion itself consists of 53 bp of unknown origin flanked by sequences from the *P* element 31-bp inverted repeats. The 53-bp sequence is not found within the *P* element or the *e(r)* gene and consists of repeats of TATGTTA interspersed with repeats of TA. Interestingly, the TATGTTA repeat is found within the *P* element inverted repeats and suggests that the 53-bp sequence may have originated from the inverted repeats during mobilization. Similar unusual sequences have been reported for other imprecise *P* element excisions and it has been proposed that they are generated from *P* element sequences during mobilization (TAKASU-ISHIKAWA *et al.* 1993).

While the revertant data are consistent with the conclusion that the *e(r)* gene had been cloned, conclusive data were provided by germ-line transformation. In these experiments, the wild-type 6.2-kb *Sal*I fragment, which contains the *P* element insertion site of $e(r)^{p1}$, was inserted into the *D. melanogaster* genome using the pCaSpeR 4 vector. Four independent insertions were isolated, one on the second chromosome and three on the third chromosome. Each rescued the $e(r)^{p1}$ mutation, demonstrating that the wild-type *e(r)* gene is contained within this 6.2-kb *Sal*I fragment.

***e(r)* transcripts:** Because $e(r)^{p1}$ is a recessive hypomorphic mutation, the *P* element insertion probably results in the reduction of the transcript levels or in the alteration in the structure of a transcript. To examine this issue, the 6.2-kb *Sal*I fragment in which the *P* element had inserted was used to probe a Northern blot of $e(r)^{p1}$ and wild-type RNA. Two transcripts of 1.0 and 1.2 kb were seen in wild-type adult males and females (Figure 5A). The 1.0-kb transcript is found in roughly equal amounts in both males and females, whereas the 1.2-kb transcript is found predominantly in females. In $e(r)^{p1}$, the amount of both transcripts is greatly reduced (Figure 5A). A *Ras2* probe was used as a control for the amount of RNA loaded per lane. Both the 1.0-kb and the 1.2-kb transcripts map to a 2.0-kb *Xba*I-*Sal*I fragment that is 3′ of the *P* element (Figure 3). These data suggest that the two transcripts seen in wild-type males and females are encoded by a single gene, *e(r)*, and that the effects of the $e(r)^{p1}$ mutation are the result of the reduction in the amounts of these transcripts.

To analyze the structure and the function of the *e(r)* gene in more detail, 17 *e(r)* cDNA clones from adult and embryo cDNA libraries were isolated and sequenced. A Northern blot using one of the cDNAs was performed and showed that both transcripts hybridized to the cDNA sequence (E. WOJCIK and S. TSUBOTA, unpublished results). This evidence eliminates the possibility that the two bands observed in the Northern analysis arise from two closely spaced independent genes.

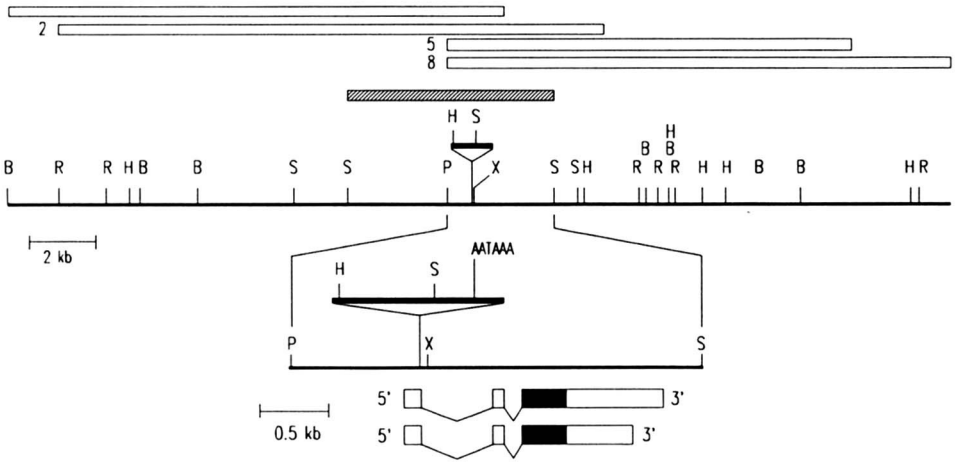

FIGURE 3.—Restriction map of the wild-type *e(r)* region. The four open boxes above the map are the wild-type λ clones that define the region. The *P* element in $e(r)^{p1}$ is shown as a solid bar above the map. The 6.2-kb *Sal*I fragment used as a probe Northern blot (Figure 5A) is shown as a striped bar above the map. The two *e(r)* transcripts that differ in their 3′ ends are shown below the map. The solid black regions in the transcripts define the amino acid coding region. *Xba*I and *Pvu*II were only mapped within the 6.2-kb *Sal*I fragment. B = *Bam*HI, H = *Hin*dIII, *P* = *Pvu*II, R = *Eco*RI, S = *Sal*I, and X = *Xba*I. There are no *Xho*I sites in the region.

5' TATTATCATAGTCCATAT CATGATGAAATAACATG GTTATGTT
ATATGTTATATATATATATGTTATATATATGTTATATATA
TATGT TATATGTTATTTCATCATG GTCCATATAGTCCGGCGTT 3'

FIGURE 4.—DNA sequence of the insertion site of the $e(r)^{p1}$ revertant, resulting from an imprecise excision. The *e(r)* sequence is given in the smaller font, with the site duplications underlined. The remnants of the *P* element terminal repeats are enclosed in boxes and flank the sequence of unknown origin.

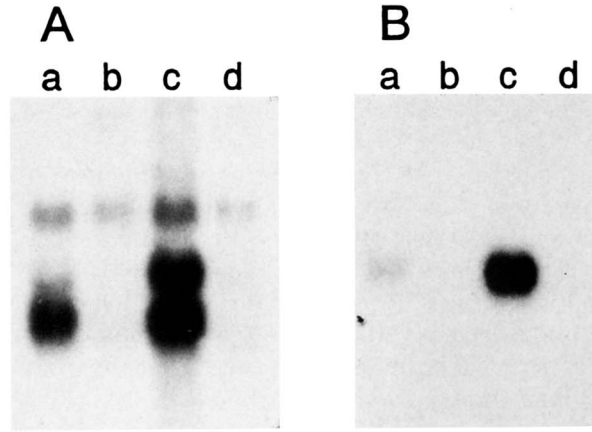

FIGURE 5.—Northern blot of the *e(r)* transcript in $e(r)^{+}$ and $e(r)^{p1}$ adults. For each lane, 7 μg of poly(A)$^{+}$ RNA were loaded. (a) $e(r)^{+}$ males, (b) $e(r)^{p1}$ males, (c) $e(r)^{+}$ females, (d) $e(r)^{p1}$ females. (A) A ^{32}P-labeled 6.2-kb *Sal*I fragment (Figure 3) was used as a probe for *e(r)* and a *Ras2* probe was used to control for RNA loading. The upper band is *Ras2*, which is approximately 1.4 kb. The two lower bands are the *e(r)* transcripts of approximately 1.2 and 1.0 kb. (B) A ^{32}P-labeled 3′ fragment from the *Cla*I site to the 3′ end of the longer cDNA (Figure 6A) was used as a probe. Only the longer transcript is detected in both wild-type males and females.

The composite cDNA sequence was compared to the DNA sequence of the corresponding genomic regions to identify exons and introns. The longest cDNA minus the poly(A)$^{+}$ tail, which was isolated from the embryo cDNA library, is 1237 bp in length (Figure 6A). This size agrees with the estimate for the size of the longer *e(r)* transcript (1.2 kb). That this is probably a full-length cDNA is supported by the fact that its 5′ end corresponds exactly to the 5′ ends of the *e(r)* transcripts as determined by primer-extension analysis (Figure 7). All of the cloned cDNAs contained the same intron/exon structure and all contained at least a portion of the first exon. Given these data and the fact that primer-extension analysis revealed only a single putative start of transcription, it can be concluded that both *e(r)* transcripts have the same 5′ ends.

From 5′ to 3′ the transcribed region contains three exons of 130, 84 and 1023 bp and two introns of 520 and 125 bp. The sequencing of the cDNAs revealed that they were all spliced identically. However, they fell into two classes which differed at the site of polyadenylation. The longer class contained the 1023-bp third exon, while the shorter class contained a 802-bp third exon. The transcript represented by the longer cDNA contains the polyadenylation signal AAUAAA 39 bp upstream of the polyadenylation site. The transcript represented by the shorter cDNA contains a weak polyadenylation signal, UAUAAA (WICKENS 1990), 47 bp upstream of the polyadenylation site. To verify that these two classes of polyadenylated cDNA clones represented the 1.2-kb and the 1.0-kb *e(r)* transcripts, a Northern blot of poly(A)$^{+}$ RNA was probed with a 3′ fragment that was specific to the longer cDNA. As expected, it hybridized only to the longer of the two *e(r)* transcripts (Figure 5B). When this filter was probed with a fragment covering the entire gene, both the larger and the smaller transcripts were detected (E. WOJCIK and S. TSUBOTA, unpublished results).

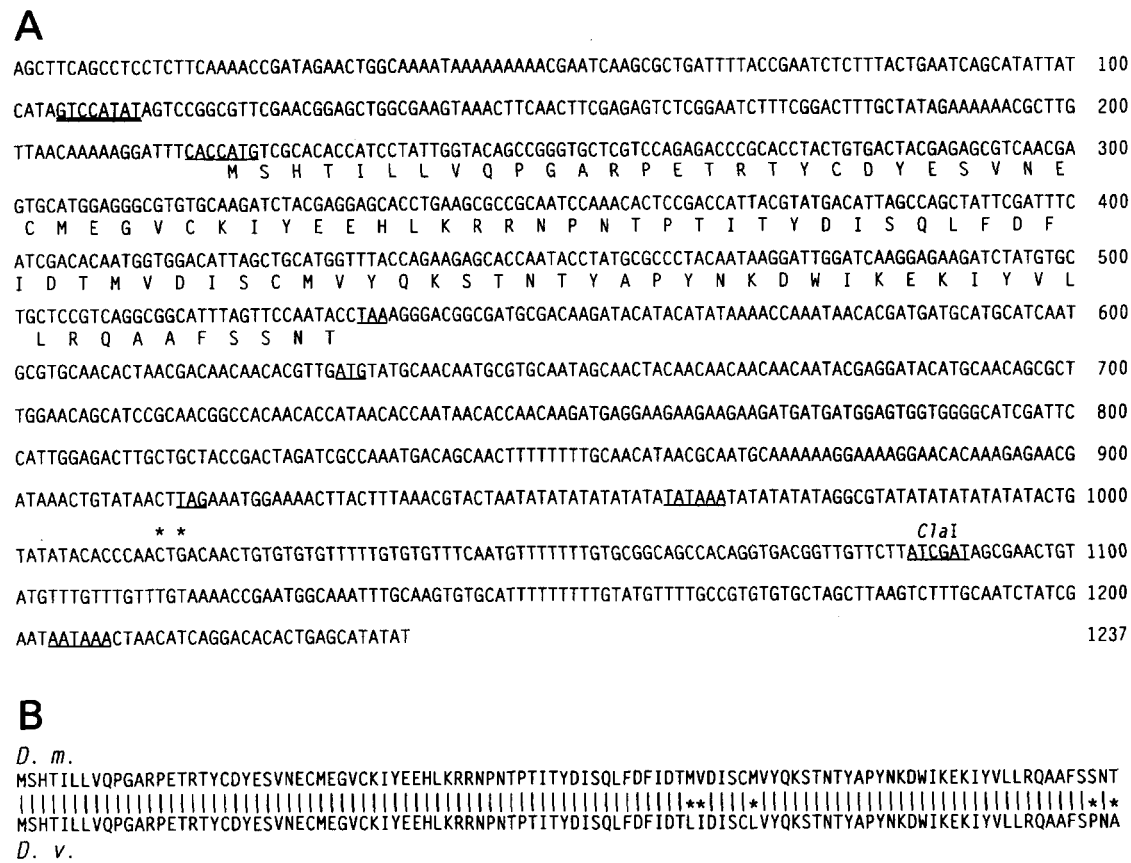

FIGURE 6.—(A) Sequence of the putative full-length *e(r)* cDNA. The consensus start of translation and the translation stop are underlined. The sequence of the putative *e(r)* protein is given. The translation start and stop of a second open reading frame are also underlined. The polyadenylation sites of the short *e(r)* transcripts are designated with asterisks and the sequences encoding the two polyadenylation signals, UAUAAA and AAUAAA are underlined. The sequence in the 5′ end that was duplicated upon insertion of the $e(r)^{p1}$ *P* element is double underlined. The *Cla*I site that was used to generate the 3′ probe is indicated. (B) Sequence of the *D. melanogaster* and *D. virilis e(r)* proteins. Asterisks designate amino acid changes.

***e(r)* protein:** An analysis for open reading frames within the cDNA sequence, using the Eugene analysis software package, produced only one sequence as the probable coding region (Figure 6A). It has a consensus Drosophila translation start site (CAVENER 1987), begins at the most 5′ ATG and has very good Drosophila codon usage throughout its length (GRIBSKOV *et al.* 1984). Another possible coding region starts downstream from this open reading frame (Figure 6A). However, it does not have a consensus Drosophila translation start sequence and has only mediocre Drosophila codon usage throughout its sequence. Given these features, it is unlikely that the second open reading frame is a real coding region.

Data to support the conclusion that the first open reading frame encodes the *e(r)* protein come from *P* element-mediated transformation. As was stated earlier, the wild-type 6.2-kb *Sal*I fragment rescues the $e(r)^{p1}$ mutation. To determine if a mutation in the first open reading frame identified by the cDNA analysis would destroy *e(r)* function, site directed mutagenesis was performed to replace the seventh and ninth codons with stop codons. The 6.2-kb *Sal*I fragment containing the mutated open reading frame was cloned into pCaSpeR 4 and used to transform *D. melanogaster*. Four different transformants containing autosomal insertions, three on the third chromosome and one on the second chromosome, were isolated. In contrast to the four wild-type insertions, all of the mutant insertions failed to rescue the $e(r)^{p1}$ mutation. Together these data indicate that the first open reading frame is the *e(r)* coding region.

The first open reading frame encodes a protein of 104 amino acids long (Figure 6). There are no striking motifs or domains which would reveal a possible function for the protein. A comparison of this sequence with those in GenBank and PIR did not reveal any striking similarities with other proteins, which could provide clues to the biochemical function of the *e(r)* protein.

To access the evolutionary conservation of the *e(r)* protein and to identify functionally important domains of the protein, the *e(r)* open reading frame from *D. virilis* was cloned and sequenced. *D. virilis* and *D. melanogaster* diverged approximately 60 million years ago (BEVERLY and WILSON 1984), which allows enough time for the accumulation of a significant number of amino acid substitutions. Proteins whose functions are known

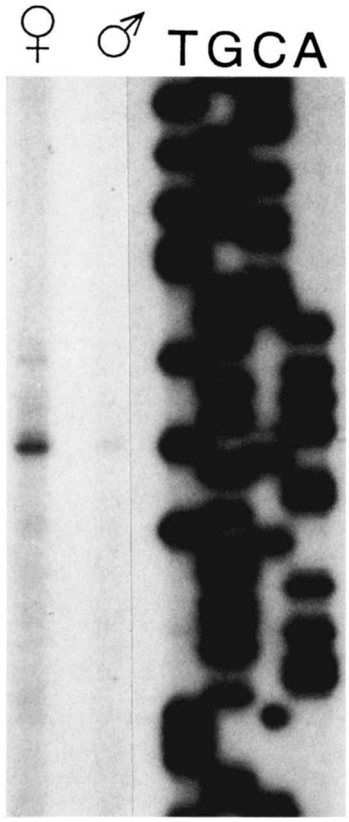

FIGURE 7.—Primer-extension analysis of the 5′ ends of the *e(r)* transcripts. Poly(A)$^+$ RNA from wild-type males and females was used in these reactions. To map the 5′ end of the transcripts, a sequencing ladder generated from a genomic DNA clone was electrophoresed next to the extension products. The same primer (5′-CGTTCGAACGCCGGACT-3′) was used for both the primer extension reaction and sequencing. Only one strong band was seen, which maps precisely to the end of the longest cDNA.

to be important in the development of *D. melanogaster* have displayed 60–80% amino acid identity with the corresponding protein in *D. virilis* (COLOT *et al.* 1988; HEBERLEIN and RUBIN 1990; KASSIS *et al.* 1986; MICHAEL *et al.* 1990; TREIER *et al.* 1989). The *e(r)* protein differs dramatically from these proteins. The *e(r)* protein from *D. virilis*, like its counterpart in *D. melanogaster*, is 104 amino acids long and is 95% identical to the *D. melanogaster* protein. (Figure 6B). Also, the amino acid substitutions that have occurred are conservative. The only other amino acid sequences that have been shown to display this high degree of sequence conservation between *D. melanogaster* and *D. virilis* are the homeobox regions of the *engrailed* (KASSIS *et al.* 1986) and *rough* (HEBERLEIN and RUBIN 1990) genes. This high degree of conservation of the *e(r)* proteins argues very strongly that the *e(r)* protein is under strict evolutionary constraints and suggests that the *e(r)* gene encodes a function of vital importance to the normal development of the fruit fly.

DISCUSSION

The *e(r)* gene encodes two transcripts which have the same 5′ ends and differ only in their 3′-untranslated regions as a result of alternative polyadenylation. The purpose for the two different transcripts is unclear. One possibility is that the longer transcript is involved in oogenesis, because it is found predominantly in adult females. In support of this hypothesis, we have shown that the longer transcript is found in the nurse cells of the developing ovaries and in preblastoderm embryos (E. WOJCIK and S. TSUBOTA, unpublished results). It seems likely that the longer transcript is deposited into the egg by the nurse cells and that the 3′ untranslated region may be involved in this process or in the localization of the transcript.

The sequence data indicate that the longer transcript is generated from the consensus polyadenylation signal, AAUAAA, while the shorter transcript is generated from a much weaker signal, UAUAAA (WICKENS, 1990). Given the fact that the ratios of the abundance of these two transcripts in adult males and females are drastically different, a mechanism must exist which regulates the polyadenylation of the *e(r)* transcripts. In the absence of any more data, we would argue that the normal polyadenylation system would use the stronger signal, AAUAAA, and that a regulatory mechanism exists to ensure the use of the weaker signal, UAUAAA, of the shorter transcript. This mechanism would be particularly prevalent in males where the shorter transcript predominates.

The analysis of the cDNA sequences revealed two possible *e(r)* open reading frames. However, the more 5′ of the two open reading frames appears to be the actual *e(r)* coding region based on its more 5′ position, a better translation initiation consensus sequence, and better codon usage. We have been able to confirm, by Western blotting analysis with antibodies to the protein encoded by the first open reading frame, that this open reading frame is translated and yields a protein of the predicted molecular weight. In addition, antibody staining and *in situ* hybridization reveal that that *e(r)* transcript and the protein encoded by the first open reading frame have identical patterns of localization in the embryo (E. WOJCIK and S. TSUBOTA, unpublished results). These data, plus the fact that a transformation construct containing stop mutations in the first open reading frame fails to rescue the $e(r)^{p1}$ mutation, leave little doubt that the first open reading frame encodes the *e(r)* protein.

The *P* element in $e(r)^{p1}$ is inserted within the first exon, 112 bp from the designated start of transcription within the non-translated leader (Figure 6A). The location of the insertion site raises the question as to why $e(r)^{p1}$ has such a reduced amount of transcript. If transcription continued unimpaired through the *P* element, then larger transcripts would be seen in $e(r)^{p1}$ flies in

amounts comparable to the wild-type *e(r)* transcripts. This is not seen. These data present the possibility that the *P* element may be having an inhibitory effect on transcription initiation or that transcription is terminating within the *P* element. The sequence of the *P* element in $e(r)^{p1}$ suggests that transcription is terminating in the *P* element. The sequence from the *Sal*I site to the 3′ end of the *P* element is identical to that of the intact *P* element (O'HARE and RUBIN 1983). Within this sequence is the normal polyadenylation signal for the *P* element, located 211 bp from the 3′ end (LASKI *et al.* 1986). Therefore, in $e(r)^{p1}$, even if transcription initiation is not suppressed, transcription should be terminated within the *P* element. This could explain the apparent lack of *e(r)* transcripts in the $e(r)^{p1}$ as any *e(r)* probe would have, at best, only 112 bp of identity with the putative $e(r)^{p1}$ transcript.

Although the *P* element in $e(r)^{p1}$ has inserted a polyadenylation site prior to the start of translation, the genetic data argue that $e(r)^{p1}$ is a hypomorph and therefore some expression of *e(r)* must be occurring. Possible mechanisms that could yield some wild-type expression in these mutants are transcription through the *P* element, the use of cryptic splice sites, and the use of cryptic transcription initiation sites. The mechanism by which an insertion in the 5′-untranslated leader sequence can produce a hypomorphic mutation remains unknown, however there are numerous cases of such mutations in other genes (SEARLES *et al.* 1986; COULTER *et al.* 1990; BEGLEY 1992).

The genetic data indicating the hypomorphic nature of $e(r)^{p1}$ beg the question as to the phenotype of the null mutation. At this time, no null mutations have been identified, so the question remains unanswered. However, given the extreme conservation in *e(r)* protein between *D. melanogaster* and *D. virilis*, we would argue that *e(r)* encodes an important function and that the null phenotype would be lethality. We would also argue that given that the *e(r)* transcript is deposited into the developing oocyte, that the lethality would be first exhibited during embryogenesis. The testing of these hypotheses await the isolation of *e(r)* null mutations.

The *e(r)* gene was identified as a possible regulator of the *r* gene. However, the available data do not present a strong argument for this possibility and the case for *e(r)* as a *r* regulator remains open. Nevertheless, the fact that $e(r)^{p1}$ enhances the mutant wing phenotype of hypomorphic *r* alleles argues that if the two genes are not interacting in some manner, then they are at least involved in some common developmental pathway. One possibility is that *e(r)* is involved in either the biosynthesis or degradation of pyrimidines. All of the genes for the enzymes in pyrimidine biosynthesis have been mapped (JARRY and FALK 1974; NORBY 1973; LASTOWSKI and FALK 1980; RAWLS 1980; RAWLS *et al.* 1981) as well as a gene involved in the degradation of pyrimidines (BAHN 1973; STROMAN 1974). Based on its map position, *e(r)* cannot be any of these genes. If *e(r)* is involved in the metabolism of pyrimidines, then its role may be regulatory in nature.

The phenotype of $e(r)^{p1}$ reveals the usefulness of using hypomorphic alleles to screen for suppressor and enhancer mutations in other interacting genes. In the presence of a wild-type *r* allele, $e(r)^{p1}$ does not display an aberrant wing phenotype and the flies appear otherwise normal. This means that $e(r)^{p1}$ would not have been detected in a screen utilizing a wild-type *r* allele. If we are correct in predicting that *e(r)* null mutations are recessive lethals, then these mutations could only be isolated in laborious screens for lethal mutations, the vast majority of which would be unrelated to *r*. Thus, the screens for suppressors and enhancers of hypomorphic mutations allows for the isolation of mutations in genes that would otherwise be difficult to identify.

This work was supported by National Institutes of Health grant GM38315 (to S.I.T.) and National Science Foundation grant DMB-9205091 (to S.I.T.).

LITERATURE CITED

AMBROSIO, L., and P. SCHEDL, 1984 Gene expression during *Drosophila melanogaster* oogenesis: analysis by *in situ* hybridization to tissue section. Dev. Biol. **105:** 80–92.

BAHN, E., 1973 A suppressor locus for the pyrimidine requiring mutant *r*. Drosophila Inf. Ser. **49:** 98.

BEGLEY, D. A., 1992 Cis- and trans-acting regulatory elements of the *rudimentary* gene in *Drosophila melanogaster*. Ph.D. Thesis, University of Michigan.

BEVERLY, S. M., and A. C. WILSON, 1984 Molecular evolution in *Drosophila* and higher dipteran. II. A time scale for fly evolution. J. Mol. Evol. **21:** 1–13.

BROTHERS, V., S. TSUBOTA, S. GERMERAAD and J. FRISTROM, 1978 *Rudimentary* locus of *Drosophila melanogaster*: partial purification of a carbamylphosphate synthase-aspartate transcarbamylase-dihydroorotase complex. Biochem. Genet. **16:** 321–332.

CAVENER, D., 1987 Comparison of the consensus flanking translational start sites in *Drosophila* and vertebrates. Nucleic Acids Res. **15:** 1353–1361.

CHOMCZYNSKI, P., and N. SACCHI, 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156–159.

COLOT, H. V., J. C. HALL and M. ROSBASH, 1988 Interspecific comparison of the period gene of *Drosophila* reveals large blocks of non-conserved coding DNA. EMBO J. **7:** 3929–3837.

COULTER, D. R., E. A. SWAYKUS, M. A. BERAN-KOEHN, D. GOLDBERG, E. WIESCHAUS *et al.*, 1990 Molecular analysis of *odd-skipped*, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. EMBO J. **8:** 3795–3804.

FALK, D. R., and D. NASH, 1974 Sex-linked auxotrophic and putative auxotrophic mutants in *Drosophila melanogaster*. Genetics **76:** 755–766.

FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:** 6–13.

GRIBSKOV, M., J. DEVEREUX and R. R. BURGESS, 1984 The codon preference plot: graphic analysis of protein coding sequences and predictions of gene expression. Nucleic Acids Res. **12:** 539–549.

HEBERLEIN, U., and G. M. RUBIN, 1990 Structural and functional comparisons of the *Drosophila virilis* and *Drosophila melanogaster* rough genes. Proc. Natl. Acad. Sci. USA **87:** 5916–5920.

JARRY, B., and D. FALK, 1974 Functional diversity within the *rudimentary* locus of *Drosophila melanogaster*. Mol. Gen. Genet. **135:** 113–122.

KARESS, R., and G. RUBIN, 1984 Analysis of P transposable element functions in Drosophila. Cell **38:** 8684–8688.

KASSIS, J. A., S. J. POOLE, K. K. WRIGHT and P. H. O'FARRELL, 1986 Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J. **5:** 3583–3589.

KINGSTON, R. E., 1987 Primer extension, p. 4.8.1–4.8.3, in *Current Protocols in Molecular Biology*, edited by F. M. AUSUBEL, R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN *et al.* John Wiley & Sons, New York.

KUNKEL, T. A., K. BEBENEK and J. MCCLARY, 1991 Efficient site-directed mutagenesis using uracil-containing DNA. Methods Enzymol. **204:** 125–139.

LASKI, F., D. RIO and G. RUBIN, 1986 Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. Cell **44:** 7–19.

LASTOWSKI, D. M., and D. R. FALK, 1980 Characterization of an autosomal rudimentary-shaped wing mutation in *Drosophila melanogaster* that affects pyrimidine synthesis. Genetics **96:** 471–78.

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster.* Academic Press, San Diego.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MICHAEL, W. M., D. L. BOWTELL and G. M. RUBIN, 1990 Comparison of the sevenless genes of *Drosophila virilis* and *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **87:** 5351–5353.

MOZER, B., R. MARLOR, S. PARKHURST and V. CORCES, 1985 Characterization and developmental expression of a *Drosophila ras* oncogene. Mol. Cell. Biol. **5:** 885–889.

NORBY, S., 1970 A specific requirement for pyrimidines in *rudimentary* mutants of *Drosophila melanogaster.* Hereditas **66:** 205–214.

NORBY, S., 1973 The biochemical genetics of *rudimentary* mutants of *Drosophila melanogaster.* I. Aspartate carbamoyltransferase levels in complementing and non-complementing strains. Hereditas **73:** 11–16.

O'HARE, K., and G. RUBIN, 1983 Structure of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34:** 25–35.

PIRROTTA, V., 1988 Vectors for P-element transformation in *Drosophila*, in *Vectors. A Survey of Molecular Cloning Vectors and Their Uses*, edited by D. B. ROBERTS. IRL Press, Oxford.

POOLE, S. J., L. M. KAUVAR, B. DREES and T. KORNBERG, 1985 The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell **40:** 37–43.

RAWLS, J. M., 1979 The enzymes for *de novo* pyrimidine biosynthesis in *Drosophila melanogaster*: their localization, properties and expression during development. Comp. Biochem. Physiol. **62B:** 207–216.

RAWLS, J. M., 1980 Identification of a small genetic region that encodes orotate phosphoribosyltransferase and orotidylate decarboxylase in *Drosophila melanogaster.* Mol. Gen. Genet. **178:** 43–49.

RAWLS, J. M., and J. FRISTROM, 1975 A complex genetic locus that controls the first three steps of pyrimidine biosynthesis in Drosophila. Nature **255:** 738–740.

RAWLS, J. M., C. L. CHAMBERS and W. S. COHEN, 1981 A small genetic region that controls dihydroorotate dehydrogenase in *Drosophila melanogaster.* Biochem. Genet. **19:** 115–127.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J., SCHARF, R. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SEARLES, L. L., A. L. GREENLEAF, W. E. KEMP and R. A. VOELKER, 1986 Sites of P element insertion and structures of P element deletions in the 5′ region of *Drosophila melanogaster RPII215.* Mol. Cell. Biol. **6:** 3312–3319.

SEGRAVES, W. A., C. LOUIS, S. TSUBOTA, P. SCHEDL, J. M. RAWLS *et al.*, 1984 The *rudimentary* locus of *Drosophila melanogaster.* J. Mol. Biol. **175:** 1–17.

SPRADLING, A., 1986 P element-mediated transformation, in *Drosophila: A Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.

STROMAN, P., 1974 Pyrimidine-sensitive *Drosophila* wing mutants: *withered* (*whd*), *tilt* (*tt*) and *dumpy* (*dp*). Hereditas **78:** 157–168.

TAKASU-ISHIKAWA, E., M. YOSHIHARA and Y. HOTTA, 1993 Extra sequences found at P element excision sites in *Drosophila melanogaster.* Mol. Gen. Genet. **232:** 17–23

TREIER, M., C. PFEIFLE and D. TAUTZ, 1989 Comparison of the gap segmentation gene *hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. EMBO J. **8:** 1517–1525.

TSUBOTA, S. I., and P. SCHEDL, 1986 Hybrid dysgenesis-induced revertants of insertions at the 5′ end of the *rudimentary* gene in *Drosophila melanogaster*: transposon-induced control mutations. Genetics **114:** 165–182.

TSUBOTA, S. I., D. ROSENBERG, H. SZOSTAK, D. RUBIN and P. SCHEDL, 1989 The cloning of the *Bar* region and the *B* breakpoint in *Drosophila melanogaster*: evidence for a transposon-induced rearrangement. Genetics **122:** 881–890.

WICKENS, M., 1990 How the messenger got its tail: addition of poly(A) in the nucleus. Trends Biochem. Sci. **15:** 277–281.

Communicating editor: V. G. FINNERTY