Molecular characterization of the zerknüllt region of the Antennapedia gene complex in Drosophila

Christine Rushlow, Helen Doyle, Timothy Hoey, and Michael Levine

Department of Biological Sciences, Fairchild Center, Columbia University, New York, New York 10027 USA

zerknüllt (zen) is unique among the 18 known homeo box genes in Drosophila since it is required for the differentiation of the dorsal–ventral pattern, and does not appear to be involved in the process of segmentation. Here we show that the zen region of the Antennapedia complex (ANT-C) consists of two closely linked homeo box genes, designated z1 and z2. The z1 and z2 transcription units show essentially identical patterns of expression during early development, which are consistent with the timing and sites of zen+ gene activity. The putative proteins encoded by z1 and z2 are highly divergent and are related only by virtue of homeo box homology. We have used P-element-mediated germ line transformation to show that z1 alone can provide zen+ gene function, suggesting that the z2 gene might be dispensable. The occurrence of closely linked homeo box genes that display similar patterns of expression is not unique to the zen locus. Such gene duplications might provide important clues to the evolution of the homeo box gene family in Drosophila and other organisms.

[Key Words: Drosophila; zerknüllt (zen), homeo box genes]

Received August 7, 1987; revised version accepted October 14, 1987.
these similarities, nucleotide sequence analyses show that the putative proteins encoded by these genes are highly divergent and are related only by homeo box homology. Several lines of evidence suggest that z1 might correspond to zen. We show that z1 alone can provide zen + gene function, since zen- embryos are rescued by a z1/P-transposon that completely lacks z2 sequences. The occurrence of two closely linked homeo box genes that display similar patterns of expression is discussed in the context of other homeo box gene duplications in Drosophila.

**Results**

The zen region of the ANT-C was isolated in a previous screen of a Drosophila DNA library for clones that cross-hybridize with the homeo box sequence of the homeotic gene Sex combs reduced (Scr) (Doyle et al. 1986). One of the clones that was obtained in this screen (called S60) maps to the 84A-B region of the third chromosome, within the limits of the ANT-C. S60 (hereafter called z1) was used to rescreen the Drosophila library to obtain flanking regions, and a composite restriction map of the interval is shown in Figure 1a. Based on Southern blot assays, two regions were found to hybridize with the Scr homeo box probe, designated z1 and z2. These two transcription units are separated by ~10 kb. Genomic DNA fragments from the z1 and z2 regions were used to screen a phage λ gt10 cDNA library, which was prepared with poly(A)+ RNA from 3- to 12 hr wild-type embryos (the library is described in Poole et al. 1985). The largest of the cDNAs that were obtained are nearly full length; a ~1.3-kb cDNA was obtained for z1, and a ~900-bp cDNA was isolated for z2 (see below). Restriction maps of these cDNAs are shown in Figure 1b, below the respective genomic DNA regions of the composite restriction map shown in Figure 1a.

Several lines of evidence suggest that the more distal
transcription unit [zl] might correspond to zen. First, the proximal breakpoint of a zen deficiency [Df(3R) LIN] falls within, or just adjacent to, the 3' end of the zl coding sequence and does not disrupt the z2 transcription unit (see Fig. 1a; vertical arrow and brackets). Second, embryos homozygous for a cytologically normal EMS-induced mutant allele [zenw36] are not stained by antibodies directed against the zl protein (C. Rushlow unpubl.). Third, a genomic DNA fragment that contains the entire zl coding sequence, but no z2 sequences, is able to rescue zen- embryos by P transformation [Fig. 1a and below]. Finally, it has been shown that embryos deficient for the entire z2 transcription unit (i.e., zl +, z2- embryos) are viable [M.A. Pultz and T.C. Kaufman, pers. comm.]. Despite these observations, we will refer to the two transcription units as zl and z2 since it is not known whether zl- embryos are viable. It is possible that either zl or z2 can provide zen + gene function when one of the genes is removed.

The zl gene rescues zen mutants

P-transformation assays were done to determine whether the zl gene could rescue zen- embryos. These experiments involved the use of a 4.5-kb XbaI–ScaI genomic DNA fragment that contains the entire zl coding sequence, as well as 1.6-kb of 5'-flanking sequences and 1.5 kb of 3' flank. The limits and location of the zl DNA that was used are shown in Figure 1a (horizontal line below the restriction map of the zen region). The 5' limit of the restriction fragment extends to the 3' terminus of bicoid, which is a maternally expressed homeo box gene that is closely linked to zen [Frigierio et al. 1986; Frohnhoefer and Nüsslein-Volhard 1986]. The zl fragment was inserted into a derivative of the Carnegie 20 transformation vector [Rubin and Spradling 1983; D. Mismer and G. Rubin, unpubl.], and injected into rosy- embryos as previously described [Rubin and Spradling 1982]. Six independent transformed lines were established (listed in Table 1) and tested for their ability to rescue zenw36/Df(3R) LIN trans-heterozygotes. Such heterozygotes display the zen null phenotype (C. Rushlow, unpub. results). zenw36 is a null allele that was induced by EMS (Wakimoto et al. 1984), and, based on Southern analyses, does not contain a deletion of the zen region larger than 50 bp (data not shown). Df(3R) LIN is a large deletion that uncovers several loci within the ANT-C, including Dfd and bic [C. Nüsslein-Volhard, unpubl. results], and breaks within the zl gene [Fig. 1a]. This deletion strongly disrupts the expression of both zl and z2 [C. Rushlow, in prep.]. Each of the five transformed lines that contains at least one copy of the P-zl transposon on the second chromosome rescues zenw36/Df(3R) LIN heterozygotes and yields adult survivors [see Table 1]. One of these lines contains only a single copy of the transposon, but nonetheless gives nearly complete rescue (72% of the expected number of zenw36/Df(3R) LIN adults were obtained). The only line that failed to rescue contains the P-zl transposon on the third chromosome, which does not segregate with zen alleles. None of the transformed lines rescues zenw36 homozygotes, suggesting that this chromosome might contain additional lethal mutations.

Sequence and organization of the zl and z2 transcripts

Northern analyses using poly[A]+ RNA from staged embryos show that zl and z2 each encode a single mRNA that display similar times of appearance during early development [Fig. 2]. The zl mRNA is ~1.3 kb in length, whereas z2 is ~0.9 kb. Both RNAs appear transiently during early embryogenesis and show peaks of expression from 2 to 3 hr after fertilization. zl and z2 transcripts are not detected after the sixth hour of embryogenesis and do not reappear during embryonic development [Fig. 2]. We have not used larval or adult extracts for Northern analyses, although temperature-shift experiments using a zen + strain suggest that zen + function is not required after early periods of embryogenesis [Wakimoto et al. 1984]. Primer extension and S1 nuclease protection analyses were done to identify the 5' ends of the zl and z2 transcripts. For these studies, poly[A]+ RNA was obtained from extracts of 2- to 3-hr embryos, since this is the period of peak zl and z2 expression. The primer extension experiments involved the use of a 20-nucleotide synthetic oligomer for zl and a 33-nucleotide oligomer for z2. The nucleotide sequence and location of the primers within the zl and z2 genes are shown in

Figure 2. Northern analysis of zl and z2 RNAs. Poly(A)+ RNA was extracted from wild-type embryo populations at 0–2, 2–3, 3–6, 6–12, and 12–24 hr after fertilization. Equivalent amounts of these RNAs were fractionated on an agarose-formaldehyde gel, transferred to nitrocellulose, and hybridized with a zl cDNA probe [a]. A single ~1.3-kb transcript was detected, which peaks in expression in 2- to 3-hr embryos. The zl hybridization signal was removed by incubating the nitrocellulose filter in 0.1 x SSC at 100°C for 3 min. The blot was then rehybridized with a z2 cDNA probe [b] and a single ~0.9-kb transcript was detected, which also peaks in 2- to 3-hr embryos.
The complete amino acid sequence of this protein is shown in Figure 5.

The nucleotide sequence of the z2 coding region is presented in Figure 6. The z2 cDNA lacks the first six nucleotides of the transcript, which were determined from the sequence of appropriate 5' genomic DNAs. The first AUG codon is located 67 bp downstream from the transcription start site and is in frame with an ORF that extends throughout most of the cDNA and has a total length of 756 bp. Translation of this ORF would result in a z2 protein of ~28 kD composed of 252 amino acid residues. The complete amino acid sequence of this protein is shown in Figure 6.

The putative z1 and z2 proteins contain bona fide homeo box sequences (homeo domains), each of which shares approximately 60% amino-acid identity with the Antennapedia class homeo domains, including those associated with Antp, ftz, and Ubx (McGinnis et al. 1984; Scott and Weiner 1984). The z1 and z2 homeo domains share more homology with each other (75%) than either one does with any other homeo domain characterized to date. However, the two genes do not share significant homologies outside the homeo box regions. The z1 and z2 homeo domains are located in the amino-terminal half of the putative proteins, which is in contrast to the more carboxy-terminal locations observed for most of the other homeo box genes that have been characterized (Laughon and Scott 1984; Poole et al. 1985; Schneuwly et al. 1986).

Localization of z1 and z2 transcripts in developing embryos

The distribution of z1 transcripts in tissue sections of wild-type embryos has been described previously (Doyle et al. 1986). To compare the normal patterns of z1 and z2 expression, serial tissue sections were hybridized with z1 and z2 cDNA probes. This analysis shows that the two genes display extremely similar spatial and temporal patterns of expression.

z1 and z2 transcripts are broadly distributed along the dorsal surface of precellular embryos and also include the anterior and posterior poles (Fig. 7, a and b). Considered in cross-section, z1 and z2 transcripts encompass a total of ~40% of the embryo’s circumference, extending from the dorsal-most surface toward more lateral regions (data not shown). By the onset of gastrulation, z1 and z2 expression is lost from the poles and becomes restricted to the dorsal-most cells, encompassing only ~10% of the embryo’s circumference (Fig. 7c, d). At this time, most of the z1 and z2 transcripts are localized within the differentiating amnioserosa. However, transcripts are also detected in portions of the dorsal ectoderm that are anterior to the presumptive amnioserosa and might correspond to the future optic lobe (Campos-Ortega and Hartenstein 1985).

Subtle differences in the z1 and z2 expression patterns can be observed during germ band elongation (Fig. 7e, f). First, z2 transcripts persist at relatively higher levels than z1 transcripts. This is consistent with Northern

Nucleotide sequence of z1 and z2

DNA templates that were used for dideoxynucleotide sequencing are shown in Figure 1b. The sequenced interval includes each of the cDNAs (see Fig. 1b), as well as the genomic regions that immediately flank the 5' and 3' ends. Genomic fragments that encompass the small introns for each gene were also sequenced. The locations of the introns on each cDNA are indicated by the vertical arrows in Figure 1b. The horizontal arrows indicate the regions that were sequenced, the solid arrows correspond to cDNA fragments, and the dashed arrows correspond to genomic fragments. The nucleotide sequence of the z1 coding region is presented in Figure 5. The z1 cDNA lacks the first five nucleotides of the transcript, which were determined from overlapping genomic DNA fragments from the 5' end. The first AUG codon is located 52 bp downstream from the transcription start site and is in frame with an open reading frame (ORF) that extends throughout most of the cDNA and has a total length of 1059 bp. If translation begins with the first AUG of this ORF, then a protein of ~39 kD consisting of 353 amino acid residues would result. The

The putative z1 and z2 proteins contain bona fide homeo box sequences (homeo domains), each of which shares approximately 60% amino-acid identity with the Antennapedia class homeo domains, including those associated with Antp, ftz, and Ubx (McGinnis et al. 1984; Scott and Weiner 1984). The z1 and z2 homeo domains share more homology with each other (75%) than either one does with any other homeo domain characterized to date. However, the two genes do not share significant homologies outside the homeo box regions. The z1 and z2 homeo domains are located in the amino-terminal half of the putative proteins, which is in contrast to the more carboxy-terminal locations observed for most of the other homeo box genes that have been characterized (Laughon and Scott 1984; Poole et al. 1985; Schneuwly et al. 1986).

Localization of z1 and z2 transcripts in developing embryos

The distribution of z1 transcripts in tissue sections of wild-type embryos has been described previously (Doyle et al. 1986). To compare the normal patterns of z1 and z2 expression, serial tissue sections were hybridized with z1 and z2 cDNA probes. This analysis shows that the two genes display extremely similar spatial and temporal patterns of expression.

z1 and z2 transcripts are broadly distributed along the dorsal surface of precellular embryos and also include the anterior and posterior poles (Fig. 7, a and b). Considered in cross-section, z1 and z2 transcripts encompass a total of ~40% of the embryo’s circumference, extending from the dorsal-most surface toward more lateral regions (data not shown). By the onset of gastrulation, z1 and z2 expression is lost from the poles and becomes restricted to the dorsal-most cells, encompassing only ~10% of the embryo’s circumference (Fig. 7c, d). At this time, most of the z1 and z2 transcripts are localized within the differentiating amnioserosa. However, transcripts are also detected in portions of the dorsal ectoderm that are anterior to the presumptive amnioserosa and might correspond to the future optic lobe (Campos-Ortega and Hartenstein 1985).

Subtle differences in the z1 and z2 expression patterns can be observed during germ band elongation (Fig. 7e, f). First, z2 transcripts persist at relatively higher levels than z1 transcripts. This is consistent with Northern

Nucleotide sequence of z1 and z2

DNA templates that were used for dideoxynucleotide sequencing are shown in Figure 1b. The sequenced interval includes each of the cDNAs (see Fig. 1b), as well as the genomic regions that immediately flank the 5' and 3' ends. Genomic fragments that encompass the small introns for each gene were also sequenced. The locations of the introns on each cDNA are indicated by the vertical arrows in Figure 1b. The horizontal arrows indicate the regions that were sequenced, the solid arrows correspond to cDNA fragments, and the dashed arrows correspond to genomic fragments. The nucleotide sequence of the z1 coding region is presented in Figure 5. The z1 cDNA lacks the first five nucleotides of the transcript, which were determined from overlapping genomic DNA fragments from the 5' end. The first AUG codon is located 52 bp downstream from the transcription start site and is in frame with an open reading frame (ORF) that extends throughout most of the cDNA and has a total length of 1059 bp. If translation begins with the first AUG of this ORF, then a protein of ~39 kD consisting of 353 amino acid residues would result. The
data that indicate a persistence of z2 RNAs in 3- to 6-hr embryos, whereas z1 RNAs essentially disappear during this time (see Fig. 2). Second, there are differences in the expression of z1 and z2 within the region that contains the presumptive optic lobe. z2 transcripts are expressed in this region for a longer period as compared with z1 (compare Fig. 7 e with Fig. f). At slightly earlier stages, high levels of both z1 and z2 transcripts are detected within the presumptive optic lobe. However, given the resolution of the in situ hybridization method, it is not clear whether z1 and z2 are expressed in exactly the same set of cells within this region. Third, z1 transcripts are detected in a subset of the pole cells within the posterior midgut (PMG), whereas z2 transcripts are not detected in the pole cells per se, but are present in a portion of the somatic cells that comprise the PMG. De-

Figure 3. Identification of the z1 and z2 transcription start sites. (a) S1 nuclease protection analysis of the z1 transcript. The location of the S1 probe used in this analysis is shown in the map below the autoradiogram in a. The single-stranded probe of 310 nucleotides was hybridized with poly[A]+ RNA from 2- to 3-hr embryos, digested with S1 nuclease, and electrophoresed in an acrylamide gel. The protected DNA fragment is 115 nucleotides long, which indicates the distance to the transcription start site (indicated by +1 in the map). [Lane 1] Molecular weight markers; [lane 2] the deoxyguanosine plus deoxyadenosine sequence of the S1 probe; [lane 3] digestion of RNA–DNA hybrids with 100 units of S1 nuclease; [lane 4] digestion with 50 units of S1 nuclease; [lane 5] digestion with 25 units of S1 nuclease. (b) Primer extension analysis of the z1 transcript. The 20-nucleotide synthetic oligomer used in this study is shown in the map below the autoradiograms. The primer-extended product is 115 nucleotide long, which is the same size as the fragment protected in the S1 experiment. (Lane 1) No RNA control; poly[A]+ RNA was not added to the primer extension reaction; [lane 2] 20 units of reverse transcriptase; [lane 3] 10 units of reverse transcriptase; [lane 4] molecular weight markers. (c) Primer extension analysis of the z2 transcript. The 33-nucleotide primer used in this study is shown in the map below the autoradiogram. The primer-extended product is 222 nucleotides. [Lane 1] Molecular weight markers; [lane 2] deoxyguanosine reaction from the dideoxy sequence of a 5' fragment from the z2 gene; [lane 3] 10 units of reverse transcriptase; [lane 4] no RNA control.
Figure 4. Molecular organization of zl and z2. These summaries are based on a comparison of genomic and cDNA sequences, as well as 5' mapping studies. Putative protein coding regions are indicated by the solid bars, whereas those regions that are transcribed but untranslated are indicated by open bars. The homeo box is cross-hatched, and introns are shown as lines. (a) The zl transcription unit is over 1340 bp in length. There is a TATA box located 25 bp upstream from the transcription start site. The gene is interrupted by one intron of 64 bp that is located between +138 and +202. The zl cDNA lacks a complete 3' end, but sequencing of appropriate genomic DNA templates (see Fig. 1b) reveals a possible polyadenylation signal at +1344. Cleavage and polyadenylation near this site would result in a ~1.3-kb mRNA, which is in agreement with the Northern data shown in Fig. 2a. (b) The z2 transcription unit is over 1 kb in length. There is a TATA box located 28 bp upstream from the transcription start site. The gene is interrupted by one intron of 67 bp that is located between +126 and +193. There is a potential polyadenylation signal at +1004 that would result in an mRNA about 900 bp in length, which is in agreement with Northern data (see Fig. 2b). Please note that since S1 nuclease protection studies have not been done with the 3' regions, it is possible that additional zl and/or z2 transcribed sequences occur further downstream.

Discussion

We have described the structure, sequence, and expression of two closely linked homeo box genes that reside within the zen region of the Antennapedia complex. The zl and z2 transcription units are separated by ~10 kb and are the smallest of the 11 homeo box genes that have been characterized in detail in Drosophila. Each gene encodes a single mRNA, which appears to specify a single protein product. These putative zl and z2 proteins are more related to each other than either one is to any other homeo box product, although they are related only by virtue of their homeo boxes and do not share other sequence homologies. The patterns of zl and z2 expression during early embryogenesis are essentially identical, and both correlate with the timing and sites of zen gene function. Several lines of evidence suggest that zl corresponds to zen and that z2 might be dispensable and not required for the viability of developing embryos. However, we cannot rule out the possibility that either zl or z2 can provide zen + gene function when one of the genes is missing.

Comparison of zl and z2 with other homeo box genes

The zl and z2 transcription units appear to be the smallest of the known homeo box genes and are only ~1350 bp and ~1010 bp in length, respectively. This is in contrast to the large sizes observed for a number of other homeo box transcription units, including the homeotic genes Antp and Ubx, both of which are over 75 kb in length (Bender et al. 1983; Garber et al. 1983; Scott et al. 1983). In general, it appears that those homeo box genes that function during early development are smaller than those that function later. Thus, the segmentation genes even-skipped, fushi tarazu, and en grailed are each less than 5 kb in length and encode proteins that show localized patterns of expression during cellularization (Carroll and Scott 1985; DiNardo et al. 1985; Frasch et al. 1987). In contrast, Antp and Ubx proteins are not detected until germ band elongation, at least several hours after the appearance of the segmentation proteins (White and Wilcox 1984; Beachy et al. 1985; Carroll et al. 1986; Wirz et al. 1986). The small size of the zl transcription unit is consistent with the early appearance of zl proteins in precellular embryos (Rushlow et al., 1987) and with its early time of function based on genetic analyses of a temperature-sensitive zen mutant (Wakimoto et al. 1984). Moreover, zl and z2 transcripts are among the first zygotically expressed

GENES & DEVELOPMENT 1273
homeo box products that are detected during early development and show dorsally restricted patterns of expression by cleavage stage 10-11 (Doyle et al. 1986; C. Rushlow, unpubl.).

Despite their close linkage, similar structures, and essentially identical patterns of expression, the z1 and z2 coding sequences are quite divergent. They do not share significant homologies outside the homeo box, although both proteins possess similar amino acid compositions. For example, both proteins possess high contents of

AAGCTCTAGATGGGTAACATGCGAGAGAGATGGCGCTGCAGTG


tCCACGCGGAAATGGGTAATGCGGAGACTGCACCATGCTGGCTGCAGTG

The nucleotide labeled +1 corresponds to the transcription start site. There is a well-conserved TATA sequence located 25 bp upstream from the transcription start site. The first AUG is located 52 bp downstream from the start site and begins an open reading frame of 353 amino acid residues. The amino acid sequence of this reading frame is shown above the nucleotide sequence. The amino acid sequences 90-149 are boxed. The vertical arrow between nucleotides 138 and 139 indicates the location of the splice junctions of a 64-bp intron. The dashed underline delineates the nucleotide sequence of the oligomer for the primer extension analysis shown in Fig. 3b, and the circles bracket the S1 nuclease probe used for the protection experiment shown in Fig. 3a.

Figure 5. Nucleotide sequence of z1. The complete nucleotide sequence of the z1 cDNA is shown. This cDNA is not full length and lacks sequences from both termini. The nucleotide sequence of these missing regions was obtained from genomic DNAs, which are underlined in the above sequence. The nucleotide labeled +1 corresponds to the transcription start site. There is a well-conserved TATA sequence located 25 bp upstream from the transcription start site. The first AUG is located 52 bp downstream from the start site and begins an open reading frame of 353 amino acid residues. The amino acid sequence of this reading frame is shown above the nucleotide sequence. The numbers in the right column refer to the amino acid residues. The homeo domain is boxed (amino acid residues 90-149). The vertical arrow between nucleotides 138 and 139 indicates the location of the splice junctions of a 64-bp intron. The dashed underline delineates the nucleotide sequence of the oligomer used for the primer extension analysis shown in Fig. 3b, and the circles bracket the S1 nuclease probe used for the protection experiment shown in Fig. 3a.
serine and alanine residues and contain short, interrupted stretches of polyalanine and polyserine. Support for an evolutionary linkage of the two genes is provided by the observation that the z1 and z2 homeo boxes are closely related and share 75%-amino-acid identity.

The z1 and z2 homeo boxes are distinct from the 16 other Drosophila homeobox genes that have been reported [reviewed in Levine and Harding 1987]. These different homeo boxes have been grouped into three classes: the Antennapedia class, the engrailed class, and the paired class. The z1 and z2 homeo boxes are slightly more related to the Antennapedia class [McGinnis et al. 1984; Scott and Weiner 1984] than to the engrailed and paired classes [Fjose et al. 1985; Poole et al. 1985; Frigerio et al. 1986]. z1 and z2 share ~60%-amino-acid identity with members of the Antennapedia class, ~50% identity with the engrailed class, and only ~40% identity with the paired class. There is a highly conserved region of the z1 and z2 homeo domains that is identical to the corresponding portions of the homeo domains associated with the different genes of the Antennapedia class. This region consists of 14 amino acid residues (residues 131-144 of the zl protein sequence shown in Fig. 5 and residues 84-97 of the z2 protein sequence shown in Fig. 6) and includes the putative DNA binding recognition helix [Laughon and Scott 1984; Desplan et al. 1985]. Therefore, we would expect the zl and z2 proteins to show similar specificities in DNA binding; both of them

![DNA sequence](image-url)
homeo domains are located in the amino-terminal half of the respective proteins. The pair-rule gene even-skipped is the only other homeo box gene that encodes a homeo domain with a similar amino-terminal location [Macdonald et al. 1986; Frasch et al. 1987].

Which gene corresponds to zen? The patterns of z1 and z2 expression correlate well with the timing and sites of zen+ gene function. The phenocritical period of zen activity was shown to occur between 2 and 4 hr after fertilization [Wakimoto et al. 1984], which corresponds to the time when there are peak levels of z1 and z2 transcripts. The complex zen− phenotype seen in advanced-stage mutant embryos ap-
pears to result from a disruption of the dorsal ectoderm during gastrulation (C. Rushlow, unpubl.). In particular, the dorsal-most ectoderm, which normally differentiates into the amnioserosa, is transformed toward a more ventral cell type in zen−. We have shown that both z1 and z2 transcripts accumulate primarily in the dorsal ectoderm of wild-type embryos and are observed in the cells that comprise the amnioserosa during germ band elongation. Moreover, z1 and z2 transcripts are also detected in the optic lobe primordium, which is another dorsally derived structure that is absent in zen− mutants (Wakimoto et al. 1984). z1 transcripts (but not z2) are also detected in a subset of the pole cells, which are an unexpected site of expression since previous genetic studies do not indicate an obvious role for zen+ activity in these cells (Wakimoto et al. 1984). A more rigorous determination of any role for zen in pole cells would involve pole cell transplantation studies using zen− embryos as donors.

It is not clear from the expression patterns whether z1, z2, or both genes correspond to zen. However, here we have shown that the z1 transcription unit, in the absence of z2 sequences, can rescue zen− embryos by P transformation (Fig. 1a; Table 1). The 4.5-kb genomic DNA fragment that was used for these studies includes only a limited amount of flanking sequences (1.6 kb of 5' and 1.5 kb of 3' flank), which nonetheless appears to contain the cis-regulatory elements required for normal function. The efficiency of rescue with P-z1 transposons integrated at diverse chromosome locations suggests that flanking sequences exert little, if any, position effects on the level of z1 expression from the transposon. Each of five different transformed lines provided efficient rescue, whereby over 50% of the expected number of "zen−" adults were observed (see legend to Table 1). These results are in contrast with rescue studies done with the homeo box gene fushi tarazu (ftz), which is also located within the ANT-C. The rescue of ftz− embryos required an extensive amount of 5'-flanking sequences (over 6 kb), and there were position effects on the expression of the ftz transposon as well as a relatively poor efficiency of rescue (Hiromi et al. 1985).

The demonstration that the z1 gene can rescue zen− embryos suggests that z1 is a critical component of zen+ gene function. Several lines of evidence suggest that the z2 gene might be dispensable. The zenw36/DF(3R) LIN mutants that were used for the rescue studies appear to be both z1− and z2−. zenw36 homozygotes show a marked underexpression of both z1 and z2 transcripts, and DF(3R) LIN homozygotes do not express z1 and show a grossly abnormal pattern of z2 expression (C. Rushlow and M. Levine, in prep.). Thus, it would appear that z1−, z2− embryos that contain a single copy of the z1 gene are fully viable. Further support for this observation stems from genetic studies which indicate that z2− embryos are viable and do not show a zen mutant phenotype (M.A. Pultz and T.C. Kaufman, pers. comm.).

It is intriguing that z2 might be dispensable and not required for normal development. The pattern of z2 expression is virtually identical to z1 and corresponds to that expected for a zen product. Moreover, z2 does not appear to be a pseudogene since the nucleotide sequence of the z2 transcription unit strongly suggests that it encodes a protein. Thus, it appears that the z2 gene has been selected at the level of protein function, even though it might not be required for normal development.

There are at least two explanations for this discrepancy. First, it is possible that the z2 gene confers a selective advantage to flies in nature that is not obvious when they are cultured in the laboratory. For example, perhaps the z1 and z2 proteins function more effectively together than does the z1 protein alone. Alternatively, it is possible that z2 is not dispensable and that either z1 alone or z2 alone can provide zen+ gene function. A critical test of this latter possibility is to determine whether embryos that are z1− and z2+ are viable. In an effort to identify such a situation, we examined the expression of z1 and z2 in embryos homozygous for zenw36, as well as in DF(3R) LIN homozygotes. Surprisingly, as indicated above, both of these zen mutants show altered patterns of z1 and z2 expression. Additional experiments, particularly involving the use of P transformation using z2 DNA sequences, will be required to resolve this issue.

There are other examples of closely linked homeo box genes that show essentially identical patterns of expression, including the segmentation genes engrailed (en) and invected (inv) (Coleman et al. 1987) and the segment polarity gene gooseberry (Bopp et al. 1986). The significance of such gene duplications is unclear. It is possible that duplicated, dispensable genes such as z2 represent members of the homeo box gene family that are in the process of acquiring novel functions.

### Table 1. Rescue of zen mutants with a P-z1 transposon

| Line | Linkage | Transposon copy number | Number of survivors | Percentage of expected survivors
|------|---------|------------------------|---------------------|-------------------------------|
| 1    | II      | 2                      | 7                   | 56                            |
| 2    | II      | ND                     | 40                  | 264±2                        |
| 3    | III     | 1                      | 0                   | —                             |
| 4    | II, III | 2                      | 17                  | 112                           |
| 5    | II      | ND                     | 13                  | 96                            |
| 6    | II      | 1                      | 14                  | 72                            |

Males carrying at least one copy of the P-z1 transposon were crossed to zenw36/FT3 TM3 Sb females. Several of the Sb+ male progeny were crossed to DF(3R) LIN e/TM3 Sb females. e Sb+ progeny were scored as survivors and correspond to zenw36/DF(3R)LIN trans-heterozygotes.

1 Of the progeny from the final cross [above], 12.5% should correspond to mutant trans-heterozygotes since only 50% of the Sb+ male progeny from the first cross contain a copy of the z1 transposon. The values in this column represent the percentage of the expected number of mutant trans-heterozygotes that actually survived to adulthood.

2 This value exceeds 100% since the transformed line used for rescue contains additional copies of the z1 transposon on the X and/or fourth chromosome.

3 There are no survivors expected from this cross since the transformed line contains a single integration on the third chromosome, which does not segregate with the zen mutant alleles.
Materials and methods

Fly strains

zenw36 is an EMS-induced zen “point” mutation that is cytologically normal [Wakimoto et al. 1984]. The deficiency Df(3R)LIN uncovers several genes within the ANT-C, including Deformed, bicoid, and zen [Frohnhofer and Nüsslein-Volhard, 1986]. Df(3R)LIN fails to complement zenw36 [C. Rushlow, unpubl.].

P transformation

Germ line transformation was done essentially as described by Rubin and Spradling [1982]. A helper P element called pr25.7WC [Karess and Rubin 1984] was used. A 4.5-kb Xbal–SacI genomic DNA fragment containing the coding region as well as 1.6 kb of 5′ sequence and 1.5 kb of 3′ sequence was inserted into the NotI site of pDM30, which is a modification of the Carnegie 20 vector (a gift from D. Mismer). ryw306 embryos were injected with 50–100 μg/ml of helper plasmid and 300 μg/ml of the z1/P transposon.

Cloned DNAs

The z1 and z2 cDNAs were isolated in a screen of a 3- to 12-hr cDNA library cloned in λgt10 [Poole et al. 1985].

Sequencing

DNA fragments were cloned into the bacteriophage M13 vectors [Messing 1983], mp18+ and mp19+ [New England Biolabs]. Sequencing was performed by the dideoxy method [Sanger 1977] using [35S]dATP [New England Nuclear]. The single-stranded z1 DNA fragment that was used for the S1 nuclease protection experiment shown in Figure 3a was sequenced by the chemical cleavage method [Maxam and Gilbert 1977].

RNA Extraction

RNA was extracted using a hot phenol–SDS method (Scott et al. 1983) from wild-type embryos at 0–2, 2–3, 3–6, 6–12, and 12–24 hr after fertilization. Poly(A)+ RNA was selected on oligo(dT)-cellulose [Sigma] columns, fractionated on an agarose–formaldehyde gel, and transferred to nitrocellulose [Maniatis et al. 1982]. Approximately 0.5 μg of poly(A)+ RNA was loaded per lane. [32P]UTP-labeled single-stranded RNA probes were synthesized by T7 transcription from z1 and z2 cDNAs [Promega]. Hybridization was performed at 60°C for 24 hr in 50% formamide, 5 x SSC, 5 x Denhardt’s, 50 mM PO4, 0.1% SDS, 100 μg/ml carrier salmon sperm DNA, 100 μg/ml carrier tRNA. The blots were washed 4 x 5 min in 2 x SSC, 0.1% SDS at room temperature and then 3 x 15 min in 0.1 x SSC, 0.1% SDS at 65°C.

Transcript mapping

Primer extension analyses were performed essentially as described by McKnight and Kingsbury [1982]. The nucleotide primers shown in Figure 3 were [32P]-labeled at their 5′ termini with polynucleotide kinase, and each hybridized with 2 μg of embryonic poly(A)+ RNA in 0.5 M NaCl, 10 mM PIPES [pH 6.4] for 6 hr at 55°C. The primer–RNA hybrid templates were extended with reverse transcriptase for 1 hr at 42°C in 50 mM Tris (pH 8.0), 10 mM DTT, 6 mM MgCl2, 24 μg/ml actinomycin D 0.5 mM deoxy-NTPs. The products were electrophoresed in a 6% polyacrylamide–8 M urea gel. The DNA probe used for S1 nuclease protection experiments was obtained from a 316-bp genomic DNA fragment. After in vitro-labeling with polynucleotide kinase, the probe was electrophoresed in a nondenaturing gel to separate the strands. Single-stranded probe was hybridized with poly(A)+ RNA for 12 hr at 37°C in 0.25 M NaCl, 30 mM NaAc, 1 mM MnSO4, 5% glycerol and digested with 25, 50, and 100 units of S1 nuclease. Digestions were done at 37°C for 30 min. The resulting products were electrophoresed in a 6% polyacrylamide–8 M urea gel.

In situ hybridization

Tissue sections were prepared as previously described [Harding et al. 1986]. 35S-Labeled antisense RNA probes were synthesized by T7 transcription from z1 and z2 cDNAs and hybridized to tissue sections as described by Ingham et al. [1985].

Acknowledgments

We thank Thom Kaufman and Christiane Nüsslein-Volhard for providing fly stocks. We also thank Allen Laughon, Mary Anne Pultz, Mark Seeger, Thom Kaufman, and Tom Kornberg for sharing unpublished results. We are indebted to Janice Fischer for help with the transformation experiments. This work was funded by grants from the National Institutes of Health, Searle Scholars Program, and the American Cancer Society.

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under accession number Y00213.

References

Beachy, P.A., S.L. Helfand, and D.S. Hogness. 1985. Segmental distribution of bithorax complex proteins during Drosophila development. Nature 313: 545–551.

Bender, W., P. Spierer, and D.S. Hogness. 1983. Chromosomal walking and jumping to isolate DNA from the Ace, rosy, and bithorax complex in Drosophila melanogaster. J. Mol. Biol. 168: 17–33.

Carroll, S.B. and M.P. Scott. 1986. Localization of the fushi tarazu protein during Drosophila embryogenesis. Cell 43: 47–57.

Carroll, S.B., R.A. Laymon, M.A. McCutcheon, P.D. Riley, and M.P. Scott. 1986. The localization and regulation of Antennapedia protein expression in Drosophila embryos. Cell 47: 113–122.

Campos-Ortega, J. and V. Hartenstein. 1985. The embryonic development of Drosophila melanogaster, pp. 31–36. Springer-Verlag, Berlin, Heidelberg.

Coleman, K.G., S.J. Poole, M.P. Weir, W.C. Soeller, and T. Kornberg. 1987. The injected gene of Drosophila: Sequence analysis and expression studies reveal a close kinship to the engrailed gene. Genes Dev. 1: 19–28.

Desplan, C., J. Theis, and P.H. O’Farrell. 1985. The Drosophila developmental gene, engrailed, encodes a sequence-specific DNA binding activity. Nature 318: 630–635.

DiNardo, S., J.M. Kuner, J. Theis, and P.H. O’Farrell. 1985. Development of embryonic pattern in D. melanogaster as revealed by accumulation of the nuclear engrailed protein. Cell 43: 59–69.

Doyle, H.J., R. Harding, T. Hoey, and M. Levine. 1986. Tran-
scripts encoded by a homeo box gene are restricted to dorsal tissues of Drosophila embryos. Nature 323: 76–79.

Fjose, A., W.J. McGinnis, and W.J. Gehring. 1985. Isolation of a homeo box-containing gene from the engrailed region of Drosophila and the spatial distribution of its transcripts. Nature 313: 284–289.

Frasch, M., T. Hoey, C. Rushlow, H. Doyle, and M. Levine. 1987. Characterization and localization of the even-skipped protein of Drosophila. EMBO J. 6: 749–759.

Garber, R.L., A. Kuroiwa, and W.J. Gehring. 1983. Genomic and cDNA clones of the homeotic gene Antennapedia in Drosophila. EMBO J. 2: 2027–2034.

Harding, K., C. Rushlow, H.J., Doyle, T., Hoey, and M. Levine. 1986. Cross-regulatory interactions among pair-rule genes in Drosophila. Science 229: 1236–1242.

Hiromi, Y., A. Kuroiwa, and W.J. Gehring. 1985. Control elements of the Drosophila segmentation gene fushi tarazu. Cell 43: 603–613.

Ingham, P.W., K.R. Howard, and D. Ish-Horowicz. 1985. Transcription pattern of the Drosophila segmentation gene hairy. Nature 318: 439–445.

Karess, R.E. and G. Rubin. 1984. Analysis of P transposable element functions in Drosophila. Cell 38: 135–146.

Laughon, A. and M.P. Scott. 1984. Sequence of a Drosophila segmentation gene: Protein structure homology with DNA-binding proteins. Nature 310: 25–31.

Laughon, A., A.M. Boulet, J.R. Bermingham, R.A. Laymon, and M.P. Scott. 1986. Structure of transcripts from the homeotic Antennapedia gene of Drosophila melanogaster. Two promoters control the major protein-coding region. Mol. Cell. Biol. 6: 4676–4689.

Levine, M. and K. Harding. 1987. Spatial regulation of homeo box gene expression in Drosophila. Oxford Survey of Eukaryotic Genes, vol. 3 Oxford University Press (in press).

Macdonald, P.M., P.N. Ingham, and G. Struhl. 1986. Isolation, structure, and expression of even-skipped: a second pair-rule gene of Drosophila containing a homeo box. Cell 47: 721–734.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, New York.

Maxam, A.M. and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. 74: 560–565.

McGinnis, W., M.S. Levine, E. Haen, S. Kuroiwa, and W.J. Gehring. 1984. A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature 308: 428–433.

McKnight, S.L., and R. Kingsbury. 1982. Transcriptional control signals of a eukaryotic protein coding gene. Science 217: 316–324.

Messing, J. 1983. A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Methods Enzymol. 101: 20–28.

Mlodzik, M. and W. Gehring. 1987. Expression of the caudal gene in Drosophila: Formation of an RNA and protein gradient during early embryogenesis. Cell 48: 465–478.

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.
Molecular characterization of the zerknüllt region of the Antennapedia gene complex in Drosophila.

C Rushlow, H Doyle, T Hoey, et al.

Genes Dev. 1987, 1:
Access the most recent version at doi:10.1101/gad.1.10.1268

References
This article cites 37 articles, 7 of which can be accessed free at:
http://genesdev.cshlp.org/content/1/10/1268.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.