Rescue of Drosophila engrailed mutants with a highly divergent mosquito engrailed cDNA using a homing, enhancer-trapping transposon

Mary Whiteley and Judith A. Kassis
Division of Cellular and Gene Therapies, Center for Biologics Evaluation and Research, Food and Drug Administration, 29 Lincoln Drive, Bethesda, MD 20892-4555, USA

SUMMARY
Specific fragments of Drosophila regulatory DNA can alter the insertional specificity of transposable elements causing them to ‘home’ to their parent gene. We used this property to insert a transposon-encoded functional coding region near a defective one and rescue a null mutation. This approach differs from homologous recombination in that the endogenous defective coding region is left in place and the genomic DNA is altered by the addition of the therapeutic transposon. We constructed a P-element-based transposon in which an engrailed cDNA from Anopheles gambiae (a mosquito) is expressed from a Drosophila engrailed minimal promoter. The promoter fragment used includes 2.6 kb of regulatory DNA that causes transposons to home to the endogenous Drosophila engrailed gene at high frequencies. We inserted this transposon onto a Drosophila chromosome that produces no functional engrailed proteins. When this transposon integrated near the engrailed promoter, adult viability was restored to engrailed mutant flies showing that the highly divergent mosquito engrailed protein can replace the Drosophila engrailed protein at all stages of development. Insertion of this transposon into the adjacent inverted gene, which is transcribed in a pattern similar to engrailed, led to only embryonic rescue, suggesting an important difference in the regulation of these two genes.

Key words: engrailed, mosquito, P-element homing, enhancer-trapping, gene therapy, Drosophila

INTRODUCTION
Complementation of a genetic defect caused by loss of a vital protein function often requires correct temporal and spatial expression of the therapeutic transgene. Although correction of the mutation by homologous recombination is ideal, it is difficult to achieve in vivo. The random incorporation into the genome of a vector containing the therapeutic gene is easier but has three significant limitations: (1) the regulatory sequences for the gene must be known and incorporated into the vector, (2) expression levels and patterns are often influenced by the insertion site, and (3) insert sizes in most vectors are limited. An alternative approach to obtaining the correct expression pattern was suggested by the discovery of enhancer trapping. In enhancer trapping, a transposon-encoded reporter gene expressed from a minimal promoter is regulated by genomic sequences flanking the insertion site (O’Kane and Gehring, 1987; Fasano and Kerridge, 1988; Bellen et al., 1989; Bier et al., 1989). The expression pattern of the reporter often mimics that of a nearby gene, because the transposon has come under the control of the same regulatory influences as the endogenous gene. This suggests that addition of a transposon containing a functional coding region to an endogenous locus that encodes a nonfunctional protein might complement the mutant phenotype even though the genomic location would be altered by the addition of a transposon. To test whether enhancer-trapping could be used to express a therapeutic gene, we have taken advantage of a method for selective insertion of a Drosophila P-element vector. Although most P-element derived vectors are thought to insert in the genome in a fairly non-selective manner, there are exceptions to this; specific fragments of regulatory DNA from the Drosophila genes engrailed (en) (Hama et al., 1990; Kassis et al., 1992), polyhomeotic (Fauvarque and Dura, 1993), and Ultrabithorax (Hudson et al., 1995) cause P-transposons to insert in the genome in a selective manner, often near the endogenous gene. This phenomenon of insertion near the parent gene has been called ‘homing’ (Hama et al., 1990). A 2.6 kb fragment of en regulatory DNA (from −2.4 kb to +188 bp) that acts as a minimal promoter for enhancer detection also causes homing of the transposon near en (Kassis, 1990; Kassis et al., 1992).

The en gene is important for many aspects of Drosophila development, including segmentation of the embryo, neurogenesis and formation of the posterior compartments in the adult cuticle (Kornberg, 1981; Morata and Lawrence, 1975). Although en mutations span a genomic region of at least 70 kb of DNA, the transcription unit occupies only about 4.5 kb (Drees et al., 1987). Presumably, the bulk of the DNA in the en locus is required for the regulation of expression in precise patterns throughout development. In addition to the en gene, the Drosophila genome contains a related gene, inverted (inv). inv is located adjacent to en, is expressed in a pattern very similar to en, contains a closely related homeodomain, and provides a function redundant to that of en (Coleman et al., 1987; Gustavson et al., 1996).

The en gene is evolutionarily conserved over a diversity of...
transgenic lines and cloned as a Xba amplified using the primers GTGTCTAGA GCGCTGACCAA TGGT- 
Next, the mosquito HI fragment. Finally 3 as an 
GA CTTGGTGCTAAA TC and cloned as a GAA TTCCGTTGA TA TGA TCGCA and CTCGGA TCCTGGTTTC-
mosome, one had an insertion on the third chromosome, one had an transgenic flies were obtained. One had an inser tion on the X chro-

 MATERIALS AND METHODS

Isolation and characterization of a mosquito (Anopheles gambiae) en gene

The mosquito en genomic clone was obtained from an Anopheles gambiae library by low stringency hybridization (20% foramide, 2x SSC, 42°C; washing conditions were 2xSSC, 0.1% SDS, 50°C) with a Drosophila en homeobox probe. Approximately 7.5 kb of sequence were generated from two EcoRI fragments cloned into M13 in two orientations. The cDNA was obtained by RT-PCR using RNA from mosquito embryos, and primers from the genomic sequence. Several clones were sequenced to ensure that no PCR errors had occurred.

Construction of P[mos-en]

P[mos-en] was prepared by sequential additions of fragments to pCaSpeR (Pirrota, 1988). The en promoter and homing fragment (from -2.407 to +212 bp) was amplified using the primers CTC-GAACCCCGTGATGACAGCA and CTCGACGTCGTTTTC- GACTTGTGTCAATC and cloned as a EcoRI-BamHI fragment. Next, the mosquito en cDNA was cloned from M13 replicative form as an BamHI fragment. Finally 3’ sequences from Drosophila en were amplified using the primers GTGTCTAGAGCGGCTACCA ATGGT- TCACA and GTGTCTAGATTTTCTAAATTATTTTTATTTTTAATTT and cloned as a XbaI fragment.

Generation of transgenic lines

Constructs were injected into homozygous y w67c2 (referred to as y w below) embryos by standard procedures with 300 ng/ml of P[mos- en] and 150 ng/ml of p25.7wc (Spradling, 1986). Approximately 400 injected embryos gave 51 fertile adults. From these fertile adults, three transgenic flies were obtained. One had an insertion on the X chromosome, one had an insertion on the third chromosome, one had an insertion on the second chromosome. The second chromosome insertion (line las1G) was located about 10 kb distal to the en genetic region. P[mos-en] in las1G was mobilized to en mutant chromosomes by two different schemes. For mobilization onto the enB86 chromosome, progeny of y w: P[mos-en]enB86: P[ryy* Δ-3]99B (which contains the P-element transposase; Robertson et al., 1988) males crossed to y w virgins were scored for a change in eye color. Flies with a change in eye color were assayed for the presence of the enB86 mutation by PCR (see below). We examined approximately 3200 flies, selected 39 flies with a change in eye color and recovered 3 insertions on the enB86 chromosome. From these three insertions, we derived lines 6B and 10B at 48A (Fig. 2B) and line 9 at 47F. For mobiliza-
tion onto the enM199 chromosome, the same general scheme was used except that the enM199 chromosome carried three visible markers (Sp BL PinB). Colored-eyed flies that were Sp BL PinB were selected. From approximately 2500 flies, we selected 11 flies and recovered 5 insertions on the enM199 chromosome. The presence of the enM199 mutation was confirmed by PCR (see below).

P[mos-en] in line 12A apparently disrupted inv expression (see Results). In order to generate a derivative of line 12A which no longer expressed mosquito en, internal deletions of P[mos-en] were obtained by crossing line 12A to flies which express the P-element transposase (Robertson et al., 1988) and selecting for loss of eye color. In one line, 12A-11, the white gene and the mosquito en gene were deleted (as determined by PCR), but the P-element ends were intact and had not moved from the 12A insertion site (as determined by inverse PCR).

Local hopping of P[mos-en] for selection of insertions that rescue adult viability

Rescue of enM199

Individual Sp Bl enM199 P[mos-en]-12A PinB/SM6a, P[ryy* Δ-3]99B males were crossed to w; enB86/SM6a, enL/PyO, or en24/PyO virgins. Vials were scored for the presence of progeny that did not carry SM6a or PyO. These were potentially flies that produced no functional Drosophila en, but were rescued by mosquito en. Such flies were obtained in about one of every ten vials (25-100 total progeny scored per vial). These flies were then further tested by crossing to a variety of en mutants and by PCR to confirm the presence of the enM199 mutation.

Rescue of enB86

Individual enB86 P[mos-en]-10B/PyO, P[ryy* Δ-3]99B males or females were crossed to w; en24/PyO flies. Vials were scored as above. Although a similar frequency of rescue was obtained, only one line is discussed here.

Detection of the enM199 and enB86 mutations

The primers used for amplification of DNA flanking the enM199 mutation were GAGGCGCGTGGATGGCCCCACCTC and TTCATGCGATCAG. PCR was for 35 cycles. Products were phenol-chloroform extracted, ethanol precipitated, resuspended, cut with HhaI, and run on 1.3% agarose gels. HhaI digestion yielded four bands in wild-type en and three bands in enM199.

The primers used for amplification of DNA flanking the enB86 mutation (a 53 bp deletion in the first exon of en; Gustavson et al., 1996) were ATGCGCCCTGGAGGATGGCCCCACCTC and GAGGCGCGTGGATGGCCCCACCTC. PCR was for thirty-five cycles (94°C for 1 minute, 72°C for 3 minutes, 55°C for 1.5 minutes) using Buffer D or J from the INVITROGEN PCR optimimerTM Kit. PCR products were separated on 1.6% agarose gels.

Localization of P[mos-en] insertion sites

The insertion sites of P[mos-en] were identified by amplifying flanking genomic DNA via inverse PCR (as described by Whiteley et al., 1992), 32P-labelling of the product, and hybridizing to lambda clones from the 48A region of the chromosome (Kuner et al., 1985). Insertions were further localized by hybridization to subclones from the region. The locations of P[mos-en] near the en promoter in lines 10B-5 and 12A-3R were determined by direct PCR with primers from P-element ends and primers from the first en exon and from upstream of the en transcription start site. In both 10B-5 and 12A-3R, a primer from the 3’ P-element end combined with primers from the first exon gave PCR products and a primer from the 3’ P-element end combined with primers from en upstream sequences gave PCR products.
**Embryo stainings**

In situ hybridization was performed essentially as described by Tautz and Pfeifle (1989). Digoxigenin-labelled probes were prepared by PCR products from the first exons of mosquito and *Drosophila en*. The antibody staining protocol was essentially as described by Patel (1994). Rabbit anti-en antibody (a gift from Charles Girdham and P.H. O’Farrell) was used at a dilution of 1:600. Peroxidase-labelled, anti-rabbit secondary antibody was obtained from Jackson laboratories and used at a concentration of 1:300. The monoclonal antibody 4D09 (Patel et al., 1989) was used at a concentration of 1:2. A Vectastain elite ABC kit (Vector Labs) was used for detection of the 4D09 antibody as described by Zhang et al. (1996).

**RESULTS**

**Isolation of a mosquito engrailed gene**

The *Anopheles gambiae* en gene was isolated from a genomic library by low stringency hybridization with a *Drosophila en* homeobox probe. Two overlapping lambda clones encompassing about 25 kb of DNA were isolated (not shown). From these clones, approximately 7.5 kb of continuous sequence was obtained including the entire mosquito en transcription unit (Fig. 1). This sequence includes about 800 bp of sequence upstream of the presumed translation start site, the entire coding region and two introns. The details of the sequence have been lodged with GenBank (Accession number U42429). The intron/exon organization of the mosquito en gene was determined by comparing the sequences of cDNAs (obtained by RT-PCR using mosquito embryonic RNA) to the genomic sequence. Like *Drosophila en*, mosquito en contains two introns, located in identical positions in the coding sequence (Fig. 1A). Intron 1 is significantly larger in mosquito en, 4.5 kb, compared with 1.1 kb in *D. melanogaster*.

The predicted protein sequence of mosquito en was determined from the cDNA. The mosquito en protein shows significant divergence from the *Drosophila* protein. The overall sequence identity is only 35% and is confined to seven domains (Fig. 1). Four of these domains, the en/inv domains, are found in both the *Drosophila* en and inv proteins and in all en-class proteins including those from *Bombyx mori*, *Tribolium* and in the mouse genes En-1 and En-2. Two domains, the en-specific domains, were previously identified as conserved between *Bombyx mori* and *Drosophila* en, but are not present in the *Drosophila* inv protein. Finally, we identified a new domain, which we refer to as ‘dipteran-specific en’ since it is found in the en gene from the dipteran mosquito and *Drosophila*, but not in the *Drosophila* inv protein.

**Fig. 1.** Comparison of mosquito and *Drosophila en* genes. (A) The transcription units (excluding the untranslated regions) for the en gene from *D. melanogaster* and the mosquito, *Anopheles gambiae* are shown. Boxes indicate the coding regions. White boxes indicate nonconserved regions, protein domains conserved in en and inv are shown in black (I-III and the homeodomain, HD), regions conserved in *en* are shown in grey (*en*1, *en*2 and dipteran-specific, DS). Introns are shown by lines connecting the boxes. *en* intron 1 is 4.5 kb in mosquito and 1.1 kb in *D. melanogaster en*. Mosquito *en* intron 2 is 170 bp. *D. melanogaster en* intron 2 is 280 bp. There is no sequence conservation outside of the coding regions. (B) Sequence comparison within the conserved domains in en proteins. The amino acid sequence of conserved domains of the en protein from *Anopheles gambiae* (mos) is shown. Dash lines indicate amino acid identity. Asterisks indicate end of coding region. mel, *Drosophila melanogaster* (Poole et al., 1985); bm, *Bombyx mori* (Hui et al., 1992); tb, *Tribolium* (Brown et al., 1994); En1, the mouse gene *En-1* (Logan et al., 1992); En2, the mouse gene *En-2* (Logan et al., 1992).
Fig. 2. The structure and insertion sites of the homing, enhancer-trapping transposon \( P[\text{mos}-en] \).

(A) \( P[\text{mos}-en] \) consists of the mosquito en cDNA coding region cloned behind the D. melanogaster en promoter, untranslated leader and homing fragment (boxed). DNA from the D. melanogaster en gene provided the 3' untranslated region and the transcription stop site (hatched box). The mini-white gene is present as a selectable marker. The direction of transcription from the en and white promoters are shown by arrows. The D. melanogaster en fragments extend from –2.4 kb to +188 bp (the fragment that leads to homing; Kassis et al., 1992) and bp +5100 to –5600 (the 3' untranslated leader and polyadenylation site). (B) \( P[\text{mos}-en] \) insertions in en mutant chromosomes. The thin line indicates DNA in polytene subdivision 48A. Arrows indicate the position of insertion of \( P[\text{mos}-en] \) in 7 lines (12A, 12A-5R, 10B-5, 11A, 6B, 10B and 19A). The names of the original insertions are shown in bold. 12A-5R was generated from a remobilization of the 12A insertion and contains the original insertion at 12A and a new insertion (*). 10B-5 was generated from a remobilization of 10B and contains the original insertion at 10B and a new insertion (*). The thick black line indicates the extent of the en gene as determined by breakpoint mutations (Kuner et al., 1985). The black boxes indicate the approximate locations of the transcription units of \( en \) and \( inv \) and three uncharacterized genes (Drees et al., 1987; Kassis, unpublished).

The only intron shown is the large \( inv \) intron. Somewhere within this intron is a 9 bp microexon (Coleman et al., 1987). The insertion of \( P[\text{mos}-en] \) in line 12A disrupts \( inv \) protein accumulation (see Fig. 3 and 4), possibly by disrupting the splicing of this microexon. 6B, 10B and 10B-5 are inserted on a chromosome which carries the mutation \( en^{b86} \). 12A, 12A-5R, 11A, and 19A are inserted on a chromosome which carries the

not in other \( en \) or \( inv \) genes. Because of the strong conservation of these domains, we believed that the mosquito en protein would be able to substitute for the Drosophila en protein. At the very least, since early mosquito development is similar to early Drosophila development, we reasoned that mosquito en would be able to substitute for Drosophila en during that stage.

**Rescue of Drosophila en mutants with a mosquito en cDNA encoded within a homing transposon**

The \( P \)-transposon (\( P[\text{mos}-en] \)) used for these experiments is shown in Fig. 2A. It contains the Drosophila en homing fragment and promoter, the coding region from the mosquito en gene, and the 3' region of the Drosophila en transcription unit. \( P[\text{mos}-en] \) was mobilized from a wild-type chromosome onto en mutant chromosomes in a transposase-producing strain (Robertson et al., 1988). Of 8 independent insertions on \( en \) mutant chromosomes, 5 were at polytene chromosome subdivision 48A (the genomic location of \( en \)) and one was at 47F. The location of the original five insertions at 48A are shown in bold in Fig. 2B. The expression pattern of mosquito en RNA was examined in all lines using in situ hybridization with a mosquito en specific probe. There was no expression of mosquito en in the two lines in which \( P[\text{mos}-en] \) was inserted outside of the 47F-48A region. In lines 6B, 10B, 11A, and 19A, mosquito en RNA is expressed in very weak \( en \)-like stripes in germ band shortening embryos (not shown). This pattern expression appears much later than Drosophila en, and there is no rescue of the \( en \) phenotype in any of these lines. In contrast, in line 12A, which has \( P[\text{mos}-en] \) inserted into the large intron of the \( inv \) gene, mosquito en RNA was expressed in a pattern similar to Drosophila en, and many aspects of the \( en \) embryonic phenotype were rescued (see below).

The Drosophila en gene is necessary for normal segmenta-

tion of the embryo. For this study we have used two \( en \) mutant alleles, \( en^{IM199} \) and \( en^{IM199} \), which contains a single base pair change which causes the protein to prematurely terminate within the homeodomain at codon 497 (Gustavson et al., 1996). \( en^{b86} \) has a 53 bp deletion at codon 75, which causes a frameshift and generates a 104 residue protein (Gustavson et al., 1996). Both mutations produce strong \( en \) phenotypes. The phenotype of an \( en^{IM199} \) embryo cuticle is shown in Fig. 3B. There are pair wise fusions of the abdominal segments, the head fails to involute and there is extensive cell death which leads to a shortened cuticle. In contrast to the severe effects of \( en \) mutations, it has been recently shown that flies with \( inv \) mutations in the \( en \) gene are homozygous viable and have no abnormalities (Gustavson et al., 1996). Despite this, \( inv-en \) double mutants have a stronger phenotype than single \( en \) mutants (Gustavson et al., 1996). During the course of these experiments, a \( P[\text{mos}-en] \) insertion occurred within the large \( inv \) intron (on an \( en^{IM199} \) chromosome, line 12A), disrupting \( inv \) expression and function, creating an \( inv-en^{IM199} \) double mutant. The evidence for this follows.

The monoclonal antibody 4D09 recognizes a conserved epitope present within the homeodomain of Drosophila en and \( inv \) and in the mosquito en protein (Patel et al., 1989). Fig. 4 shows 4D09 antigen in wild-type (A,B), \( en^{IM199} \) (C,D), and in a derivative of line 12A, 12A-11, which contains an internal deletion of the \( P[\text{mos}-en] \) transposon and no longer produces mosquito en protein (see Materials and Methods). In wild-type embryos, 4D09 recognizes both \( en \) and \( inv \). In \( en^{IM199} \) embryos, the \( en \) protein is truncated and the 4D09 antibody recognizes only the \( inv \) protein. This is detected strongly in the hindgut and nervous system in homozygous mutant embryos (Fig. 4C,D). In 12A-11 embryos, 4D09 also sees only the \( inv \) protein, due to the presence of the \( en^{IM199} \) mutation and the
deletion of the mosquito en cDNA in this line. 12A-11 embryos make very little 4D09 antigen compared to \( \text{en}^{\text{IM199}} \) embryos (compare Fig. 4C,D with 4E,F), suggesting that very little inv protein is made in these embryos. Likewise, the cuticles of 12A-11 embryos are more disrupted than \( \text{en}^{\text{IM199}} \) embryos (compare Fig. 3C and 3B). Embryos are shorter and segment fusions are more severe, consistent with the observation that \( \text{en-inv} \) double mutants have a more severe phenotype than single \( \text{en} \) mutants. The 12A-11 phenotype is less severe than that seen for a deletion of \( \text{en-inv} \) region (Gustavson et al., 1996), indicating it is not a complete loss of function. This is consistent with the low level of inv protein present in these mutants.

In 12A homozygotes, which carry the mutation \( \text{en}^{\text{IM199}} \) and an intact \( P[\text{mos-en}] \), 4D09 could recognize both mosquito en and the Drosophila inv proteins if both were expressed. In 12A homozygotes late in embryogenesis, 4D09 antigen is detected in the posterior spiracles, in the last abdominal stripe, and only very weakly in the hindgut (Fig. 4H, compare with a wild-type embryo of the same stage, Fig. 4G). This suggests that the level of inv protein is very low in these embryos, similar to its low level in 12A-11 embryos.

**Mosquito en in line 12A rescues the \( \text{en} \) embryonic phenotype**

Mosquito en RNA in line 12A was first evident at the beginning of gastrulation (Fig. 5A), a few minutes after Drosophila en could be detected. Throughout embryonic development, mosquito en continued to be expressed in a pattern similar to Drosophila en (Fig. 5B,C), however its expression was weaker than Drosophila en, and some stripes contained single cells or groups of cells which showed no expression (see arrows in Fig. 5B,C). The difference in expression level between Drosophila en and mosquito en was particularly evident when examining the protein level with the monoclonal antibody 4D09. As stated above, 4D09 recognizes mainly the mosquito en protein in line 12A, and both inv and en in wild-type embryos. Fig. 5D shows a wild-type and a homozygous 12A embryo collected and stained at the same time. The pattern of expression is remarkably similar in the two embryos, but the 12A homozygote has a much lower level of expression and some stripes are narrower than wild-type en stripes (see arrows). In addition, while ectodermal en stripes are maintained throughout embryonic develop-

### Table 1. Percentage rescue of \( \text{en} \) mutants to viable adults by \( P[\text{mos-en}] \)

<table>
<thead>
<tr>
<th>Mutant</th>
<th>12A*</th>
<th>12A-5R</th>
<th>10B-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{en}^{\text{B96}} )</td>
<td>0</td>
<td>91†</td>
<td>100‡</td>
</tr>
<tr>
<td>( \text{en}^{\text{IM199}} )</td>
<td>ND</td>
<td>86†</td>
<td>105†</td>
</tr>
<tr>
<td>( \text{en}^{\text{LA10}} )</td>
<td>0</td>
<td>59†</td>
<td>ND</td>
</tr>
<tr>
<td>( \text{en}^{\text{CX1}} )</td>
<td>0</td>
<td>77†</td>
<td>ND</td>
</tr>
<tr>
<td>( \text{inv}^{\text{30 en}} )</td>
<td>85†</td>
<td>ND</td>
<td>89†</td>
</tr>
<tr>
<td>( \text{Df}(2R)\text{enE} )</td>
<td>75§</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>( \text{Df}(2R)\text{enX31} )</td>
<td>ND</td>
<td>46§</td>
<td>85†</td>
</tr>
</tbody>
</table>

ND, not done

*Line 12A rescues many aspects of the \( \text{en} \) embryonic phenotype (see Results).

†Wings had slight defects in the posterior compartments (similar to those shown in Fig. 7B,C). Both sexes fertile.

‡The \( \text{en}^{\text{IM199}} \) chromosome was the parent chromosome for 12A and 12A-5R. 12A-5R/\( \text{en}^{\text{IM199}} \) flies were only obtained after a recombination event separated an additional lethal present on this chromosome.

§Wings were severely disrupted in the posterior compartment (Fig. 7D,E). Both sexes sterile.

The \( \text{en}^{\text{B96}} \) chromosome was the parent chromosome for 10B-5. \( \text{en}^{\text{B96}}/10B-5 \) flies were obtained only after a recombination event separated an additional lethal present on this chromosome.

Methods: Males or females from the \( P[\text{mos-en}] \) lines listed on top were crossed to males or females from the \( \text{en} \) mutants listed on the left. All \( \text{en} \) mutations and \( P[\text{mos-en}] \) insertions are kept over the second chromosome balancers \( \text{Sm6A} \) or \( \text{CyO} \). Because of this, only one-third of the progeny are expected to be 12A, 12A-5R or 10B-5\( \text{en} \) mutation. 100% indicates that one-third of the progeny were of the genotype listed on the top genotype listed on the left. At least 100 progeny were examined for each cross.
development in wild-type embryos, no ectodermal stripes are detected late in development in 12A embryos (except the last stripe, see Fig. 4H vs. 4G).

We examined the effect of mosquito en in line 12A on two different en mutant phenotypes. Functional en protein is required to maintain en ectodermal stripes throughout embryonic development. Thus, in an en point mutant, where en transcript is made, but functional en protein is not, en RNA expression is normal early in development, but fades at about 6.5 hours after egg laying (AEL) (Heemskerk et al., 1991). Fig. 6 shows en RNA expression in a wild-type, en^M199 mutant, and a 12A embryo at about 6.5 hours AEL. The mosquito en protein in line 12A is able to restore Drosophila en ectodermal stripes in these embryos. The cuticles of homozygous 12A embryos were also examined. Homozygous 12A cuticles looked normal (Fig. 3D) except for some missing denticles, mostly in the first row of each segment (detail not shown). This denticle loss may result from the lack of mosquito en expression in cells within some stripes, or to the absence of mosquito en stripes late in development. Despite the normal appearance of the cuticles, most homozygous 12A embryos fail to hatch and 12A could not complement any of the en alleles tested (Table 1). Since the lethality of homozygous 12A embryos could be due to another embryonic lethal present on the same chromosome, we examined the

---

**Fig. 5.** Distribution of mosquito en in 12A embryos. (A-C) Mosquito en RNA. (A) Stage 6, (B) Stage 9, (C) Stage 13. All embryos are anterior left, dorsal up. Arrows point to stripes which contain cells that do not express mosquito en. (D) Wild-type (lower) and 12A (upper) embryos at stage 12 of development stained with the monoclonal antibody 4D09. In wild-type embryos, 4D09 detects en and inv proteins. In 12A embryos, 4D09 detects mainly mosquito en proteins (see text). Arrows point to a stripe which is narrower in 12A embryos.

**Fig. 4.** Distribution of en and inv proteins as detected with the monoclonal antibody 4D09. (A,B) Wild-type; (C,D) en^M199; (E,F) 12A-11 embryos; stage 14, anterior left. (A,C,E) Ventral view, (B,D,F) dorsal view. (G) Wild-type and (H) 12A embryos, stage 15, anterior left, dorsal view. Arrows point to the hindgut.
lethal period of 12A/enB86 embryos. 12A/enB86 embryos hatch but die sometime during the larval period. The cuticles of 12A/enB86 embryos look similar to those from 12A homozygotes (data not shown).

Mosquito en rescues Drosophila en mutants to viable adults

In order to obtain P[mos-en] lines that rescue en mutants to viable adults, we took advantage of the phenomenon of local hopping (Tower et al., 1993; Zhang et al., 1993). When P-elements transpose, they often insert nearby in the genome. P[mos-en] in line 12A was mobilized, and chromosomes that produced viable adults over en null mutations were selected. Six chromosomes were obtained. Five of these carry deletions and rearrangements of DNA flanking the 12A insertion site; these will be described elsewhere. The sixth chromosome, 12A-5R, contains two insertions of P[mos-en], one at the original 12A site and a second insertion 300 bp upstream of the en transcription start site. Mosquito en expression in line 12A-5R was indistinguishable from Drosophila en expression in wild-type embryos in both pattern and level (as judged by reaction with the 4D09 antibody; data not shown). The cuticles of 12A-5R embryos look normal, showing that the mosquito en protein can fully rescue this phenotype when correctly expressed. Finally, the P[mos-en] insertions on chromosome 12A-5R were able to complement many different en alleles: enB86, enLA10 and enIM199 (which contain stop codons within the encoding regions; Gustavson et al., 1996); breakpoint mutations enCX1 (which breaks in the first en intron or exon) and en24 (which breaks 20 kb upstream of the en transcription unit). Flies obtained from these crosses look normal, except for minor defects in the posterior compartment of the wings (Fig. 7B,C), and were fertile. In addition, 12A-5R is able to survive over en-inv double mutations including Df(2R)enE, a 50 kb deletion that removes the entire en transcription unit and the 3’ end of the inv transcription unit (Gustavson et al., 1996). 12A-5R/Df(2R)enE flies have wings with severe defects in the posterior compartments (Fig. 7D,E) and both sexes are sterile.

The parental chromosome for 12A-5R carries the en mutation

---

**Fig. 6.** en RNA expression is restored in 12A embryos. (A) Wild-type, (B) enIM199 and (C) 12A, stage 11 embryos. Drosophila en RNA was detected with a probe to the first exon which does not react with inv or mosquito en RNA. All embryos are oriented anterior left, dorsal up. The embryos in B and C are slightly older than in A, and expression in the ventral nervous system is evident just below the ectoderm. 12A embryos were collected from 12A/11enCyO parents. 11enCyO is a balancer chromosome which contains a P[en-lacZ] gene inserted in the wingless gene. Embryos were first stained for en RNA, then for lacZ using an antibody to β-galactosidase. Homozygous 12A embryos were identified by the lack of anti-β-galactosidase staining.

---

**Fig. 7.** Wing phenotypes of mosquito en-rescued flies. All defects occur within the posterior compartment. (A) Wild-type wing, (B,C) 10B-5/Df(2R)enE wings. Most common phenotypes seen are lack of posterior crossvein (B, arrow), a defect in vein IV at the anterior crossvein (B, arrowhead) and a split in the end of vein V (C, arrow). (D,E) 12A-5R/Df(2R)enE. Two examples of severely disrupted wing veins seen in posterior compartment of flies of this genotype.
en IM199, which contains a single base change within codon 497 changing a CAG to TAG (Gustavson et al., 1996) causing the loss of a HhaI site that can be easily detected after amplification of flanking DNA using PCR. DNA samples from adult flies of the genotypes 12A-5R and 10B-5R, which contains wild-type en, 12A-5R/en IM199 and 12A-5R/Df(2R)enE, en IM199 contains a mutation which removes a HhaI site present in this PCR fragment. Four HhaI sites are present in the wild-type en gene, and three are present in the mutant en IM199 gene. en IM199/Sm6a flies have 5 bands arising from both wild-type and mutant en. In contrast, 12A-5R/en IM199 and 12A-5R/Df(2R)enE have only three bands generated from the mutant en. The number of bp in each fragment is displayed on the left. (B) PCR amplification of DNA surrounding the 53 bp deletion in en IM199. PCR products from DNA of adult flies of the genotypes OregonR (wild-type), en IM199/Sm6a (which contains wild-type en), and 10B-5Df(2R)enX31 is shown. A molecular mass marker (123 bp ladder, GIBCO-BRL) is shown in the left lane.

Local hopping was also used to generate P[mos-en] insertions on a chromosome that contained the mutation en B86. In this case, local hopping of P[mos-en] in line 10B gave rise to a chromosome with the original 10B insertion and an insertion very near the en transcription start site (10B-5). 10B-5 made no Drosophila en, as evidenced by no reaction with an antibody directed against the N-terminal region of en (Fig. 9); in contrast, a strong reaction with 4D09 indicates that mosquito en proteins were made at a high level (not shown). 10B-5 rescued adult viability of all en mutations tested (Table 1), including en-inv double mutations and a deletion, Df(2R)enX31, that removes about 200 kb of DNA (Kuner et al., 1985). In addition, 10B-5 homozygotes survived and, though sterile, looked wild-type. en B86 contains a 53 bp deletion of DNA within the first exon of en. PCR of DNA flanking this deletion confirmed that no wild-type Drosophila en gene was present in 10B-5/Df(2R)enX31 adults (Fig. 8B). The wing phenotypes observed in 10B-5/Df(2R)enE flies (Fig 7B,C) were less severe than those observed in 12A-5R/Df(2R)enE (Fig. 7D,E), indicating a more complete rescue with 10B-5 than with 12A-5R. We believe that the inv gene is inactivated in 12A-5R, due to the presence of the 12A insertion and that the intact inv gene on the 10B-5 chromosome leads to more complete rescue.

The 10B-5 chromosome rescued both Df(2R)enE and Df(2R)enX31, while 12A-5R only rescued the former (Table 1). Both deletions lead to a complete loss of en and inv proteins, however, Df(2R)enE leaves 40 kb of en regulatory DNA intact. We suggest that en regulatory DNA on the Df(2R)enE chromosome can activate the expression of P[mos-en] in 12A-5R by a transvection phenomenon (Judd 1988), leading to rescue. This regulatory DNA is not present on the Df(2R)enX31 chromosome.
DISCUSSION

This study yielded three principle findings. First, gene expression by enhancer-trapping is precise enough to restore viability to Drosophila en mutants. Second, the highly divergent mosquito en protein is able to substitute for the Drosophila en protein during all stages of development. Finally, our results suggest that inv can play a role in proper development. This conclusion is consistent with that reached by Gustavson et al. (1996).

Rescue of Drosophila en mutants using enhancer-trapping and P-element homing

The en gene of Drosophila is a member of the segment-polarity class of developmental genes. It is expressed throughout development, in a dynamic pattern, and as much as 70 kb of DNA may contribute to its proper expression pattern. Although some of the sequences necessary for correct regulation have been identified, only partial rescue of the en phenotype has been achieved (Smith and Jaynes, 1996). In this report we have restored adult viability to two different lethal en mutants by combining the techniques of enhancer-trapping and P-element homing.

A fragment of en DNA has been identified which acts as both a minimal promoter for enhancer-trapping and causes P-elements to ‘home’ to polytene chromosome bands 47F/48A, the location of the endogenous en gene and the related gene inv (Hama et al., 1990; Kassis et al., 1992). Although the mechanism for homing is unknown, we hypothesize that proteins bound to en regulatory DNA present in the transposon interact with proteins bound to the endogenous en locus, concentrating the transposon in that region of the genome. Transposition then occurs locally. In this study we used the vector P[mos-en], which contains the mosquito en cDNA cloned behind the en fragment which contains a minimal promoter and causes P-element homing. We obtained eight P[mos-en] insertions in 48A and one in 47F. Like previous studies on P-element homing, we note that insertion at 48A is not site-specific, but occurs over a large region: insertions are distributed over about 200 kb. In all nine lines, mosquito en expression resembled Drosophila en at some stage of development. Thus, we suggest that this 200 kb region represents a chromosomal domain: an en promoter in P[mos-en] present anywhere in this domain can interact with at least some en enhancers. Mosquito en was not expressed in an en-like pattern when P[mos-en] was inserted elsewhere in the genome.

Although it was easy to obtain insertions of P[mos-en] into polytene chromosome band 48A, rescue of the en phenotype was only achieved when P[mos-en] was inserted in the inv or en gene. Further, the insertion in the inv gene led only to embryonic rescue. Complete rescue to adulthood was only achieved by insertions of P[mos-en] just upstream of the en promoter. We suggest that there may be constraints on promoter-enhancer interactions in vivo which only allow accurate, high level expression when the transposon-encoded promoter is located in the same place as the endogenous promoter. We are currently trying to move P[mos-en] to different locations within the en locus to test this hypothesis.

Rescue by enhancer-trapping could be tried for a few other genes in Drosophila. DNA from both the polyhomoecotic and the Ultrabithorax loci have been shown to mediate P-element homing (Fauvarque and Dura, 1993; Hudson et al., 1995). In addition, the same en fragment that causes transposons to home to polytene chromosome band 48A also seems to direct P-transposons to the escargot locus (Whiteley et al., 1992; Kassis, 1994). Genes which are ‘hot-spots’ for P-element insertion are also candidates for this approach. While a large number of lines might need to be screened to obtain the line of interest, it should be noted that insertion at some sites in the genome will not be allowed because mis-expression will lead to embryonic lethality.

Rescue of Drosophila en mutants by the mosquito en protein

The en gene encodes a transcription factor with a highly conserved homeodomain. D. melanogaster contains two en-class genes, en and inv, which are expressed in similar patterns during development (Coleman et al., 1987; Gustavson et al., 1996). Some insects such as the silk moth Bombyx mori and the honeybee also contain two en-related genes (Hui et al., 1992; Walldorf et al., 1989), whereas the more primitive insects, grasshopper and Tribolium contain only one en-related gene (Patel et al., 1989; Brown et al., 1994). Sequence comparisons between the Drosophila and the Bombyx genes have identified 4 domains shared by inv and en, two en-specific domains and one inv-specific domain (Hui et al., 1992). Vermbrate species also have two en-related genes (Mouse En-1 and En-2 are shown in Fig. 1B), which contain only those domains shared by inv and en. en/inv domain I has recently been shown to mediate active transcriptional repression and is also present in some other homeodomain-containing proteins (Smith and Jaynes, 1996). We have sequenced the en gene from a mosquito, Anopheles gambiae. We chose the mosquito because, though evolutionarily very divergent, early mosquito development is similar to early Drosophila development and we reasoned that the function of the en protein would be conserved. We believe that the gene we have cloned is the true en homologue since it contains the two en-specific domains identified by Hui et al. (1992) but not the inv-specific domain. Although we did not identify an inv homologue in this work, our genomic Southern suggest that the Anopheles gambiae genome contains two en-related genes (data not shown). In addition to the domains previously identified, one additional domain was conserved between mosquito en and Drosophila en which we refer to as Dipteran specific.

The mosquito en protein functions remarkably well in Drosophila. Table 1 shows that two rescue chromosomes obtained, 12A-5R and 10B-5, rescue adult viability of many different en mutants, including Df(2R)enE, a deletion which removes both en and inv. 10B-5 rescued more completely, yielding fertile 10B-5/Df(2R)enE adults with only slight defects in the wing veins of the posterior compartment. Further, 10B-5 survives as a homozygote and has phenotypically wild-type wings. However, the mosquito en protein probably does not function in Drosophila as well as the Drosophila en protein. There are no defects in the wings of wild-type en/Df(2R)enE flies. However, since we have not rescued a Drosophila en mutant with a transposon-encoded Drosophila en gene, we cannot rule out the possibility that the transposon-encoded gene does not function as well as the endogenous gene because of slight differences in expression pattern or level.

Our result on the conservation of en function is consistent
with the results of Hanks et al. (1995) who reported that the mouse En-1 protein can be completely substituted by the mouse En-2 protein, even though the proteins from these genes share only 55% amino acid identity. The Drosophila and mosquito en proteins share only 35% identity, emphasizing the flexibility of sequences outside of the conserved domains. Given these results, it is interesting to speculate about whether a mouse en homologue could function in Drosophila. Examination of the conservation of the en/inv domains between Drosophila, mosquito and mouse (Fig. 1B) suggests that a mouse en protein may not function as well in Drosophila as the mosquito protein does. Although the homeodomain is highly conserved between these three diverse organisms, the other domains are much more highly conserved between Drosophila and mosquito, perhaps reflecting more common function of the Drosophila and mosquito en proteins.

There have been several reports of rescue of various Drosophila developmental mutants with highly divergent proteins (Muller et al., 1995; Lutz et al., 1996; Xue and Noll, 1996). In one extreme example, Xue and Noll (1996) report that the Drosophila gooseberry and the mouse Pax3 proteins can partially substitute for the Drosophila protein, paired. Remarkably, two copies of a Drosophila gooseberry transgene can restore adult viability to some embryos with the lethal mutation paired. They suggest that it is the evolution of new cis-regulatory regions rather than the divergence of protein coding regions which has led to the diversification of gooseberry, paired and Pax3 function. A similar conclusion was reached by Hanks et al. (1995) on the ability of En-2 to replace En-1 in mice. Our results reinforce the view that changes in cis-regulatory regions may play a bigger role in evolution than protein divergence, at least for homeodomain-containing proteins. In the course of these experiments, we obtained a number of insertions of P[mos-en] which cause dominant defects in the adult cuticle. We presume these defects are caused by expression of mosquito en in tissues where Drosophila en is not normally expressed. Thus, the en protein has acquired new activities, by the acquisition of a new cis-regulatory regions via enhancer-trapping.

**engrailed and invaded**

The inv gene has recently been shown to be dispensable for viability of D. melanogaster, yet en-inv double mutants have a more severe phenotype than en mutants (Gustavson et al., 1996), suggesting that inv can play a role in development. In this study we generated a P[mos-en] insertion into the large intron of inv (in line 12A), creating an inv mutation on an en\[^{IM99}\] chromosome. inv protein levels are severely reduced in 12A embryos (Fig. 4H). Deletion of the mosquito en cDNA from line 12A led to the creation of an inv\[^{en}\] double mutant chromosome (line 12A-11). The cuticle phenotype of this double mutant is more severe than en\[^{en}\], but less severe than that reported for Df(2R)enE, a deletion of inv and en. In fact, some inv protein is made in 12A-11 embryos (Fig. 4E,F). It is not known how the P[mos-en] insertion in the inv intron disrupts inv expression. In situ hybridization of 12A embryos with a probe to the first exon of inv did not show any decrease in the transcript levels (data not shown). The inv transcript contains a 9 bp microexon at an unknown location within the large intron (Coleman et al., 1987). It is possible that P[mos-en] in 12A disrupts the splicing of this exon, and that either there is a minor class of inv transcript which does not contain this exon or that a low level of splicing can occur in 12A and 12A-11 embryos.

The presence of an intact inv gene seems to contribute to rescue by the mosquito en gene. Line 12A-5R, which contains P[mos-en] inserted on the en-inv double mutant chromosome described above, survives over all single en mutants tested, as well as over the double mutant Df(2R)enE. However, 12A-5R/Df(2R)enE flies have wings with severely disrupted posterior compartments and are sterile. Another rescue line, 10B-5, which contains an intact inv gene, also survives over Df(2R)enE. However, 10B-5/Df(2R)enE flies are fertile and have only minor defects in the posterior compartment.

en and inv are reported to be expressed in similar patterns (Gustavson et al., 1996), yet P[mos-en] in line 12A only rescued the en mutant phenotype in embryos. This suggested to us that 12A expression might differ from en expression, and that this might lead to larval lethality. We examined the expression of inv in wild-type embryos using a probe to the first exon of inv and compared it to mosquito en expression in 12A embryos and to Drosophila en expression in wild-type embryos (data not shown). We noted both differences and similarities. Mosquito en RNA in line 12A was evident earlier than inv, at about the same time as Drosophila en, inv expression was not evident until about 30 minutes later. This difference is probably due to a difference in transcript sizes: the en transcript extends over about 25 kb and would take considerably longer to be transcribed than the 4.5 kb en mRNA and the 1.8 kb mosquito en mRNA. Like mosquito en in line 12A, we note that inv stripes in wild-type embryos seem to be narrower than Drosophila en stripes, and that sometimes gaps occur within inv stripes (data not shown). Thus, although similar, en and inv expression are not identical. We believe there may be other subtle differences in inv and en expression which explain why P[mos-en] in line 12A cannot rescue en mutants to viable adults. Likewise, the difference in inv and en expression patterns may explain why inv cannot substitute for en during Drosophila development.

We thank Elizabeth Gustavson and Tom Kornberg for sharing information on the sequence of en mutations prior to publication, for en mutant flies, and for interesting discussions; Charles Girdham and Pat O’Farrell for the anti-en antibody; Nipam Patel for the monoclonal antibody 4D09; Louis Miller and members of his laboratory for providing Anopheles gambiae embryos and adults; Pat Romans for A. gambiae DNA; Ward Odenwald for examining the nervous systems of 12A embryos; Jim Jaynes for examining the cuticles of homozygous 12A embryos; and Lesley Brown, Jim Jaynes, Ward Odenwald, Pete Mathers, Jim Kennison, Diane Mucci, and Philip Noguchi for comments on this manuscript. The mosquito en genomic clone was isolated by Rebecca Clayton. Finally, we thank Philip Noguchi for continuing to support our Drosophila work at CBER. This work was supported in part by a grant on gene therapy from the Cystic Fibrosis Foundation.

**REFERENCES**


Bier, E., Vaessin, H., Shepherd, K., Lee, K., McCall, K., Barbel, S.,