Expression of *engrailed* Proteins in Arthropods, Annelids, and Chordates

Nipam H. Patel,† Enrique Martin-Blanco,† Kevin G. Coleman,† Stephen J. Poole,† Michael C. Ellis,† Thomas B. Kornberg,† and Corey S. Goodman*†

*Howard Hughes Medical Institute
†Department of Biochemistry and Biophysics
University of California
Berkeley, California 94720

$\text{tDepartment of Biochemistry and Biophysics}$
$\text{University of California}$
$\text{San Francisco, California 94143}$

Summary

*Engrailed* is a homeobox gene that has an important role in Drosophila segmentation. Genes homologous to *engrailed* have been identified in several other organisms. Here we describe a monoclonal antibody that recognizes a conserved epitope in the homeodomain of *engrailed* proteins of a number of different arthropods, annelids, and chordates; we use this antibody to isolate the grasshopper *engrailed* gene. In Drosophila embryos, the antibody reveals *engrailed* protein in the posterior portion of each segment during segmentation, and in a segmentally reiterated subset of neuronal cells during neurogenesis. Other arthropods, including grasshopper and two crustaceans, have similar patterns of *engrailed* expression. However, these patterns of expression are not shared by the annelids or chordates we examined. Our results provide the most comprehensive view that has been obtained of how expression patterns of a regulatory gene vary during evolution. On the basis of these patterns, we suggest that *engrailed* is a gene whose ancestral function was in neurogenesis and whose function was co-opted during the evolution of segmentation in the arthropods, but not in the annelids and chordates.

Introduction

Early pattern formation in Drosophila organizes the blastoderm into a series of repeated units and uniquely specifies each of these units. It is controlled by a hierarchy of maternal-effect, segmentation, and homeotic genes (reviewed by Ingham, 1988). One of these genes is *engrailed*. In Drosophila, *engrailed* is essential during several developmental phases, and it has characteristics common to both segmentation and homeotic genes. For instance, *engrailed* mutant embryos do not segment normally (Kornberg, 1981; Nasrallah and Wieschaus, 1980). Other aspects of the *engrailed* mutant phenotype are suggestive of homeotic function (Garcia-Blanco, 1975; Lawrence and Morata, 1976), of a role in organizing the preblastoderm (Karr et al., 1985), and, in imaginal posterior compartments, of a role in maintaining compartment and segment borders (Morata and Lawrence, 1975, Kornberg, 1981; Lawrence and Struhl, 1982). One further aspect of *engrailed* function is its expression during neurogenesis in a segmentally reiterated subset of neuroblasts and neurons (DiNardo et al., 1985; Brower, 1986). Thus, *engrailed*, like other segmentation genes (Doe et al., 1988a, 1988b), appears to play multiple roles during development: during organization and growth of epidermal derivatives and during neurogenesis.

During segmentation, the Drosophila *engrailed* protein is produced in the cells of the posterior region of each body segment (DiNardo et al., 1985; Karr et al., 1989). Consistent with its putative regulatory role, the *engrailed* protein contains a homeodomain (Fjose et al., 1985; Poole et al., 1985), localizes to nuclei (DiNardo et al., 1985; Karr et al., 1989), and has the capacity to bind to DNA in vitro (Desplan et al., 1988). The Drosophila *engrailed* homeobox sequence is distinctive, but it is not unique. The *inverted* gene, juxtaposed to *engrailed* on the Drosophila chromosome and expressed in an almost identical pattern during segmentation, has a homeobox sequence that is strikingly similar (52/61 amino acids) to that of *engrailed* (Coleman et al., 1987). In addition, *inverted* has regions closely related to *engrailed* upstream (23/26 identical amino acids) and downstream (26/30 amino acids) of its homeobox, and in a region near the amino terminus (812 amino acids). Thus Drosophila has a second *engrailed* gene, *inverted*, the function of which is not known.

A number of *engrailed* genes have been identified in other organisms. For example, *honey bee* (D. W. H. and W. Gehring, personal communication), mouse (En-1 and En-2, Joyner et al., 1985; Joyner and Martin, 1987), chicken (Darnell et al., 1986), zebrafish (Fjose et al., 1988), and human (Poole et al., 1989; Logan et al., 1989) contain two *engrailed* genes, and single *engrailed* genes have been found in sea urchin (Dolecki and Humphreys, 1988), leech (Weisblat et al., 1988), grasshopper (this report), and nematode (A. Kamb and T. B. Kornberg, unpublished data). The homeodomain sequences of these genes are similar to Drosophila *engrailed* and, where sequence analysis extends sufficiently, to regions immediately upstream and downstream of the homeodomain as well. For the mouse *En-1* gene, most of the coding sequence has been obtained, and in addition to the conserved extended homeodomain region (69/99 of the residues are conserved), it contains the region near the amino terminus conserved in Drosophila *engrailed* and *inverted* (Freimann and Martin, personal communication). These two conserved regions, the small amino-terminal line and the extended homeodomain, are the only sequences conserved among these three genes. Conservation of the homeobox regions of the other vertebrate and invertebrate *engrailed* genes is comparable. Since the sequence of the *engrailed*-type homeodomains is distinctive and unlike any of the other homeodomain sequences known (Scott et al., 1989), and since it is in every known case surrounded by a con-
Figure 1. Expression of engravilled Proteins in Arthropods, Annelides, and Chordates
Staining of Drosophila (A), grasshopper (B), crayfish (C), leech (D), zebrafish (E, G, H), and chick (F, I) embryos with the 4D9 MAb and HRP immunocytochemistry. In all three arthropods (A, B, C), engrailed is expressed in the posterior region of each segment; in the other organisms examined (D-I), engrailed does not appear to play a role during metameric development (for example, arrowhead in F marks the developing somites in the chick.)
Evolution of engrailed Expression

957

served region, we consider this extended engrailed homeodomain region a signature of engrailed genes.

Proteins that share extensive sequence similarity have structural and functional identity in their conserved regions. Since the conserved portion of the engrailed proteins is in the putative DNA-binding domain, we assume that the biochemical function of these proteins as transcription regulators has been conserved. Given the strong correlation between patterns of engrailed expression and function in the Drosophila epidermis and the varied developmental programs of the different animals in which engrailed genes have been identified, it is of interest to determine if engrailed has a conserved pattern of expression indicating a common, and possibly ancestral, developmental function.

In the studies described here, a monoclonal antibody reagent (MAb 4D9) was found to recognize specifically engrailed proteins in a variety of species. Use of the 4D9 MAb has allowed us to show that engrailed proteins can be found in many organisms throughout the animal kingdom, including annelids, chordates, and arthropods (and other phyla not described here). In all arthropods examined, engrailed is expressed during segmentation in the posterior portion of each segment, and subsequently in a subset of neuronal cells during neurogenesis. Among the members of the three phyla described here, expression of engrailed during neurogenesis is the only conserved feature. We conclude that an ancestral function of engrailed may have been in neurogenesis, and its other functions in arthropods may represent a more recent addition, not shared by the annelids and chordates.

Results

Monoclonal Antibody 4D9

Lines of myeloma cells producing monoclonal antibodies directed against E. coli–derived engrailed and invected proteins were generated. Among these cell lines, two (4D9 and 4F11) recognized epitopes conserved between the invected and engrailed proteins. 4D9 and 4F11 antibodies bound to nuclei of Drosophila cell lines that express engrailed or invected. These cell lines were derived from a Schneider 2 (S2) cell line that was transfected with plasmids carrying either engrailed or invected cDNAs under HSP-70 promoter control (Gay et al., 1988). Both monoclonals are effective histological reagents that localize engrailed and invected proteins in Drosophila embryos (Figure 1A), and the patterns they generate are consistent with previous studies (DiNardo et al., 1985; Karr et al., 1989). Cross-reaction of 4D9 or 4F11 with other antigens in embryos is not detectable. The two MAbs also recognize both engrailed and invected proteins in Western blot assays (data not shown). Due to its ability to recognize engrailed in a wide variety of organisms (Figure 1), MAb 4D9 was characterized further.

To identify the epitope recognized by the 4D9 MAb, various portions of the engrailed protein were produced in E. coli, and these were used in Western blot assays. Whereas the complete engrailed protein reacted with MAb 4D9, the protein deleted of the homeodomain does not (Figure 2). The engrailed homeodomain, produced as a fusion protein linked to the N-terminal portion of TRP-LE (see Experimental Procedures) is recognized by the MAb. Portions of the engrailed homeodomain were produced in the same way, and reaction of these constructs with 4D9 indicates that the epitope includes the 14 amino acids, 35–48, of the homeodomain (Figure 2). This determination was confirmed with a synthetic peptide representing residues 36–50, which reacts directly with MAb 4D9 on dot blots and which blocks binding of MAb 4D9 to Drosophila embryos.

Eighty one homeobox containing genes have been isolated from a variety of organisms (Scott et al., 1989). Comparing the sequence of the peptide recognized by 4D9 with the homologous portion of these other genes reveals that this region is highly variable and that the sequence of engrailed residues 36–50 is unique. Even among the known engrailed genes, some of the residues in this region of the homeobox vary. Amino acid 36 (serine in Drosophila engrailed) is alanine in the murine, human, chicken, zebrafish, and sea urchin genes. Amino acid 37 (serine) is glycine in invected, glutamine in the murine, human, chicken, and zebrafish, and lysine in the sea urchin and leech genes. Amino acid 40 (glycine) is serine in the murine and human, threonine in the sea urchin, and asparagine in the leech gene. With the exception of invected, amino acid 44 (alanine) is serine in all of the other genes. Residues 46–48 are invariant among the engrailed-like genes, and are also highly conserved among other homeobox genes.

This comparative sequence information further limits the size of the MAb 4D9 epitope, since 4D9 recognized some but not all of the engrailed genes. For instance, when expressed in Drosophila S2 cells, a chicken engrailed protein reacts with 4D9, but the mouse En-1 protein does not. This result suggests that residues 36, 37, and 44 have little influence on antibody binding, since the murine and chicken genes have identical amino acids at these positions. This result also implicates residue 40 as critical to antibody binding, and suggests that serine at this position
A. engrailed Protein

The structure of the Drosophila engrailed protein is portrayed with the homebox and other regions of conservation indicated (A). MAb 4D9 binds to Western blots to the intact protein (+), but not to protein lacking the homeodomain (−). MAb 4D9 also binds to fusion proteins (B) and to synthetic peptides (D) containing residues 36–48 of the homeodomain, and to engrailed proteins in chicken (references denote source of sequence information for residues 36–48; Darnell et al., 1986), leech (Wedeen, Price, and Weisblat, personal communication), zebrafish (Fjöse et al., 1988), and grasshopper embryos (Poole et al., 1989) engrailed protein, or to protein in sea urchin embryos (Dolecki and Humphreys, 1988), in which residue 40, a glycine, is instead serine or threonine. Nor does it bind to a synthetic peptide (D), in which residue 40, a glycine, is instead serine.

Figure 2. Localization of the MAb 4D9 Epitope

The 4D9 staining pattern in Drosophila is consistent with previous data: it is nuclear and includes only those cells known to express engrailed. Second, MAb 4D9 recognizes the protein product of the Drosophila engrailed and invected genes and a chicken engrailed gene when expressed in E. coli. Third, the MAb 4D9 epitope is in a variable region of the homeodomain that is unique to the known engrailed genes. We have obtained no evidence that it cross-reacts with homeodomain proteins other than those with sequence homology to engrailed.

MAb 4D9 Recognizes the Grasshopper engrailed Protein

We used MAb 4D9 to screen a grasshopper (Schistocerca americana) embryo 1g11 cDNA library. Five antibody-positive plaques were isolated from a screen of 4.0 x 10⁶ recombinants. All five phages contain inserts that hybridized at moderate stringency to a Drosophila engrailed DNA fragment.
homeobox probe. The largest cDNA was 2.6 kb and consisted of two EcoRI fragments of 1.1 and 1.5 kb. Hybridization experiments using a Drosophila engrailed homeobox probe indicated that an engrailed-type homeobox was present in the 1.1 kb fragment. To confirm the identity of the cDNA, we carried out tissue in situ hybridization analysis and sequenced the homeobox region of the cDNA.

Both strands of both EcoRI fragments were used to generate RNA probes. Only one probe from each fragment hybridized to RNA in sections of grasshopper embryos, and the patterns observed for each were identical. In sections of grasshopper embryos at 35% of embryonic development (Bentley et al., 1979), striped patterns of hybridization were localized to the ectodermal cells in the posterior portion of each segment (Figure 3). The grasshopper cDNA also reveals a pattern of RNA distribution in the nervous system that matches the protein distribution revealed by MAb 4D9. Unfortunately, the in situ technique cannot resolve patterns at the level of single neurons, as is possible with immunohistochemistry (see below). However, the patterns obtained with the two techniques are in general agreement. Both antibody and in situ analyses indicate that the majority of neuronal expression is underneath the region of ectodermal expression. In addition, there is a distinct cluster of neurons positioned in the anterior part of the neuromere that is observed with both methods (Figures 3B and 3D). Thus, the grasshopper cDNA reveals a pattern of RNA distribution that closely correlates with the protein distribution as seen with MAb 4D9.

A portion of the 1.1 kb fragment was sequenced, and as suggested by its hybridization properties, this fragment contains a homeobox (Figure 4A) and encodes an engrailed-like homeodomain (Figure 4B). Several important features are revealed by comparing this sequence to the Drosophila engrailed and invected genes. Foremost, the amino acid sequence ELGLNEAQIKI, previously identified as the residues that make up the epitope recognized by the 4D9 MAb (see above), is present within the grasshopper homeodomain. Additionally, the nucleotide sequence of this grasshopper homeobox is much closer to that of the Drosophila engrailed and invected genes than to the homeobox sequence from any other known Drosophila gene. Nevertheless, the grasshopper sequence (G-en) cannot unambiguously be determined to be closer to either the Drosophila engrailed (D-en) or the invected (D-inv) sequences. Within the homeodomain, D-en and D-inv differ at nine of 61 positions; G-en also differs by nine residues from either D-en or D-inv. Comparing all three sequences, 13 positions within the homeodomains show discordance. At one or these locations, all three genes contain a different amino acid. At four positions, D-en and D-inv share the same amino acid sequence; at four other positions, G-en and D-en share the same residues; and at the remaining four residues, G-en and D-inv are identical. Thus these three sequences appear to be equidistant from one another.

The relationship of G-en to the Drosophila genes is not resolved by examining the 75 bp of G-en sequence upstream of that shown in Figure 4, since no way was found to align it with the D-en or D-inv sequences at either the nucleotide or amino acid level. At the beginning of the alignment that was found, a stretch of 17 amino acids (upstream homology region) is conserved between G-en, D-en, D-inv, and the two mouse genes, En-1 and En-2 (Joyner and Martin, 1987). This is followed by the amino acids Arg-Ser in both G-en and D-inv, which, in the D-inv gene, are known to be introduced by a six-nucleotide microexon (Coleman et al., 1987). Thus we would suspect the existence of the same microexon in the G-en gene, but do not know if alternative transcripts exist without this predicted microexon. The sequence Pro-Arg-X-Arg-X-X-Lys follows in all three insect genes and in both mouse genes. After these amino acids, there is a stretch of 5–7 amino acids before the beginning of the homeodomain, except in D-inv, in which there are 28.

To examine more rigorously the relatedness of these three sequences, we employed the method of evolutionary parsimony analysis (Lake, 1987). We applied the analysis to the G-en sequence shown in Figure 2A and to the Drosophila engrailed and invected sequences (Poole et al., 1985; Coleman et al., 1987), disregarded regions where gaps were inserted for alignment, and used the mouse sequences (Joyner and Martin, 1987) as the out groups (Lake, 1987). This analysis (see Experimental Procedures) confirms that G-en is equidistant from both of the Drosophila genes.

We have also used the 1.1 kb homeobox-containing fragment of the G-en cDNA to probe genomic Southern blots (Figure 5). Under standard conditions of stringent stringency, this probe is able to detect both engrailed and invected in Drosophila genomic digests. However, only a single band of hybridization in grasshopper genomic digests is observed even under conditions of reduced stringency, suggesting that this is the sole grasshopper engrailed gene.

Expression of engrailed in Arthropods

Grasshopper (Schistocerca americana)

We examined grasshopper embryos ranging from the beginning of gastrulation to organogenesis stages (13% to 60% of development [Bentley et al., 1979]). Patterns of engrailed protein expression are, in most respects, remarkably similar to those of Drosophila embryos. Expression appears shortly after the start of gastrulation in a series of stripes in the developing thorax and then extends both anteriorly into the head and posteriorly into the abdomen; these stripes are in the posterior part of every segment (Figure 1B). During neurogenesis, a subset of neuroblasts, ganglion mother cells, and neurons also expresses engrailed protein (Figure 6E). Both antibody and in situ analyses indicate that the majority of neuronal expression is underneath the region of ectodermal expression. In addition, there is a distinct cluster of neurons positioned in the anterior part of the neuromere that is observed with both methods (Figures 1B and 1D). Thus, the grasshopper cDNA reveals a pattern of RNA distribution that closely correlates with the protein distribution as seen with MAb 4D9. Expression in the ectodermal stripes becomes weak-
er after about 35% of development, but is still detectable at all stages examined. The principal difference between the patterns of en- 
grailed expression in Drosophila and in grasshopper is the order in which the segmental stripes appear. In Drosoph- 
ilia, the stripes appear in an order characterized by a rapid anterior-to-posterior gradient in which even-numbered 
stripes appear slightly before odd-numbered stripes (Fig-
ure 6A). In grasshopper, segments are added with a 
rostral-caudal polarity and the stripes of en- 
grailed expression follow, forming one at a time. In grasshopper, no transient pair-rule patterns appear (compare Figures 6A 
and 6B). (For further details, see Patel et al., 1989.)

Crayfish and Lobster (Procambarus clarki and 
Hymenoptera americana)

Crayfish embryos were examined from the time of early 
budding through the period of neurogenesis (equivalent to 25% through 45% of grasshopper development). As in grasshopper, the abdominal stripes of en- 
grailed protein are added one at a time (Figure 6C) and 
are clearly localized to the posterior region of each segment in the abdomen (A) and extends along the posterior margin of each leg (C). During neurogenesis, en- 
grailed expression is confined to the progeny of one of 
the teloblast lineages. Not all of the progeny of this partic-
ular teloblast are en- 
grailed-positive, but rather only a small 
group of progeny in a continuous band appear transiently 
to express en- 
grailed protein. At late stage 8, by which time 
the two germinal bands have come together, expression 
occurs in two rows of bilaterally symmetric cells; these rows are positioned at the dorsal edge of the germinal plate. There 
is one 4D9-positive cell per hemisegment, and there are 
about 12 segments containing these stained cells. At this 
stage there is also a series of large, flat nuclei positioned 
around the yolk sac that also stain with the 4D9 MAb.

Expression of en- 
grailed in Annelids

Leech (Helobdella triserialis)

We first observe staining with the 4D9 MAb in leech emb-
yos during the phase of teloblast divisions (between late 
stage 7 and early stage 8, Stent et al., 1982). At this stage, en- 
grailed expression is confined to the progeny of one of 
the teloblast lineages. Not all of the progeny of this partic-
ular teloblast are en- 
grailed-positive, but rather only a small 
group of progeny in a continuous band appear transiently 
to express en- 
grailed protein. At late stage 8, by which time 
the two germinal bands have come together, expression 
occurs in two rows of bilaterally symmetric cells; these rows are positioned at the dorsal edge of the germinal plate. There 
is one 4D9-positive cell per hemisegment, and there are 
about 12 segments containing these stained cells. At this 
stage there is also a series of large, flat nuclei positioned 
around the yolk sac that also stain with the 4D9 MAb.

Nervous system expression of en- 
grailed in the locust be-

Figure 3. Tissue In Situ Hybridization Analysis of the Grasshopper engrailed Gene

Double exposure bright-field and dark-field photographs of frontal (A, B, C) and parasagittal (D) sections of 35% grasshopper embryos. Anterior 
is up in all panels. Sections were hybridized with an RNA probe from the 1.5 kb non-homeobox-containing EcoRI fragment of the G-en cDNA. Position 
of the second thoracic leg is indicated in (B) and (C). Hybridization (revealed by accumulation of silver grains that appear red in these photographs) 
is clearly localized to the posterior region of each segment in the abdomen (A) and extends along the posterior margin of each leg (C). Hybridization 
is also seen in the posterior portion of each neuromere (B, D), and a cluster of neurons in the anterior region also contains the transcript (arrows 
in A and D). These patterns closely match those seen with the 4D9 MAb. Scale bars (A, B) 136 μm; (C) 110 μm; (D) 40 μm.
stage 9, most, if not all, nuclei in the second suboesophageal ganglion stain with MAb 4D9 and, in addition, there are also a number of positive nuclei in the anterior region of the third suboesophageal ganglion and four stained nuclei in the fourth suboesophageal ganglion (Figure 1D).

There are also strongly stained nuclei extending ventrally from the second suboesophageal ganglion to the ectoderm and occasionally a 4D9-positive nucleus in the ectodermal layer. At this stage, staining at the dorsal edges of the germinal plate forms a series of arcs of weakly positive cells extending out dorsally.

Oligochaeta (Eisenia fetida)

We examined oligochaete embryos during the period of neurogenesis; this period corresponds to stage 9 of leech development. In this annelid, however, the pattern of MAb 4D9 staining is quite different from that observed in the leech. Expression of en grailed occurs in a segmentally reiterated pattern of neuronal nuclei in the developing central nervous system (data not shown). This is in marked contrast to the regionalized pattern observed in the central nervous system of leech embryos at the same stage.

Expression of en grailed in Chordates

Zebrafish (Brachydanio rerio)

Zebrafish embryos were examined from a period before gastrulation until formation of the most posterior somites. As in the other chordates (see below), expression was first observed when the neural tube forms (Figure 1E) in the posterior mesencephalon/anterior metencephalon (Figure 1H). As development proceeds, a deep indentation separates these two regions of the central nervous system, and expression clearly extends across this morphological boundary.

A short time after somitogenesis begins, additional staining was observed in three to four nuclei in each somite (Figure 1G). These nuclei are positioned in the most ventral and anterior parts of the somite and close to the spinal cord. No expression in the somites is present prior to or during the time that the somites become delineated. Rather, as the wave of somitogenesis proceeds posteriorly, staining lags about four to five somites behind the most newly formed somite. It appears that these stained nuclei correspond to a particular set of muscle cells (C. Kimmel, personal communication). A second wave of additional somite staining follows this original somite expression, and lags behind the first by about 10 somites. These additional 10–15 nuclei are stained less intensely than the original 3–4 nuclei and are loosely packed around the developing spinal cord. These nuclei are probably a subset of the sclerotome. The nervous system staining in the posterior mesencephalon/anterior metencephalon is maintained through the oldest embryos examined.

Frog (Xenopus laevis)

Xenopus embryos were examined from the early gastrula period (stage 10; Nieuwkoop and Faber, 1956) to near the end of somitogenesis (stage 30). At stage 20, immediately following fusion of the neural folds, MAb 4D9 staining was observed in three to four nuclei in each somite (Figure 1H). As development proceeds, a deep indentation separates these two regions of the central nervous system (data not shown). This is in marked contrast to the regionalized pattern observed in the zebrafish, this region corresponds to the posterior mesencephalon/anterior metencephalon. As more nuclei begin to express en grailed protein, staining also extends ventrally around the neural tube. At stage 25 we noted a few positive nuclei far anterior in the region of the prosencephalon. Staining in the posterior mesencephalon/anterior metencephalon persisted through the oldest stages examined, and no expression was observed in the somites or spinal cord.

Chick (Gallus domesticus)

Chick embryos were examined from stage 7 (one somite; Hamburger and Hamilton, 1951) to stage 16 (25 somites). We did not observe any MAB 4D9 staining until approximately stage 9. At this time, just prior to the fusion of

Figure 5. Grasshopper Has Only One en grailed Gene

Genomic southern blots of Drosophila (lanes 1 and 2) and grasshopper (lanes 3, 4, and 5) DNA digested with EcoRI (lanes 1, 2, 4, and 5) or HindIII (lane 3) and hybridized with a Drosophila en grailed homeobox probe (lane 1) or a grasshopper en grailed homeobox probe (lanes 2–5) at high (lanes 1–4) or reduced (lane 5) stringency. In lane 1, a Drosophila en grailed probe reveals EcoRI bands at 4.0 and 1.1 kb that correspond to the invected and en grailed genes, respectively. The grasshopper en grailed homeobox probe hybridizes to these same two Drosophila genes even under high-stringency conditions (lane 2). When grasshopper genomic DNA is probed with the grasshopper en grailed homeobox probe under high-stringency conditions, hybridization is seen to a single 4.8 kb band in an EcoRI digest (lane 4) and a single 5.5 kb band in a SalI digest (not shown). No additional bands are seen in any of the digests under reduced-stringency conditions (lane 5). As Experimental Procedures are sufficiently low to allow one Antennapedia class homeobox to detect other Antennapedia class homeoboxes (McGinnis et al., 1984). Arrowheads mark positions of various DNA size standards (given in kb).
Evolution of engrailed Expression

Figure 6. Expression of engrailed Proteins during Arthropod Segmentation and Neurogenesis

Expression of engrailed proteins during segmentation in Drosophila (A), grasshopper (B), and crayfish (C), and during neurogenesis in Drosophila (D), grasshopper (E), and crayfish (F), as revealed by the 4D9 MAb and HRP immunocytochemistry. (A) This cellular blastoderm has virtually completed the formation of segmentally reiterated stripes, although the formation of odd-numbered stripes lags behind the even-numbered stripes (arrowheads). (B, C) The grasshopper and crayfish embryos, however, generate engrailed stripes one at a time (arrowheads) as their germ bands extend caudally. In grasshopper and crayfish, the abdominal stripes form without apparent pair-rule patterning. (E, F) During the period of neurogenesis, a subset of neuroblasts, ganglion mother cells, and neurons expressed engrailed protein; this pattern appears quite similar in Drosophila (A), grasshopper (B), and crayfish (C). Calibration bar: (A) 90 μm; (B) 200 μm; (C) 110 μm; (D) 40 μm; (E) 50 μm; (F) 40 μm.

In the neural folds, engrailed expression appears in a scattering of nuclei on the dorsal surface of a rostral region of the neural tube. As development progresses, increasing numbers of nuclei in this region stain, and by stage 11, it is clear that the region of expression is the posterior mesencephalon and anterior metencephalon (Figure 1F). MAb 4D9 staining begins dorsally, but moves ventrally at about the time of cranial flexure. The borders of the MAb 4D9 staining are well defined and clearly extend to both sides of the mesencephalon/metencephalon invagination.
grailed homeodomains. It stains Drosophila embryos in a staining pattern in grasshopper and other arthropods and a chicken engrailed gene when expressed in E. coli. The product of the Drosophila engrailed and invected genes, therefore generally similar to both zebrafish and Xenopus although, in contrast to zebrafish, no staining was detected in the chick somites.

Discussion

Analysis of pattern formation in the Drosophila embryo has led to the discovery of a hierarchy of genes that controls the generation and specification of the body segments. Two aspects of these studies are relevant here. First, the regions in which these segmentation genes are expressed in the embryo correlate in most cases with the areas in which their functions are required. Therefore, in Drosophila, it is clear that their patterns of expression are intimately related to their functions. Second, many of these genes encode proteins that have homeodomain sequences, and similar homeobox genes have been isolated in vertebrates. The patterns of expression of some of these vertebrate proteins have been described, and the presence of a homeodomain suggests a function in transcriptional regulation. For many of these proteins, however, their developmental roles and their relationships to the Drosophila homeodomain proteins are unclear. Given the extraordinary sequence conservation among these regulatory proteins, it is important to determine how their functions and roles are related, and to understand how they evolved. Fortunately, the expression and evolution of one of these segmentation genes, engrailed, can be readily studied.

The MAb 4D9 binds to a conserved region in the homeodomain of the engrailed protein and recognizes this epitope in animals from many phyla (Figure 1). These include Drosophila, a more primitive insect (grasshopper), and two crustaceans (crayfish and lobster), annelids from two different groups (a leech and an oligochaete), and arthropods are thought to have diverged over 600 million years ago, yet engrailed genes are present in both phyla and their protein products are recognized by MAb 4D9.

Evidence that MAb 4D9 recognizes engrailed proteins in these organisms is strong. Its epitope is unique to engrailed homedodomains. It stains Drosophila embryos in a manner consistent with the known patterns of engrailed expression, indicating that it binds selectively to Drosophila engrailed proteins. Furthermore, it has been used to isolate a grasshopper engrailed gene and, moreover, its staining patterns in grasshopper and other arthropods are similar to Drosophila. The antibody recognizes the protein product of the Drosophila engrailed and invected genes and a chicken engrailed gene when expressed in E. coli. MAb 4D9 also stains chicken, zebrafish, and frog with patterns that have significantly greater resolution than, but are consistent with, patterns of expression previously detected by in situ hybridization of zebrafish (Njølstad and Fjøs, 1989), chicken (Gardner et al., 1988), and mouse (Davis et al., 1988; Davidson et al., 1988; Davis and Joyner, 1988). As noted in the Introduction, conservation of a sequence of almost 100 amino acid residues in and around the homeodomain of engrailed genes is an indication of the functional relatedness of their protein products. This extended sequence similarity justifies their common designation as engrailed genes and suggests that they are true evolutionary homologs. MAb 4D9 is therefore a probe that can be used specifically to detect engrailed proteins in distant related animals.

While the evolutionary conservation of the engrailed homeodomain indicates a conserved biochemical function as a transcriptional regulator, the patterns of expression of engrailed protein in different animals suggest that engrailed has several developmental roles, only some of which have been conserved. Within the arthropods we examined the pattern is consistent: engrailed is expressed in the posterior portion of each metamere during segmentation and in a segmentally reiterated subset of cells during neurogenesis. Interestingly, no other organism examined expresses engrailed in developing metameres. All do share one feature, namely, expression during neurogenesis. This has important implications for comparison of metameric development in different phylogenetic lineages and for determination of how, during their evolution, different developmental programs utilized regulatory genes such as engrailed.

In young leech embryos, transient patterns of expression in the progeny of a particular teloblast lineage are followed later by expression in the developing nervous system. Neuronal expression is regional (within only a few anterior segmental ganglia) and is not segmentally reiterated. In contrast, engrailed is expressed in the developing nervous system of an oligochaete in a segmentally reiterated subset of neurons. Therefore, these annelids feature engrailed expression in the developing nervous system in quite different patterns.

The three chordates examined in this study (zebrafish, Xenopus, and chick) have similar patterns: engrailed is expressed in a specific region of the developing neural tube (posterior midbrain and anterior hindbrain), but not in developing metameres. This pattern of expression is reminiscent of the leech patterns, and recalls the regional expression of homoeotic genes in the developing Drosophila nervous system (reviewed by Doe and Scott, 1988).

In zebrafish, engrailed is expressed in a region of the developing neural tube and, a short time after somitogenesis begins, in three to four nuclei in each somite. However, it is likely that engrailed is not involved in the process of segmentation in zebrafish, because the expression of engrailed in the somites occurs after the delineation of the somites. Thus, in addition to its role in neurogenesis, engrailed may also be involved in specifying the fate of a segmentally reiterated subset of mesodermal cells. When comparing the patterns of expression of different animals, it is relevant that many organisms have two engrailed genes, and MAb 4D9 may not recognize the products of both genes. For instance, MAb 4D9–positive mesodermal cells seen in zebrafish were not seen in Xenopus or chick.
and this may reflect recognition by MAb 4D9 for the products of only one of the two engrailed genes in chick and Xenopus and both engrailed genes in zebrafish.

Clearly, different phyla have different patterns of engrailed expression. If we can infer function from pattern of expression, then it is only in the arthropods that engrailed plays a major role during the process of segmentation. It is possible, of course, that the engrailed function was initially involved in chordate metamertization, but that this role was lost during evolution. However, the most parsimonious hypothesis to account for these staining patterns is that engrailed was never involved in chordate metamertization, and that an ancestral function of engrailed was to specify cell fate within the nervous system. Some present groups of organisms appear to use engrailed to specify a region of the developing nervous system (e.g., in the leech and in chordates), whereas other organisms appear to use engrailed to specify a reiterated subset of neurons within the developing nervous system (e.g., in an oligochaete and in arthropods).

The differences between the patterns of engrailed expression in the annelids and in arthropods lead to the conclusion that arthropods are not closely related to annelids. This argues against the popular notion that metamcerism in arthropods evolved from annelids (Anderson, 1973). Our data are consistent with the evolutionary analysis of the sequences of 18S rRNA (Field et al., 1989), which also suggested that arthropods and annelids are not closely related. Moreover, our data help to clarify further another evolutionary question: the relationship of crustaceans and insects. Previous work on the pattern of neuronal cells (Thomas et al., 1984), on embryology (Weygolt, 1979), and on comparative anatomy (Boudreaux, 1979) of crustaceans and insects suggested that these two groups of arthropods share a relatively recent common ancestor. The similar pattern of engrailed expression in these two classes is consistent with this hypothesis. This is in contrast to the arguments of others (Anderson, 1979, 1982; Manton, 1977; Schramm, 1986), who asserted that crustaceans and insects represent unrelated groups sharing certain genes toward "arthropodization."

Our analysis of a grasshopper engrailed gene has interesting implications with respect to the evolution of the engrailed gene. Whereas Drosophila contains two engrailed genes, engrailed and invected (Coleman et al., 1987), only a single gene was found in grasshopper, and analysis of its sequence indicates that it is equally related to the two Drosophila genes. This arrangement may have been generated in either of two ways. First, the two Drosophila genes may have arisen by duplication some time in the insect lineage after the last ancestor common to both Drosophila and grasshopper. This hypothesis suggests that the two genes in mouse (En-1 and En-2; Joyner and Martin, 1987), chicken (Darnell et al., 1986), zebrafish (Fjose et al., 1988), and human (Poole et al., 1989; Logan et al., 1989) also arose as an independent duplication of an ancestral engrailed gene early in the chordate lineage. This is supported by the observation that the mouse genes are more similar to each other than either is to the Drosophila engrailed or invected genes. The same conclusion has been suggested by Dolecki and Humphreys (1988), after finding a lone engrailed gene in sea urchin. The second possibility is that the duplication occurred before the split of the lines leading to Drosophila and grasshopper, and then one of the genes was later lost in the grasshopper lineage. The lone grasshopper gene may appear to be equidistant from the two Drosophila genes because of certain functional constraints on the G-on gene product.

Sequence analysis has also been used recently as a way of addressing some of the unresolved questions of phylogenetic relationships, and thus of supplementing previous morphological, embryological, and paleontological data (e.g., Field et al., 1988). The phylogenetic patterns of expression of regulatory genes such as engrailed present another way of using molecular data to examine phylogenetic relationships. Previous studies have examined slow changes in highly conserved, housekeeping molecules like 18S rRNA (Field et al., 1988). In contrast, the engrailed gene is a regulatory gene whose patterns of expression are conserved within groups of closely related organisms, and yet vary dramatically between more distantly related groups. The nature of these differences may reflect the evolutionary radiation of different phylogenetic lines.

In summary, our results suggest that engrailed is a gene that predates the divergence of arthropods, annelids, and chordates; had an ancestral function controlling cell fate during neurogenesis; and was co-opted for the process of segmentation during the evolution of the arthropods but not the annelids and chordates. Furthermore, the data support the hypothesis that metameric development evolved independently enough in these three different phylotactic lineages (annelid, arthropod, and chordate) that a regulatory gene that plays a crucial role during segmentation in one phylum (engrailed in arthropods) does not appear to play the same role in the other two.

**Experimental Procedures**

**Generation of Monoclonal Antibodies**

**Full-length engrailed protein and the C-terminal two-thirds of the invected protein were generated in E. coli with the T7 polymerase expression system (Studier and Moffett, 1986). BALB/c mice were immunized with either protein by intraperitoneal injections; each injection contained about 100 µg of either engrailed or invected protein. Primary injections (300 µl) consisted of protein suspended in 150 µl PBS and 100 µl complete adjuvant. Mice were given three boosts of 100 µg protein in incomplete adjuvant at approximately 2 week intervals. Three days before fusion, a final injection of 50 µg was administered. Spleen cells were fused with NS-1 myeloma cells (Kohler and Milstein, 1975; Oil and Herzenberg, 1980). Hybridoma supernatants were screened on Schneider 2 cell lines that express either engrailed or invected protein under HSP-70 promoter control (Gay et al., 1988), and hybridoma lines of interest were isolated and recloned by single-cell cloning. The MAb 4D9 was derived from a mouse injected with invected protein; the MAb 4F11 was derived from a mouse injected with engrailed protein. The 4D9-producing line has been deposited with the American Type Culture Collection and is available on request.

**Fusion Proteins**

The homeobox fusion protein clone (pMNIRIST) was produced from the plasmid pEH18 BIH, which contains DNA from the BamHI site at nucleotide 1405 to the EcoRI site at nucleotide 2017 of the clone C2.1 (Poole et al., 1985). The BamHI-FnuRI fragment was sequentially mutage
mixed in two positions by site-directed mutagenesis employing two synthetic primers: first, by introducing a UGA codon after the serine at position 512, and second, by introducing an in-frame EcoRI site (Glu Phe) and a Met before the homeodomain. This clone, pEMRIST, was verified by the dideoxy sequencing method using single-stranded DNA. From pEMRIST, a 450 bp EcoRI fragment was cloned into the expression plasmid pMNCV1 to generate pmNRIST and to engineer the production of a protein that includes 106 amino acids of E. coli TplP LE protein and 60 amino acids of the untagged homeodomain.

To produce a protein with only the terminal 26 amino acids of the homeodomain, a linker for EcoRI-Met (12-mer) was introduced at the Pvul site located at the middle of the homeobox (between amino acids 486 and 487) on the plasmid pEMRIST. An EcoRI band of ~330 bp was included residues from amino acid 487 (Leu) to the end of the homeodomain. Plasmids pRlBG and pBGST, which generate the amino-terminal 47 amino acids of the homeodomain (to amino acid 500 [ile]) and the carboxy-terminal 14 amino acids of the homeodomain (from amino acid 500 [ile]), respectively, were produced by introducing an EcoRI-Met linker (10-mer) at the BgIII site located at amino acid 500 in pEMRIST. The EcoRI bands, of approximately 150 and 300 bp, were cloned into pMNCV1. The protein produced by pRlBG includes a tail of five carboxy-terminal amino acids (Met Asn Ser His Val STOP).

Fusion proteins were induced by adding 3-iodoacyclic acid (20 μg/ml) to a mid-log culture at OD 600 = 0.7, and expression proceeded for 4-60 min. Drosophila embryos were fixed for 45 min in a 1:1 mixture of heptane and PEM-FA (0.1 M PIPES [pH 6.95], 2.0 mM EGTA, 1.0 mM MgSO4, 1.0% Triton X-100). Chicken, crayfish, and grasshopper embryos were dissected in PBS and PEM-FA and fixed for 45-60 min. After fixation, chick, crayfish, grasshopper, Xenopus, leech, and oligochaete embryos were washed three times in PBS for a total of 30 min and then transferred to PBT (1x PBS, 0.2% BSA, 0.1% Triton X-100).

Histochemistry

All embryos were stained and viewed as whole-mount preparations. Drosophila embryos were fixed for 45 min in a 1:1 mixture of heptane and PEM-FA (0.1 M PIPES [pH 6.95], 2.0 mM EGTA, 1.0 mM MgSO4, 3.7% formaldehyde). The aqueous layer was removed and replaced with an equal volume of methanol to devetilinize the embryos (Mitchison and Sedat, 1983). The embryos were washed three times with methanol and then transferred to PBT (1x PBS, 0.2% BSA, 0.1% Triton X-100).

Chicken, crayfish, and grasshopper embryos were dissected in PBS and then transferred to the PEM-FA fixation for 45-60 min. Zebrafish, Xenopus, leech, and oligochaete embryos were dissected directly in PEM-FA and fixed for 45-60 min. After fixation, chick, crayfish, grasshopper, Xenopus, leech, and oligochaete embryos were washed three times in PBS for a total of 30 min and then transferred to PBT.

At this point, the protocol is identical for all organisms. Embryos were washed twice for 30 min each in PBT and then incubated in PBT plus N (PBT plus 5% normal goat serum, Gibco) for 30 min. An equal volume of Mab 4D4 supernatant was added to the PBT plus N and the embryos were incubated overnight at 4°C. Embryos were washed three times for 5 min and four times for 30 min in PBT. Embryos were again incubated in PBT plus N for 30 min and then goat antiserum (Jackson Immunoresearch Labs) was used at a concentration of 1:200.

Tissue In Situ Hybridization

In situ analysis was done using 35S RNA probes and the protocols established for Drosophila tissue by Akam (1983) and Hafen and Levine (1986). Sequencing

Sequencing was done from Bluescript (Stratagene) subclones using the dideoxy method (Sanger et al., 1977).

Evolutionary Parsimony Analysis

The analysis of the phylogenetic relationships among the five en- grated class genes (mouse En+ and En-2, Drosophila engrailed and invected, and grasshopper engrailed) was performed on the alignable parts of the sequences (273 nucleotides), using evolutionary par- simony (Lake, 1987). We chose this method because (first), it is based on inherent vector algebraic properties of nucleotide sequence evolu- tion; (second), it yields the correct phylogenetic tree under circum- stances where parsimony or distance matrix analysis may fail; and (third), it offers a robust statistical test of the results.

<table>
<thead>
<tr>
<th>En-1 A En-2 B</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-invariant Chi square P&gt; Y1, Z-invariants</td>
</tr>
<tr>
<td>D-inv, D-en 18 12.5 0.005 not significant</td>
</tr>
<tr>
<td>D-inv, G-en 10 4.5 0.05 not significant</td>
</tr>
<tr>
<td>D-en, G-en 14 9.8 0.005 not significant</td>
</tr>
</tbody>
</table>

Acknowledgments

For supplying animals, we thank the members of the following labs: L. Reichardt (chick), J. Gerhart and R. I. Horvitz (Xenopus), J. Kuehwa (zebrafish), J. Weiss (moth), J. Wine (crayfish), A. Hertz (lobster), D. Bentley (grasshopper), and D. Weisblat (leech and oligochaete). We therefore wish to thank these for the pMNCV1 expression plasmid: V. Paragas, Z. Traquina, K. Jepson-Innes, G. Ghiselli, and P. Pazinino for

were cleaned in glycerol. Photographe were taken using Nomarski optics on either a Zeiss Axioshot or a Nikon compound microscope.
Evolution of engrailed Expression

967

technical assistance; A. Sidow for performing the evolutionary par-
simony analysis; C. Wedeen, D. Price, and A. Hemmati Brivanlou for
help with dissections and interpretations of staining patterns; S. Smiley, A. Kolodkin, C. Kimm, A. Joyner, and G. Lehner for helpful
discussions; S. Smiley and K. Sander for thoughtful suggestions on
the manuscript; and P. Nilssand, A. Fjose, C. Wedeen, D. Price, and D.
Weeks for communication of results prior to publication. Supported by
postdoctoral fellowships from the C. S. I. C. (to E. M. B.) and the Weingart Foundation (to K. C.), by a predoctoral fellowship from
the National Science Foundation (to M. C. E.), and by grants from the na-
tional Institutes of Health (to T. B. K. and C. S. G.) and the Howard
Hughes Medical Institute (to C. S. G.).

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 March 10, 1989; revised June 27, 1989.

References

Akam, M. (1983). The localization of Ultrabithorax transcripts in Dro-
sophila tissue sections. EMBO J. 2, 2075–2084.


thropods. In Arthropod Phylogeny (San Francisco: Van Nostrand
Reinhold).


Bentley, D., Keshishian, H., Shankland, M., and Torian-Raymond, A.
(1979). Quantitative staging of embryonic development of the grass-

Boucheaux, H. B. (1979). Arthropod Phylogeny: With Special Refer-
ence to Insects (New York: Wiley Interscience).

discs. EMBO J. 5, 2649–2656.


Coleman, K. G., Poole, S. J., Weir, M. J., Scoffier, W. C., and Kornberg,
T. B. (1987). The inverted gene of Drosophila: sequence analysis and
expression studies reveal a close kinship to the engrailed gene. Genes
Dev. 1, 13–28.

homeobox clone with homology to the Drosophila engrailed homeo box.

sequence similarity to Drosophila engrailed is expressed during
the development of the neural tube and vertebral in the mouse. Develop-
ment 104, 305–316.


Expression of the homeobox-containing gene En-2 delineates a spe-

Desplan, C., Theis, J., and O’Farrell, P. H. (1988). The sequence speci-

D’Nardo, S., Kumer, J. M., Theis, J., and O’Farrell, P. H. (1986). Develop-
ment of embryonic pattern in D. melanogaster as revealed by accumu-

in the developing nervous system of Drosophila. Trends Neurosci. 11,
101–106.

Expression and function of the segmentation gene fushi tarazu during

ronal fate by the Drosophila segmentation gene evan-skipped. Nature
333, 376–378.


DNA restriction endonuclease fragments to high specific activity. Anal.
Biochem. 132, 6–13.

Field, K. G., Olsen, G. J., Lane, D. J., Giovannoni, S. J., Ghiselin,

box-containing gene from the engrailed region of Drosophila and the

Fjose, A., Eiken, H. G., Nilssand, P. R., Molven, A., and Nordvik, I.
(1988). A zebrafish engrailed-like homeobox sequence expressed dur-

Garcia-Bellido, A. (1976). Genetic control of wing disc development

Gardner, C. A., Darnell, D. K., Poole, S. J., Ordahl, C. P., and Barald,
of the early embryonic chick nervous system. J. Neurosci. Res. 21,
429–437.

grailed protein is phosphorylated by a serine-specific kinase. Nucl.
Acids Res. 16, 6637–6647.

Hafen, E., and Levine, M. (1986). The localization of RNAs in Drosoph-
lla tissue sections by in situ hybridization. In Drosophila: A Practical
49–78.


sequence homology to the Drosophila engrailed gene: expression during

Joyner, A. J., Kornberg, T., Coleman, K. G., Cox, N. R., and Martin,

cus of D. melanogaster provides an essential zinc finger function in precur-

grailed protein in early Drosophila embryos. Development 105, 605–
612.


segment formation in Drosophila. Proc. Natl. Acad. Sci. USA 78,
1095–1099.


tation gene: protein structure homology with DNA-binding proteins.

Lawrence, P., and Muroya, O. (1976). Comparisons in the wing of Dro-

phenotype of Drosophila. EMBO J. 1, 927–933.

Chromosomal localization of the human homolog of -coating genes

Cell

968


