Cell, Vol. 43, 29–37, November 1985, Copyright © 1985 by MIT

0092–8674/85/110029–09 $02.00/0

Expression during Embryogenesis of a Mouse Gene with Sequence Homology to the Drosophila *engrailed* Gene

Alexandra L. Joyner,* Thomas Kornberg,†‡ Kevin G. Coleman,‡ David R. Cox,†‡ and Gail R. Martin*†

* Department of Anatomy

† Division of Genetics

‡ Department of Biochemistry and Biophysics

University of California, San Francisco

San Francisco, California 94143

Summary

Regions of the mouse and human genomes with strong homology to the Drosophila *engrailed* gene have been identified by Southern blot analysis. One mouse *engrailed*-like region, Mo-en.1, has been cloned and partially sequenced; homology with the *engrailed* gene is localized to a 180 bp *engrailed*-like homeo box and 63 nucleotides immediately 3′ to it. The protein sequence this region can encode includes 81 amino acids, of which 60 (75%) are identical with those of the putative translation product of the corresponding *engrailed* sequence. These data suggest that Mo-en.1 represents a mouse homolog of a gene of the Drosophila *engrailed* gene complex. Mo-en.1 has been mapped to chromosome 1, indicating it is not linked to other homeo box sequences that have been mapped in the mouse genome. Analysis of poly(A)⁺ RNA extracted from teratocarcinoma cells and whole mouse embryos demonstrates that the conserved homeo box region of Mo-en.1 is expressed differentially during mouse embryogenesis.

Introduction

Attempts to identify genes that regulate mammalian development have been hindered by the long gestation periods of mammalian embryos, their relatively small numbers, and their inaccessibility during development in utero. These problems are minimized in studies of insect development, and genetic experiments in the fly Drosophila melanogaster have identified a number of loci with a basic role in the control of pattern formation. The recent discovery of a region of significant homology between some of these genes in Drosophila and sequences in mammalian DNA (McGinnis et al., 1984a) has suggested that genes that control mammalian development may be identified and cloned by using sequences from Drosophila genes as probes.

The Drosophila genes that are known to regulate early embryonic development each fall into one or more of three general categories: maternal-effect genes, which are expressed during oogenesis and specify the structure and spatial coordinates of the egg (Nusslein-Volhard, 1979); segmentation genes, which determine the number and polarity of the body segments (Nusslein-Volhard and Wieschaus, 1980); and homeotic genes, which specify segment identity (Ouweneel, 1976). Several of the genes in the latter two categories share a conserved 180 bp DNA sequence that has been termed the "homeo box" (McGinnis et al., 1984a, 1984c; Scott and Weiner, 1984). The presence of the homeo box sequence in these genes suggests that they may derive from a common ancestor and perhaps carry out their respective functions by similar mechanisms.

The Drosophila homeo-box-containing genes that have been identified thus far can be grouped into two classes. The smaller class is represented only by the two genes (Poole et al., 1985; Fjose et al., 1985) in a region, the *engrailed* gene complex (EN-C), that includes the *engrailed* gene (Garcia-Bellido and Santamaria, 1972). The larger class includes at least seven genes (Regulski et al., 1985) in the *Antennapedia* gene complex (ANT-C; Kaufman et al., 1980) and the *Bithorax* gene complex (BX-C; Lewis, 1978). There are several features that distinguish the two classes of genes. EN-C is located on the right arm of chromosome 2 and ANT-C and BX-C are on the right arm of chromosome 3. The homeo boxes found in the EN-C genes are interrupted by an intervening sequence, whereas those in the ANT-C/BX-C genes are not. The polypeptides encoded by the homeo boxes found in the two EN-C genes share approximately 87% amino acid homology; the homeo boxes in the ANT-C/BX-C genes also share 75% or more amino acid homology; however, there is only approximately 50% amino acid homology between homeo boxes from genes of the two classes. Thus, homeo boxes themselves can be generally classified as EN-C or ANT-C/BX-C on the basis of their degree of amino acid sequence homology with members of a given class. The most striking difference between the genes in the two classes is that the two genes in the EN-C show DNA sequence and putative amino acid homology outside their homeo boxes. This unique region of homology can encode 31 amino acids immediately 3′ to the homeo box. In contrast, no homology outside the homeo box has yet been found among any of the ANT-C/BX-C homeo-box-containing genes of Drosophila.

Several different DNA segments containing homeo boxes have been isolated, using ANT-C/BX-C homeo box sequence probes, from the genomes of the frog (Carrasco et al., 1984; Muller et al., 1984), mouse (McGinnis et al., 1984b; Colberg-Poley et al., 1985; Hauser et al., 1985), and man (Levine et al., 1984). The putative amino acid sequence encoded by each of these vertebrate homeo boxes shares at least 70% homology with all members of the ANT-C/BX-C class of homeo box. In the mouse genome the ANT-C/BX-C homeo box sequences appear to reside in two clusters, one on chromosome 11 that contains at least two, and probably more, homeo boxes (Joyner et al., 1985; Rabin et al., 1985; Hauser et al., 1985) and one on chromosome 6 that contains at least three homeo boxes within a 40 kb region (McGinnis et al., 1985b; D. Duboule, personal communication). A cluster of at least three homeo boxes that appears to have homology

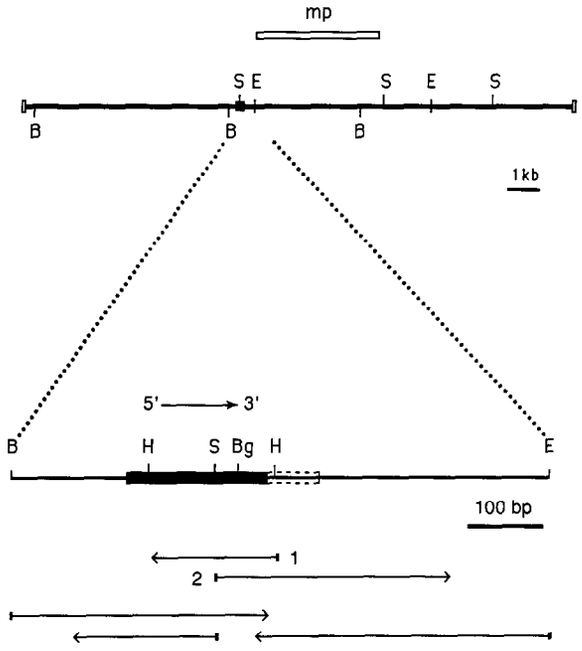


Figure 1. Restriction Map of λ Mo-en.1

The upper thick horizontal line represents the mouse DNA insert in the λ clone, with open rectangles marking the ends of the insert, and the lower horizontal line is an enlargement of the Eco RI/Bam HI restriction fragment containing the homeo box. The arrow indicates the direction of transcription of the homeo box. The homeo box is shown as a filled-in box and the 3′ region of homology between *engrailed* and Mo-en.1 is delineated by a dashed rectangle. The fragments of the subclones in M13 phage that were sequenced are shown below the Eco RI/Bam HI fragment as thin lines with arrowheads indicating the direction of sequencing. The open box marked mp represents the subcloned fragment used as a probe in the mapping studies, and Southern and Northern blot analyses. The M13 clones marked 1 and 2 were used to construct the probes for the Southern and Northern blot studies shown in Figures 4 and 6. B, Bam HI; S, Sst I; E, Eco RI; H, Hae III; Bg, Bgl II.

with the one found on mouse chromosome 11 maps to a region on human chromosome 17 (Joyner et al., 1985; Rabin et al., 1985; Hauser et al., 1985). Recent experiments have also demonstrated that these homeo box sequences reside in transcribed regions of the genome. RNAs containing several of the ANT-C/BX-C homeo box segments have been found to be differentially expressed during frog (Carrasco et al., 1984; Muller et al., 1984) and mouse embryonic development (Hauser et al., 1985), and during differentiation of mouse (Colberg-Poley et al., 1985) and human (Hauser et al., 1985) teratocarcinoma cells. Vertebrate representatives of the smaller EN-C class of homeo box have not been previously reported.

There is increasing evidence to support the concept that genes with homologous functions in vertebrates and invertebrates may share common DNA sequences despite the vast evolutionary distance that separates them. For example, the sequence homology between bovine and Drosophila rhodopsin has been exploited in the isolation of the gene from the fly genome (O'Tousa et al., 1985; Zucker et al., 1985), and the identification of the human and hamster genes for RNA polymerase was facilitated by the use of Drosophila sequences (Ingles et al., 1983). The studies presented here were designed to identify genes in mammals that control pattern formation in a manner analogous to that of the Drosophila *engrailed* gene by identifying mammalian genes with sequence homology to the Drosophila EN-C genes. Using a Drosophila *engrailed* gene cDNA clone as a probe, two regions of the mouse genome and a region of the human genome were found to share homology with sequences in the Drosophila *engrailed* gene; and one of the homologous mouse sequences, Mo-en.1, was cloned. As a first step in determining the function of this mouse *engrailed*-like sequence, we determined the sequence of the region of *engrailed* homology, the chromosomal location of Mo-en.1, and its temporal pattern of expression during mouse embryonic development.

Results

Isolation of Mouse Genomic Clones with Sequence Homology to the Drosophila *engrailed* Gene

To determine whether the mouse genome contains sequences homologous to those in the Drosophila *engrailed* gene, Southern blot analysis was carried out using mouse genomic DNA digested separately with four different restriction endonucleases. The Southern blot was probed under conditions of low stringency with a 1.4 kb Drosophilia *engrailed* cDNA clone that contains a homeo box (see Experimental Procedures). Two prominent restriction fragment bands hybridized to this probe in each of the four mouse DNA digests (data not shown, but see Figure 4). To explore the possibility of a relationship between these homologous sequences and those that can be detected with an ANT-C/BX-C homeo box probe, the same Southern blot was rehybridized to a probe, H.1, that contains the human Hu1 homeo box, which is a member of the ANT-C/BX-C class (Joyner et al., 1985). Six to eight restriction fragments were detected in each DNA digest; none of these comigrated with the two detected with the *engrailed* probe (data not shown).

To isolate the DNA sequences with *engrailed* homology, a mouse genomic library was screened under conditions of low stringency using the 1.4 kb *engrailed* cDNA probe. Five clones that cross-hybridize with the cDNA probe were isolated. Southern blot analysis, using the *engrailed* cDNA probe, of restriction endonuclease digests of one of these clones, termed Mo-en.1, indicated that it represents one of the two mouse genomic regions containing *engrailed* homology. The *engrailed*-homologous region in Mo-en.1 localizes to a 700 bp Bam HI/Eco RI fragment. A partial restriction map of the Mo-en.1 clone, which was further analyzed as described below, is shown in Figure 1.

Sequence Analysis of the Mo-en.1 *engrailed*-Homologous Region

The nucleotide sequence of the Mo-en.1 region that cross-hybridized with the *engrailed* probe was determined. Figure 2 shows 313 nucleotides of Mo-en.1, which include a homeo box, 37 bases of 5′ sequence, and 96 bases of 3′ sequence ending at a putative translation stop codon. To facilitate comparison with the homologous portion of the Drosophila *engrailed* cDNA, its sequence is also shown

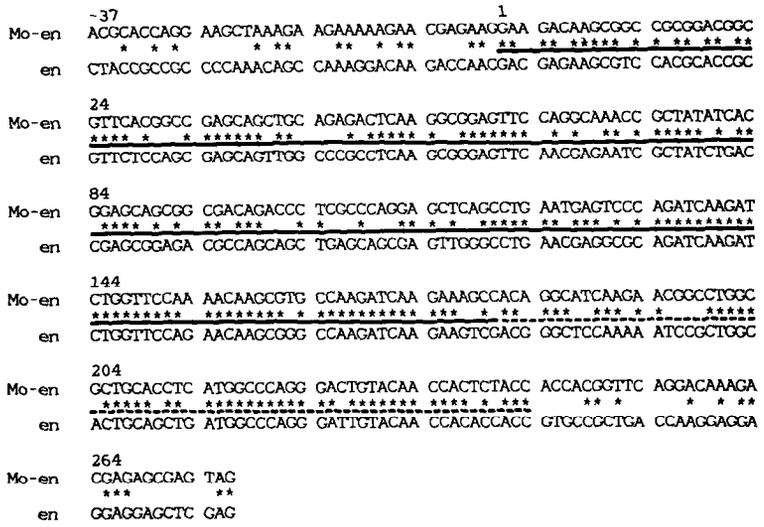


Figure 2. Nucleotide Sequence Comparison of the Mo-en.1 and *engrailed* Homeo Box Regions Showing the Homology within and 3′ to the Homeo Box

Direction of transcription is from left to right. Sequences of both strands of Mo-en.1 were obtained by the dideoxy chain terminating method. The sequence of the *engrailed* cDNA was determined previously (Poole et al., 1985). The nucleotide numbered 1 is the first nucleotide of the homeo box, and the homeo box is *underscored by a solid line. The region of homology 3′ to the homeo box is underscored by a dashed line.* Asterisks indicate nucleotides that are identical in Mo-en.1 and *engrailed*. The last three nucleotides in Mo-en.1 are the putative termination codon for translation of the open reading frame containing the Mo-en.1 homeo box. Mo-en, Mo-en.1; en, *engrailed*.

and the two sequences are aligned to achieve the best match of their respective homeo box sequences. No homology was found over a 60 bp stretch immediately 5′ of the homeo box, nor over a 160 bp stretch beginning 63 bp 3′ of the homeo box; but the homeo box (delineated by a solid line) and the 63 bases immediately 3′ of it (delineated by a dashed line) share approximately 73% homology. Although the Mo-en.1 homeo box sequence also shares approximately 65%–75% nucleotide homology with ANT-C/BX-C homeo boxes from Drosophila and vertebrates, the putative amino acid sequence of the Mo-en.1 homeo box is much more closely related to those of the EN-C genes than to those of the ANT-C/BX-C genes (see below).

The homologous mouse and Drosophila genomic regions differ in that Mo-en.1 encodes a continuous open reading frame whereas the homeo boxes of the two EN-C genes are each interrupted by an intervening sequence. Comparison of the amino acid sequences of the putative homeo box peptide domains specified by Mo-en.1 and *engrailed*, as well as the homologous sequence from the other EN-C gene, *invected* (previously called *engrailed*-related; Poole et al., 1985), reveals that the Mo-en.1 homeo box is more closely related to the EN-C than to the ANT-C/BX-C class of homeo boxes. This is illustrated in Figure 3, in which a comparison of the putative amino acid sequences of the homeo box regions of *Mo-en.1*, *engrailed*, *invected*, and *Antennapedia* are shown. If only identical amino acids are considered, the Mo-en.1 homeo box segment shares 73% homology with the homeo boxes of each of the EN-C genes, and only 52% homology with the homeo box of the *Antennapedia* gene. A similar degree of amino acid homology (45%–53%) exists between Mo-en.1 and every other member described thus far of the ANT-C/BX-C class of homeo box from either Drosophila or vertebrates. Furthermore, of the 31 amino acids 3′ to the homeo box that are highly conserved between *engrailed* and *invected*, 17 of the first 21 amino acids immediately 3′ to the homeo box are also conserved in Mo-en.1 (Figure 3). In contrast, in the region flanking the homeo box there is no such homology between Drosophila and vertebrate genes containing homeo boxes of the ANT-C/BX-C class.

Conservation of the EN-C Homeo Box Region in the Mouse and Human Genomes

These data indicate that the *engrailed* homology in Mo-en.1 includes a homeo box and sequences 3′ to it. To determine whether these same sequences are also present in the second mouse genomic region detected with the *engrailed* cDNA probe, Southern blots of mouse genomic DNA were hybridized under conditions of moderate stringency to a Mo-en.1-derived probe, a-s, specific for the homeo box region (Figure 4). This probe, which extends from the Sst I site in the middle of the homeo box to approximately 50 bp 3′ of the homeo box, hybridized to two regions in mouse genomic DNA that were indistinguishable from those detected with the 1.4 kb *engrailed* cDNA probe. Thus, the second mouse genomic region, Mo-en.2, also contains sequences homologous with the conserved EN-C region found in Mo-en.1.

To determine whether sequences in the EN-C homeo box region also are conserved in the human genome, restriction-enzyme-digested human genomic DNA was probed with the a-s probe. At least one restriction fragment in each digest hybridized to the Mo-en.1 homeo box region probe (Figure 4). To determine whether this human region contains additional homology to Mo-en.1, a similar Southern blot was hybridized to a single-copy probe, mp, derived from Mo-en.1, which lacks any of the sequences conserved in Drosophila (data not shown). This probe, which extends from an Eco RI site 360 bp 3′ of the homeo box to a Sst I site approximately 4 kb 3′ of the homeo box, detected a single band in each human DNA digest, and these bands were found to comigrate with the most prominent bands detected with the a-s probe. Thus, the human genome contains a region homologous to Mo-en.1 that includes both a homeo box region and 3′-flanking sequences.

Assignment of Mo-en.1 to Mouse Chromosome 1

The Mo-en.1 locus was assigned to a mouse chromosome

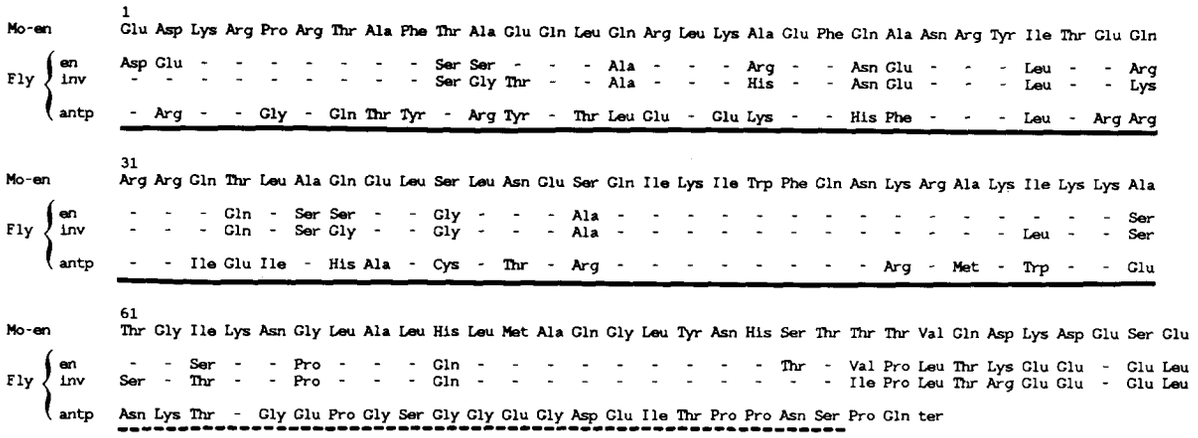


Figure 3. Comparison of the Putative Amino Acid Sequences of the Mo-en.1, EN-C, and *Antennapedia* Homeo Box Regions

The first amino acid shown is the first amino acid of the homeo box. The homeo box is underscored by a solid line, and the 3′ region of homology between Mo-en.1, *engrailed*, and *invected* is underscored by a dashed line. For each amino acid a dash indicates that it is identical with the amino acid found in the corresponding position in Mo-en.1. The last amino acid shown is the one at which the 3′ homology between *engrailed* and *invected* ends. The putative open reading frame for Mo-en.1 ends one codon past the amino acids shown, and *engrailed* ends 10 codons past the sequence shown. The sequences of the *engrailed* and *invected* cDNAs were determined by Poole et al. (1985). The *Antennapedia* sequence shown is from Scott and Weiner (1984) and A. Laughon, R. Laymon, and M. Scott (personal communication). Mo-en, Mo-en.1; en, *engrailed*; inv, *invected*; antp, *Antennapedia*.

using a panel of nine Chinese hamster–mouse somatic cell hybrids that segregate mouse chromosomes. The mouse chromosome complement of each hybrid clone was determined by isozyme and/or karyotype analysis as previously described (Cox et al., 1982). These nine hybrid cell lines together contain the full complement of mouse chromosomes except for chromosomes 11 and Y (Figure 5). Southern blots of mouse and Chinese hamster genomic DNA digested with Eco RI and hybridized with the Mo-en.1 mapping probe, mp, showed strong hybridization to the expected 5 kb band in mouse DNA and faint hybridization to two smaller bands in Chinese hamster DNA. An analysis of the segregation of Mo-en.1 and mouse chromosomes in the hybrid cell clones showed concordant segregation of Mo-en.1 with chromosome 1 and discordant segregation with each of the other mouse chromosomes (Figure 5), thus assigning Mo-en.1 to mouse chromosome 1. Furthermore, since the 5 kb mouse Mo-en.1 Eco RI fragment is not detectable in DNA from hybrid clone III-16, we conclude that Mo-en.1 maps distal to band C2 on chromosome 1.

Regulated Expression of Mo-en.1 in Differentiating Teratocarcinoma Cells and Mouse Embryos

The strong conservation of coding capacity over a 243 bp region of Mo-en.1 and the genes in the Drosophila EN-C suggests a conservation of function for this region of the proteins. We would therefore predict that this region is transcribed and translated during mouse embryogenesis, as it is during embryonic development in Drosophila. As a first step in testing this hypothesis, RNA from mouse teratocarcinoma cells at several stages of differentiation in vitro and from normal mouse embryos between 9.5 and 17.5 days of gestation was analyzed for Mo-en.1 sequences.

The PSA-1 teratocarcinoma stem cell that was used provides a model system for the peri-implantation stages of mouse embryogenesis and thus makes possible the study of embryonic cells at stages at which it would otherwise not be feasible to obtain experimental material (Martin, 1980). Undifferentiated PSA-1 cells, which appear to be equivalent to the inner cell mass of the mouse blastocyst, can be stimulated to differentiate by allowing them to form aggregates in suspension culture. These cell aggregates subsequently develop into structures known as embryoid bodies (Martin and Evans, 1975a). After 6–8 days of suspension culture these embryoid bodies closely resemble the fetal portion of mouse embryos at 4.5–6.5 days of gestation (stage 1). After approximately 10 days or more in suspension the embryoid bodies form "balloon-like" structures that have some features in common with embryos at approximately 7.5 days of gestation (stage 2) (Martin et al., 1977). The formation of many of the cell types in mid-gestation embryos (9.5–11.5 days), such as keratinizing epithelium, cartilage, striated muscle, and nerve, can be obtained by allowing the embryoid bodies to attach to a tissue culture surface and culturing them for an additional 7 days or more (stage 3) (Martin and Evans, 1975b).

Poly(A)$^+$ RNA was extracted from undifferentiated PSA-1 cells and from PSA-1 cells at each of the three stages of differentiation, and was analyzed using the Northern blot technique under conditions of high stringency. A 3.1 kb message was detected with a probe for the predicted sense strand of the homeo box (s probe, Figure 6), only in RNA from differentiated PSA-1 cells. This Mo-en transcript appears to be most highly expressed when the PSA-1 cells are at stages 2 and 3 of differentiation. A band of weak intensity of approximately 2 kb was also detected in RNA from cells at all stages of differentiation and from undifferentiated PSA-1 cells. No signal was detected when a similar blot was hybridized with a single-stranded probe complementary to the antisense strand of the Mo-en.1 homeo box region (a-s probe, see Figure 4), although in-

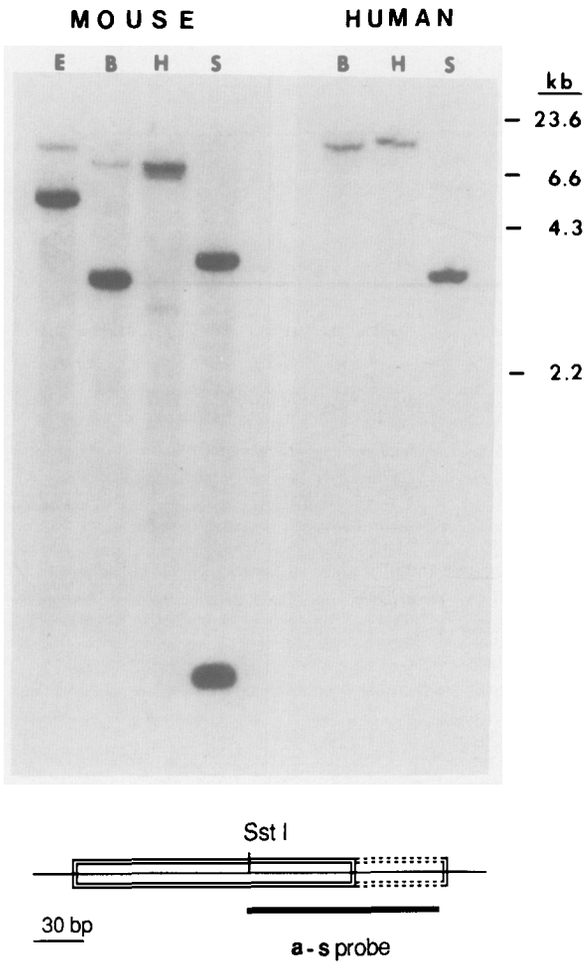


Figure 4. Mouse Genomic DNA Contains Two Copies of the Conserved EN-C Region and Human DNA Contains Conserved Mo-en.1 Sequences

Ten micrograms of mouse and human genomic DNAs were digested with the enzymes indicated, separated by electrophoresis on a 1% agarose gel, transferred to GeneScreen and irradiated with UV light. The blot was hybridized to the Mo-en.1 homeo box probe, a-s, diagrammed below the autoradiogram, and washed under moderate stringency conditions. The rectangle outlined by a double solid line in the diagram of the probe delineates the Mo-en.1 homeo box and the rectangle outlined by the dashed double line delineates the 3′ region of homology between Mo-en.1 and the genes of EN-C. B, Bam HI; E, Eco RI; H, Hind III; S, Sst I.

tact mRNA was clearly present in all the lanes as demonstrated by rehybridization of the blot with a β-actin probe (data not shown).

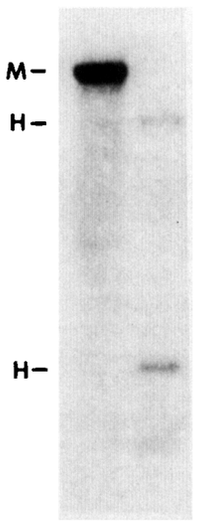
Since the s probe detected both Mo-en.1 and Mo-en.2 DNA under conditions of DNA–DNA hybridization parallel to those used in the Northern blot experiment (data not shown), it is conceivable that the s probe is hybridizing to transcripts from the Mo-en.1 and/or Mo-en.2 loci. To determine whether the 3.1 kb and 2 kb Mo-en RNAs observed are transcribed from Mo-en.1, a similar RNA blot was hybridized to the double-stranded mp 3′-flanking probe, which hybridizes to a single-copy sequence in the mouse genome. The mp probe hybridized to 3.1 and 2 kb transcripts indistinguishable from those observed with the s probe, and the former was expressed only in differentiated cells. In addition, a 3.5 kb band was present at all stages of differentiation and in the undifferentiated PSA-1 cells (data not shown). The origin of this RNA is not known, but it may derive from a separate transcription unit or may represent the product of differential splicing of the homeobox-containing transcription unit.

To determine whether Mo-en.1 is expressed at later stages of development, poly(A)$^+$ RNA was extracted from mouse embryos at 9.5 through 17.5 days of gestation and Northern blot analysis was carried out. In all the embryo RNA samples, a prominent 3.1 kb transcript, which appears to be most highly expressed between 10.5 and 12.5 days of gestation, was detected with the s probe (Figure 6). A weakly hybridizing 2 kb band was also detected in all the embryo samples. There was no detectable hybridization when a similar blot was hybridized with the a-s probe, whereas subsequent hybridization of this blot with a β-actin probe demonstrated that intact RNA was present in all lanes (data not shown). Following hybridization of similar blots to the mp probe, bands indistinguishable from those detected in RNA from differentiated PSA-1 cells were observed in all embryo RNA samples (data not shown).

Discussion

This study was initiated on the premise that if a homolog of the Drosophila *engrailed* gene exists in the mouse genome, it should be possible to identify and isolate it by virtue of conserved DNA sequences. We have identified a mouse gene, Mo-en.1, that can code for a protein that includes 81 amino acids sharing 75% homology with sequences in the two genes in EN-C. These 81 amino acids include a 60 amino acid homeo box domain that is significantly more homologous to homeo boxes of the EN-C genes than to any other homeo box described thus far, and 21 amino acids that lie 3′ to the homeo box. The finding that Mo-en.1 contains a homeo box demonstrates that it is one of a growing number of possible vertebrate homologs of the Drosophila genes that control pattern formation. The conserved sequence 3′ to the homeo box further identifies Mo-en.1 as a possible homolog of a gene in EN-C. Interestingly, our data provide evidence that EN-C gene homologs also may exist in the human and hamster genomes. In addition, a chicken genomic clone has been isolated that contains DNA sequences highly homologous to the same EN-C gene sequences that are conserved in Mo-en.1 (D. Darnell and C. Ordahl, personal communication).

Although the existence in Mo-en.1 of conserved EN-C gene sequences, particularly those found outside the homeo box, suggest that the gene we have isolated is a homolog of a Drosophila EN-C gene, this point will not be certain until it is known that the Mo-en.1 gene carries out a function in the mouse that is analogous to that performed by the EN-C genes in Drosophila. The *engrailed* gene is involved in the processes that subdivide the insect embryo into separate developmental units called compartments. Each segment of the fly is composed of an an-



Clone	Mo-en	Mouse Chromosome																		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	X
I-3A-3	+	+	+		+									+		+			+	+	
I-7B-4	+	+	+								+				+	+			+	+	
III-14	+	+			+		(+)	+		+						+	+			+	+
I-3A-2	-		+		+			+	+				+			+	+		+	+	
II-3A	-		+	+	+		+		+		+				+	+	+	+	+		+
III-12	-					+	+						+	+	+	(+)		+			+
III-16	-	[+]	+	+		+	+	+	+	+			+	+	+	(+)	+	+	+	+	+
III-23	-		[+]	+					+				+		+	(+)			+		[+]
VI-25	-		[+]	+			+				(+)						[+]				+

Figure 5. Chromosomal Mapping of Mo-en.1 Using Somatic Cell Hybrids

The autoradiogram of the Southern blot shows mouse (left) and Chinese hamster (right) genomic DNA digested with Eco RI and hybridized to the mapping probe, mp, shown in Figure 1. A similar Southern blot of Eco RI digests of genomic DNA from nine mouse-hamster somatic cell hybrids was hybridized to the mp probe. The results are summarized in the table shown on the right. The presence of mouse chromosomes was determined by isozyme analysis of hybrid cell extracts and/or by karyotype analysis. (+) indicates that a chromosome was present by isozyme analysis but not visualized by karyotype analysis. [+] indicates that only part of the chromosome is present by karyotype analysis. Cell line III-16 contains only a portion of chromosome 1 with the material distal to band 1C2 deleted. Cell line III-23 contains a rearrangement of chromosomes 2 and X, and clone VI-25 contains a 2;16 translocation. M, mouse; H, hamster.

terior and a posterior compartment (Garcia-Bellido et al., 1973), and in mutants that lack *engrailed* function, posterior compartments develop abnormal patterns and the compartment and segment borders are not maintained (Lawrence and Morata, 1976; Kornberg, 1981; Lawrence and Struhl, 1982). The discovery that only posterior compartment cells are affected by mutations at the *engrailed* locus has suggested that this gene may function as a binary developmental switch to "select" a posterior developmental pathway (Garcia-Bellido, 1975). Consistent with this hypothesis is the recent demonstration by in situ hybridization that *engrailed* is expressed only in the cells of the posterior compartments (Kornberg et al., 1985; Fjose et al., 1985). The *invected* gene has been recently identified as a transcribed gene that shares considerable sequence homology with, and is a neighbor of, the *engrailed* gene (Poole et al., 1985). No mutations are known that would indicate its developmental role.

In view of the biological insights that have been gained by studying mutant alleles of the *engrailed* gene, as well as other genes that control pattern formation in Drosophila, it is clear that the study of mutant alleles of the Mo-en.1 locus would help to illuminate both the function of this gene and the developmental processes in which it is involved. As a first step toward determining whether mutant alleles of Mo-en.1 exist we have used somatic cell hybrids to assign Mo-en.1 to a mouse chromosome. The data demonstrate that it maps to chromosome 1, and therefore is physically separate from the clusters of mouse ANT-C/BX-C homeo-box-containing genes that map to chromosomes 6 and 11. Several mutations, including splotch, dominant hemimelia, and loop-tail, that have been mapped to chromosome 1 are lethal when homozygous and have phenotypes suggestive of the types of developmental abnormalities that might result from lesions in an *engrailed*-like function (Green, 1981). Mapping studies in progress will help to determine whether any of these mutations are alleles of the Mo-en.1 locus.

In the Drosophila embryo, the *engrailed* gene shows a cell- and stage-specific pattern of expression that is displayed as a zebra-like array of bands of *engrailed*-expressing and nonexpressing cells. Our data show that Mo-en.1 encodes a major 3.1 kb homeo-box-containing transcript that is expressed by differentiating PSA-1 teratocarcinoma cells. Given that the differentiating PSA-1 cells are an in vitro model for embryonic cells in the early postimplantation stages of development, these results imply that the 3.1 kb transcript, containing the Mo-en.1 homeo box region, is first detectable between 4.5 and 6.5 days of development. We also demonstrate that this 3.1 kb transcript continues to be expressed throughout embryogenesis, with a peak of expression between 10.5 and 12.5 days of gestation. A second, less abundant 2 kb homeo-box-containing transcript is expressed at all stages without apparent regulation. Although the available data do not provide any means of drawing specific parallels between the expression of Mo-en.1 and the genes in the EN-C, the fact that this mouse gene is expressed differentially during embryogenesis is at least consistent with the idea that it may be involved in the control of embryonic development.

In considering the ways in which Drosophila and mouse *engrailed*-like genes might have similar functions, one difficulty is that the fundamental similarities between development in higher invertebrates and vertebrates are not obvious. If a parallel can be drawn between the segments

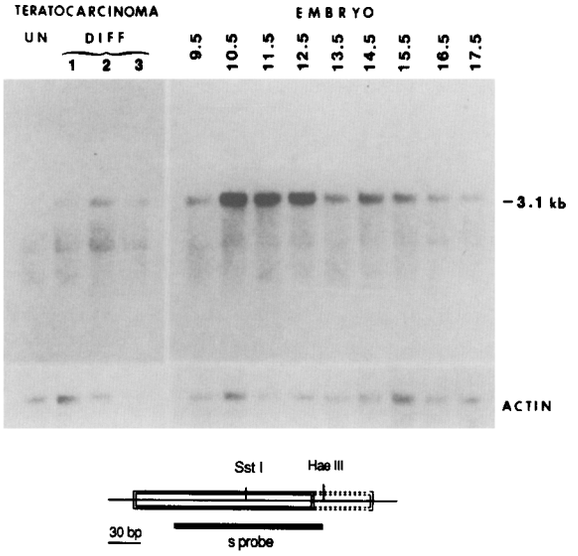


Figure 6. Expression of the Mo-en.1 Homeo Box in Teratocarcinoma Cells and Mouse Embryos

Northern blots of 5 μg of poly(A)⁺ RNA extracted from PSA-1 teratocarcinoma cells maintained in the undifferentiated state, from PSA-1 cells at three stages of differentiation (see text), and from whole mouse embryos at the days of gestation indicated were hybridized and washed under conditions of high stringency to a 140 bp homeo-box-containing single-stranded sense probe, s probe (diagrammed). The hybridized probe was subsequently stripped and the membranes were hybridized to a β-actin probe. See Figure 4 for a description of the diagram of the probe.

that constitute the Drosophila body plan and the repeating structures found in the vertebrate embryo, the somites (discussed by Hogan et al., 1985), then the function of a vertebrate *engrailed* gene might be to specify the distinction between the anterior and posterior portion of each vertebrate somite. There is evidence to suggest that genes with such a function do exist in vertebrates. Keynes and Stern (1984) showed that whereas motor and sensory axons normally grow through the anterior half of each successive somite, rotating the somitic mesoderm 180° about an anterior-posterior axis before somite segmentation has been completed results in axon growth through the posterior (formerly anterior) half of each somite. Apparently the commitment in the somite cells that leads to axon guidance preceded the experimental graft and was retained despite position change. Such operationally defined manifestations of determination suggest different developmental commitments in the anterior and posterior portions of the somites. Vertebrate homologs of the EN-C genes might play a role in the control of such somite cell specification in the mouse, and if so, they should be expressed in the vertebrate somites in a position-dependent manner. Further studies of Mo-en.1 expression and function in developing mouse embryos should help to determine whether it plays such a role and may provide new insights into the relationships between the processes that control development in higher vertebrates and invertebrates.

Experimental Procedures

All experimental procedures were carried out according to the methods described by Maniatis et al. (1982), unless otherwise specified.

DNA Probes

All probes were radiolabeled to a specific activity of at least 2×10^8 cpm/μg, using a nick-translation kit (Bethesda Research Laboratories) unless otherwise specified. The Drosophila *engrailed* probe was a 1.4 kb *engrailed* cDNA fragment extending from the Eco RI site that is approximately 300 bp 3′ of the homeo box to a site approximately 900 bp 5′ of the homeo box (Poole et al., 1985). The human ANT-C/BX-C homeo box probe, H.1, is a 1.7 kb Hind III fragment that contains the human Hu1 ANT/C-BX-C homeo box (Joyner et al., 1985). The chicken β-actin probe was kindly provided by Dr. D. Cleveland (Cleveland et al., 1980).

The mapping probe, mp, is a 3.6 kb Eco RI/Sst I fragment derived from Mo-en.1, which extends 3′ from an Eco RI site 360 bp 3′ of the homeo box. The mp fragment was subcloned from λ Mo-en.1 into Eco RI and Sst I digested SP65 plasmid DNA (Promega Biotech, Madison, Wisconsin). The single-stranded antisense probe, a-s, was approximately 140 bp in length and was derived from the M13 clone 2 shown in Figure 1. The single-stranded sense probe, s, was approximately 120 bp in length and was derived from the M13 clone 1 shown in Figure 1. The probes were made according to the methods in Church and Gilbert (1984) by primer extension under conditions of limiting radiolabeled ^{32}P-dCTP, and the labeled fragments of appropriate sizes were isolated after electrophoresis on a 5% acrylamide, 7 M urea gel.

Isolation of *engrailed*-Homologous Mouse Genomic DNA

A recombinant λ EMBL3A (Frischauf et al., 1983) library, kindly provided by Drs. J. Vogel and R. Goodenow, was screened with the 1.4 kb *engrailed* cDNA fragment. The library contains 12-20 kb fragments of mouse genomic DNA that had been partially digested with Sau 3A. Replica nitrocellulose filters were hybridized for 16 hr at 64°C in 6× SSC, 10× Denhardt's, 0.1% SDS, and 0.1% sodium pyrophosphate. The filters were washed two times for 30 min each in 2× SSC, 0.1% SDS, at 50°C followed by one wash in 2× SSC at 50°C.

Southern Blot Analysis

Recombinant λ DNAs (1 μg) or total genomic DNAs (10 μg) isolated from human and mouse tissues or from hamster-mouse somatic cell hybrid clones were cleaved with restriction enzymes, electrophoretically fractionated in 1% agarose gels, and transferred to nitrocellulose or GeneScreen (New England Nuclear) in 10× SSC. Subsequent to transfer, the nitrocellulose filters were baked according to the methods described by Southern (1975), whereas the GeneScreen nylon membranes were exposed to UV light, according to the methods of Church and Gilbert (1984) in order to cross-link the DNA to the nylon membranes. The Southern blots of mouse genomic DNA or recombinant λ clone DNA on nitrocellulose filters were hybridized to the Drosophila *engrailed* probes or H.1 using the following "low stringency" conditions: hybridizations were done at 37°C for 48 hr in 5× SSC, 5× Denhardt's solution, 250 μg/ml sonicated boiled salmon sperm DNA, 50 mM $NaPO_4$ (pH 7), 0.1% SDS, and 45% deionized formamide, and the filters were washed as described for the phage replica filters. The Southern blots of mouse, human, and somatic cell hybrid genomic DNA on GeneScreen membranes were hybridized with the Mo-en.1-derived probes using the following "moderate stringency" conditions: the hybridization conditions were as described by Church and Gilbert (1984) with the addition of 15% formamide, and the filters were washed twice at 50°C for 30 min, in 2× SSC and 1% SDS followed by one wash in 0.2× SSC at 50°C. In addition, the conditions of "high stringency" that were parallel to the Northern blot hybridization conditions were the same as the moderate conditions described above except that the $NaPO_4$ concentration was lowered to 0.2 mM and an additional wash at 65°C in 0.2× SSC was carried out.

Nucleotide Sequence Analysis

All DNA sequences were obtained by the dideoxynucleotide chain terminating technique (Sanger et al., 1977) using the universal 15-mer

primer (Pharmacia). Mo-en.1 fragments for sequencing were cloned into M13 vectors (Messing, 1983; Pharmacia), and the recombinant phages that were subsequently sequenced were identified by hybridizing replica filters of bacterial host cells (JM101) that had been transfected with the appropriate ligation mixtures to radioactively labeled 1.4 kb *engrailed* cDNA fragment under the low stringency conditions used to screen the mouse genomic library. DNA for sequencing was then isolated from cross-hybridizing recombinant phages (Sanger et al., 1980). All sequences shown were determined by independently sequencing both strands. For each M13 clone, only the length of DNA that was sequenced is shown in Figure 1.

Northern Blot Analysis

Total cellular RNA was isolated from PSA-1 cells or random bred mouse embryos (100 9.5 day embryos, 110 10.5 day embryos, 30 11.5 day embryos, and 10 to 20 embryos each from 12.5 to 17.5 day embryos) using essentially the guanidinium thiocyanate/$CsCl_2$ method of Chirgwin et al. (1977). The PSA-1 cells were cultured as described by Martin et al. (1977). Mouse embryos were obtained by mating mice of various strains. The day on which the vaginal plug was detected was termed 0.5 days of gestation. Embryos at the designated stages were dissected from the implantation site and separated from the extraembryonic membranes. One gram of tissue or pellets of 10^8 cells were added to 5 ml of buffer (4 M guanidinium thiocyanate, 100 mM Tris [pH 7.5], 0.1 M β-mercaptoethanol) and homogenized using a Polytron homogenizer (Brinkman Instruments) for 1 min, then 0.5 ml of 5% Sarcosyl was added, and the solution was mixed and then centrifuged at 10,000 rpm for 10 min. The mixture was sometimes stored at −70°C prior to layering 25 ml of the solution onto 15 ml of 5.7 M $CsCl_2$, 100 mM EDTA in SW27 tubes and centrifugation in an SW27 rotor at 25,000 rpm for 24 hr at 20°C. The resulting clear pellets were resuspended in 4 M urea, 100 mM Tris (pH 7.5), 5 mM EDTA, and the RNA mixture was extracted twice with phenol:chloroform (1:1), extracted once with chloroform, and precipitated with ethanol. The RNA was dissolved in H_2O and loaded onto oligo(dT)-cellulose columns (Collaborative Research), and poly(A)$^+$ was eluted. RNA was separated on 1% agarose-formaldehyde gels and blotted onto GeneScreen in 20× SSC as described by Maniatis et al. (1982), except that the running buffer was 0.5 M Hepes (pH 7.0), 10 mM EDTA, 50 mM Na-acetate, 6.6% formaldehyde, and the GeneScreen membranes were irradiated with UV light according to Church and Gilbert (1984). The hybridizations using single-stranded probes were carried out at 65°C for 24 hr in 1% bovine serum albumin, 0.2 M $NaPO_4$ (pH 7.2), 1 mM EDTA, 7% SDS, and 45% formamide. The filters were washed twice in 2× SSC, 1% SDS at 65°C for 30 min and once in 30% formamide, 0.2× SSC at 65°C for 30 min. The hybridizations using the mp or β-actin probes were carried out using the conditions described for the "moderate" hybridization and washing of Southern blots. The GeneScreen filters were stripped of hybridized probes by washing in 75% formamide, 0.1% SDS, 0.2× SSC, at 65°C for 30 min.

Acknowledgments

We are grateful to Drs. Michael Lovett and Patricia Ashley for helping to provide mouse-hamster somatic cell DNAs, and Drs. Thomas Cooper, George Church, Stephen Poole, Barry Drees, and Leslie Lock for helpful advice and encouragement. We thank Marianne Gallup and Joy Graesser for excellent technical assistance.

This work was supported by grants from the National Institutes of Health to G. R. M. and T. K.; A. L. J. was supported by a postdoctoral fellowship from the M. R. C. of Canada.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 9, 1985

References

Carrasco, A. E., McGinnis, W., Gehring, W. J., and DeRobertis, E. M. (1984). Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. Cell *37*, 409–414.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. Proc. Natl. Acad. Sci. USA *81*, 1991–1995.

Cleveland, D. W., Lopata, M. A., MacDonald, R. J., Cowan, N. J., Rutter, W. J., and Kirschner, M. W. (1980). Number and evolutionary conservation of α- and β- tubulin and cytoplasmic β- and γ-actin genes using specific cloned cDNA probes. Cell *20*, 95–105.

Colberg-Poley, A. M., Voss, S. D., Chowdhury, K., and Gruss, P. (1985). Structural analysis of murine genes containing homeo box sequences and their expression in embryonal carcinoma cells. Nature *314*, 713–718.

Cox, D. R., Sawicki, J. A., Yee, D., Appella, E., and Epstein, C. J. (1982). Assignment of the gene for β_2-microglobulin (*β2m*) to mouse chromosome 2. Proc. Natl. Acad. Sci. USA *79*, 1930–1934.

Fjose, A., McGinnis, W. J., and Gehring, W. J. (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.

Frischauf, A.-M., Lehrach, H., Poustka, A., and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. *170*, 827–842.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. In Cell Patterning, Ciba Foundation Symposium 29, new series (Amsterdam: Elsevier), pp. 161–182.

Garcia-Bellido, A., and Santamaria, P. (1972). Developmental analysis of the wing disc in the mutant *engrailed* or *Drosophila melanogaster*. Genetics *72*, 87–104.

Garcia-Bellido, A., Ripoll, P., and Morata, G. (1973). Developmental compartmentalization of the wing disc of *Drosophila*. Nature New Biol. *245*, 251–253.

Green, M. C., ed. (1981). Genetic Variants and Strains of the Laboratory Mouse. (Stuttgart: Gustav Fischer Verlag).

Hauser, C. A., Joyner, A. L., Klein, R. D., Learned, T. K., Martin, G. R., and Tjian, R. (1985). Expression of homologous homeo-box-containing genes in differentiated human teratocarcinoma cells and mouse embryos. Cell *43*, this issue.

Hogan, B., Holland, P., and Schofield, P. (1985). How is the mouse segmented? Trends Genet. *1*, 67.

Ingles, C. J., Biggs, J., Wong, J. K.-C., Weeks, J. R., and Greenleaf, A. L. (1983). Identification of a structural gene for a RNA polymerase II polypeptide in *Drosophila melanogaster* and mammalian species. Proc. Natl. Acad. Sci. USA *80*, 3396–3400.

Joyner, A. L., Lebo, R. V., Kan, Y. W., Tjian, R., Cox, D. R., and Martin, G. R. (1985). Comparative chromosome mapping of a conserved homeo box region in mouse and human. Nature *314*, 173–175.

Kaufman, T. C., Lewis, R., and Wakimoto, B. (1980). Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homoeotic gene complex in polytene chromosome interval 84A-B. Genetics *94*, 115–133.

Keynes, R. J., and Stern, C. D. (1984). Segmentation in the vertebrate nervous system. Nature *310*, 786–789.

Kornberg, T. (1981). *engrailed*: a gene controlling compartment and segment formation in Drosophila. Proc. Natl. Acad. Sci. USA *78*, 1095–1099.

Kornberg, T., Siden, I., O'Farrell, P., and Simon, M. (1985). The *engrailed* locus of *Drosophila*: in situ localization of transcripts reveals compartment-specific expression. Cell *40*, 45–53.

Lawrence, P. A., and Morata, G. (1976). Compartments in the wing of *Drosophila*: a study of the *engrailed* gene. Dev. Biol. *50*, 321–337.

Lawrence, P. A., and Struhl, G. (1982). Further studies of the *engrailed* phenotype in *Drosophila*. EMBO J. *1*, 827–833.

Levine, M., Rubin, G. M., and Tjian, R. (1984). Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila. Cell *38*, 667–673.

Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. Nature *276*, 565–570.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning:

A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Martin, G. R. (1980). Teratocarcinomas and mammalian embryogenesis. Science *209*, 768–776.

Martin, G. R., and Evans, M. J. (1975a). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies *in vitro*. Proc. Natl. Acad. Sci. USA *72*, 1441–1445.

Martin, G. R., and Evans, M. J. (1975b). Multiple differentiation of clonal teratocarcinoma stem cells following embryoid body formation in vitro. Cell *6*, 467–474.

Martin, G. R., Wiley, L. M., and Damjanov, I. (1977). The development of cystic embryoid bodies *in vitro* from clonal teratocarcinoma stem cells. Dev. Biol. *61*, 230–244.

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A., and Gehring, W. J. (1984a). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell *37*, 403–408.

McGinnis, W., Hart, C. P., Gehring, W. J., and Ruddle, F. H. (1984b). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. Cell *36*, 675–680.

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., and Gehring, W. J. (1984c). A conserved DNA sequence in homoeotic genes of the *Drosophila* Antennapedia and bithorax complexes. Nature *308*, 428–433.

Messing, J. (1983). New M13 vectors for cloning. Meth. Enzymol. *101*, 20–78.

Muller, M. M., Carrasco, A. E., and DeRobertis, E. M. (1984). A homeobox-containing gene expressed during oogenesis in Xenopus. Cell *39*, 157–162.

Nusslein-Volhard, C. (1979). Maternal effect mutations that alter the spatial coordinates of the embryo of *Drosophila melanogaster*. In Determinants of Spatial organization, S. Subtelny and I. R. Konigsberg, eds. (New York: Academic Press), pp. 185–211.

Nusslein-Volhard, C., and Weischaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. Nature *287*, 795–801.

O'Tousa, J. E., Baehr, W., Martin, R. L., Hirsh, J., Pak, W. L., and Applebury, M. L. (1985). The Drosophila ninaE gene encodes an opsin. Cell *40*, 839–850.

Ouweneel, W. J. (1976). Developmental genetics of homeosis. Adv. Genet. *18*, 179–248.

Poole, S. J., Kauvar, L. M., Drees, B., and Kornberg, T. (1985). The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell *40*, 37–43.

Rabin, M., Hart, C. P., Ferguson-Smith, A., McGinnis, W., Levine, M., and Ruddle, F. H. (1985). Two homeo box loci mapped in evolutionarily related mouse and human chromosomes. Nature *314*, 175–178.

Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M., and McGinnis, W. (1985). Homeo box genes of the Antennapedia and bithorax complexes of Drosophila. Cell *43*, this issue.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. *143*, 161–178.

Scott, M. P., and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc. Natl. Acad. Sci. USA *81*, 4115–4119.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Zucker, C. S., Cowman, A. F., and Rubin, G. M. (1985). Isolation and structure of a rhodopsin gene from D. melanogaster. Cell *40*, 851–858.