Homologues of the *engrailed* gene from five molluscan classes

C.G. Wray[a], D.K. Jacobs[a,*], R. Kostriken[b], A.P. Vogler[c], R. Baker[d], R. DeSalle[d]

[a]*Department of Biology, University of California, 621 Circle Drive South, Los Angeles, CA 90095-1606, USA*
[b]*Department of Cell Biology and Anatomy, New York Medical College, Valhalla, NY 10595, USA*
[c]*Department of Entomology, The Natural History Museum, Cromwell Road, London, SW7 5BD, UK*
[d]*American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA*

Received 1 March 1995

Abstract **We used the polymerase chain reaction (PCR) to amplify, clone, and sequence 10 *engrailed* homeodomains from 8 species in the five major molluscan classes, including the serially organized chiton (Polyplacophora) lineage. The *Drosophila melanogaster* gene *engrailed* (*en*) is one of several genes involved in embyonic segment polarity determination. Studies of *engrailed* sequence and expression in molluscs are of interest due to questions regarding the evolution and homology of segmentation in these taxa. Nucleotide and deduced amino acid sequence comparisons reflect evolutionary conservation within helices of the *en* homeodomain and ancient divergences in the region 3′ to the homeodomain.**

Key words: engrailed; Mollusc; Homeodomain; Homology; Segmentation; Evolution

1. Introduction

Sequences similar to the homeodomain containing, segment-polarity gene *engrailed* (*en*) of *Drosophila* have been reported from numerous bilaterian taxa (Table 1). Several of the major deuterostome groups, including the tetrapods, bony fish, agnathans and an echinoderm possess *engrailed* homologues. The protostoma are represented by *engrailed* genes from numerous arthropods, two annelids as well as a brachipod. The mollusca represent an ancient and divergent group within which *engrailed*-like genes have not been charaterized. The flatworms (i.e. *Schistosoma*), thought to be the most primitive bilaterian metazoans, are known to have an *engrailed* homologue. *engrailed* sequences from the Cnidaria and Porifera (sponges), now recognized as basal lineages of the metazoan phylogenetic tree [16], are not known.

Comparison of the sequence and expression of *engrailed* genes in the protostomes is particularly interesting. In several protostome phyla the evolution of serial repetition and segmentation (metamerism) remains controversial [17]. Annelid and arthropod expression studies strongly suggest that *engrailed* is a major gene involved in the development of metamerism. During the early development of several arthropods, *engrailed* is expressed in the posterior portion of each segment primordium [18]. In *Drosophila*, the *engrailed* gene interacts with other segment-polarity genes such as *wingless* and *patched* and is expressed in the iterated series of parasegments that form the segmental registry in early development [19–21]. In leech development, *engrailed* is first expressed in iterated rows of cells during germband extension [22]. We have undertaken a study of the *engrailed* sequence in the molluscs to investigate the potential homology and phylogenetic significance of metamerism amongst the protostomes.

Here we report DNA sequence of the *engrailed* gene from eight molluscs representing the five major classes in the phylum. A 232 base pair segment including the *engrailed* homeodomain and an *engrailed* specific region 3′ to the homeodomain was amplified using the polymerase chain reaction (PCR), cloned and subsequently sequenced.

2. Materials and methods

Genomic DNA was prepared from the following species: Scaphopoda, *Dentalium eboreum*, *Cadulus fusiformes*; Bivalvia, *Transannella tantilla*, *Placopecten magellanicus*, *Crassostrea virginica*; Cephalopoda, *Nautilus pompilius*; and Gastropoda, *Ilyanassa obsoleta* using proteinase K digestion followed by CTAB purification [23]. Purified lysates were extracted first with chloroform (due to CTAB step) and subsequently with phenol/chloroform and chloroform. In the chiton (Polyplacophora, *Lepitochiton caverna*) a cesium-chloride gradient was used for nucleic acid purification.

Initial amplification primers for PCR (en-3 and en-5) [24] were used to amplify only the *engrailed* homeodomain region from chiton DNA. Subsequently, a primer specific to a 5′ portion of the chiton sequence (5′-CTTCGTCTAAATGAGTCTCA-3′) was synthesized and used in combination with a degenerate primer that recognizes the *engrailed* specific region 3′ to the homeodomain (5′-TGRTTRTANARNC-CYTGNGCCAT-3′ negative strand; MAQGLYN). The *engrailed* genes for the rest of the taxa were obtained using two degenerate amplification primers (5′-GACAAGCGRCCDMGVACVGCNTT-3′; KRPRTAF; at the 5′ end of the homeodomain) and the 3′ *engrailed* specific primer MAQGLYN mentioned above. Using these primers a 232 base pair target region was amplified in a Perkin-Elmer 480 thermal cycler. Each DNA template required slightly different cycling parameters; however, a general cycle of denaturation, 45 s at 94°C, primer annealing, 45 s at 50°C and extension, 45 s at 72°C for 40 cycles commonly yields amplification of the target segment. *engrailed* amplification is most successful when the template DNA is preheated to the denaturation temperature prior to the addition of the polymerase and the amplification cocktail (hotstart PCR). After electrophoresis and selection of appropriate length amplification products on ethidium bromide stained 1% agarose gels, target fragments were cloned into the pCR-2000 vector (Invitrogen). Following transformation, selected clones were grown overnight and prepared by alkaline lysis mini-preparation [25]. Clones were digested with *Eco*RI and insert sizes were assayed on 1% agarose gels. Multiple clones for each species were sequenced in both directions with Sequenase 2.0 (United States Biochemical) after sodium hydroxide denaturation [26].

3. Results and discussion

Using PCR amplification with degenerate primers ten *engrailed*-like sequences were identified. Fig. 1 exhibits the nucleotide and amino acid sequence of the *engrailed*-like clones for the molluscs sequenced in this study. All clones approximately the length of an intron free *engrailed* target fragment (232 bp) were sequenced in order to search for multiple copies of the

*Corresponding author.

SSDI 0014-5793(95)00372-X

A

```
CHITON                  AGT GAC AGT CAG TTG GAA CGC CTG AAA AAA GAA TTT GAC AGT TCA CGA TAC CTC AGT GAA GCA
SCAPHOPOD Cadulus       TCC A-- G-A --- C-C --T --- --T C-- C-T --G --C --A -CC AGC --T --- --G -CA --- CAG
SCAPHOPOD Dentalium     TCA A-- GAA --- C-C --T --A T-A C-G GTG --G --C --A GC- -GT A-G --- --- -CG --- ACC
BIVAVLE Transennella    -CA ACA G-A --- C-C CG- AAA --- --- G-T --- --C --G GC- AAC AAA --T --- TCC --G AAC
BIVALVE Placopecten     -CG A-- G-C --A --A C-- --T --- --G CGC --- --- --A GAA -GT --C --T --T -C- --G CAG
BIVALVE Crassostrea     -CA ACA G-C --A C-C C-G A-A --A --G TCG --G --- --G GAG AAC -AT --- --- -CA --G AAG
CEPHALOPOD Nautilus A   -CA AG- G-C --- C-- C-G TAC --- --G --G --G --- --A GAA -GC --C --- --G -CG --G -AC
CEPHALOPOD Nautilus B   -CA TCG G-G --- C-- C-G --G --- -GG -GG --G --- --G GCG GGG AAA --- --G -CC --- -AT
GASTROPOD Ilyanassa A   -CC AG- GAA --- C-A TCT --T --- --G CG- --G --- --- GAG -GT --C --- --G -CG --G A-G
GASTROPOD Ilyanassa B   -CC AG- GAA --- C-A TCT --T --- --G CG- --G --- --A GAG -GT --T --- --G -CG --G A-G
DROSOPHILA en           TCC AG- GAG --- --- -CC --- --T --G CGG --G --C A-- GAG AAT --- --T CTG -CC --G CGG
LEECH ht-en             AGC -G- GA- --- C-- -CG A-G T-- --G CGT --- --C AG- GAG AAC AA- --- --G -CG --G CAG
MOUSE EN-1              ACG GCC GAG --- C-- C-G AGA --C --G GCG --G --C C-G GCA AAC --C --T A-- -CG --G CAG

CHITON                  AAA AGA CAG GAT TTA GCC AGA GAA CTT CGT CTA AAT GAG TCT CAA ATA AAA ATT TGG TTC CAA
SCAPHOPOD Cadulus       CG- --- AGA A-- C-C AGC -TG --- --G GAC --- TCG --A A-A --- --- --- --A --- --- ---
SCAPHOPOD Dentalium     CG- C-- A-A --C --- T-- GAC --G --G AAG --C TCG --A A-T --- --T --- --C --- --- --G
BIVAVLE Transennella    -G- C-- --- C-A C-G --G CAT --- --G AAC T-- --C --A --A --G --- --- --C --- --T --G
BIVALVE Placopecten     -G- C-- -TA --- C-C --G CAG --G --C AAC T-G -CA --A G-- --- --T --- --A --- --- --G
BIVALVE Crassostrea     -GG --G TTA --A C-- T-- GAG --- --C AAA --G TCA --- --- --- --C --- --A --- --- ---
CEPHALOPOD Nautilus A   CGC C-C -G- A-G C-G --T C-G --G --G A-C --- -GC --- G-- --G --C --- --C --- --- --G
CEPHALOPOD Nautilus B   CGC C-G --A ACA C-G --- -AG --G --G G-- --G --C --- --C --G --C --- --C --- --T --G
GASTROPOD Ilyanassa A   CGG --G -GA C-C C-G --G GCG --G --- G-G --G -CG --- --C --G --C --G --C --- --- --G
GASTROPOD Ilyanassa B   CGG --G -GA C-C C-G --G GCG --G --- G-G --G -CG --- --C --G --C --G --C --- --- --G
DROSOPHILA en           -G- C-C --- --G C-G A-- --C --G T-G G-C --G --C --- G-G --G --C --G --C --- --- --G
LEECH ht-en             -GG --- ACA TG- C-G --G -AG --- --G AAC T-G --C --- AGC --G --C --- --C --- --- --G
MOUSE EN-1              CGG C-- --- ACC C-C --- CAG --G --C A-C --G --- --- --C --G --C --G --C --- --- ---

CHITON                  AAT AAA CGA GCC AAA TTG AAA AAA TCA ACC AGT GGT CGA ACA xxx GGA CTA GCG CTA CAT TTG
SCAPHOPOD Cadulus       --C --G --- --A --G C-- --G --G A-C GG- G-- --A TCC -GC xxx -AC --A --- -A- --- C-C
SCAPHOPOD Dentalium     --C --- A-- --A --- C-A --G --- A-- GGA --C AAC A-C C-- CCG TCT --G --- TTG --C C--
BIVAVLE Transennella    --- --- A-- --T --- A-A --G --- -GT -A- G-G -A- AA- -AT xxx --- --G --- A-G G-A C--
BIVALVE Placopecten     --C --G --- --T --G A-A --G --- --G GTG GCG CCC A-- -AT xxx ACT --C --T --C AGC C-T
BIVALVE Crassostrea     --C --- --T --T --- A-C --G --- --- --G G-A --C AA- -AC xxx AC- T-G --T T-G A-A ---
CEPHALOPOD Nautilus A   --C --- --C --A --- A-- --G --- AAT -G- GTC ACA AA- -AT xxx CGC T-G --A --T --C ---
CEPHALOPOD Nautilus B   --C --G --- --A --- A-- --- --- GT- GGA G-A ATA AA- -AT xxx CT- --G --C A-G --C C--
GASTROPOD Ilyanassa A   --C --G --G --- --G A-C --G --G --G T-G G-C -TG AAG -AC xxx -AG --G --C A-G --G C--
GASTROPOD Ilyanassa B   --C --G --G --- --G A-C --G --G --G T-G G-C -TG AAG -AC xxx -AG --G --C A-G --G C--
DROSOPHILA en           --C --G --G --- --G A-C --G --G --G --G G-C TCC AA- -AT xxx CCG --G --A --G --G C--
LEECH ht-en             --C --G A-G --- --G A-- --G --G G-G -GT G-C -TG AAG -AT xxx CAG T-G --T --G --A C-C
MOUSE EN-1              --C --G --T --- --G A-C --G --- G-C --A G-C ATC AAG -AC xxx --C --G --- --G --C C-C
```

B

```
                        1         10        20        30        40        50        60
                        |         |         |         |         |         |         |
CHITON                  SDSQLERLKEEFDSSRYLSEAKRQDLARELRLNESQIKIWFQNKRAKLKKSTSGRT GLALHL
SCAPHOPOD Cadulus       -NE--D--QH--ET----T-QR-RN-SM--D-S-T--------------TGG-SS D--Q--
SCAPHOPOD Dentalium     -NE--D--QV--EAC---T-TR-K--SD--K-S-T--------------TG-NSPPS-----
BIVALVE Transennella    TTE--RK--D--EANK----NR--Q--H--N---------------I--CNGDKN ---ME-
BIVALVE Placopecten     TND--Q---R--EEC---T-QR-L---Q--N-T-A-----------I---VAP-N T---S-
BIVALVE Crassostrea     TTD--Q---S--EENH--T-KR-LE-SE--K-S-------------I----G-KN T---K-
CEPHALOPOD Nautilus A   TSD--QY-----EEC---T-DR-RK-----S-S-A-----------M--NSVTKN R-----
CEPHALOPOD Nautilus B   TSE--Q--RR--EAGK--T-DR--T--K--G---------------M--VGGIKN L--M--
GASTROPOD Ilyanassa A   TSE--S---R---EC---T-TR-RH--A--G-T-------------I---SGVKN E--MQ-
GASTROPOD Ilyanassa B   TSE--S---R--EEC---T-TR-RH--A--G-T-------------I---SGVKN E--MQ-
Schistosoma             TVP--K--SQ--EKN---D-LR-KK--T--D-R---V----------T--ASGAQN C-----
Artemia                 TAE--S---H--NEN---T-RR--------G-H-N--------N--------SGQKN P---Q-
Tribolium               -GA--A---E--AEN---T-RR--Q-SAQ-G---A-----------LI--ASGTKN P---Q-
DROSOPHILA en           -SE--A---R--NEN---T-RR--Q-SS--G---A-----------I----GSKN P---Q-
LEECH ht-en             TGD--A---R--SENK--T-QR-TC--K--N---------------M--ASGVKN Q---Q-
MOUSE EN-1              TAE--Q---A--QAN--IT-QR--T--Q--S---------------I--A-GIKN ------
```

Fig. 1. (A) Nucleotide sequences (internal to primers) from 10 molluscan *engrailed*-like genes cloned after PCR amplification. Mollusc sequences are aligned to *Drosophila en*, mouse *En-1* and leech *ht-en* for reference. Dashes indicate identity to top line of sequence. × indicates gap in the alignment. (B) Deduced amino acid sequences. Homeodomain residues are underlined in top sequence, α-helices are overlined. Boxed glycine residues designate those sequences/proteins which are or could be successfully recognized by the mAb 4D9 antibody. Additional metazoan amino acid sequences are included for reference.

gene. Wagner et al. [27] point out that amplification bias of one gene cognate over another (PCR selection) can be driven by differential thermal stability of primer–template duplexes in different gene cognates. As a consequence of this possible bias, as well as the presence of introns in *engrailed* genes in several taxa, the sequences reported may not include all the *engrailed*-like gene sequences in the taxa surveyed. In addition, after numerous attempts we have been unable to obtain *engrailed* genes from the squid, *Loligo* (Cephalopoda).

Two copies of *engrailed*-like genes were isolated for the cephalopod, *Nautilus pompilius*. These genes differ from one another at 19 amino acid positions, 13 of which are within the homeo-

domain. This large difference between the two *Nautilus* genes suggests the presence of two copies of the *engrailed* gene. In fact the differences between the two *Nautilus* genes are substantially larger than those observed in the same homeodomain region of the fly (*engrailed* vs. *invected*) [5] or mouse (*en*-1 vs. *en*-2) [1] where 8 and 7 amino acid differences, respectively, are found. The only other intraspecific variation observed occured in the Gastropod, *Ilyanassa obsoleta*. In *Ilyanassa*, variation is evident in two third positions, one of which results in a change from aspartate to glutamate at position 21 of the homeodomain. Based upon sequence comparison, this amino acid substitution, near the end of the first helix, does not appear to be significant. The minor variation in *Ilyanassa* may be allelic in nature and may not signify the presence of two genes. Single copies of the gene were isolated from the scaphopods, bivalves, and the chiton. The *Dentalium* (Scaphopod) *engrailed*-like clones exhibit an unusual additional codon in the variable region downstream of helix 3 and 4. Sequence data from both strands of all clones documents this unusual feature of the *Dentalium* sequence.

Comparison of the molluscan *engrailed* nucleotide sequence indicates that significant intraclass and interclass divergence has accumulated since the Cambrian appearance of this phylum. As expected, nucleotide substitution rates are variable in relation to codon position: 59% of first positions, 45% of second positions, and 98% of third positions exhibit variation within the molluscs. Across the molluscs sampled, silent substitution occurs at 20% of first positions and 49% of third positions. Levels of sequence divergence within the three bivalves is comparable to levels within the molluscs as a group (Fig. 1A).

Amino acid substitutions within the molluscs are similar to the variation in the other phyla sequenced to date. Within the molluscs we observe no variation in the third helix, the major groove binding 'recognition' portion of the homeodomain. The only variation in the 4th helix occurs at position 48 (Fig. 1B) where isoleucine, leucine, and methionine all occur. Variation in the rest of the homeodomain is constrained to positions that typically show substitution in other known *engrailed* homeodomains. In the region obtained by our PCR approach the greatest variation occurs between the end of the homeodomain and the 3′ *engrailed* specific region. This portion of the gene appears to be slightly more divergent in molluscs than within the vertebrates or insects. Higher divergence within the group may reflect the greater antiquity of many of the branching events within the mollusca.

The sequence data we have obtained can demonstrate the presence of the appropriate epitope for antibody studies. Of the *engrailed* antibodies the epitope of the mAb 4D9 [8] monoclonal antibody, generated from the *Drosophila* protein, has been characterised. The mAb 4D9 eptitope is localised to the turn sequence between helix 2 and 3 of the homeodomain [8]. When a glycine is present in the middle of the turn sequence (position 31 in Fig. 1B) the mAb 4D9 antibody is known to bind successfully [8,9]. When asparagine, serine or threonine occur at this position antibody binding does not occur [8,22]. The presence of a glycine at this position in the snail and in one of the two *Nautilus* sequences suggest that these taxa may be fruitfully examined using the mAb 4D9 antibody. The serine in the second *Nautilus* sequence suggests that binding is unlikely. The other molluscs contain amino acids in the turn region that have not been previously examined using the mAb 4D9 antibody.

Table 1
Taxonomic diversity of *engrailed* sequences [1–15]

Common name		No. gene cognates	No. independent clones surveyed	Accession no.
	Deuterostoma			
Human	*Homo sapiens*	2		L12698-700
Mouse	*Mus musculus*	2		Y00201
Chicken	*Gallus gallus*	2		L12694
Frog	*Xenopus laevis*	2		X59123-24
Zebrafish	*Brachydanis rerio*	3		X59125-26
Lamprey	*Lampetra planeri*	1		X59122
Hagfish	*Myxine glutinosa*	2		X59120-21
Sea urchin	*Tripneustes gratilla*	1		M19709
	Protostoma			
Fly	*Drosophila melanogaster*	2*		K03055-58
Moth	*Bombyx mori*	2*		M64335-36
Bee	*Apis mellifera*	2		M29490
Beetle	*Triboleum castaneum*	1		NA
Grasshopper	*Schistocerca americana*	1		M29262
Brine shrimp	*Artemia franciscana*	1		X70939
Leech	*Helobdella triserialis*	1		X58692
Polychaete	*Ctenodrilus serratus*	1		NA
Nematode	*Caenorhabditis elegans*	1		NA
Brachiopod	*Terebratulina retusa*	1		X62688
Flatworm	*Schistosoma mansoni*	1		M85305
	Mollusca (this study)		No. independent clones surveyed	
Chiton	*Lepitochiton caverna*	1	4	U21675
Scaphopod	*Cadulus fusiformis*	1	2	U23153
Scaphopod	*Dentalium eboreum*	1	2	U23154
Clam	*Transannella tantilla*	1	4	U23212
Oyster	*Crassostrea virginica*	1	6	U23214
Scallop	*Placopecten magellanicus*	1	6	U23213
Nautilus	*Nautilus pompilius*	2	8	U23431 U21857
Mud snail	*Ilyanssa obsoleta*	1	6	U23432-33

*Gene cognates represent *engrailed* and *invected*.

Acknowledgements: We thank Amelie and Rudy Scheltma for assistance in collecting *Ilyanassa*, Dave Lindberg and Marta de Maintenon for assistance in collecting *Transanella*, Charles Marshall and Bruce Runnegar for providing *Cadulus* tissue, and Timothy Collins for providing *Dentalium* DNA. This work was supported by NASA exobiology Grant 3312 to D.K.J.

References

[1] Logan, C., Hanks, M.C., Noble-Topham, S., Nallainathan, D., Provart, N.J. and Joyner, A.L. (1992) Dev. Genet. 13, 345–358.
[2] Joyner, A.L., Kornberg, T., Coleman, K.G., Cox, D.R. and Martin, G.R. (1985) Cell 43, 29–37.
[3] Holland, P.W.H. and Williams, N.A. (1990) FEBS lett. 277, 250–252.
[4] Dolecki, G.J. and Humphreys, T. (1988) Gene 64, 21–31.
[5] Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) Cell 40, 37–43.
[6] Hui, C. Matsuno, K., Ueno, K. and Suzuki, Y. (1992) Proc. Natl. Acad. Sci. USA 89, 167–171.
[7] Fleig, R. (1990) Roux's Arch. Dev. Biol. 198, 467–473.
[8] Patel, N.H., Martin-Blanco, E., Coleman, K.G., Poole, S.J., Ellis, M.C., Kornberg, T.B. and Goodman, C.S. (1989) Cell 58, 955–968.
[9] Brown, S.J., Patel, N.H. and Denell, R.E. (1994) Dev. Genet. 15, 7–18.
[10] Manzanares, M. Marco, R. and Garesse, R. (1993) Development 118, 1209–1219.

[11] Wedeen, C.J., Price, D.J. and Weisblat, D.A. (1991) FEBS Lett. 279, 300–302.
[12] Dick, M.H. and Buss, L.W. (1994) Mol. Phylo. Evol. 3, 146–158.
[13] Holland, P.W.H., Williams, N.A. and Landfear, J. (1991) FEBS Lett. 2, 211–213.
[14] Burglin, T.R., Finney, M., Coulson, A. and Ruvkun, G. (1989) Nature 341, 239–243.
[15] Webster, P.J. and Mansour, T.E. (1992) Mech. Dev. 38, 25–32.
[16] Wainwright, P.O., Hinkle, G., Sogin, M.L. and Stickel, S.K. (1993) Science 260, 340–342.
[17] Eernisse, D.J., Albert, J.S. and Anderson, F.E. (1992) Syst. Biol. 41, 305–330.
[18] Patel, N.H., Kornberg, T.B. and Goodman, C.S. (1989) Development 107, 201–212.
[19] Kornberg, T., Siden, I., O'Farrell, P.H. and Simon, M. (1985) Cell 40, 40–53.
[20] Fjose, A., McGinnis, W.J. and Gehring, W.J. (1985) Nature 313, 284–289.
[21] DiNardo, S., Kuner, J.M., Theis, J. and O'Farrell, P.H. (1985) Cell 43, 59–69.
[22] Wedeen, C.J. and Weisblat, D.A. (1991) Development 113, 805–814.
[23] Clark, C.G. (1992) in: Protocols in Protozoology (Lee, J.J. and Soldo, A., eds.) pp. D3.1–D3.2, Allen Press.
[24] Kamb, A., Weir, M., Rudy, B., Varmus, H. and Kenyon, C. (1989) Proc. Natl. Acad. Sci. USA 86, 4372–4376.
[25] Sambrook, J. Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press.
[26] Toneguzzo, F., Glynn, S., Levi, E., Mjolsness, S. and Hayday, A. (1988) BioTechniques 6, 460–469.
[27] Wagner, A., Blackstone, N., Cartwright, P., Dick, M., Misof, B., Snow, P., Wagner, G.P., Bartels, J., Murtha, M. and Pendleton, J. (1994) Syst. Bio. 43, 250–261.