

## Genomic organization and developmental pattern of expression of the *engrailed* gene from the brine shrimp *Artemia*

Miguel Manzanares, Roberto Marco and Rafael Garesse\*

Departamento de Bioquímica UAM, Instituto de Investigaciones Biomédicas CSIC, Facultad de Medicina, Universidad Autónoma de Madrid, C/ Arzobispo Morcillo 4, Madrid 28029, Spain

\*Author for correspondence

### SUMMARY

We report the isolation and characterization of an *engrailed* gene in the crustacean *Artemia franciscana*. The *Artemia* gene spans a genomic region of 15 kilobases and the coding sequence is interrupted by two introns. It appears to be the only gene of the *engrailed* family present in the *Artemia* genome. The predicted *engrailed*-like protein is 349 amino acids long and contains several domains including the homeodomain, well conserved when compared to other proteins of the *engrailed* family. Based on sequence comparisons we have detected, in the *Artemia engrailed* protein, several features which are in common with the *Drosophila* and *Bombyx engrailed* proteins. It also has some features specific for *invected* proteins. Therefore, this gene appears to have diverged from an ancestral gene common to both the *engrailed* and *invected* insect genes. Whole-mount *in situ* hybridization experiments show that the expression of this gene in postembryonic development of *Artemia* is restricted to

the posterior part of at least the thoracic and maxillar segments. The pattern is generated sequentially from a growth zone organized in columns of cells close to the caudal region of the larvae. Cell proliferation in the growth zone follows an interspersed pattern without evidence of early lineage restrictions. The *engrailed* expression is detected in the growth zone before any segmentation is visible and continues to be expressed in a posterior location in the segments that are morphologically defined. Initially expressed in isolated cells, it spreads into rows broadening to two-three cells as segments mature. The evidence presented here is compatible with the hypothesis that intercellular signaling mechanisms are in part responsible of the early activation of selector genes.

Key words: *Artemia*, Crustacea, *engrailed*, homeobox, segmentation

### INTRODUCTION

Segmentation is a general characteristic of the body plan of vertebrates and many invertebrate species. The process has been extensively studied in *Drosophila* using a combination of genetic and molecular approaches that have allowed the identification and characterization of a large group of genes directly involved in pattern formation (reviewed in Akam, 1987; Ingham, 1988). Initially, the anteroposterior and dorsoventral axes are established through the action of maternal products asymmetrically localized in the egg (St. Johnson and Nüsslein-Volhard, 1992). Their function is to define large and specific regions in the embryo in which a cascade of hierarchically organized zygotic genes are induced, becoming progressively involved in the definition of a more refined pattern of subdivisions. The process of early specification of segments is very rapid in *Drosophila*, where most of the metameric organization, although not morphologically visible until late in gastrulation, is already defined during the blastoderm stage, as indicated by the striped expression of segmentation genes.

The majority of the segmentation genes identified in

*Drosophila* encode transcription factors in accordance with their implied regulatory function. Their gene products contain domains acting as sequence-specific DNA binding sites that have been conserved during evolution. One of these motifs, the homeodomain, is a 183 bp sequence encoding an helix-turn-helix domain present in a multitude of genes involved in pattern formation (Gehring, 1987; Scott et al., 1989). *engrailed*, a segment polarity gene (Akam, 1987), is involved in the specification of posterior compartments in the ectodermal layer of the different segments. It contains a homeobox of a specific subclass, also present in *invected*, a closely linked gene of similar expression pattern and unknown function (Coleman et al., 1987; Poole et al., 1985). By means of their sequence similarity to the *engrailed* gene, two or more homeobox containing genes of the *engrailed* subclass have been identified in several organisms, ranging from closely related insects, *Bombyx mori* (Hui et al., 1992) and *Apis mellifera* (Walldorf et al., 1989), to vertebrates, *Xenopus laevis* (perhaps 4 genes; Hemmati-Brivanlou et al., 1991; Holland and Williams, 1990), zebrafish (3 genes, Ekker et al., 1992; Fjose et al., 1992; Holland and Williams, 1990), hagfish (Holland and

Williams, 1990), chicken (Logan et al., 1992), mouse (Joyner et al., 1985; Joyner and Martin, 1987; Logan et al., 1992) and man (Logan et al., 1992). On the other hand, in grasshopper (Patel et al., 1989a), leech (Wedeen et al., 1991), sea urchin (Dolecki and Humphreys, 1988), brachiopods (Holland et al., 1991) and perhaps lamprey (Holland and Williams, 1990) only one *engrailed*-like gene seems to be present.

The *engrailed* gene has a dual function in insects, as deduced from the expression pattern and mutational analysis in *Drosophila*. Early in embryogenesis it is expressed in the posterior compartment of each segment defining its anterior limit (DiNardo et al., 1985; Kornberg, 1981; Lawrence and Struhl, 1982; Morata and Lawrence, 1975). Later on, it is also expressed in specific subsets of neuroblasts and neurons in each segment, where it is thought to play an important role in neural specification (Brower, 1986; DiNardo et al., 1985). The pattern of expression of the *engrailed* protein has been examined in several organisms representing different phyla (Davis et al., 1991; Hemmati-Brivanlou and Harland, 1989; Patel et al., 1989a, 1989b). An early segmentally repeated pattern is found only in arthropods and annelids (Wedeen and Weisblat, 1991). Vertebrates also exhibit segmentally iterated expression in some tissues (Davis et al., 1991; Patel et al., 1989a), but it appears long after morphological segmentation has been established, so *engrailed* is not thought to play a role in vertebrate segmentation. On the other hand, expression in the developing nervous system has been found in all phyla examined. These data suggest that the primitive function of the *engrailed* gene was in neural specification or determination, already playing this role in organisms preceding the protostome and deuterostome divergence. In the deuterostome branch, namely, in echinoderms and chordates this neural function is maintained, while during the evolution of protostomes this gene acquired a new role in segmentation in a common ancestor of annelids and arthropods. Interestingly, a segmentation gene of the pair-rule class, *even-skipped*, which is also involved in neurogenesis, has a function in segmentation of *Drosophila* but not in the grasshopper *Schistocerca americana* (Patel et al., 1992). This could fit in a scenario where different genes involved in nervous system development are gradually coopted for a function in segmentation as new ways of generating pattern emerge: *engrailed* after the protostome/deuterostome divergence, and later *even-skipped* after the branching of higher and lower insects in the uniramian radiation. The study of the evolution of the *engrailed* gene family will provide insight into the links existing between development, genetics and evolutionary processes (Holland, 1990).

The anostracan crustacean *Artemia* offers an interesting system with certain properties advantageous for the study of segmentation. *Artemia* embryogenesis occurs continuously in the female ovisac or separated in two stages when diapausic cysts, arrested at an early gastrula stage, are laid by the animals. In contrast to other arthropods where segmentation is completed during embryogenesis, the larva (nauplius) of *Artemia* hatches with only two or three incompletely developed cephalic segments plus the telson. The rest of the segments are added progressively during a long postembryonic developmental period of about 2 weeks, in

which segments are generated from a growth zone existing between the last segment and the telson. In addition to the originally present cephalic segments, one mandibular, two maxillar, eleven thoracic, two genital and six abdominal segments are added sequentially through larval and juvenile stages (Schrehardt, 1987). The progressive nature of segmentation in *Artemia* is visible not only in the production of new segments, but in their maturation. At a certain stage all segments, of for example the thorax, are at different stages of development. In this way at the same time, one can observe a whole progression of segmentation in a single animal. Furthermore, since it occurs during post-embryogenesis, when the organism is feeding and growing, experimental manipulation of development by alteration of the nutritional and environmental conditions is possible (Hernandorena and Marco, 1991). Therefore, even if segmentation and homeotic genes play equivalent roles in *Artemia* and *Drosophila* there must be adjustments in the regulatory gene networks, which could provide further insight into how the activation and maintenance of the different gene expressions, characteristic of every segment, are produced in *Drosophila* and *Artemia*. In fact, although extremely rapid, there is evidence that even in *Drosophila* the activation of pair-rule and segment polarity genes also occurs progressively during blastoderm formation (Karr et al., 1989; Weir and Kornberg, 1985). For these reasons, we have started to study the process of segmentation in *Artemia*, cloning and characterizing homeobox containing genes. In this report we present the isolation and characterization of the *engrailed* gene, including its spatial pattern of expression in relation to the development of this organism.

## MATERIAL AND METHODS

### Animals

*Artemia franciscana* diapausic cysts were obtained from San Francisco Bay Brand. Cysts were developed in the laboratory for the desired period of time (all stages are given from the time of cyst activation) in 0.25 M NaCl at 30°C as previously described (Batuecas et al., 1988). *Artemia parthenogenetica diploidica* cysts were collected from La Mata lagoon in Torrevejeja, Alicante (Spain), kindly provided by Dr F. Amat. Staging is according to Schrehardt (1987).

### Library screening and isolation of clones

*Artemia franciscana* cDNA gt11 libraries of 40 hours of development (kindly provided by Dr L. Sastre; Palmero et al., 1988) were screened with a *Drosophila engrailed* cDNA probe (a generous gift of Dr T. Kornberg) that includes the homeobox and flanking sequences (Poole et al., 1985).  $1.5 \times 10^6$  plaques were transferred in duplicate to nitrocellulose filters, prehybridized in 40% formamide,  $6 \times$  SSC, 1% SDS,  $5 \times$  Denhardt and 100  $\mu$ g/ml of denatured salmon sperm DNA at 42°C and then hybridized overnight in the same conditions with the *Drosophila* probe labeled with [ $^{32}$ P]dCTP at a concentration of  $10^6$  cts/minute per ml (specific activity greater than  $10^7$  cts/minute per  $\mu$ g). Filters were washed in  $4 \times$  SSC/0.5% SDS at room temperature and 37°C followed by a wash in  $2 \times$  SSC/0.5% SDS at 55°C and autoradiographed with an intensifying screen for 4 days at -70°C. Positive clones were purified by two additional rounds of screening, the phages amplified and the inserts subcloned in the Bluescript vector (Stratagene) using standard protocols (Sambrook et al., 1989).

An *Artemia franciscana* EMBL-3 genomic library ( $2.5 \times 10^6$  pfu representing 5 genomic equivalents, also provided by Dr L. Sastre; Escalante and Sastre, 1993) was screened under high stringency conditions using, as probes, various fragments of *Artemia* cDNA clones and specific fragments of genomic clones for walking. Nitrocellulose filters were hybridized overnight in 7% SDS/0.25 M sodium phosphate buffer pH 7.2 at 65°C, washed in  $1 \times$  SSC/0.5% SDS at 68°C and autoradiographed with an intensifying screen for 4 hours at -70°C. Purified genomic clones were analyzed by multiple restriction enzyme digestions and fragments of interest subcloned in Bluescript.

### Southern analysis

Genomic DNA was extracted from newly hatched nauplii as described by Cruces et al. (1981). 15 µg of DNA were digested with each enzyme, electrophoresed on a 0.8% agarose gel, transferred to Zeta Probe-GT membrane (BioRad) and hybridized to a 373 bp (base pairs) *Artemia* genomic fragment that includes the homeobox, following the manufacturers instructions.

### DNA sequencing

The cDNA sequence was obtained in both directions using a shotgun strategy as described by Bankier et al. (1987), and partial genomic sequences were obtained using specific cDNA primers. M13mp (18 and 19) and Bluescript clones were sequenced with the chain termination method (Sanger et al., 1977) using Sequenase™ (USB) and polyacrylamide gradient gels (Biggin et al., 1983) or Taq polymerase and automatic sequencing (Applied Biosystems), following the manufacturers instructions. Sequences were analyzed using the programs developed by Staden (1986) and the GCG programs of the University of Wisconsin (Devereux et al., 1984) on a Digital Vax computer.

### Whole-mount in situ hybridization

We have used a protocol that includes several modifications to previously published ones (Hemmati-Brivanlou et al., 1990; Tautz and Pfeifle, 1989). *Artemia* of the desired stage were fixed for 2 hours at room temperature in a 1:1 mixture of growth medium and freshly made 8% paraformaldehyde in PBS. Nauplii were taken through increasing concentrations of methanol in 4-5 steps: in each one (over 5 minutes) half of the volume of the mixture was replaced with methanol, then washed three or four times with methanol and stored at -20°C. The nauplii to be stained were taken through a similar procedure to replace the methanol with PBT (PBS/0.1% Tween 20). To allow a proper penetration of reagents through the cuticle, nauplii were briefly sonicated (5-7 seconds at an amplitude of three microns in a Soniprep 150-MSE immersion tip sonicator), washed twice in PBT, treated with 50 µg/ml of proteinase K in PBT for 20 minutes, washed with 2 mg/ml glycine in PBT for 5 minutes and twice in PBT. After refixing with 4% paraformaldehyde in PBT for 1 hour at room temperature, the following washes (5 minutes each) were carried out: 5 times in PBT; once in 1:1 PBT/Hyb solution (50% formamide,  $5 \times$  SSC,  $5 \times$  Denhardt, 0.1% Tween 20, 100 µg/ml denatured salmon sperm DNA) and finally once in Hyb solution. After 1 hour of prehybridization at 45°C, a nick-translated digoxigenin labeled *Artemia* probe was added at 1-5 µg/ml and allowed to hybridize overnight at 45°C. Nauplii were washed at 45°C with a series of Hyb solution/PBT (4:1, 3:2, 2:3, 1:4) and twice with PBT. Anti-DIG antibodies (Boehringer Mannheim Genius kit) that had been absorbed overnight at 4°C against fixed nauplii were added at a 1:2000 dilution and incubated for more than 1 hour at room temperature. They were then washed four times for 20 minutes in PBT and twice for 10 minutes in 0.1 M NaCl/50 mM MgCl<sub>2</sub>/0.1 M Tris-HCl (pH 9.5)/0.1% Tween 20. Staining was developed in the last buffer with the color substrates NBT/X-phosphate provided with the kit. Developing time was usually 2-3 hours at room temperature. Specimens were mounted

in 80% glycerol in PBS and observed and photographed with a Zeiss Axiophot microscope.

### Confocal microscopy

*Artemia*, fixed as described above, were stained with 1 µg/ml ethidium bromide or 5 µg/ml acridine orange and mounted in glycerol-propylgallate and observed in a Zeiss confocal microscope using a Helium/Neon or an Argon laser.

## RESULTS

### Cloning of *Artemia engrailed* cDNA

A cDNA library from 40 hour *Artemia* nauplii (stage L4) was screened under low stringency conditions using a probe which contains the *Drosophila engrailed* homeobox (see Material and methods). Several positive clones were isolated, all carrying inserts of around 1.3 kb (kilobases). Fig. 1 presents the complete nucleotide and predicted amino acid sequence of the longest clone. The sequence reveals that it corresponds to an *Artemia* gene of the *engrailed* class. It is 1272 nucleotides long and contains an open reading frame of 1074 bp, encoding a presumptive protein with an homeodomain of the *engrailed* class near the carboxy terminus. The ATG triplet at nucleotide position 26 of the cDNA sequence codes for the presumptive initiator methionine since it is preceded in the same reading frame by several termination codons (deduced from sequences located 5' to the cDNA as found in genomic clones, see below). The next in-frame methionine is located 47 amino acids downstream, at nucleotide position 167. Between these two methionines there is an amino acid domain conserved in the amino-terminal region of *Drosophila* and *Bombyx engrailed* proteins (see below). Thus the second methionine is unlikely to be the initiator of translation. Although the sequence flanking the first ATG is not similar to the consensus defined for other organisms (Cavener, 1987), this lack of conservation has been found in the sequences flanking initiator codons in other *Artemia* genes (listed in Marco et al., 1991). The first in-frame termination codon is located at position 1073. The predicted *Artemia engrailed* protein is therefore 349 amino acids long with a deduced relative molecular mass of  $39 \times 10^3$ .

The ATG initiation codon is preceded by 25 nucleotides of 5' untranslated sequence and the termination codon is followed by a 3' untranslated region of 197 nucleotides where no canonical polyadenylation signal has been found. The size of the message is approximately 1.5 kb as determined by northern analysis (data not shown), and therefore the characterized clone although containing all the coding sequence does not correspond to a full length cDNA.

### Sequence analysis of the *Artemia engrailed* protein

Based on sequence comparisons between *engrailed* and *invected* genes from insects and *engrailed* genes from vertebrates, four domains of sequence similarity have been identified in the *engrailed* protein family (Ekker et al., 1992; Hemmati-Brivanlou et al., 1991; Hui et al., 1992; Logan et al., 1992): the homeodomain and three additional motifs located in the amino terminal (I), central (II) and carboxy-

M G S A I F E P G P L S L L N L A C S N L T E R Y D G P S P L S	32
GAATTCCTCGAACTTAGTTGTGTGATGGGCGAGTGCTATTTTTGAACCCGGACCATTAATTCATTGCTTAACCTCCGCTGTTTCGAAC TTGAC TGA AAGATAC GATGG ACC AAGTC CACTAAG	120
A S T P G P S P D R P G S A T M S S P L S S P T G I S Y Q S L L S G I L P A A M	72
TGCTAGTACACCCAGG ACCGTCCCCGATAGACCTGGGGTCTGCAAC AATGTCATCACCAATTATC ATCCC CG ACCGGGATTCATAC CAGTC ATTGTTATCTGGTATACTAC CAGCAGCTAT	240
F P Y G Y P P V G Y M Y P T G F P T L A A I Q S G H L A F R Q L V P T L P F M T	112
GTTTC CATATGGATA TCC TC CAG TGGGCTATATGTACC CAACTGGGTTTC CCACTTTGCGGGC CATAC AATCAGGACACTTAGCC TTTTCGCGAGC TAG TTCCAAC ACTTC CTTTAAATAC	360
V K S S E G Q V K E V V S T Q S Q K K P L A F S I D S I L R P D F G K E T N E V	152
AGTTAAATCTTCAGAGGG AC AAG TG AAGAGGTTAG TTAGC ACACAATC TC AAAAG AARCC TCTAGCGTTTCTATAGACAGTATATTACGCC TG ATTTTGGC AAGGA AACAAATGAAGT	480
K R R R H A S P H R E E P K K K V Q Y I E Q M K K K E E I K E E A R T E S R L S S	192
TAAAGCTGC CATGCTAG TCC CAC ATAGAGAAGA AC CGAAG AAAAAGTCC AATAC ATTGAGCA AA TGA AAAAG AAGAAG AGATC AAG GAAGAAG CGCGG ACGGAGAG CC GTTGTG CATC	600
S S K D S V P D N D K I N P P L P P E A S K W P A W V F C T R Y S D R P S S G R	232
GAGTAGCA AAGATTC TGTCCTG ATAATGATAAAATTAATCCACC TCTTCCCTGGAAGCCTC TAAATGGCCAGCCTGGTCTTTTGTAC TCG ATATT CAGATAGGCC TTCGTCAGGTAG	720
S P R C R R M K K D K A I T P	272
GAGTCCTCGATGTCG CCG AATGAAAAGAGATAAAGCCATAACCCTGACG AAAAAGGCC AAG AACTG CATTTAC TGC GAGC AACTG AG TCG CCA AACCAC GAATTTATG A AATTCG	840
▲	
Y L T E E K K Q D L A K E L G L H E N Q I K I W F Q N N R A K L K K S S G Q K M	312
ATACC TAAACC GAACG TCG AC GAC AGGATTTAGCG GGG AATTAGGCCTTC ATG AAAATCAGATTA AATA TGG TTTCAAAATATAGAGC TAAATGAAAATTC TTC TG GGC AAAAG AA	960
P L A L O L M A O G L Y N H S T I P T E D D E D D E I S S T S L Q A R I E *	349
TCCCCTGC CACTC CA ACTAATGG CAC AAGGCTTTAC AAC CATTC AAC AATACCTACAGAAGATG ATG AAGATGATGAAAATTAGC TCAAC GTC AC TGC AAGCTAGGATTS AATGAGAGCT	1080
TCATTCGTATCTTTTCG CATCATCTTGTCTTTGAGGTG ATTAC AATGTTTGTGCC AAA TTTTGT ATATG TACATAATCC TG TATATAATTTGATGC TTTAC ATG AC TGTTCGATTTTATC	1200
TCCGAACGAAGTATTTAAATG TATTTCC ATG TTTGTTTGT C AATAC ATGTC TATGT ATATC GCGGAATTC	1272

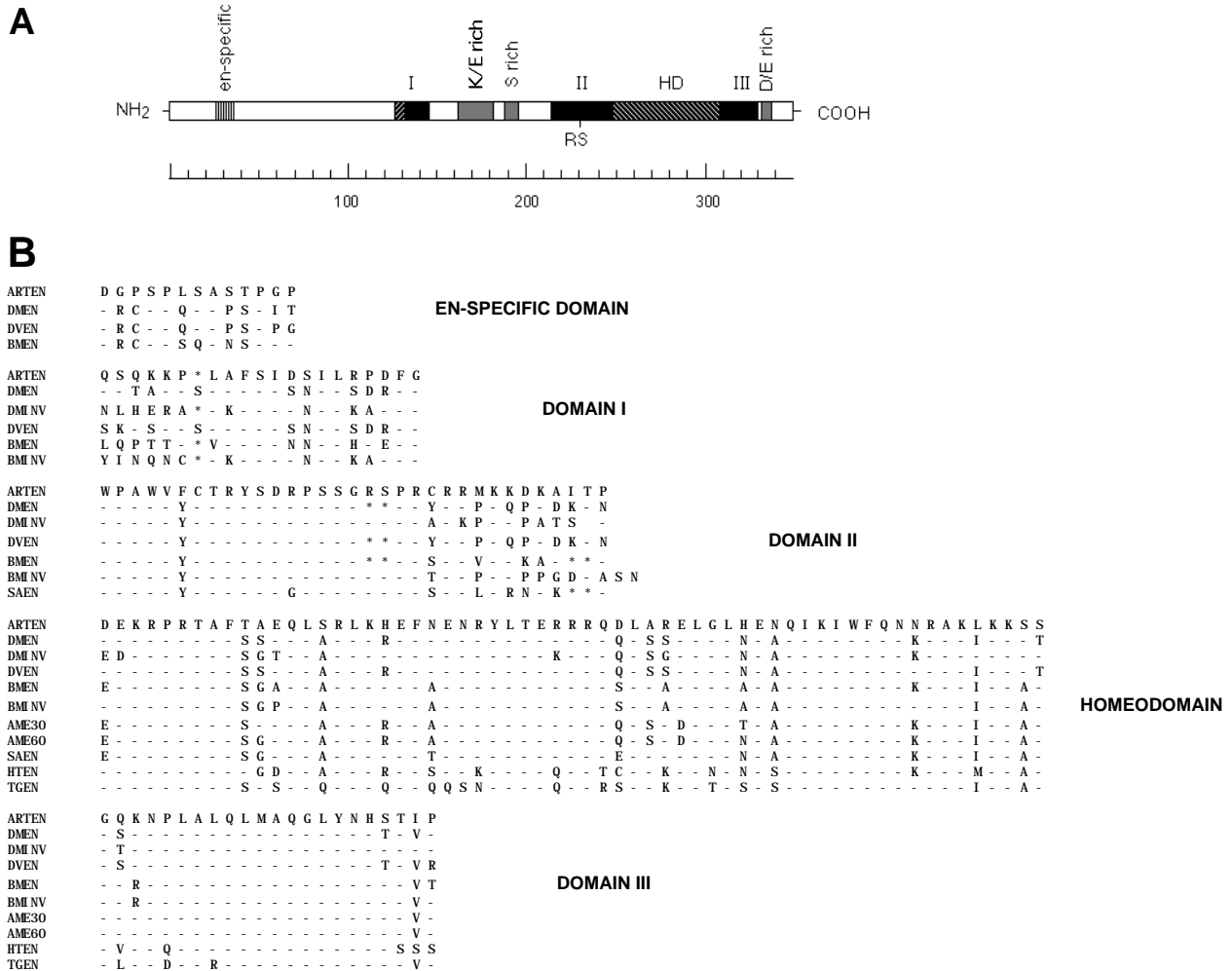
**Fig. 1.** Nucleotide sequence of the *Artemia engrailed* cDNA and predicted amino acid sequence. The suggested initiator methionine is at nucleotide position 26 (see text) and the first in-frame termination codon is denoted by an asterisk. The homeodomain is boxed. Other conserved domains are underlined and are, from the 5' to the 3' end: *engrailed*-specific, domain I (less conserved), domain I, domain II and domain III. Intron sites are marked with an arrowhead. This sequence has been entered in the EMBL data base under the accession number X70939.

terminal (III) regions of the protein, the last two flanking the homeodomain. All four domains are conserved in the *Artemia* sequence as indicated in Fig. 2A,B. The homeodomain is approximately 80% identical to the *Drosophila engrailed* and *invected* sequences; changes are located in the more variable positions, with some exceptions. For example, in the phylogenetically well conserved 14 amino acid epitope recognized by the monoclonal antibody mAb 4D9 (residues 282-295 within the homeodomain in Fig. 1; Patel et al., 1989a) there is an asparagine to histidine change (position 289) that is surely critical for antibody recognition. In accordance with this finding we have systematically failed to obtain staining in *Artemia* using this antibody. All critical residues for DNA-protein interactions are strictly conserved (Kissinger et al., 1990).

Domain I comprises 14 amino acids that are highly conserved in the proteins from the *engrailed* family of insects (Hui et al., 1992) and vertebrates (Ekker et al., 1992; Hemmati-Brivanlou et al., 1991; Logan et al., 1992). In *Artemia*, this region is preceded by a stretch of six amino acids showing a lower but still recognizable similarity with *Drosophila engrailed* sequences (Fig. 2B), but not with the *Drosophila invected* nor the *Bombyx engrailed* and *invected* proteins. Domain II spans a region of 33 amino acids preceding the homeodomain, in which the *Artemia* sequence contains an arginine-serine (RS) insertion with respect to *Drosophila engrailed* (Fig. 2B). This doublet is present in *Drosophila* and *Bombyx invected* genes where it is encoded by a microexon six bp long (Coleman et al., 1987; Hui et

al., 1992), and has been claimed to be a hallmark for *invected* proteins. This genomic organization is not conserved in *Artemia* (see below). Interestingly this motif is present in the only *engrailed* gene identified to date in the short germ band insect *Schistocerca americana* (Patel et al., 1989a). Following the extremely conserved domain III, at the carboxy-terminal end of the protein, there is a short stretch rich in aspartic and glutamic acid (D/E rich region, Fig. 2A), also present in *Bombyx engrailed* and similar to the poly-glutamic stretch present in *Drosophila engrailed* and *invected* and *Bombyx invected*. This conserved highly acidic region could be involved in transcriptional activation (Ptashne, 1988).

Hui et al. (1992) have defined *engrailed* and *invected* specific domains by comparison of *Drosophila* and *Bombyx* sequences. These regions are not conserved in characterized members of the family from vertebrates (Ekker et al., 1992; Hemmati-Brivanlou et al., 1991; Logan et al., 1992) and there is no information of full length sequences from other invertebrates. A region closely related to the most N-terminal *engrailed*-specific domain can be identified in the *Artemia engrailed* sequence (see Fig. 2). In *Drosophila* and *Bombyx* the domain starts in the initiator methionine and spans 15 amino acids. In the *Artemia engrailed* protein the first four amino acids are not present and the region is located 25 amino acids downstream from the initiator methionine. The *Artemia engrailed* protein is shorter than the insect *engrailed* and *invected* polypeptides and does not include the second *engrailed*-specific motif found in



**Fig. 2.** Organization of *Artemia engrailed* protein domains and conservation between invertebrates. (A) Schematic representation of the protein with the different domains indicated. Domain I is subdivided into a first stretch of less conserved residues (stippled) and a more conserved part (filled). Position of the arginine-serine doublet (RS) and regions enriched in certain amino acids are also indicated (see text for details). (B) Comparison of predicted protein sequences from *Artemia* and other available invertebrate *engrailed* and *invected* genes. Identical residues are indicated by a dashed line; the asterisks show gaps introduced to maximize similarity; the symbol in domain II comparison indicates a stretch of 22 residues specifically present in the *Drosophila invected* product. Position of amino acid residues with respect to Fig. 1 are: 26-36 (*engrailed*-specific), 127-146 (domain I), 215-247 (domain II), 248-308 (homeodomain) and 309-330 (domain III). Abbreviations are as follows: ARTEN, *Artemia franciscana engrailed*; DMEN, *Drosophila melanogaster engrailed* (Poole et al., 1985); DMINV, *Drosophila melanogaster invected* (Coleman et al., 1987); DVEN, *Drosophila virilis engrailed* (Kassis et al. 1986); BMEN and BMINV, *Bombyx mori engrailed* and *invected* (Hui et al., 1992); AME30 and AME60, *Apis mellifera* clones E30 and E60 (Walldorf et al., 1989); SAEN, *Schistocerca americana engrailed* (Patel et al., 1989a); HTEN, *Helobdella triserialis engrailed* (Wedden et al., 1991); TGEN, *Tripneustes gratilla engrailed* (Dolecki and Humphreys, 1988).

*Drosophila* and *Bombyx*. Furthermore, the *invected*-specific sequence shared by the *Drosophila* and *Bombyx invected* proteins is not found in *Artemia*.

In addition to these domains, the *Artemia engrailed* protein presents several additional features schematically summarized in Fig. 2A. Between the domains I and II there are two regions, one especially rich in lysine and glutamic acid (K/E rich region) and another one rich in serine (S rich region). In general the sequence is rich in proline residues (10%) especially the N-terminal first third (16%), as it is also found in other proteins of the *engrailed* family. No other overall similarity is present in the sequence as shown by

multiple alignment with all *engrailed* family sequences present in the databases.

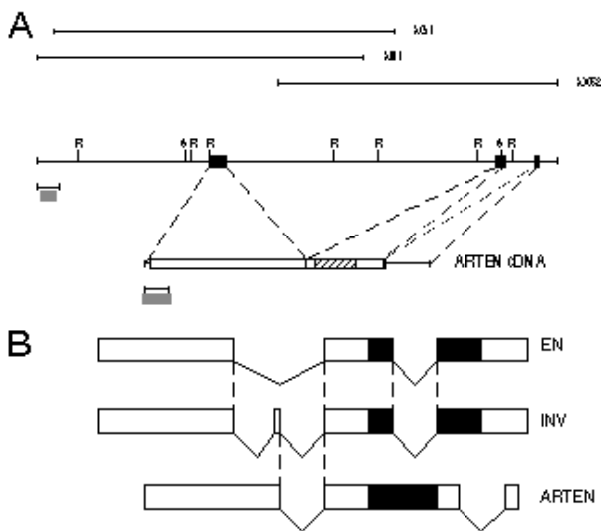
### Genomic organization of the *Artemia engrailed* gene

Using probes from the 5' and 3' regions of the *Artemia engrailed* cDNA, we screened an *Artemia* genomic library as described in Material and methods. Several recombinant clones were isolated and further 5' and 3' genomic probes were used for walking. In total we have isolated overlapping phages covering a region of 25 kb. We determined the organization of the *Artemia engrailed* gene by Southern blot

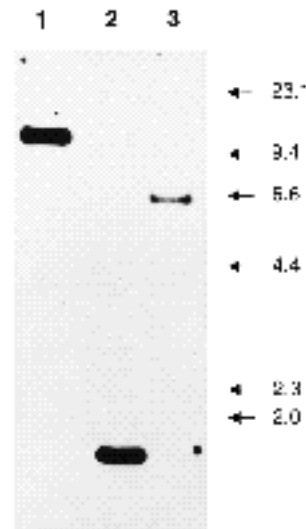
experiments and sequence analysis using cDNA specific primers at intron-exon boundaries. It is distributed in three exons and expands at least 15 kb (Fig. 3). The first exon includes the arg-ser doublet that is encoded in a separate microexon in insect *invected* genes.

The first intron begins at a position corresponding to the right boundary of the *invected* microexon and the first *engrailed* intron of insects. This position is conserved in all *engrailed* genes whose intron/exon organization has been determined (Coleman et al., 1987; Dolecki and Humphreys, 1988; Fjose et al., 1992; Hui et al., 1992; Joyner and Martin, 1987; Logan et al., 1992; Poole et al., 1985). No other intron position is conserved in *Artemia*. The second intron is located near the end of the coding region at position 1065 (Fig. 1), a novel intron position in the *engrailed* family, and is 1.2 kb in length. A summary of intron/exon organization compared to insect *engrailed* and *invected* genes is presented in Fig. 3B.

To estimate the number of genes of the *engrailed* family present in *Artemia*, Southern blots of *Artemia* genomic DNA were hybridized under high stringency conditions with a probe derived from a genomic clone that comprises the homeobox and 3 conserved regions (Fig. 4). It reveals a single prominent band of the expected size (Fig. 4, lane 2), suggesting that *Artemia* contains only one *engrailed* gene. In fact in the blot not only the DNA of *Artemia franciscana* but also that of another eurasian brine shrimp species, *Artemia parthenogenetica diploidica*, is included. Data obtained in our laboratory shows that these species diverged



**Fig. 3.** Organization of *Artemia engrailed* gene. (A) The thick line in the middle shows the genomic region. Exons are indicated as filled boxes. Thin lines in the upper part represent clones spanning the region. On the lower part of the panel there is a schematic representation of the cDNA where the coding region is represented by a box, exon limits marked and the homeobox stippled. (B) Schematic representation of the intron position in *Drosophila engrailed* (EN) and *Bombyx engrailed* (INV) and *Artemia engrailed* (ARTEN) genes. The position of the homeobox is indicated by a filled box and intron/exon boundaries located at identical positions are indicated by dashed lines. The size of exons are not drawn to scale. R, *EcoRI*; S, *SalI*.

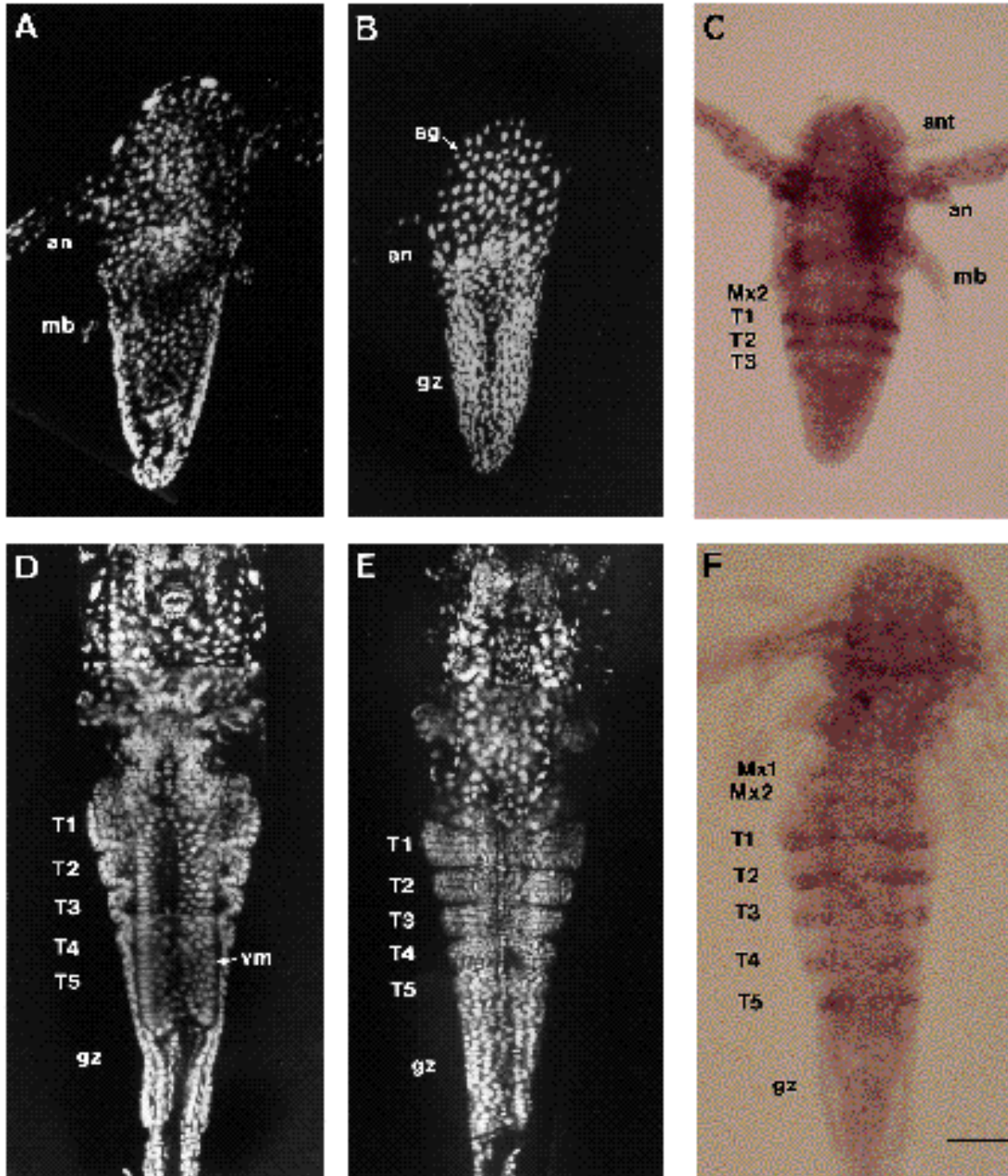


**Fig. 4.** Southern blot of *Artemia* genomic DNA hybridized to a probe that contains the *engrailed* homeobox under high stringency conditions. Lanes 1 and 2, *Artemia franciscana* DNA digested with *Bam*HI and *Eco*RI respectively. Lane 3, *Artemia parthenogenetica diploidica* DNA digested with *Bam*HI. The sizes in kb of  $\lambda$ -*Hind*III digest molecular mass markers (New England Biolabs) are shown on the right.

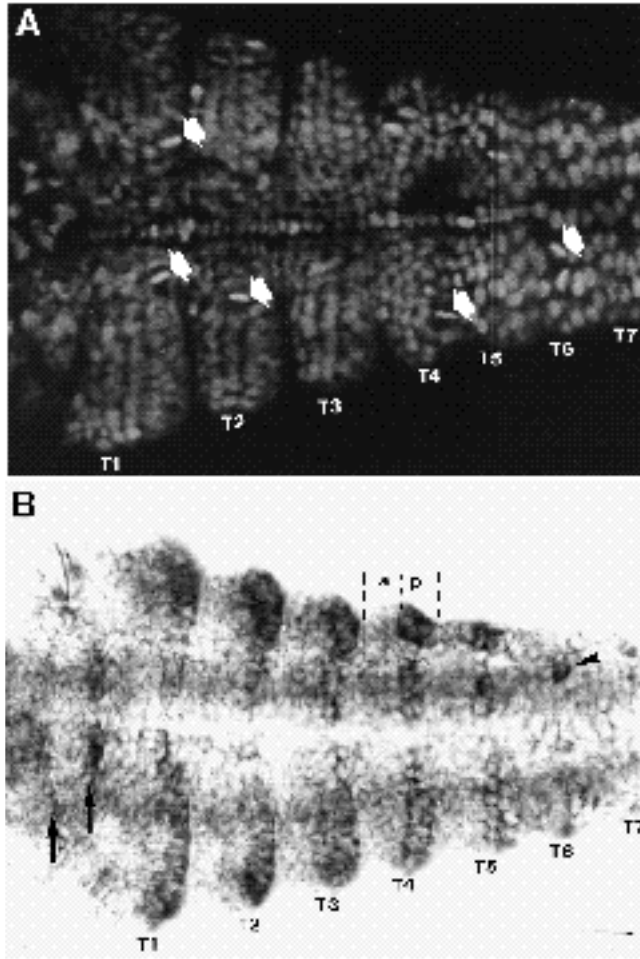
a long time ago (more than 50 million years; Perez et al., 1993). In accordance with this divergence, the Southern blot of *Artemia parthenogenetica* shows a different restriction pattern (Fig. 4, lanes 1 and 3) than *Artemia franciscana* and the intensity of hybridization is much less than when using the DNA homologous to the probe. Again only one band is detected at high stringency. Under low stringency conditions or after overexposure of the filters hybridized under high stringency conditions, five or six additional bands are visible (data not shown). The intensity of these extra bands is similar, probably indicating the presence of other *Artemia* homeobox-containing genes. Only if genes of the *engrailed* family in *Artemia* are more divergent and numerous than their counterparts in other organisms could these bands correspond to additional genes of the *engrailed* class.

#### Developmental expression of the *Artemia engrailed* gene

Transcription of the *engrailed* gene is detected in northern blots (not shown) shortly after resumption of development of the arrested gastrulae, and then is expressed continuously until the whole process of segment formation has been completed. We have examined the spatial pattern of expression by using a DIG-labeled *Artemia engrailed* probe. In separate animals we have visualized the cell arrangement by confocal imaging of nuclear staining in whole animals. Two stages of development were studied: the naupliar stage L1 (Fig. 5A-C) where only a few well developed cephalic structures are present (corresponding to 20 hours of development), and the more advanced metanaupliar stage L4 (Figs 5D-E, 6) where four thoracic segments are visible (50 hours of development). *Artemia* postgastrular development proceeds in the absence of cell division until hatching of the



**Fig. 5.** Expression of *engrailed* in developing *Artemia*. Larvae were stained with acridine orange (A, B) or ethidium bromide (D, E) to visualize nuclei by confocal microscopy in order to monitor cell arrangement. (C, F) In situ hybridization with an *engrailed* DIG-labeled probe to a different set of animals. Two different optical sections of a dorsal view of a L1 stage nauplius are shown, at an internal level (A) where the developing gut and the surrounding ectoderm can be seen, and at the epidermal level (B). Cells present in the growth zone (gz) do not have any distinct organization and no segment primordia are visible. The polyploid nuclei of the larval salt gland (sg) are clearly discernible from the diploid nuclei of the protruding cone, where segments will be formed. (C) *engrailed* expression at this stage is strongest in what will be the first two thoracic segments (T1 and T2). An incomplete stripe corresponding to the second maxillary segment (Mx2) has appeared on the right side of the animal. A nascent stripe for the third thoracic segment (T3) is beginning to form at the rear end. (D-F) In a ventral view of a L4 metanauplius three thoracic segments have differentiated (E) as well as the visceral mesoderm (vm) that surrounds the gut (D). Ordered cell rows and the formation of the inter segmental grooves give an idea of how advanced the differentiation of each segment is. Cephalic structures are still made up in part by larval polyploid cells. D and E correspond to two optical sections of the same animal. Five stripes of thoracic *engrailed* expression are visible as well as the two maxillary stripes (F). The last *engrailed* stripe that has appeared (T5) corresponds to a region where no obvious cell arrangement has taken place. ant, antennulae; an, antennae; mb, mandible. Scale bar: A-C, 46  $\mu$ m; D-F, 73  $\mu$ m.



**Fig. 6.** Ventral view of L4 metanauplii stained with ethidium bromide (A) or the DIG-labelled probe (B). (A) The dynamic pattern of cell arrangement is visible. The columnar organization in presumptive segments T4-T6 changes gradually to a row arrangement in the more mature segments (T1-T4). Scattered mitoses are present all over the field. Some of them are shown with white arrows. This is a higher magnification of the same animal than in Fig. 6E. (B) *engrailed* expression is detectable only in the posterior compartment of each segment. The fourth thoracic segment is delineated to show segmental and compartmental limits. The two arrows indicate the maxillary stripes that are partially out of focus in this photograph. The five first thoracic stripes (T1-T5) are formed and accumulation of transcripts is seen in the distal end of the most mature segments where the appendages will form. The arrowhead in T6 shows an isolated cell expressing *engrailed*. The T7 stripe is beginning to appear. a, anterior compartment; p, posterior compartment. Scale bar, 28  $\mu$ m.

nauplius (Olson and Clegg, 1978). The approximately 5000 cells already present in the dormant gastrula organize and differentiate into the naupliar cephalic structures, including three pairs of appendages (the antennulae, the antennae and the mandible) and the salt gland (Fig. 5A-C). The remaining cephalic and postcephalic structures will be formed from the group of about 2000 undifferentiated cells that protrude in the form of a cone from the posterior sector of the nauplius

(Fig. 5A-C). These cells, as well as other cells populating the interior of the differentiated cephalic structures are endowed with smaller, diploid nuclei, while the larval specific structures (appendages and salt gland) are made up by polyploid cells (Fig. 5B). These structures will be replaced by the definitive adult organs developed from the groups of diploid precursor cells remaining inside the cephalic structures. The cone can be considered as a morphogenetic field with a growth zone from which segments are formed. Analysis of mitotic pattern in the growing zone of the *Artemia* cone shows that mitotic figures occur interspersed all over the field (Fig. 6A). Field growth seems to occur mostly by cell intercalation, although contributions of cell lineage cannot be totally ruled out. This is in contrast to the development of malacostracans such as crayfish or *Dyastylis rathkei* where segmentation is an embryonic process and where cell growth is achieved by the asymmetric division of a parallel row of teloblastic cells (Dohle and Scholtz, 1988).

Early during naupliar development (stage L1) *engrailed* transcripts accumulate in two stripes (Fig. 5C) of otherwise undifferentiated cells (Fig. 5A-B). They correspond to what will be the posterior cells of the first and second thoracic segments. The cells are aligning themselves in columns leaving a separation at the middle line of the dorsal and ventral sides (Fig. 5B). In accordance with this separation the stripes do not completely surround the cone, they are interrupted in the middle lines both at the dorsal and ventral sides (Fig. 5C). A little later during this stage, an extra anterior band appears, corresponding to the second maxillary segment, and an incomplete band corresponding to the third thoracic segment becomes visible (Fig. 5C).

In L4 stage metanauplii, three thoracic segments are clearly visible in a ventral view of the animal (Fig. 5D-E); two or three additional ones are forming. The thoracic appendages develop by the proliferation and differentiation of the more lateral cells, while cells in the middle of the ventral side will form the neuromeres. In the T1-T3 segments, five or six rows of cells can be distinguished, the two central ones being the more orderly (Figs 5E, 6A). The *engrailed* stripes are sharper at their posterior border, mainly because of the overlap with the clear establishment of the intersegmental groove (Figs 5D-F, 6B). In addition, the *engrailed* stripes are wider and stronger in the lateral groups of cells where the thoracic appendages are going to be formed (Fig. 6B). *engrailed* stripes in thoracic segments four, five and even six are appearing. It can be seen that in the growing zone *engrailed* transcription is increased or turned on in single cells (arrowhead in Fig. 6B) and eventually spreads laterally into stripes, initially one cell wide but soon broadening into two to three cells as the segments develop (Fig. 6B). Thus, the *engrailed*-expressing cells make up roughly one third of the total segment, similar to what has been described in other arthropods. The organization of the growing zone in columns switches to rows as segments are formed (Figs 5E, 6A). Finally, in L4 metanauplii, *engrailed* is also expressed in two stripes located between the head and the first thoracic segment. They correspond to the two maxillary segments that develop reduced appendages in the adult animal. Unlike the thoracic stripes, they remain one cell wide throughout those larval stages that



have been examined and their formation is delayed with respect to the first thoracic stripes. This order of *engrailed* stripe appearance resembles what has been described for *Schistocerca* (Patel et al., 1989b).

## DISCUSSION

In this study we have described the molecular characterization of an *engrailed*-like gene of the crustacean *Artemia franciscana*. Analysis of the protein sequence encoded in the *Artemia engrailed* gene reveals the presence of four conserved domains already described for the proteins of the *engrailed* family in insects and vertebrates (Ekker et al., 1992; Hui et al., 1992; Hemmati-Brivanlou et al., 1991; Logan et al., 1992). These domains are extremely well conserved within arthropods and show a lower but still significant similarity with their vertebrate counterparts. The *Artemia* sequence presents some specific characteristics of both the *engrailed* and *invected* proteins. One is an amino acid motif located at the amino-terminal end of the *Drosophila* and *Bombyx engrailed* proteins but not in the *invected* proteins. Another is the amino acid doublet arg-ser that has been found only in the *invected* products, and therefore has been considered a hallmark for *invected* proteins. The presence of this doublet in the single *engrailed* gene from the short germ band insect *Schistocerca americana* (Patel et al., 1989a) raises the possibility of an ancestral arthropod gene that would have had these mixed identities of *engrailed/invected* genes. Finally, the *Artemia engrailed* protein has an overall amino acid composition that fits well with that found in other proteins of the *engrailed* family, including a serine-rich region that is probably the target of phosphorylation by protein kinases, a process that has been clearly demonstrated in the *Drosophila engrailed* protein (Gay et al., 1988).

The *Artemia engrailed* gene spans a region of 15 kb and is organized in three exons and two introns. In *Drosophila* and *Bombyx*, *engrailed* genes also have three exons whereas the *invected* genes have four, due to the presence of an additional six nucleotide microexon encoding the amino acids arg-ser. Although this amino acid doublet is conserved in the *Artemia* sequence, it is located at the end of the first exon, and is not encoded in a separate microexon. Interestingly, the position of only one of these introns is conserved in *Artemia*. The second intron located in the homeobox, well conserved in *Drosophila* and *Bombyx engrailed* and *invected* genes is absent in *Artemia*, a situation also found in vertebrates and echinoderms. This occurs also in the hymenopteran insect *Apis mellifera*, where two different genomic clones containing only the homeodomain coding exon of *engrailed*-like proteins have been identified (Walldorf et al., 1989). The absence of the intron that interrupts the homeobox could indicate that in hymenopterans there is a genomic organization intermediate between crustaceans (*Artemia*), and higher insects (lepidopteran, *Bombyx*, and dipteran, *Drosophila*). More sequence information is needed to determine whether these two genes from *Apis* are true homologues of *engrailed* and *invected* or whether they represent an intermediate situation in the evolution of the *engrailed* class of genes in insects. The

second intron of the *Artemia engrailed* gene is located in the 3' region, a novel situation in the genes of the *engrailed* family that could represent a final event in the divergence of the crustacean gene.

*Artemia* has only one gene with a well conserved homeobox of the *engrailed* family as supported by the following evidence. Southern blot analysis under high stringency conditions detects only one prominent fragment in both *Artemia franciscana* and *Artemia parthenogenetica*. In accordance with this result, repetitive screenings of genomic and cDNA libraries have yielded several clones that in all cases corresponded to the same gene. Northern blot analysis reveals the presence of a single message of 1.5 kb (not shown), compatible with the size of isolated cDNAs. It is true that under low stringency conditions, several additional fragments of similar intensities are visible by Southern blot analysis, but they probably correspond to genes with more diverged homeoboxes. The possibility of a large family of diverged *engrailed*-like genes in *Artemia* seems unlikely but cannot be ruled out.

In conclusion, several findings support the identification in *Artemia* of a gene of the *engrailed* family that shares some characteristics present in *engrailed* or *invected* genes from higher insects. Neither the genome organization nor the sequence analysis allows a closer relationship with either of the two genes to be deduced. Therefore, we suggest that the *Artemia* gene has some of the characteristics of the original gene of the common arthropod ancestor of insects and crustacean. Although disputed in the past, recent data support the view of a monophyletic origin of the arthropods (Ballard et al., 1992; Shear, 1992). In this context, a series of events such as the appearance of a new splice site before the arg-ser motif, duplication of the region, and subsequent loss of the microexon in one of the genes would lead to the actual situation in higher insects. As discussed above, traits of the evolution of *engrailed* genes can also be found by examining other insects such as *Schistocerca* and *Apis*. Therefore, our results support the actual view that gene duplication occurred relatively late during evolution and independently in the vertebrate and insect lineages, originating from a primitive gene with some characteristics of the present-day *Artemia* gene. Nevertheless, it is difficult to extend this discussion to ancestors preceding the protostome/deuterostome divergence. The *engrailed* genes of vertebrates and echinoderms are similar to the arthropod homologues in location of the first intron or conservation of sequence domains, but a common ancestor should lack certain characteristics such as the arg-ser doublet that would have appeared before the arthropod diversification. Alternatively, this motif may have been lost early in the deuterostome lineage. Additional full length sequences of *engrailed* genes from different phylogenetic groups will be needed to complete the description of the lineage and origin of this gene family.

Since both *engrailed* and *invected* genes have identical expression domains in *Drosophila* and no phenotypic effects of *invected* gene mutations have been reported, it seems that their biochemical function is quite close and partially redundant. The expression pattern of single *engrailed* genes in *Schistocerca* and *Artemia* argue that the segmental function of conferring posterior compartment identity is also

evolutionarily conserved. Therefore, it is possible that a single gene could play the role of both the *engrailed* and *invected* in lower insects and crustaceans.

The *Artemia engrailed* gene is transcribed throughout the developmental stages that we have examined. Its pattern of expression supports a role for the *engrailed* gene in selecting the genetic address typical of posterior compartments during segmentation. The timing and place of appearance of the *engrailed* stripes is similar to that in grasshopper (Patel et al., 1989b), in that the first segments do not show a strict anterior-posterior correlation. The first stripes to appear are those corresponding to the thoracic segments T1 and T2. Next, the second maxillary, third thoracic and first maxillary stripes form in this order. Then the rest of the stripes of thoracic segments are formed in a sequential fashion up to T6-T7 (the older stages examined in this work). The expression of *engrailed* in the cephalic structures has not been determined due to their complexity and the resolution of the method used. Nevertheless, the studies in *Artemia* confirm a complex behavior in the appearance of *engrailed* stripes in arthropods. The differences in the order of appearance in the various organisms examined (Karr et al., 1989; Patel et al., 1989b; Fleig, 1990; this study) could reflect a species-specific mode of regulation of the *engrailed* gene.

Expression of *engrailed* has been studied in another crustacean, the crayfish, where parasegmental limits are defined, but not clearly set up, by genealogical units (Dohle and Scholtz, 1988; Patel et al., 1989b). The mode of cell division and growth in this class of crustaceans (malacostracans), where segment formation is an embryonic process, indicates a great component of cell lineage in the generation of the field that then will be segmented (Dohle and Scholtz, 1988): ectodermal cells are derived from precursor cells (ectoteloblasts) in a repeatable and defined pattern. In anostracans, the class to which *Artemia* belongs, there is no evidence of this ordered cell division but of a undifferentiated growth zone where segments bud off as generalized cell division takes place (Anderson, 1967; Schram, 1986). In accordance with this, the activation of *engrailed* in *Artemia* appears to be induced in the middle of the growth zone not in a particular clonal group of cells: first in isolated cells, later spreading laterally to form a stripe, and then widening from one to two or three cells. This event needs recruitment of cells to express *engrailed* as well as clonal transmission, because the progression of *engrailed* expression is seen in early stages (nauplius stage L1) where little cell division takes place; even blocking it does not affect normal development (Olson and Clegg, 1978). Recent experiments in *Drosophila* have shown that the expression of *engrailed* in the early embryo is not clonally propagated but depends on a particular cellular environment and only later is associated with the determination state of cells (Ingham and Martinez Arias, 1992; Vincent and O'Farrell, 1992). The setting up of the expression pattern of *engrailed* in *Artemia* is thus more easily explained by hypotheses based on intercellular interactions that are regularly built up in the growth zone than on other hypothesis based on clonal production of a group of cells with heritable expression of *engrailed*. It will be important to identify and study other segmentation genes in *Artemia* such as *wingless* to complete the characterization of the process of segment determination. Furthermore, the

recent identification of *Hox* genes (Averof and Akam, 1993) and the possible orthologue of the gap gene *Krüppel* (Sommer et al., 1992) in *Artemia* will improve the knowledge of segmentation and pattern formation in crustaceans.

We wish to thank Tom Kornberg for the gift of the *Drosophila engrailed* clone, Francisco Amat for providing the *Artemia parthenogenetica* animals, and Leandro Sastre for the gift of *Artemia* libraries. An essential part of this work was done at Michael Akam's lab in Cambridge, which M. M. visited while supported by an EMBO Short Term Fellowship. We appreciate greatly the help of everyone at Cambridge and especially Michael Akam and Michael Averof for encouragement and thoughtful comments on the manuscript. We also thank Ernesto Sanchez Herrero, Leandro Sastre and Jaime Renart for critical reading of the manuscript. Antonio Fernandez gave invaluable help in preparing photographs and Mari Carmen Moratilla excellent technical assistance. Zeiss-Iberica (Madrid) and the Centre of Molecular Biology (CSIC, Madrid) kindly let us use their confocal microscopes. We are very grateful to Manuel Calleja for constant support and discussion. M. M. was the recipient of a predoctoral fellowship from the Ministerio de Educación y Ciencia of Spain. This work was supported by grant no. PB87-0208 from the DGICYT (Ministerio de Educación y Ciencia, Spain) and by grant no. ESP91-627 from the Plan Nacional del Espacio (Spain).

## REFERENCES

- Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**, 1-22.
- Anderson, D. T. (1967). Larval development and segment formation in the branchiopod crustaceans *Limnadia stanleyana* King (Conchostraca) and *Artemia salina* (L.) (Anostraca). *Aust. J. Zool.* **15**, 47-91.
- Averof, M. and Akam, M. (1993). *HOM/Hox* genes of *Artemia*: implications for the origin of insect and crustacean body plans. *Curr. Biol.* **3**, 73-78.
- Ballard, J. W. O., Olsen, G. J., Faith, D. P., Odgers, W. A., Rowell, D. M. and Atkinson, P. W. (1992). Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. *Science* **258**, 1345-1348.
- Bankier, A. T., Weston, K. M. and Barrell, B. G. (1987). Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. *Methods in Enzymology* **155**, 51-93.
- Batuecas, B., Garesse, R., Calleja, M., Valverde, J. R. and Marco, R. (1988). Genome organization of *Artemia* mitochondrial DNA. *Nucl. Acids Res.* **16**, 6515-6529.
- Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983). Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **50**, 3963-3965.
- Brower, D. (1986). *engrailed* gene expression in *Drosophila* imaginal discs. *EMBO J.* **5**, 2649-2656.
- Cavener, D. R. (1987). Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucl. Acids Res.* **15**, 1353-1361.
- Coleman, K. G., Poole, S. J., Weir, M. P., Soeller, W. C. and Kornberg, T. B. (1987). The *invected* gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes and Dev.* **1**, 19-28.
- Cruces, J., Sebastián, J. and Renart, J. (1981). Restriction mapping of the rRNA genes from *Artemia* larvae. *Biochem. Biophys. Res. Comm.* **98**, 404-409.
- Davis, C. A., Holmyard, D. P., Millen, K. J. and Joyner, A. L. (1991). Examining pattern formation in mouse, chicken and frog embryos with an *En*-specific antiserum. *Development* **111**, 287-298.
- Devereux, J., Haeblerli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.
- DiNardo, S., Kuner, J. M., Theis, J. and O'Farrell, P. H. (1985).

- Development of embryonic pattern in *Drosophila melanogaster* as revealed by accumulation of the nuclear *engrailed* protein. *Cell* **43**, 59-69.
- Dohle, W. and Scholtz, G.** (1988). Clonal analysis of the crustacean segment: the discordance between genealogical and segmental borders. *Development* **104Supplement**, 147-160.
- Dolecki, G. J. and Humphreys, T.** (1988). An *engrailed* class homeobox gene in sea urchins. *Gene* **64**, 21-31.
- Ekker, M., Wegner, J., Akimenko, M. A. and Westerfield, M.** (1992). Coordinate embryonic expression of three zebrafish *engrailed* genes. *Development* **116**, 1001-1010.
- Escalante, R. and Sastre, L.** (1993). Similar alternative splicing events generate two sarco/endoplasmic reticulum Ca-ATPase isoforms in the crustacean *Artemia franciscana* and in vertebrates. *J. Biol. Chem.* (in press).
- Fjose, A., Njølstad, P. R., Nornes, S., Molven, A. and Krauss, S.** (1992). Structure and early embryonic expression of the zebrafish *engrailed-2* gene. *Mech. Dev.* **39**, 51-62.
- Fleig, R.** (1990). *Engrailed* expression and body segmentation in the honeybee *Apis mellifera*. *Roux's Arch. Dev. Biol.* **198**, 467-473.
- Gay, N. J., Poole, S. J. and Kornberg, T. B.** (1988). The *Drosophila engrailed* protein is phosphorylated by a serine-specific protein kinase. *Nucl. Acids Res.* **16**, 6637-6647.
- Gehring, W. J.** (1987). Homeo boxes in the study of development. *Science* **236**, 1245-1252.
- Hemmati-Brivanlou, A. and Harland, R. M.** (1989). Expression of an *engrailed*-related protein is induced in the anterior neural ectoderm of early *Xenopus* embryos. *Development* **106**, 611-617.
- Hemmati-Brivanlou, A., Frank, D., Bolce, M., Brown, B., Sive, H. and Harland, R. M.** (1990). Localization of specific mRNAs in *Xenopus* embryos by whole-mount in situ hybridization. *Development* **110**, 325-330.
- Hemmati-Brivanlou, A., de la Torre, J., Holt, C. and Harland, R. M.** (1991). Cephalic expression and molecular characterization of *Xenopus En-2*. *Development* **111**, 715-724.
- Hernandorena, A. and Marco, R.** (1991). Heat induced developmental uncoupling of mesoderm from ectoderm and endoderm germ layers derivatives during *Artemia* postembryonic segmentation. *Roux's Arch. Dev. Biol.* **200**, 300-305.
- Holland, P. W. H.** (1990). Homeobox genes and segmentation: co-option, co-evolution, and convergence. *Sem. Dev. Biol.* **1**, 135-145.
- Holland, P. W. H. and Williams, N. A.** (1990). Conservation of *engrailed*-like sequences during vertebrate evolution. *FEBS Lett.* **227**, 250-252.
- Holland, P. W. H., Williams, N. A. and Lanfear, J.** (1991). Cloning of segment polarity gene homologues from the unsegmented brachiopod *Terebratulina retusa* (Linnaeus). *FEBS Lett.* **291**, 211-213.
- Hui, C., Matsuno, K., Ueno, K. and Suzuki, Y.** (1992). Molecular characterization and silk gland expression of *Bombyx engrailed* and *invected* genes. *Proc. Natl. Acad. Sci. USA* **89**, 167-171.
- Ingham, P. W.** (1988). The molecular genetics of embryonic pattern formation of *Drosophila*. *Nature* **335**, 25-34.
- Ingham, P. W. and Martinez Arias, A.** (1992). Boundaries and fields in early embryos. *Cell* **68**, 221-235.
- Joyner, A. L. and Martin, G. R.** (1987). *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes and Dev.* **1**, 29-38.
- Joyner, A. L., Kornberg, T., Coleman, K. G., Cox, D. R. and Martin, G. R.** (1985). Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila engrailed* gene. *Cell* **43**, 29-37.
- Karr, T. L., Weir, M. D., Ali, Z. and Kornberg, T.** (1989). Patterns of *engrailed* protein in early *Drosophila* embryos. *Development* **105**, 605-612.
- Kassis, J. A., Poole, S. J., Wright, D. K. and O'Farrell, P. H. O.** (1986). Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. *EMBO J.* **5**, 3583-3589.
- Kissinger, C. R., Liu, B., Martín-Blanco, E., Kornberg, T. and Pabo, C. O.** (1990). Crystal structure of an *engrailed* homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590.
- Kornberg, T.** (1981). *engrailed*: a gene controlling compartment and segment formation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **78**, 1095-1099.
- Lawrence, P. and Struhl, G.** (1982). Further studies on the *engrailed* phenotype of *Drosophila*. *EMBO J.* **1**, 827-833.
- Logan, C., Hanks, M. C., Noble-Topman, S., Nallainathan, D., Provart, N. J. and Joyner, A. L.** (1992). Cloning and sequence comparison of the mouse, human, and chicken *engrailed* genes reveal potential functional domains and regulatory regions. *Dev. Genet.* **13**, 345-358.
- Marco, R., Garesse, R., Cruces, J. and Renart, J.** (1991). *Artemia* molecular genetics. In *Artemia biology* (ed. R. A. Browne, P. Sorgeloss and C. N. A. Trotman), pp. 1-19. Boca Raton, Florida: CRC Press.
- Morata, G. and Lawrence, P. A.** (1975). Control of compartment development by the *engrailed* gene in *Drosophila*. *Nature* **255**, 614-617.
- Olson, C. S. and Clegg, J. S.** (1978). Cell division during the development of *Artemia salina*. *Wilhelm Roux's Arch. Dev. Biol.* **184**, 1-13.
- Palmero, I., Renart, J. and Sastre, L.** (1988). Isolation of cDNA clones coding for mitochondrial 16S ribosomal RNA from the crustacea *Artemia*. *Gene* **68**, 239-248.
- Patel, N. H., Martín-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B. and Goodman, C. S.** (1989a). Expression of *engrailed* proteins in arthropods, annelids and chordates. *Cell* **58**, 955-968.
- Patel, N. H., Kornberg, T. B. and Goodman, C. S.** (1989b). Expression of *engrailed* during segmentation in grasshopper and crayfish. *Development* **107**, 201-212.
- Patel, N. H., Ball, E. E. and Goodman, C. S.** (1992). Changing role of *even-skipped* during the evolution of insect pattern formation. *Nature* **357**, 339-342.
- Perez, M. L., Valverde, J. R., Batuecas, B., Amat, F., Marco, R. and Garesse, R.** (1993). Speciation in the *Artemia* genus: mitochondrial DNA analysis of bisexual and parthenogenetic brine shrimps. *J. Mol. Evol.* (in press).
- 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.
- Ptashne, M.** (1988). How eukaryotic transcriptional activators work. *Nature* **335**, 683-689.
- Sambrook, J., Fritsch, E. F. and Maniatis, T.** (1989). *Molecular cloning: a laboratory manual* (2nd edition). Cold Spring Harbor, New York: Cold Spring Harbor Press.
- Sanger, F., Nicklen, S. and Coulson, A. R.** (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- Schram, F. R.** (1986). *Crustacea*. New York: Oxford University Press.
- Schrehardt, A.** (1987). A scanning electron-microscope study of the post-embryonic development of *Artemia*. In *Artemia Research and its Applications* vol. 1 (eds. P. Sorgeloss, D. A. Bengston, W. Dedeir and E. Jaspers), pp. 5-32. Wetteren, Belgium: Universa Press.
- Scott, M. P., Tamkun, J. W. and Hartzell III, G. W.** (1989). The structure and function of the homeodomain. *Biochim. Biophys. Acta* **989**, 25-48.
- Shear, W. A.** (1992). End of the 'uniramia' taxon. *Nature* **359**, 477-478.
- Sommer, R. J., Retzlaff, M., Goerlich, K., Sander, K. and Tautz, D.** (1992). Evolutionary conservation of zinc-finger domains of *Drosophila* segmentation genes. *Proc. Natl. Acad. Sci. USA* **89**, 10782-10786.
- St Johnston, D. and Nüsslein-Volhard, C.** (1992). The origin of pattern and polarity in animal development. *Cell* **68**, 201-219.
- Staden, R.** (1986). The current status and portability of our sequence handling software. *Nucl. Acids Res.* **14**, 217-231.
- Tautz, D. and Pfeifle, C.** (1989). A non radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals a translational control of the segmentation gene *hunchback*. *Chromosoma* **98**, 81-85.
- Vincent, J.-P. and O'Farrell, P. H.** (1992). The state of *engrailed* expression is not clonally transmitted during early *Drosophila* development. *Cell* **68**, 923-931.
- Walldorf, U., Fleig, R. and Gehring, W. J.** (1989). Comparison of homeobox-containing genes of honeybee and *Drosophila*. *Proc. Natl. Acad. Sci. USA* **86**, 9971-9975.
- Wedden, C. J. and Weisblat, D. A.** (1991). Segmental expression of an *engrailed*-class gene during early development and neurogenesis in an annelid. *Development* **113**, 805-814.
- Wedden, C. J., Price, D. J. and Weisblat, D. A.** (1991). Cloning and sequencing of a leech homologue to the *Drosophila engrailed* gene. *FEBS Lett.* **279**, 300-302.
- Weir, M. P. and Kornberg, T.** (1985). Patterns of *engrailed* and *fushi tarazu* transcripts reveal novel intermediate stages in *Drosophila* segmentation. *Nature* **318**, 433-439.