3

emx-Am, a cnidarian ortholog of the *Drosophila* gene *empty spiracles*

3.1. INTRODUCTION

The most visible modification in the evolution of metazoan body plans was the organisational transition from radial to bilateral symmetry that accompanied the creation of a second body axis. The molecular bases for patterning and specification of some D/V and A/P axes of bilaterians are well understood, but the evolutionary origins of the major body axes are equivocal (Willmer, 1990). This uncertainty is, in part, due to the unresolved relationship between the major longitudinal axes of the Cnidaria and Bilateria. Despite the initial assumption that the O/A axis of cnidarians corresponded to the A/P axis of bilateral animals (Brusca and Brusca, 1990; Ruppert and Barnes, 1991), given that each represents the major axis of symmetry in the Cnidaria and Bilateria respectively, recent data cast doubt on this idea. The *cnox-2Am* gene, a *Gsx* ortholog in the cnidarian *A. millepora*, is expressed in a restricted subset of neurons along the O/A axis that is reminiscent of the D/V-restricted expression of orthologous genes in the developing fly and vertebrate CNS (Hayward et al., 2001). Although no clear Hox orthologs have yet been found in non-bilateral animals, members of other key components of the bilaterian A/P patterning systems have been cloned from cnidarians (Mokady et al., 1998; Muller et al., 1999; Schummer et al., 1992; Smith et al., 1999). Amongst these is an ortholog of *Drosophila empty spiracles (ems)* that has been isolated from the hydrozoan, *Hydractinia symbiolongicarpus*. *Cn-ems* is expressed in the 'head' of the adult polyp (Mokady et al., 1998) and thus may be involved in O/A axis patterning. In more elaborate animals, the *ems/Emx* gene family is involved in regionalisation and specification of anterior structures and the CNS. Given its apparent role in O/A development in *Hydractinia*, it is possible that its function in anterior patterning is an ancestral feature of this gene family. However, it is during embryonic stages that *ems/Emx* genes are involved in patterning the A/P axis in bilaterians. Cnidarian embryonic data are required for more informative zootype-like comparisons with higher animals.

In order to clarify relationships between the major body axes of the Cnidaria and Bilateria, the *A. millepora* ortholog of the *Drosophila ems* gene, *emx-Am*, was characterised. With its accessible stages of embryogenesis, *A. millepora* is a more appropriate organism than *Hydra* in which to study cnidarian orthologs of genes involved in early developmental patterning pathways. Comparative analyses of the expression domains of this gene in *A. millepora* will help promote an understanding of the ancestral functions of this gene family and the evolutionary origins of the major bilaterian axes.

3.1.1. The *ems*/*Emx* Gene Family

Empty spiracles (ems) is a cephalic gap gene that was first identified in *Drosophila* during screens for zygotic patterning mutations (Cohen and Jürgens, 1990; Cohen and Jürgens, 1991; Jürgens et al., 1984). It encodes a homeodomain-containing transcription factor that is required for correct head formation, brain regionalisation and tracheal system development in the fly (Dalton et al., 1989; Walldorf and Gehring, 1992). A second orthologous gene, *E5*, has also been identified in *Drosophila* but its function is unclear (Dalton et al., 1989). A head-specific pattern of expression is initiated at the syncytial blastoderm stage in a single circumferential stripe at the anterior pole of the embryo (Dalton et al., 1989; Walldorf and Gehring, 1992). Like other gap genes, it is under the control of maternal positional information (Dalton et al., 1989; Grossniklaus et al., 1994; Walldorf and Gehring, 1992) and is expressed in cells fated to form a number of head structures including the antennal sense organs. As gastrulation begins, *ems* expression is detected in the developing cephalic region. An additional *ems* expression domain is also detected in bilateral patches of ectodermal and neural cells in all trunk segments (Dalton et al., 1989; Walldorf and Gehring, 1992), where it is essential for specific interneurons to establish correct axon pathways in the posterior brain and ventral nerve cord (VNC) (Hartmann et al., 2000). Mutations of *ems* produce a gap-like phenotype in the anterior CNS, which leads to a deletion of the deutocerebral and tritocerebral neuromeres of the embryonic brain, and loss of cuticular and cephalic sensory structures as well as the filzkörper (filters present in the tracheal tube primordial) (Cohen and Jürgens, 1990; Hirth et al., 1995). *ems* mutations also lead to axogenesis defects in the VNC (Hartmann et al., 2000).

In the mouse, two *ems*-related genes (*Emx1* and *Emx2*) have been described. Both

genes are expressed in the dorsal telencephalon and olfactory regions of the embryonic brain (Simeone et al., 1992a; Simeone et al., 1992b), which correspond to the *ems*-expressing neuromeres of the developing fly brain and anterior sensory structures (Cohen and Jürgens, 1990; Hirth et al., 1995). Results of mutational analyses of *Emx1* and *Emx2* in the mouse are much weaker than the dramatic deletions observed in the fly. *Emx1*$^{-/-}$ mutants display minor defects in the corpus callosum in only some animals (Qui et al., 1996), while the majority of *Emx2*$^{-/-}$ mice die at birth due to defects in the urogenital system, a secondary site of *Emx2* expression in the mouse (Miyamoto et al., 1997). These mutant mice also display minor brain abnormalities that include the deletion of the dentate gyrus and a reduction in the size of the hippocampus (Pellegrini et al., 1996; Yoshida et al., 1997). It is possible that a derived functional redundancy operating between the mouse *Emx* genes is responsible for the mild defects observed in mutants. Double mutant experiments would be useful to determine the combined action of these genes.

Genes orthologous with mouse *Emx1 and Emx2* have been characterized in the zebrafish, *Xenopus* and chicken (Bell et al., 2001; Morita et al., 1995; Pannese et al., 1998). A third vertebrate *Emx* gene class was recently identified in the dogfish *Scyliorhinus canicula* (*ScEmx1*, *ScEmx2* and *ScEmx3*), suggesting a duplication of the ancestral gene in the vertebrate lineage produced the three extant *Emx* classes (Derobert et al., 2002). Although only two zebrafish *Emx* genes are known, one of these (*Emx1*; see Figure 3.4) clearly corresponds to *Emx3* from the dogfish; this gene type is present in pufferfish but not in any tetrapod (Derobert et al., 2002), so it seems likely that the third Emx class arose within the Chondrichthyes / Osteichthyes after their divergence from the line leading to the tetrapods. All gnathostome *Emx* genes are expressed in the developing dorsal telencephalon but also show additional patterns of expression in other cephalic regions. Tetrapod *Emx2* and *ScEmx3* genes are also expressed in the dorsal prosencephalon and (along with *ScEmx2*) in the olfactory placodes (Bell et al., 2001; Derobert et al., 2002; Pannese et al., 1998; Simeone et al., 1992a; Simeone et al., 1992b). The mouse *Emx2* gene also functions in the developing urogenital system (Miyamoto et al., 1997) as do all three dogfish *Emx* genes (Derobert et al., 2002), but a similar function is lacking in the fly, suggesting this is a derived characteristic of the vertebrate genes. Like its gnathostomal homologs, the lamprey *ems*/*Emx* gene, *Ljemx*, is expressed in the developing brain once major subdivisions become apparent within

the dorsal telencephalon (Myojin et al., 2001). Unlike all other *Emx* genes that have been studied, *Ljemx* and *ScEmx2* also have mesodermal expression domains. While the lamprey gene is expressed throughout the entire paraxial mesoderm, expression of the dogfish gene is restricted to the cephalic paraxial mesoderm (Derobert et al., 2002; Myojin et al., 2001). It is possible that this expression domain was a primitive vertebrate characteristic lost in the gnathostomes.

While conserved patterns of expression of *ems*/*Emx* genes in *Drosophila* and vertebrate embryos have led to the suggestion that the *ems*/*Emx* gene family is part of an evolutionarily conserved genetic cascade that is involved in formation and regionalisation of the developing CNS, the extent to which this is true is still unclear. In fact, data for the *ems*-related gene *Hremx* from the ascidian *Halocynthia roretzi* (a urochordate) do not support this hypothesis. Although some A/P patterning mechanisms in the CNS appear to be conserved between ascidians and vertebrates, (Wada et al., 1998; Wada et al., 1996), *Hremx* is not expressed in the larval CNS (Oda and Saiga, 2001). Expression of *Hremx* was detected in the anterior tip of the trunk and in the lateral tail epidermis (Oda and Saiga, 2001), a pattern reminiscent of *Drosophila ems* expression in the bilateral patches of epidermal cells within the trunk segments (Dalton et al., 1989). Given that ascidians are atypical in many aspects of their developmental biology (Katsuyama et al., 1995; Locascio et al., 1999), it is likely that the absence of CNS expression reflects secondary loss. In terms of ancestral roles within the chordates, the Amphioxus *Emx* genes, *AmphiEmxA* and *AmphiEmxB*, are likely to be informative, but expression data are not yet available (Minguillon et al., 2002; Williams and Holland, 2000).

In 1998, Mokady *et al* demonstrated that *Cn-ems*, from the cnidarian *Hydractinia symbiolongicarpus*, is associated with the formation of the anterior animal pole, but is not expressed in the developing nervous system. Expression of *Cn-ems* in the adult polyp is limited to the endodermal epithelial cells (or digestive cells) of the hypostome in a radially organised pattern of longitudinal stripes. Endodermal expression of *ems*/*Emx* has not been reported for any other organism. During hypostome formation in metamorphosing planulae, stage-specific differences in levels of *Cn-ems* expression were observed, but the spatial distribution of *Cn-ems* transcripts was not examined (Mokady et al., 1998). While these findings suggest *ems* orthologs were associated with

head structures before the evolution of bilateral symmetry, they do not provide clues to the ancestral origins of their function in nervous system development. Analysis of *Cn-ems* spatial and temporal expression during early embryogenesis, which has not been reported, would be more appropriate for zootype-like comparisons with bilateral animals.

3.1.2. Statement of Goals

The main aim in characterising the *Acropora* ortholog of the *ems* gene, *emx-Am*, is to clarify the relationship between the major longitudinal axes of radially and bilaterally symmetrical organisms. Data for *cnox2Am*, the *Acropora Gsx* ortholog, suggest that the O/A axis of the planula may correspond to the D/V axis of bilateral animals (Hayward et al., 2001). Although cnidarians lack convincing equivalents of Hox genes, other genes known to play major roles in specifying the A/P axis are present, including *Emx* orthologs. By investigating the role of *emx-Am* during embryonic development, the hope is to better understand the evolution of axis-specification systems. In addition, because *Acropora* represents the basal cnidarian class, the Anthozoa, examination of the expression pattern of *emx-Am* is likely to be important in understanding the evolution of function in this gene family. In particular, it may answer the question of whether a role in nervous system patterning and development is ancestral.

Mutational analyses of the fly and mouse genes have helped determine the functional role of *ems*/*Emx* genes in early metazoan development, but the regulation of this gene family is less clear. Only a small number of regulators of *ems*/*Emx* activity have been described, and relatively few targets of Ems/Emx transcription factors are known. For this reason, I proposed to identify potential targets of *Acropora* Emx-Am, using the yeast-one hybrid system to screen both coral and *Drosophila* reporter libraries. A coral reporter library was constructed with *Acropora* genomic DNA and clones containing possible target regions of the Emx-Am homeodomain, identified in the yeast-one hybrid screen, were isolated from the *Acropora* genomic library. However, in the absence of a complete *Acropora* genome sequence, characterisation of potential regulatory regions was a time-consuming process. With the benefit of a completely sequenced *Drosophila* genome, it would have been possible to assign positive clones from the fly library to candidate target genes. Ultimately, however, the fly reporter library was not supplied,

and the screen for potential binding sites was limited to the coral reporter library.

3.2. RESULTS

3.2.1. The *emx-Am* cDNA

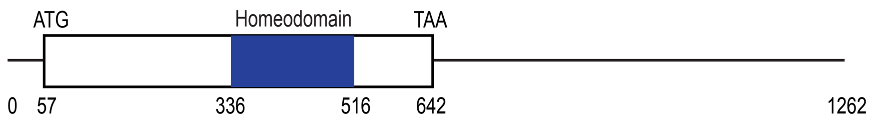
The *emx-Am* cDNA clone was isolated before the start of this PhD project, from a λZAP-II *Acropora millepora* library generated using mRNA from pre-settlement stage embryos at approximately 120 h post-fertilisation. A 168 bp PCR product generated by Ms. Heather Dodd using degenerate homeobox primers (F: 5'-AARCC NAARATHCGNACNGCNTWY-3'; R: 5'-YTTNGTNCGNCGRTTYTGRAACCA-3') was used by Ms. Dodd and Dr. David Hayward (ANU) to screen the library without success. Subsequent screening of the library by Ms. Patricia Pontynen led to the identification of a single positive clone from a screen of 4.2 x 10^5 clones. Upon basic BlastX analysis (Altschul et al., 1990), the cDNA clone originally isolated as ems-6 showed high sequence identity within the homeobox region of the *Drosophila ems* sequence, and is henceforth referred to as *emx-Am*.

The nucleotide sequence and conceptual translation of *emx-Am* cDNA are presented in Figure 3.1. The nucleotide sequence of the *emx-Am* cDNA is 1262 bp, comprising 57 bp of 5'- untranslated region, 585 bp of open reading frame and 620 bp of 3' untranslated region. Conceptual translation of the open reading frame (ORF) of the *emx-Am* cDNA results in a putative protein, Emx-Am, of 195 amino acids. Translation initiation of the Emx-Am protein is predicted to occur at the first encoded methionine at base pairs 57-59. While no upstream in frame termination codons are present in the cDNA sequence, two consecutive in-frame stop codons are present in the genomic locus, located 72 bp upstream of the start of the cDNA.

3.2.2. The Emx-Am protein

The homeodomain in the Emx-Am is located in the N-terminal half of the protein, at amino acid residues 94 to 153. The Emx-Am homeodomain is clearly a member of the Ems/Emx class. Overall, the Emx-Am homeodomain most closely resembles that of amphioxus AmphiEmxA and several vertebrate Emx-1 orthologs (>81% identity; see Figure 3.2A). Relative identity with vertebrate Emx-2 orthologs and other invertebrate sequences (including the *Hydractinia* sequence) was somewhat lower (Figure 3.2A).

A



B

```
                                                            M  H  A  Y  N  S  S  Q  S  S  F  I  N  S      14
1    AGAAGACCACAGCTCTCTAACCACAGGCTCAAACTGCAGCTACCTTTTGGACCGAAGATGCACGCGTACGACAGCTCTCAGTCGTCCTTCATCAATTCCA

     M  Y  R  P  N  S  T  I  Q  P  P  N  A  S  Q  S  Y  A  S  Y  K  S  Q  Y  E  Y  P  M  A  V  R  F  D  S   48
101  TGTATCGACCTAACTCTACAATTCAGCCTCCAAACGCATCACAAAGCTACGCTAGCTACAAATCGCAAACAGAGTACCCAATGGCTGTAAGATTCGATTC

      H  P  L  H  S  L  S  A  Y  N  S  G  S  F  S  F  Y  P  C  A  S  S  H  R  Y  V  H  D  L  N  C  E  Y    81
201  GCATCCCTTACACAGCCTTTCGGCCTACAACTCAGGTTCTTTTTCATTCTATCCGTGCGCCTCAAGTCATCGATATGTACACGATTTAAACTGCGGCTAC

      T  S  G  A  L  H  S  G  F  V  A  C  R  R  P  K  R  I  R  T  A  F  T  P  T  Q  L  L  H  L  E  N  A   114
301  ACTTCCGGTGCGCTCCACAGTGGTTTTGTAGCATGCCGACGACCAAAACGAATACGAACCGCCTTCACACCCACTCAGCTTTTGCATTTGGAAAATGCGT
                                   G
       emx.F                                              emx.R

     F  E  K  N  H  Y  I  V  G  T  E  R  K  Q  L  A  S  F  L  N  L  S  E  T  Q  I  K  V  W  F  Q  N  R  R   148
401  TCGAAAAGAACCATTACATTGTCGGAACGGAGCGCAAACAACTCGCGTCCTTTCTTAATCTCTCCGAGACTCAAATCAAGGTCTGGTTTCAGAACCGGAG

       T  K  W  K  R  Q  Q  A  E  E  K  A  A  A  Q  P  T  K  S  S  S  A  G  M  S  I  V  E  R  H  N  L  N   181
501  AACCAAATGGAAACGTCAACAAGCAGAAGAGAAAGCTGCGGCTCAACCCACGAAAAGTTCCTCTGCTGGAATGAGCATAGTTGAACGCCATAACTTGAAC

      T  G  Q  M  S  K  E  E  S  D  H  I  V  *                                                              195
601  ACTGGTCAAATGTCGAAGGAAGAGAGCGACCACATTGTGTAATTGTGTGTGCATCAGAAGTTTTCCAGAGATTGCGATTTTAAGGACACTAGTGTAGGCT

701  TCCAATCCGAACAGAGAAACGCAGCGTAGAACATCAAACTTCACATGTACTGCTTTCCAGTCATAAAACCTCCGCCTGAGACACAAGACCAGAACTTCTC

801  CTAGCTGACATTGACGAAATTAAATACCGAATATCTTAAAGGAAAAGCAGAGTACACAGTATAAAACTGCATTCAGGCGACTTTTCAAATTTCCTACAAA

901  ATGTATTTCCCTCAGGTTGTAAAATAGAAAAGCTTGTATTCAATGTATCGAGGGAAACACTCAGTAAAAAAGACAAAAGATTAAAAGTACTATATTCCGC

1001 ATGGAATTTGGTTAGACTAGAAAAATGTAGAACAAAATATGTAGCCTTCAACACCCTGAAAACTGAGACTGAGGGTGGCTGAAAATTGTTAATAAAGTAA

1101 TCTGACAGAAAATAAAAATTCGGAGTAAAGAACGCTTCATGCAACTTCTTCAAAGAGACGAAACTCGAAATACGGATGCTAGAGAAATTGTAAGAAAAAT

1201 GGAATGCTTTGCCAAGATATGAAATAAACGATTTTACAAACTGTAAAAAAAAAAAAAAAAAA
```

C

Exon 1	Intron	Exon 2
TCCACAGTG	gtaagatctc...(381bp)...ctctctgcag	GTTTTGTA

Figure 3.1: The *emx-Am* cDNA. (A) Schematic representation of the *emx-Am* cDNA. The single black line indicates the non-coding regions of the cDNA, the open box indicates the coding region and the homeodomain region is indicated by the blue box. The start (ATG) and stop (TAA) codons are shown at the top. The numbers beneath the box represent the nucleotide sequence. (B) The nucleotide sequence of the *emx-Am* cDNA and predicted amino acid sequence of the protein. The homeodomain residues are highlighted in blue. An asterisk indicates the 3' in-frame stop codon. The intron position is shown by a red triangle. A single base pair polymorphism between the cDNA and genomic sequence is indicated at position 330 (A/G). Numbers on the left side represent the nucleotide sequence; numbers on the right side represent the amino acid sequence. Primers (emx.F and emx.R) used in the real-time quantitative PCR are shown by arrows beneath the nucleotide sequence. The *Sph*I (GCATGC) and *Hind*III (AAGCTT) restriction sites are boxed. (C) The intron-exon boundary of the single intron in *emx-Am*. The GT-AG consensussplice sites are indicated in red. The total size of the intron is shown in brackets

□

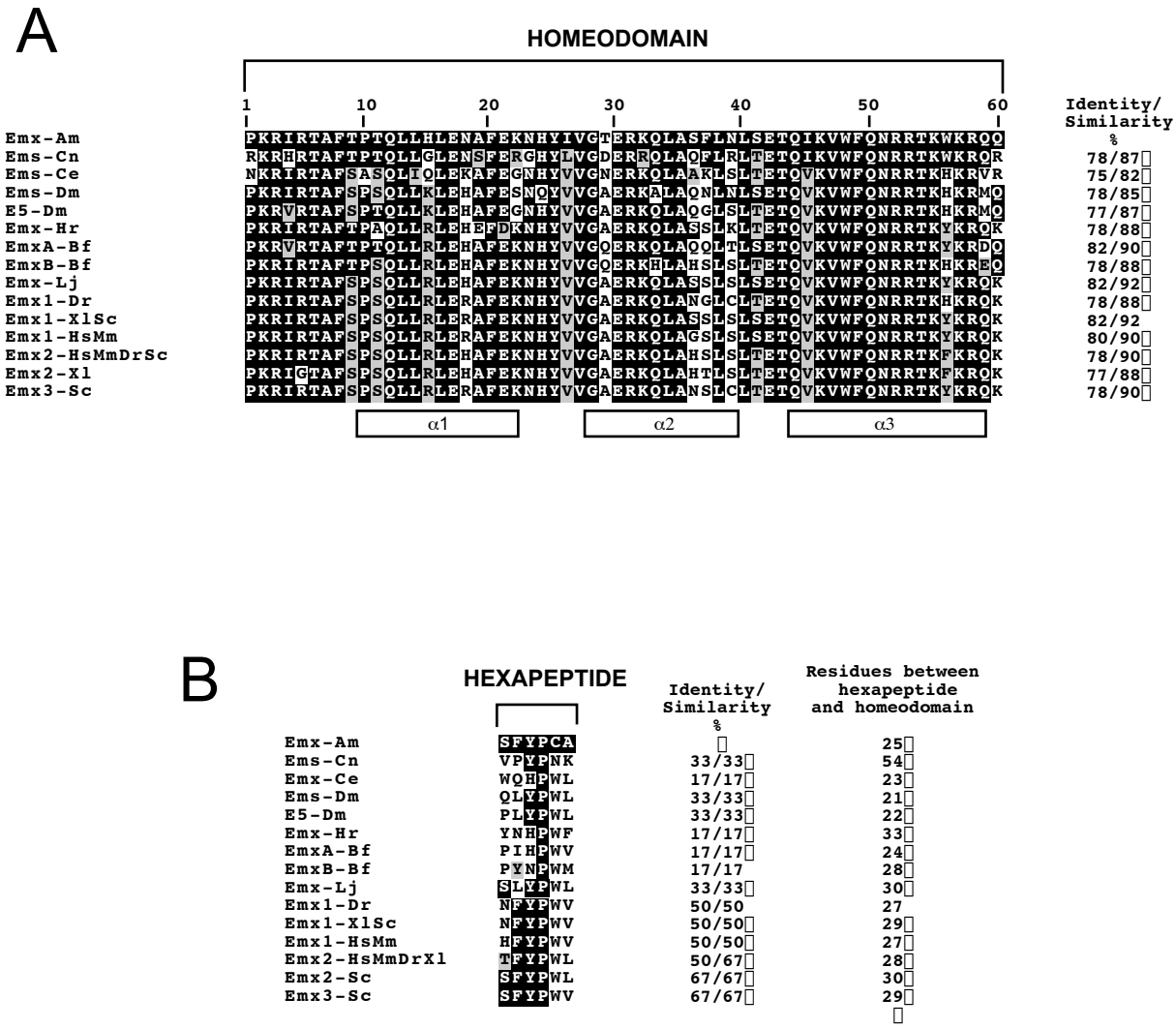


Figure 3.2: A comparison of the Emx-Am homeodomain with other Ems/Emx family members.
(A) Aligned (Clustal W; Thompson et al., 1994) amino acid sequences of the Emx-Am homeodomain with that of other Ems/Emx family members, boxshaded (Boxshade server) to indicate residues shared with Emx-Am homeodomain. Identical residues are shaded darkly; conserved substitutions are lightly shaded. Numbering above the alignment refers to the homeodomain sequence. Secondary structures within this motif are indicated below the alignment with the three alpha helices represented by open boxes. The column to the right of the alignment indicates the overall identity and similarity of each homeodomain with the Emx-Am motif. (B) Alignment (Clustal W) of the amino acid sequences of the hexapeptide motifs of the Ems/Emx family. Identical residues are shaded darkly; conserved substitutions are lightly shaded. The species names (and Genbank Accession No.s or reference) are abbreviated as follows: Am, *Acropora millepora*; Lj, *Lampetra japonica* (BAB13506); Xl, *Xenopus laevis* (Pannese et al., 1998); Sc, *Scyliorhinus canicula* (dogfish)(Emx1: AAM78422, Emx2: AAM78421, Emx3: AAM78420); Bf, *Branchiostoma floridae* (EmxA: AAF76327, EmxB: AAK93792); Hs, *Homo sapiens* (EMX1: Q04741, EMX2: Q04743); Mm, *Mus musculus* (Emx1: Q04742, Emx2: Q04744); Dr, *Danio rerio* (zfEmx1: I51736, zfEmx2: I51737); Hr, *Halocynthia roretzi* (AB055144); Dm, *Drosophila melanogaster* (ems: S22708; E5: Flybase); Cn, *Hydractinia symbiolongicarpus* (CAA72534); Ce, *Caenorhabditis elegans* (NP_491746).

Outside the homeodomain there is relatively little sequence conservation between ems/Emx proteins (Figure 3.3). One exception to this is the hexapeptide motif, containing a conserved YPW core sequence, which is located 21-33 amino acids N-terminal of the homeodomain in many ems/Emx proteins (Figure 3.2B). The hexapeptide motif (canonical sequence IYPWMK) is characteristic of many homeodomain proteins (Bürglin, 1994), including the Ems/Emx class, and is involved in complex formation between DNA, Hox proteins and their co-factors (Chang et al., 1995; Johnson et al., 1995; Knoepfler and Kamps, 1995; Neuteboom et al., 1995; Phelan et al., 1995). Specific residues within the hexapeptide are required for cooperative binding in some Hox proteins; the phenylalanine, tryptophan and methionine at positions 2, 4 and 5 respectively in Hoxb-8 and Hoxb-4 hexapeptides are essential for interaction with Pbx proteins (Neuteboom et al., 1995; Shen et al., 1996).

It is unlikely that either the *Hydractinia* or *Acropora* sequences contain hexapeptides (Figure 3.2B); in both cases the closest match contains only two residues (YP; SFYPCA and VPYPNK respectively) in common with the Ems/Emx hexapeptide consensus (QLYPWL/V) and the absence of a tryptophan (W) residue is likely prevent hexapeptide function (Peltenburg and Murre, 1996; Shen et al., 1996).

3.2.3. Evolutionary relationships amongst the *ems/Emx* gene class

The Emx-Am homeodomain is a clear member of the ems/Emx family. Maximum likelihood phylogenetic analyses based on the ems/Emx homeodomain protein sequences were carried out in MolPhy version 2.3 (Adachi and Hasegawa, 1996) using the Dayhoff substitution matrix. The optimal tree topography is shown in Figure 3.4. Phylogenetic analyses demonstrate that the Emx-Am homeodomain is more closely related to its chordate counterparts than it is to its cnidarian ortholog, Cn-ems. The derived nature of the Cn-ems sequence can be seen clearly in this maximum likelihood phylogenetic analysis, as indicated by the length of the Cn-ems branch (Figure 3.4). The early divergence of both cnidarian sequences is well supported (100% bootstrap), while the chordate sequences are clustered in a single clade. Within the chordate clade, the ascidian sequence (Hremx) is basal to the three vertebrate Emx classes identified in previous analyses (Derobert et al., 2000; Derobert et al., 2002). The position of the cephalochordate sequences is unclear; in the analyses shown these sequences cluster away from the other chordate sequences, but bootstrap support around this point is low.

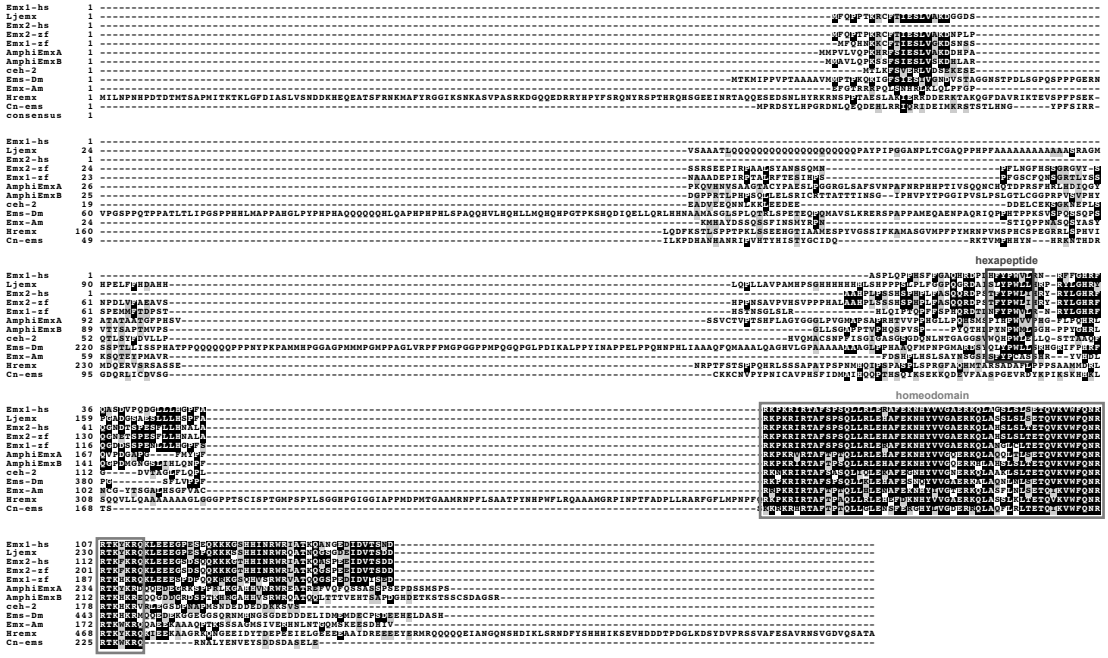


Figure 3.3: A comparison of the complete Ems/Emx proteins from representative organisms.
Aligned (Clustal W; Thompson et al., 1994) amino acid sequences of the entire Emx-Am protein sequence with that of other Ems/Emx family members, boxshaded (Boxshade Server) to indicate residues shared within Ems/Ems proteins. Identical residues are shaded black, conserved substitutions are shaded grey. Numbering in the first column refers to the protein sequence of each organism. The conserved homeodomain is boxed in blue. The semi-conserved hexapeptide motif is boxed in pink. The species names (and Genbank Accession No/reference.) are abbreviated as follows: Am, *Acropora millepora*; Lj, *Lampetra japonica* (BAB13506); Amphi, *Branchiostoma floridae* (EmxA: AAF76327, EmxB: AAK93792); hs, *Homo sapiens* (EMX1: Q04741, EMX2: Q04743); zf, *Danio rerio* (zfEmx1: I51736, zfEmx2: I51737); Hr, *Halocynthia roretzi* (AB055144); Dm, *Drosophila melanogaster* (ems: S22708; E5: Flybase); Cn, *Hydractinia symbiolongicarpus* (CAA72534); ceh-2, *Caenorhabditis elegans* (NP_491746).

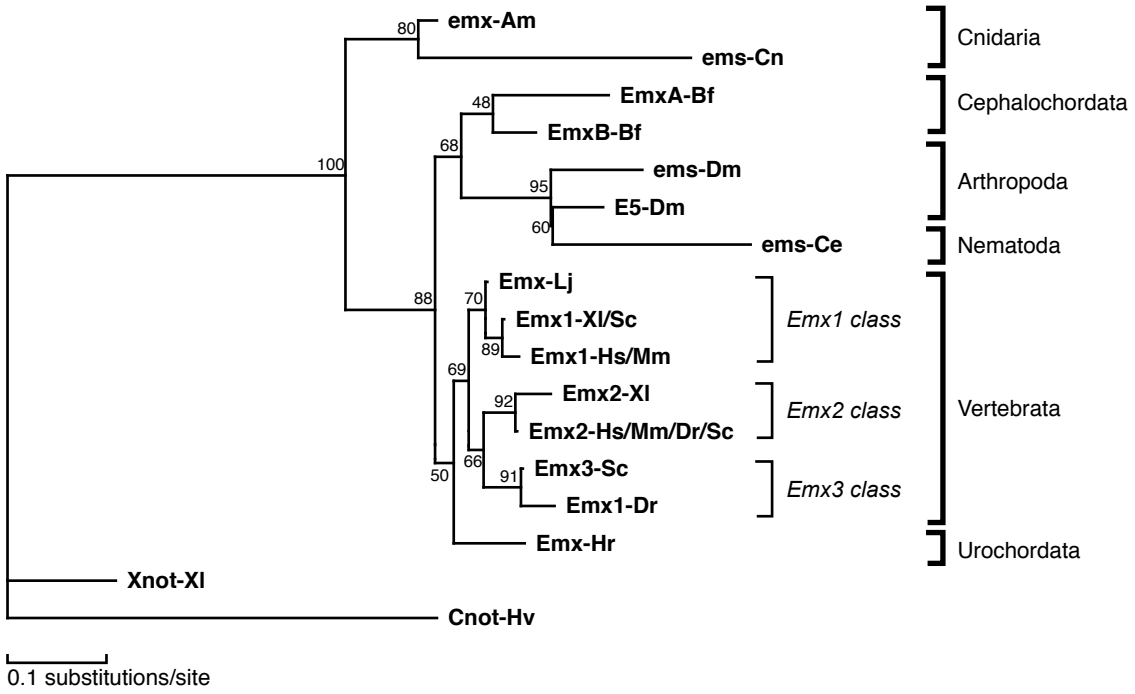


Figure 3.4: How is Emx-Am related to other Ems/Emx proteins? Emx-Am and related homeodomains from Fig 3.2 were subjected to phylogenetic analyses using MolPhy (Adachi and Hasegawa, 1996). The tree shown is a result of Maximum Likelihood distance analyses; numbers against the branches indicate the percentage of 1000 bootstrap replicates supporting the topology. The species names (and Genbank Accession No/reference.) are abbreviated as follows: Am, *Acropora millepora*; Bf, *Branchiostoma floridae* (EmxA: AAF76327, EmxB: AAK93792); Ce, *Caenorhabditis elegans* (NP_491746); Cn, *Hydractinia symbiolongicarpus* (CAA72534);. Dm, *Drosophila melanogaster* (ems: S22708; E5: Flybase); Dr, *Danio rerio* (zfEmx1: I51736, zfEmx2: I51737); Hr, *Halocynthia roretzi* (AB055144); Hs, *Homo sapiens* (EMX1: Q04741, EMX2: Q04743); Lj, *Lampetra japonica* (BAB13506); Mm, *Mus musculus* (Emx1: Q04742, Emx2: Q04744); Sc, *Scyliorhinus canicula* (dogfish)(Emx1: AAM78422, Emx2: AAM78421, Emx3: AAM78420); Xl, *Xenopus laevis* (Pannese et al., 1998). The two outgroups used in this analysis (Xnot-Xl CAA79629 ; Cnot-Hv CAB88387) are members of the Not subfamily of Antennapedia-type superfamily of homeodomain proteins.

The two *Drosophila* and single *C. elegans* sequences form a distinct and well-supported clade, clearly indicating that duplications within the chordate and *Drosophila* lineages were distinct.

3.2.4. Expression patterns of *emx-Am*

To better understand possible roles in development, spatial and temporal patterns of expression of *emx-Am* for embryonic, planula and early polyp stages were examined.

3.2.4.1. Temporal patterns of expression

Temporal expression of *emx-Am* was analysed by probing a northern blot containing mRNA from coral embryos of various developmental stages and also by quantitative PCR. For northern analyses, the radioactive probe was a 1262 bp fragment excised from the plasmid using *Xho*I and *Eco*RI, which contained the complete *emx-Am* cDNA including the 3' untranslated sequence. Hybridisation revealed a single band of approximately 1.3 kb (Figure 3.5C), corresponding well to the size of the *emx-Am* cDNA clone (1262 bp). The *emx-Am* message was first detected at low levels post-gastrulation, between 36 h and 48 h into development, after which the message was detected continuously until immediately prior to settlement, peaking at around 72 h (when the pear stage should predominate).

The temporal pattern of *emx-Am* expression was also examined using quantitative PCR on an SDS 7700 (Applied Biosystems) real-time PCR instrument. Primers used to generate a 68 bp amplicon were designed to span the intron position so that any genomic contamination would be immediately apparent. The locations of the primers (emx.F and emx.R) used in this analysis are indicated in Figure 3.1. Levels of expression were quantified relative to those of *integrin-β*, which is expressed at a uniform level throughout development (Hayward et al., unpublished). In the case of quantitative PCR, the developmental stages examined were selected based on embryonic morphology rather than time into development, in an attempt to sample the full spectrum of embryonic stages. Figure 3.5A summarises *emx-Am* expression levels relative to *integrin-β*. The raw data used to construct this histogram are given in Table 3.1.

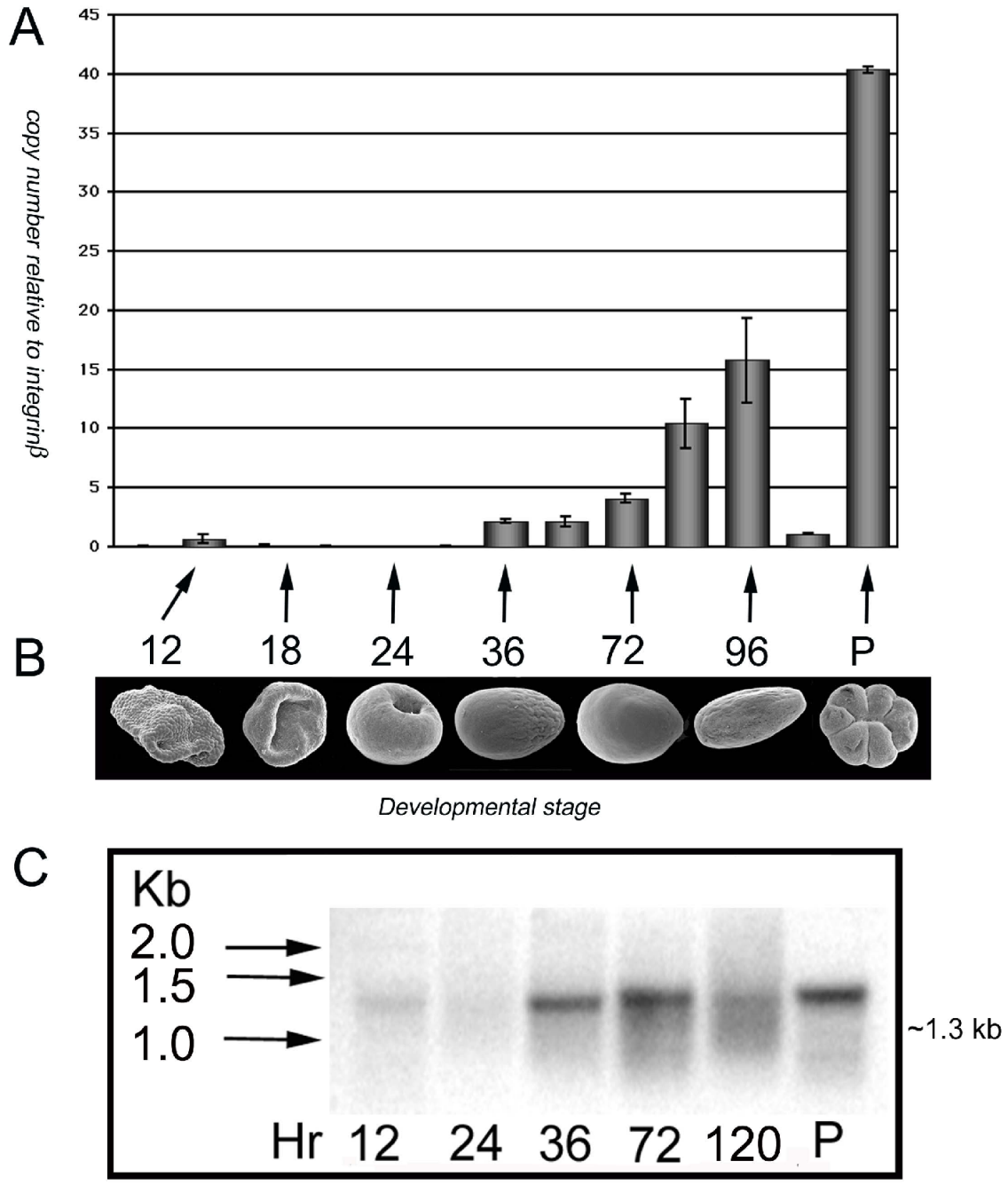


Figure 3.5: Temporal patterns of *emx-Am* expression during development.
(A) *emx-Am* expression levels during development, expressed relative to the levels of the ubiquitously expressed *integrin*-β. Numbers on the y-axis indicate the copy number of *emx-Am* relative to the copy number of *integrin*-β. Thirteen stages were examined that corresponded to the stage indicated by arrows beneath. Not all stages are represented; only those stages also examined in the northern blot are indicated. (B) Electron micrographs depicting the relative morphologies of coral embryos and larvae from the same stages of development used in the real-time PCR analysis and mRNA extraction for northern blotting (not to scale). (C) Hybridisation pattern observed after probing a northern blot of mRNA extracted from *Acropora millepora* fertilised eggs (E), embryos of various ages (numbers indicate hours post fertilisation) and post-settlement polyps (P) with a radioactively-labelled fragment of the emx-Am cDNA. Equal amounts of RNA, as measured spectrophotometrically, were loaded in each lane.

Table 3.1. Temporal patterns of *emx-Am* expression during development

Numerical data from RT-PCR of *emx-Am* expression. The morphological appearance of each of developmental stages is described in the column one. The mRNA copy number for both *emx-Am* and *integrin-b* for each stage are listed with an average figure and standard deviation. Levels of *emx-Am* expression relative to *integrin-b* were determined in the final column and used to construct the graph in Fig 3.5A

Developmental Stage	*emx-Am* copy number	*integrin-β* copy number	emx-Am normalised to integrin-β
Egg	0.240	1.7	
	0.079	1.8	
	0.140	-	
Average	0.153±0.081	1.75±0.071	0.087±0.046
Prawn Chip	0.74	0.59	
	0.24	0.79	
	0.38	0.61	
Average	0.453±0.258	0.663±0.110	0.683±0.408
Donut	0.051	1.5	
	0.370	1.4	
	0.260	-	
Average	0.227±0.162	1.45±0.071	0.156±0.111
Post Donut	0.064	0.33	
	0.030	0.32	
	-	-	
Average	0.047±0.024	0.325±0.007	0.095±0.048
Sphere with deep pore	0.049	7.0	
	-	5.9	
	-	5.5	
Average	0.049±0.000	6.13±0.777	0.008±0.0001
Sphere with pore closing	0.17	2.8	
	0.13	2.7	
	0.09	-	
Average	0.13±0.04	2.75±0.071	0.047±0.014
Pre-Pear 1	10	4.4	
	11	5.2	
	-	-	
Average	10.5±0.707	4.8±0.566	2.187±0.177
Pre-Pear 2	29	15	
	38	16	
	-	16	
Average	33.5±6.364	15.6±0.577	2.147±0.411
Pears	140	36	
	160	37	
	-	37	
Average	150±14.142	36.6±0.577	4.098±0.363
Late Pear	130	12	
	100	10	
	-	-	
Average	115±21.213	11±1.414	10.454±2.101
Planulae	180	14	
	240	15	
	-	11	
Average	210±42.426	13.3±2.08	15.789±3.576
Pre-settlement	18	15	
	20	18	
	-	18	
Average	19±1.412	17±1.732	1.117±0.095
Post-settlement	42	1.1	
	42	0.98	
	-	-	
Average	42±0.00	1.04±0.084	40.38±0.263

Each time point represents the mean of at least two independent assays; analyses were performed in triplicate, but outlying data points varying by more than 10% from the mean were excluded from the analyses, as recommended by ABI SDS Compendium 7700 – Version 3.0 (Applied Biosystems).

emx-Am expression was first detected at a low level (approximately 2.5 copies per copy of *integrin-β*) as the gastral pore of the sphere was closing (around 30 h into development) and continued until the early planula stage. At this stage, *emx-Am* message levels reached a peak level approximately 275-fold that of *integrin-β*. Immediately preceding planula settlement, *emx-Am* expression dropped dramatically, but the gene is also clearly expressed at high levels in post-settlement stage polyps.

3.1.1.2. Spatial patterns of expression

The spatial expression patterns of *emx-Am* were analysed by in situ hybridisation of embryos and planulae larvae at various developmental stages. In collaboration with Dr. Eldon Ball (RSBS, ANU), in situ hybridisation experiments were carried out using embryonic, pre-settlement and post-settlement stages. A digoxigenin (DIG)-labelled riboprobe, generated from linearised plasmid containing the complete *emx-Am* cDNA sequence, was used in the hybridisation reaction. In situ hybridisation results are consistent with northern and quantitative PCR data and reveal *emx-Am* expression first appears in pear stage planulae when the morphologically distinct oral-aboral (O/A) axis becomes apparent (Figure 3.6 A-D). Expression of *emx-Am* during embryogenesis was detected in a subset of trans-ectodermal cells, presumed to be neurons, that were restricted in their distribution to the central and aboral regions of the planula larvae (Figure 3.7). Cells containing this message were always absent from the oral end (Figure 3.6). As development continued, planulae became extended and spindle-like in appearance and were then competent to settle on the substratum. At this stage, *emx-Am* expressing cells were very rare, and were only detected along the sides and aboral end of the planulae (Figure 3.6E).

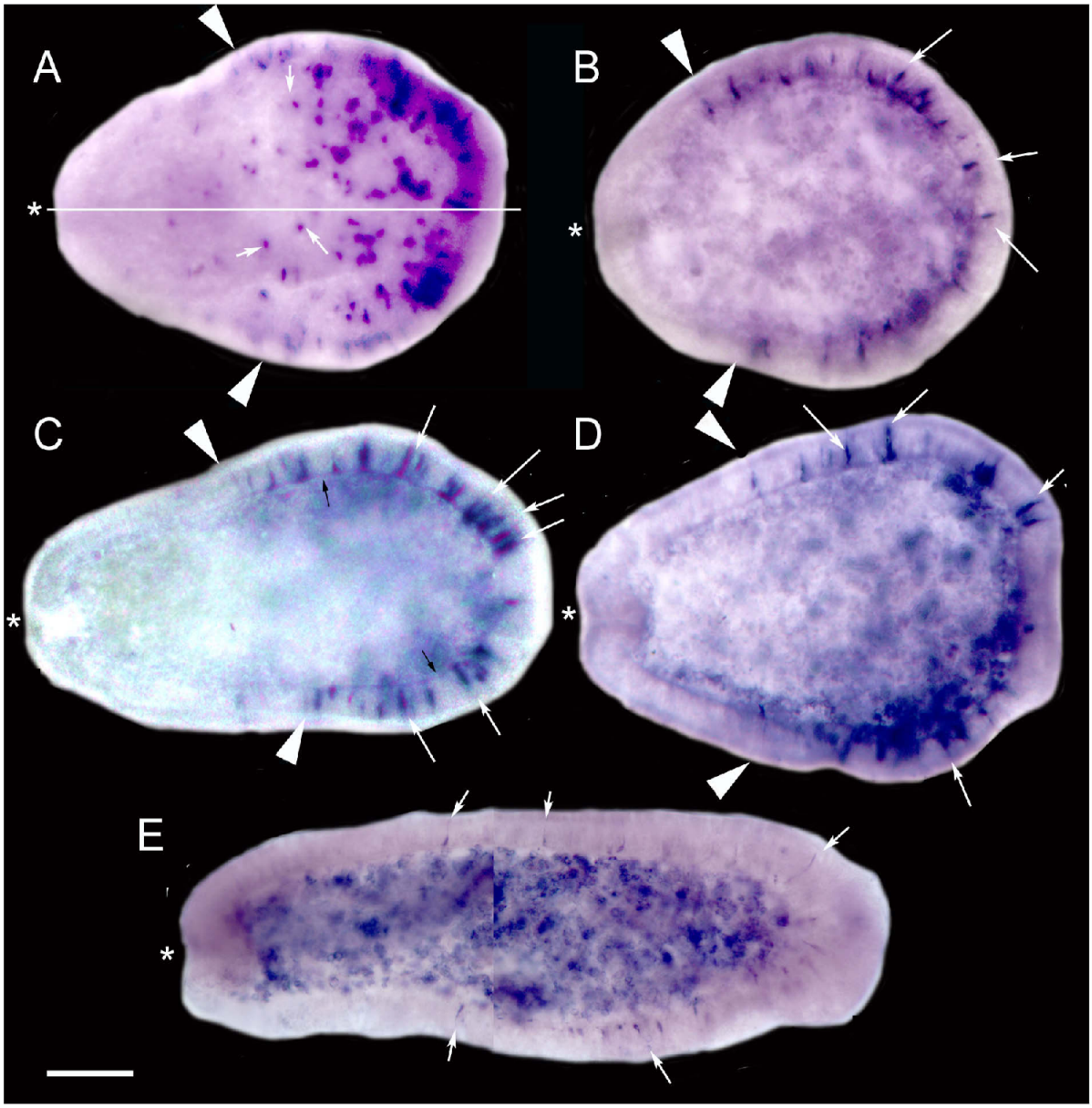


Figure 3.6: Localisation of *emx-Am* mRNA during embryonic and planulae development.
In situ hybridisation pattern observed when pear-stage and pre-settlement stage coral larvae were probed with a DIG-labelled *emx-Am* probe. (A) Whole-mount pear-stage embryo. The white line indicates the longitudinal plane of section of pear-stage embryos in (B-D). (E) Pre-settlement stage larva. All larvae are orientated such that the oral end is to the left and the oral pore is indicated by an asterisk. In all larvae, the limit of staining in the ectoderm is indicated by white arrowheads. Small white arrows indicate the location of cells expressing *emx-Am* within the ectoderm. Staining of cellular projections in the basement membrane is highlighted by black arrows. (Scale bars = 100mm)

Two distinct cell types expressing *emx-Am* were observed in these stages. The first cell type was bi- or tripolar with the nucleus and bulk of its cytoplasm located close to the basement membrane (Figure 3.7A). Fine cellular projections extended toward the surface and also extended along the basement membrane in some cells. The second type was thin and bipolar with a centrally located nucleus (Figure 3.7B). Its spindle-like appearance was apparent in some cases when its cellular projections extended in both directions towards the surface and basement membrane.

During settlement, planula larvae shortened and thickened along the oral-aboral axis then flattened onto the substratum. Tentacle formation initiated on the upper surface as calcification began. Although *emx-Am* staining in post-settlement tissue was again highly specific, in this case it was much more difficult to generalise about patterns of distribution. The *emx-Am* expressing cells were generally located within the ectoderm on the "mouth-bearing" surface of the developing polyp but no clear regional distribution pattern could be interpreted. Likewise, no clear correlation was observed between the developmental stage after settlement and the density of staining cells. Cells in settled polyps containing the *emx-Am* message were not observed in close proximity to the central pore (Figure 3.8) and were generally, but not exclusively, absent from the basal (previously aboral) surface ectoderm (Figure 3.9). In comparison to the surface ectoderm, the basal ectoderm of post-settlement polyps was thicker and the cell morphology more columnar in appearance (see Figure 3.9 A,D). Some *emx-Am* expressing cells were present on the basal surface (Figure 3.9 B,C), but none were detected in this region when the cells were columnar in appearance. This observation is consistent with the cells being neurons as the pad of cells on which the polyp develops has been described as consisting of a uniform sheet of calicoblast cells (Vandermeulen, 1975). The morphology of cells staining positive for *emx-Am* post-settlement was much more variable than in pre-settlement planulae. At least two of the distinct cell types containing the *emx-Am* message in post-settlement polyps appeared to correspond to those detected in pre-settlement embryos. The first of these had the bulk of the cytoplasm and nucleus located near the basement membrane with a single, thin cellular projection towards the surface with a bulb-like body at the apical end (Figure 3.10A). The second cell type was bipolar with a centrally located nucleus and two cellular projections extending towards the surface and basement membrane (Figure 3.10B).

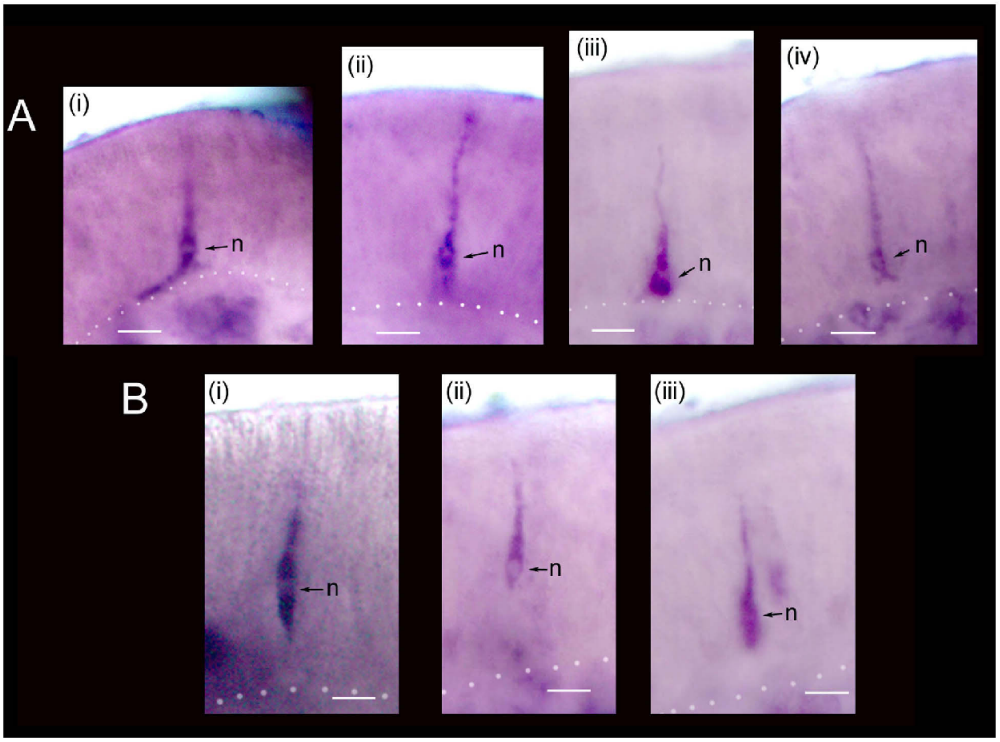


Figure 3.7: Two cellular morphologies of cells expressing *emx-Am* in pear and pre-settlement stage planulae. (A) Cell type A with basal nuclei (n) and cellular projections extending across the ectoderm (i-iv). In (i) projections are observed extending along the basement membrane. (B) Cell type B with nuclei (n) located midway across the ectoderm and projections extending in both directions, toward the basement membrane and surface ectoderm. The line of white dots in all photographs marks the location of the basement membrane separating the ectoderm from the endoderm. (Scale bars = 10m)

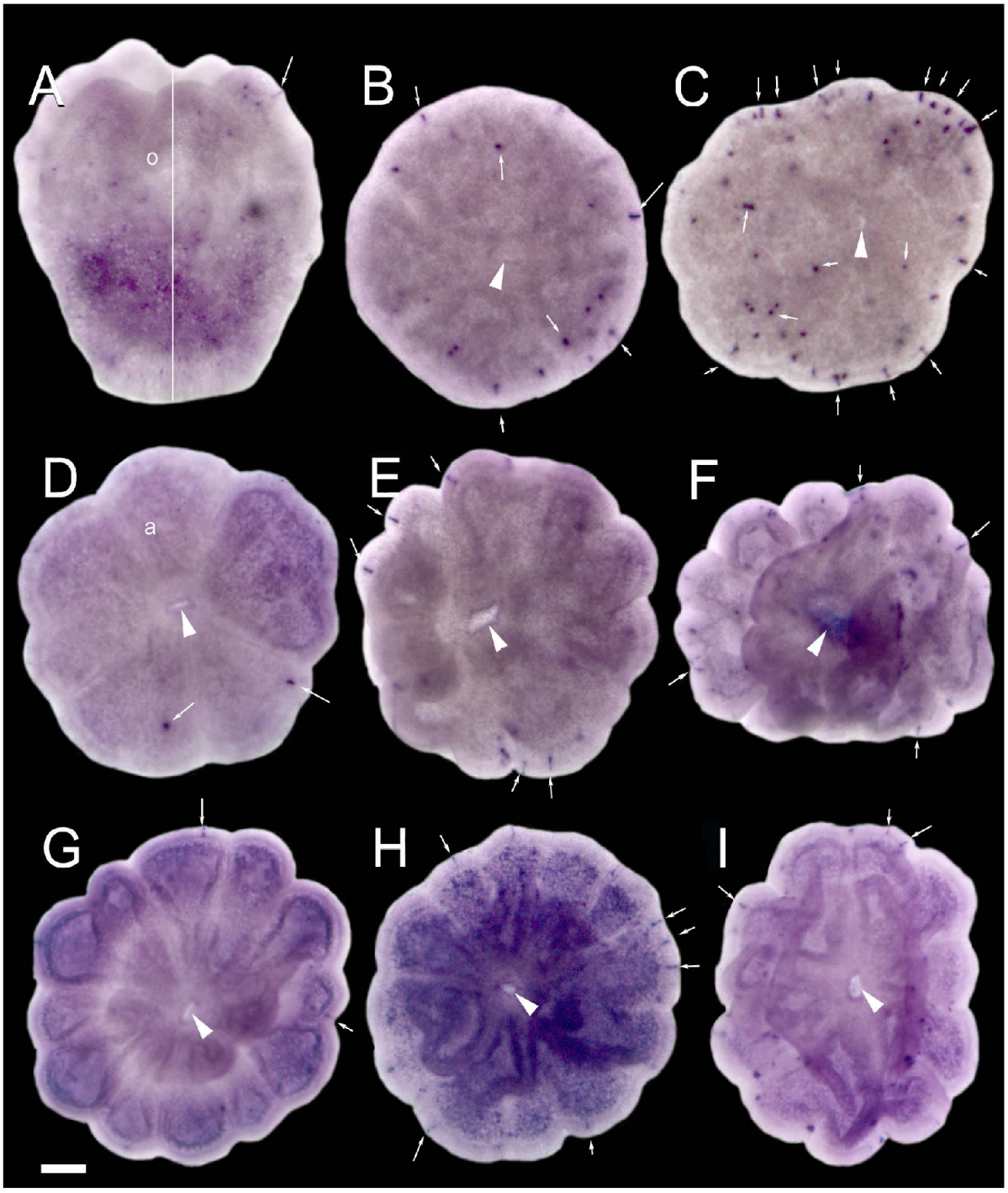


Figure 3.8: Spatial patterns of *emx-Am* expression in whole mounts of developing polyps post settlement. In situ hybridisation pattern observed when post-settlement stage coral polyps were probed with a DIG-labelled *emx-Am* probe. (A-I) Whole-mount post-settlement stage polyps at various stages of development. The white line indicates the longitudinal plane of section of post-settlement stage polyp in Figure 3.9A. All polyps are orientated such that the oral surface is up. In all polyps, small white arrows indicate the location of strongly expressing cells within the ectoderm. Large white arrowheads indicate the location of the oral pore. (Scale bars = 100 mm)

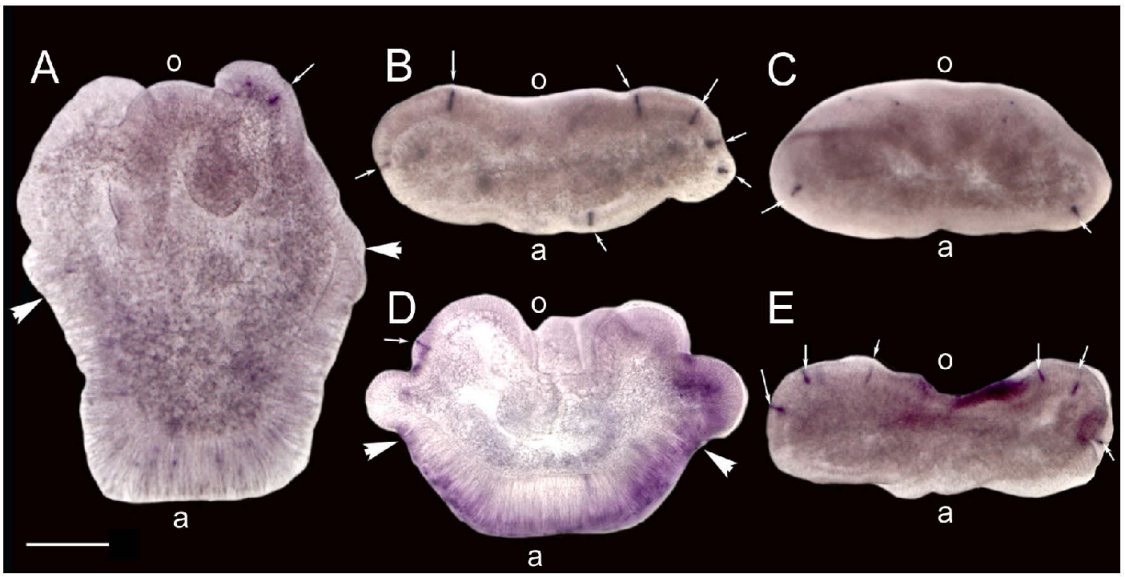


Figure 3.9: Spatial patterns of *emx-Am* expression in sectioned post-settlement polyps. In situ hybridisation pattern observed when longitudinally sectioned post-settlement stage coral polyps were probed with a DIG-labelled *emx-Am* probe. (A-E) All polyps are orientated such that the oral surface is up. In all polyps, small white arrows indicate the location of strongly expressing cells within the ectoderm. Large white arrowheads indicate the boundaries of the basal surface where cellular differentiation is observed. (Scale bars = 100 mm)

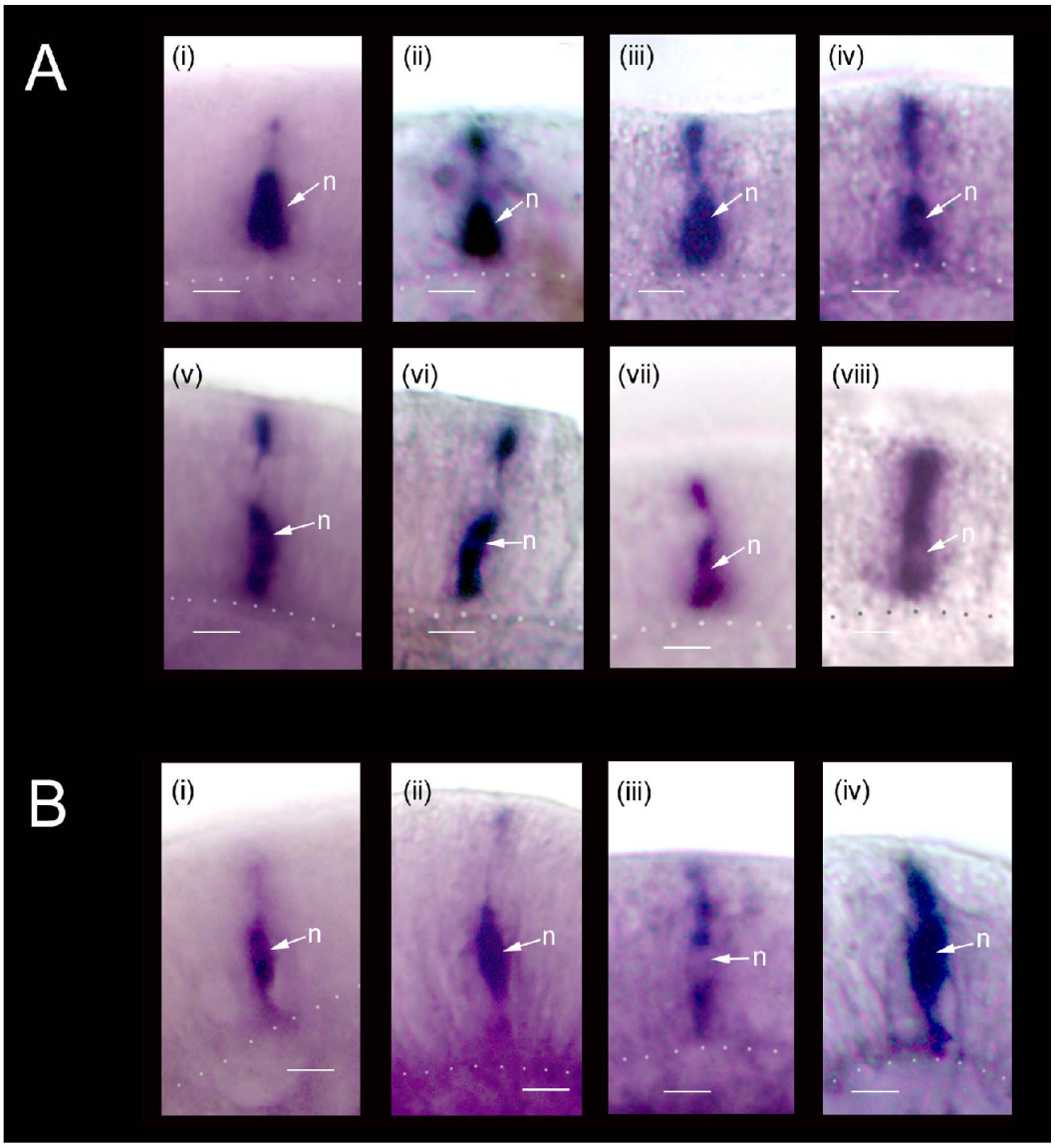


Figure 3.10: Two cellular morphologies of cells expressing emx-Am in post-settlement polyps.
(A) Cell type A with basal nuclei (n) and thin cellular projections extending across the ectoderm. (B) Cell type B with nuclei (n) located midway across the ectoderm and projections extending in both directions, toward the basement membrane and surface ectoderm. The line of white dots in all photos marks the location of the basement membrane separating the ectoderm from the endoderm. (Scale bars = 10m)

3.1.5. Location of the nerve network in coral development

While there is no known general molecular marker for cnidarian neural tissue, extensive neural networks in a number of cnidarians have been visualised using an antibody against RFamide (a neuropeptide ending in Arg-Phe-amide) (reviewed in Grimmelikhuijzen and Westfall, 1995). In order to examine the distribution of *emx-Am* expressing cells relative to the entire nervous system and determine whether these cells are neurons, pre- and post-settlement material was stained with antibodies to RFamide (primary antibody was rabbit-α-RFamide at 1:500; secondary antibody was goat-α-rabbit-biotin at 1:200; and tertiary antibody was streptavidin-Cy5 at 1:200). In longitudinally bisected pre-settlement planulae, nervous system staining was apparent throughout the ectoderm, spreading out in an extensive nerve network along the basement membrane (Figure 3.11 A-C). In post-settlement polyps, a similar nerve network was apparent in early polyps and appeared to increase in density with development (Figure 3.11 D-F). The *emx-Am* expressing cells appeared less abundant than RFamide-staining cells that are also present at the oral end of the planulae, indicating it was only a subset of neurons that expressed the *emx-Am* gene. Morphologically RFamide-positive cells in pre- and post-settlement tissue (Figure 3.11 G-I) were indistinguishable from those containing the *emx-Am* message (Figure 3.7).

3.1.6. Localisation of the Emx-Am protein in situ

Morphologically, the cells in which *emx-Am* is expressed in pre-settlement stages appear to be neurons. However, unequivocal proof of this identity would probably be best achieved by co-localisation with the message for a known neural marker. Unfortunately, no such message is presently available. Co-localisation techniques which combine in situ and antibody technology are problematic as each technique requires different treatment of the embryos that can interfere with other technique. For example, the chemical treatments required for in-situ hybridisation experiments tends to destroy any epitopes available for antibody staining experiment and in converse, antibodies often do not cope well with the chemicals involved in in-situ hybridisation. In an effort to overcome these problems, a polyclonal antibody was generated against an Emx-Am recombinant fusion protein as a step toward a double immunolocalisation technique.

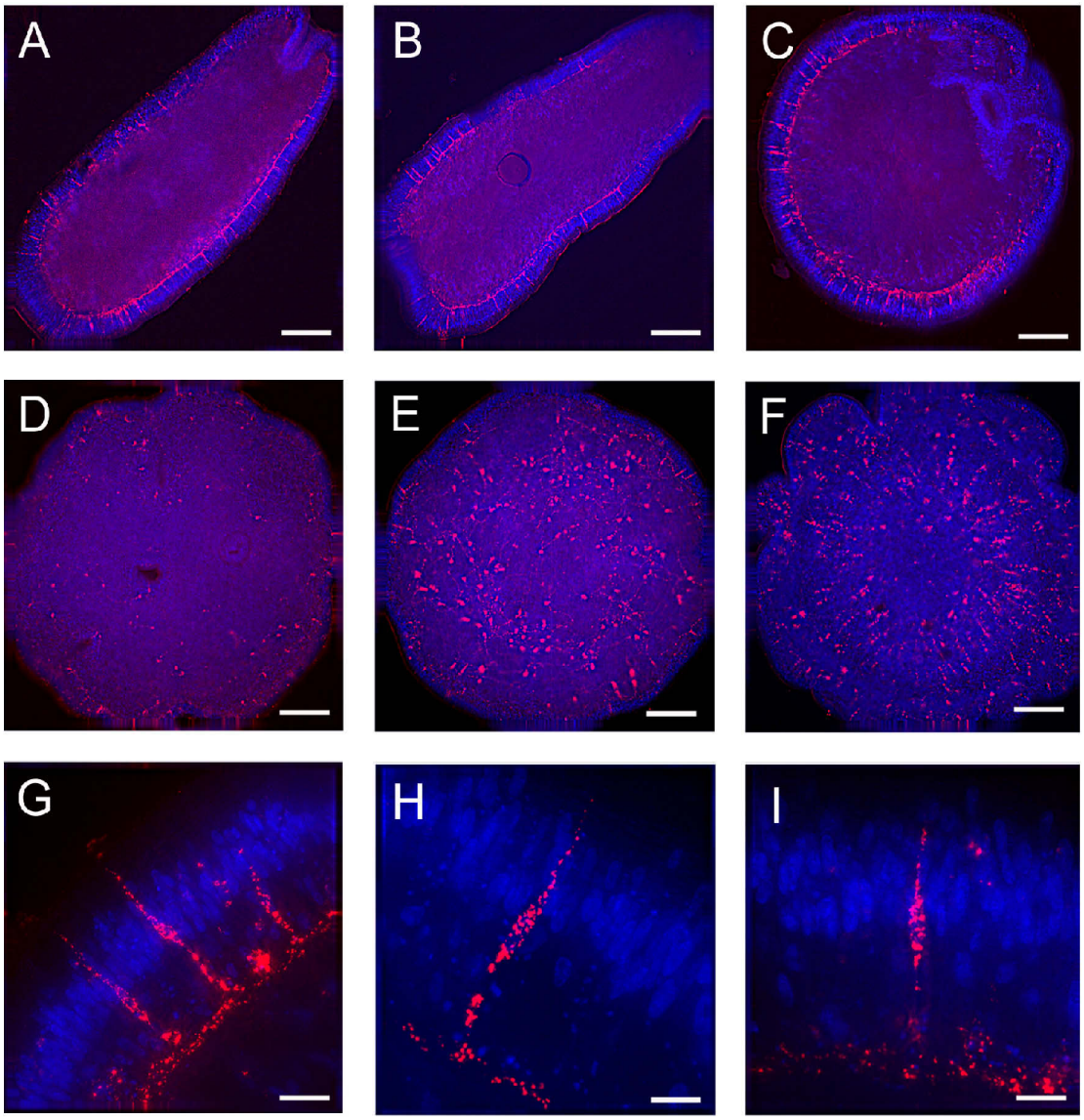


Figure 3.11: The nervous system of *Acropora millepora* in pre- and post-settlement tissue.
RFamide-expressing neurons of *A. millepora* planulae and post-settlement polyps stained with an antibody to cnidarian RFamide. Nerve cells were visualised using a fluorescent secondary antibody to both the primary anti-RFamide antibody (Cy-5; red) and a DAPI stain (blue) to visualise DNA. (A-B) Longitudinally sectioned pre-settlement planulae showing the extent of the neural network throughout the ectoderm of the larvae. (C) An early post-settlement polyp sectioned longitudinally showing the distribution of the nervous system in the ectoderm. (D-F) Whole-mount post-settlement polyp of increasing morphological complexity, demonstrating the increasing development of the nervous system following morphogenesis. (G-I) Nerve cells stained with anti-RFamide under high magnification. A nucleus is located centrally with cellular projections to both the surface and basement membrane. Projections along the basement membrane connecting cells are also apparent. Images were generated using a Delta Vision deconvolution microscope and between 8 to 20 Z-series optical sections (0.1—0.2 ∝m resolution per section). Scale bars in A-F = 100 ∝m ; scale bars in G-I = 10 ∝m

The recombinant protein contained the homeodomain region, including two amino acid residues N-terminal of this motif, and the complete (39 residue) C-terminal region. The Emx-Am HD + C recombinant fusion protein was expressed using the QIAexpressionist system (Qiagen) with a N-terminal histidine tag to allow purification of the fusion protein over a nickel column. A 595 bp cDNA fragment used in this expression construct was excised using the *Sph*I and *Hin*dIII restriction sites present within the cDNA sequence (see Figure 3.1).

This fragment was directionally cloned into the corresponding sites of the pQE-30 expression vector and sequenced to confirm that the insert was present and cloned in-frame with the His-tag. The predicted size of this fusion protein was 111 amino acids, including eight vector-encoded histidine residues, and its estimated weight was approximately 13 kDa.

The recombinant fusion protein was expressed in the M15 *E. coli* strain. Small-scale expression of three different colonies was carried out in 10 ml cultures and the lysates analysed by SDS-PAGE. In each sample a protein band of approximately 14 kDa was present, however after purification over a nickel affinity column two additional bands (8 kDa and 27 kDa) were also present and could not be eliminated (Figure 3.12A). It was estimated that they contributed to approximately 10% of the sample. However, the protein preparations were considered to be sufficiently pure for the purpose of raising antibodies. Purification of the fusion protein was only possible by solubilisation under denaturing conditions in 8 M urea.

Polyclonal antibodies were generated in rabbits using four immunisations of 250 μg of this recombinant protein. The specificity of the antibody was examined by probing a western blot of *E. coli* lysates with the unfractionated serum (Figure 3.12B). No significant cross reactivity was observed with other *E. coli* proteins (see non-induced lane, Figure 3.12B). The antibody titre was tested by indirect ELISA, and, under the conditions employed, it was determined that as little as 1 ng of antigen could be detected at a 1 in 4000 dilution of the primary antibody (Figure 3.12C).

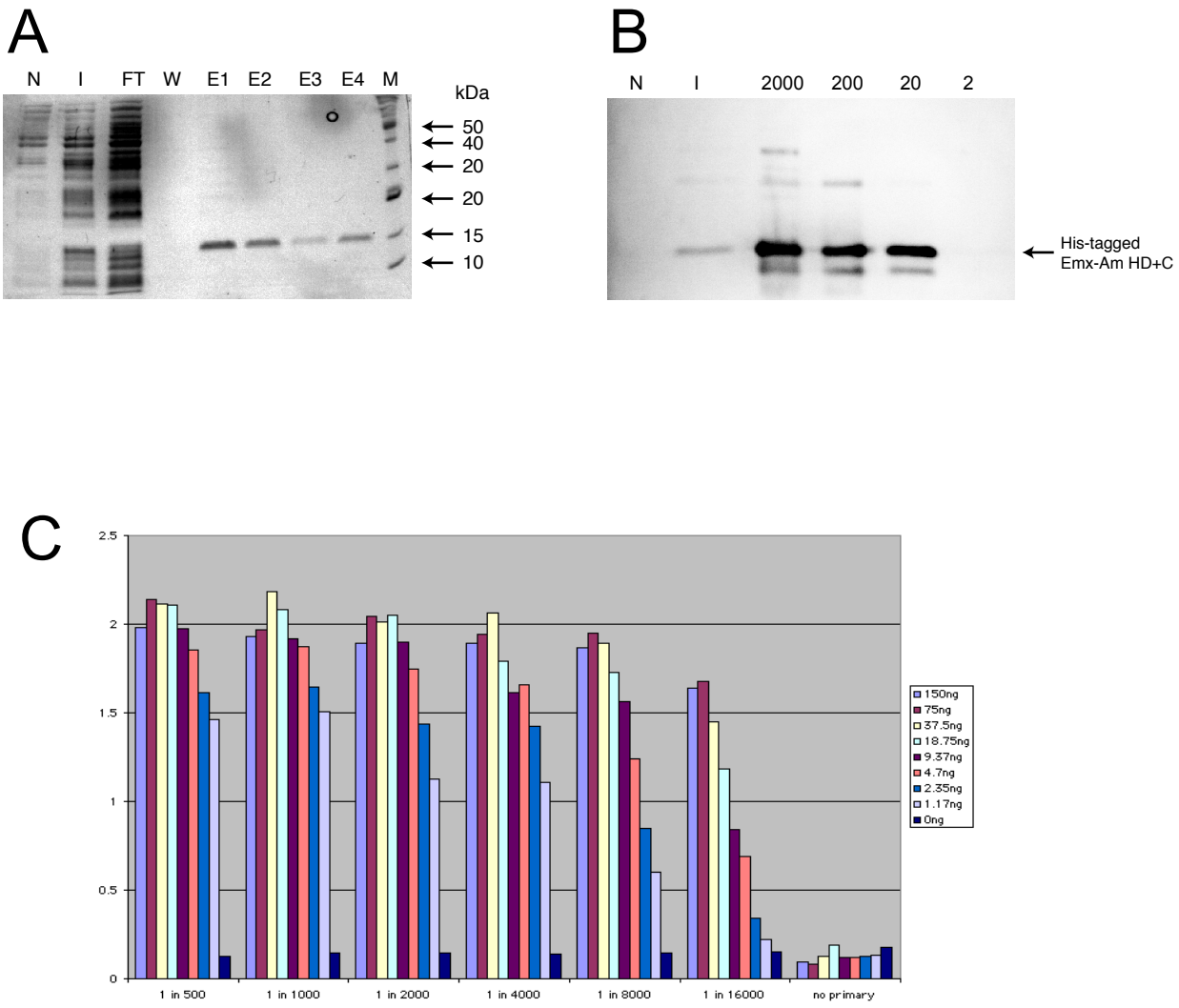


Figure 3.12: Expression of a recombinant EmxAm-HD fusion protein for raising polyclonal antibodies. (A) SDS-PAGE analysis of the recombinant His-tagged Emx-Am HD+C terminal fusion protein. The band at approximately 15 kDa represents the protein of interest. Additional bands were also detected in some purification samples. NI = non-induced bacterial cell lysate; I = induced bacteria cell lysate; FT = flow through nickel affinity column during purification step; W = last wash sample through column before fusion protein was eluted; E1-E4 = sequential eluate of fusion protein from the nickel column; M = marker (B) Western blot analysis to determine the specificity of the polyclonal antibody raised against the His-tagged EmxAm HD+C fusion protein. N = non-induced sample; I = induced sample; numbers above remaining columns indicate the amount of recombinant protein sample loaded in each well. The antiserum was not specific to the recombinant protein as it also recognised other bacterial proteins co-purified with the fusion protein. (C) ELISA to determine the sensitivity of the antiserum containing the polyclonal EmxHD antibody, for antibody staining protocols. As little as 1ng of recombinant His-tagged-EmxAm-HD+C fusion protein was detected by the antiserum at a 1 in 4000 dilution.

Embryos corresponding to stages in which the *emx-Am* mRNA had been detected were used in antibody staining experiments. The Emx-Am polyclonal antibody was incubated with pear stage coral embryos at various dilutions (1:4000; 1:2000; 1:1000; 1:100), and using both enzyme-linked and fluorescent secondary antibodies. Attempts were also made to detected the Emx-Am antibody using a secondary antibody linked to biotin and a streptavidin-Cy5 tertiary antibody. However no specific pattern could be detected using any of the methods applied. Only background staining was observed using either secondary antibodies labelled with either horseradish peroxidase or fluorophores.

In an effort to reduce background contamination, the antibody was affinity-purified by incubation with a blot of electrophoretically separated fusion protein as an affinity matrix. Three millilitres of antiserum was affinity purified using 10 μg of Emx-Am HD +C fusion protein. Indirect ELISA assays indicated the affinity-purified antibody could detect as little as 4 ng of protein at 1:500 dilution (Figure 3.13B). The specificity of the purified antibody was analysed using three western blots, each containing (i) ~150ng of GST-Emx-Am HD fusion protein (see Figure 3.13C and section 3.2.10); ii) ~150ng of His-tagged Emx-Am HD + C-terminal fusion protein; (iii) cell lysate from prawn chip stage embryos, and (iv) cell lysate from pre-settlement stage embryos. Figure 3.13A demonstrates the specificity of the purified antibody for the Emx-Am homeodomain. The depleted fraction of the antiserum bound to many non-specific proteins in both cell lysate fractions, while both the first and second eluted fractions containing affinity-purified antibody were specific for the Emx-Am fusion protein, binding to two different fusion proteins containing the Emx-Am homeodomain.

Despite the application of a wide variety of immunohistochemical techniques (and the affinity-purified Emx-Am antibody), including varied concentrations of primary antibody with different combinations of secondary/tertiary antibodies linked to enzymes or fluorophores, I was unable to detect a specific distribution pattern for the Emx-Am protein. At most, only a diffuse background staining pattern was observed. General antibody staining methods in the coral are still to be optimised; to date successful antibody staining in *Acropora* embryos has only been achieved using a few primary antibodies.

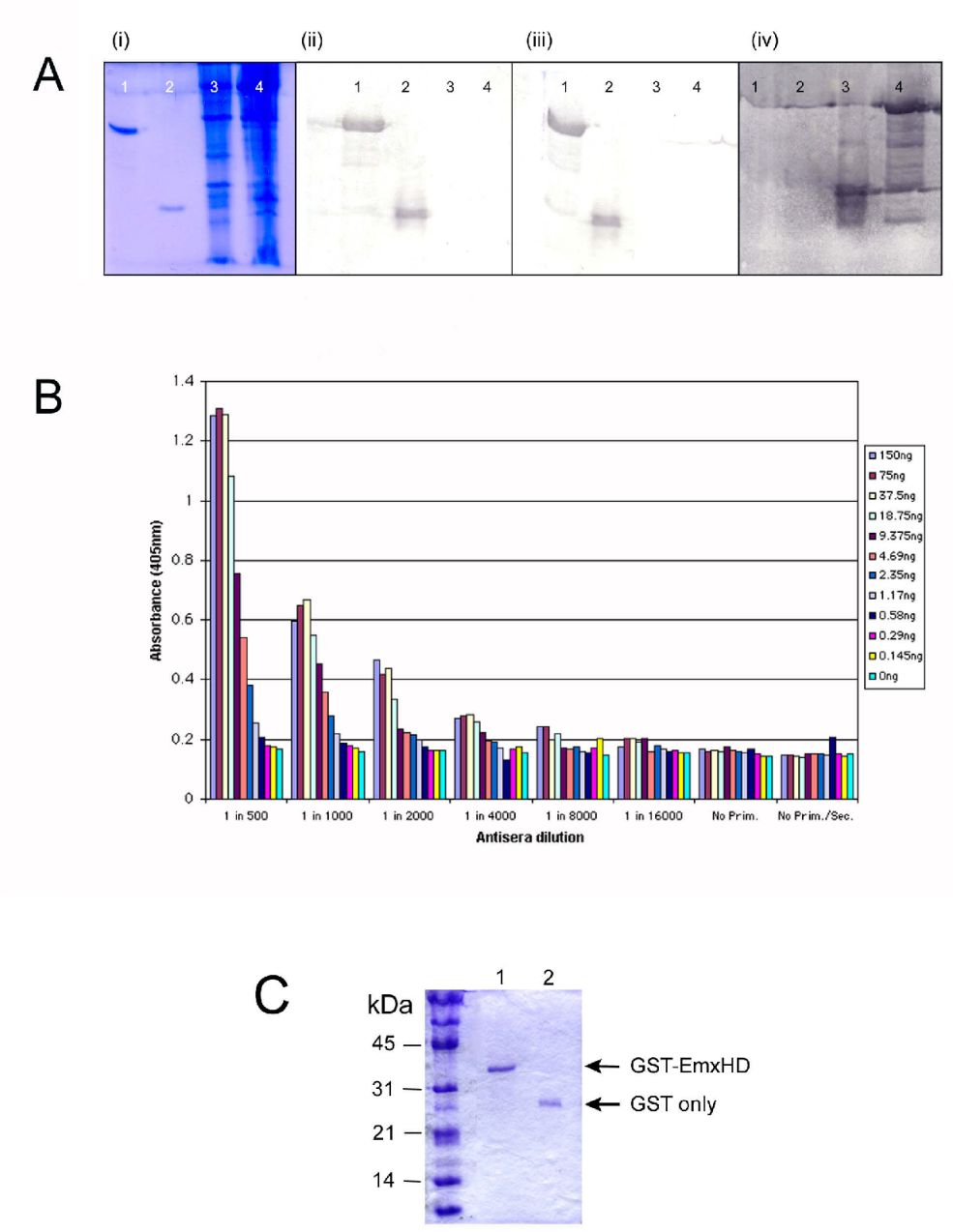


Figure 3.13: Affinity purification of the Emx-Am homeodomain antibody and expression of a GST-EmxHD fusion protein. (A) Western blot analysis to determine the specificity of the affinity purified EmxHD antibody. (i) SDS-PAGE analysis of: (lane 1) GST-EmxAm-HD fusion protein; (lane 2) His-tagged EmxAmHD+C fusion protein; (lane 3) cell lysate from prawn chip stage Acropora embryos; (lane 4) cell lysate from pear stage Acropora embryos. (ii-iv) Western blots of identical gel seen in (i) probed with (ii) primary eluate of antibody from affinity purification; (iii) secondary eluates of antibody from affinity purification; (iv) depleted fraction of antiserum after affinity purification. Affinity purified antibody is specific for the Emx-Am HD, but the sensitivity is too low to see if the antibody recognises a protein in either cell lysate. (B) ELISA to determine the sensitivity of the affinity-purified EmxHD antibody for antibody staining protocols. As little as 2ng of recombinant GST-EmxAm-HD fusion protein can be detected by the antibody at a 1 in 500 dilution. (C) SDS-PAGE analysis of: (1) GST-EmxHD recombinant fusion protein (~37 kDa) and (2) non-recombinant GST moiety (~27 kDa).

3.1.7. The *emx-Am* genomic locus

The structure of the emx-Am genomic locus was determined in order to establish the extent to which this has been conserved in evolution. A genomic clone, λEMX1, containing the *emx-Am* locus was isolated from an *A. millepora* λ-GEM12 genomic library by Dr. Julian Catmull before the start of this project. As part of this project, I employed the primer walking strategy to completely sequence λEMX1, including the 1625 bp *emx-Am* genomic locus, 2.3 kb of 5'- upstream region and 14.7 kb of 3'- downstream region (Figure 3.14). The longest ORF in this sequence was 957 bp and is located approximately 8 kb downstream of *emx-Am*. A BlastX database (Altschul et al., 1990) search of this ORF showed the highest identity (e –04) was with human bullous pemphigoid antigen (Genbank Accession No. NP 056363). In the human, this gene is located on chromosome 6p19-p11, but is not in the vicinity of human Emx1 (2p14-p13) or Emx2 (10q26.1).

Approximately 4 kb of upstream promoter sequence was obtained from two overlapping genomic clones: λEMX1 (2.3 kb) and λEMX2 (1.7 kb), which was isolated using a 200 bp PCR product amplified from the 5'- end of the λEMX1 genomic clone. Sequencing of the genomic locus revealed that the *emx-Am* gene contains a single intron of 381 bp located 20 bp 5'- of the homeobox, which possesses donor and acceptor splice junctions that conform to the GT-AG rule (Figure 3.1C).

3.1.8. Characterisation of the *emx-Am* promoter region

Potential regulatory regions of gene transcription are located upstream, in close proximity to the 5'- end of the first exon of a gene. Without the benefit of functional studies, possible regions of regulation and transcription factor binding sites may only be identified by comparing the sequence of the elements identified and verified in other organisms. Sequencing of the λEMX1 and λEMX2 genomic clones yielded 4 kb of sequence upstream of the cDNA sequence, of which 3 kb was used in further analyses. Comparison of this sequence with the BlastX (Altschul et al., 1990) database identified no ORF and a BlastN database (Altschul et al., 1990) search showed no identity with any other promoter sequence.

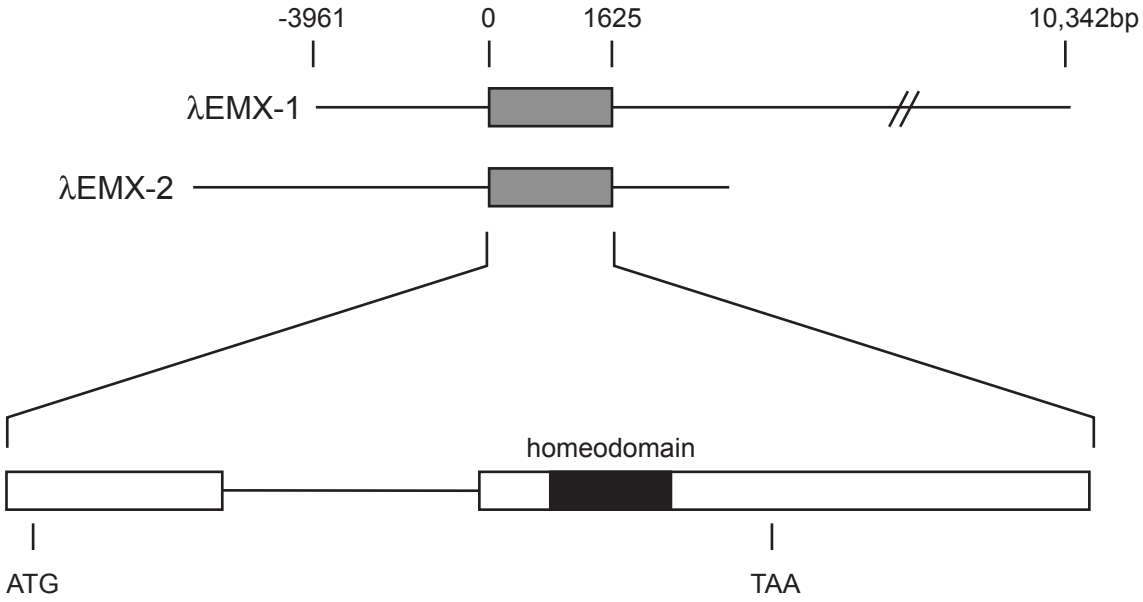


Figure 3.14: Genomic organisation of the *emx-Am* locus.
Schematic representation of the two overlapping genomic clones, λEMX-1 and λEMX-2. The shaded boxes indicate the exon-containing regions. Numbers refer to the nucleotide sequence of the genomic clones with zero defined as the 5'-end of exon 1. The expanded view shows the relative sizes of intron and exons with start and stop codons marked. The homeodomain is represented by a labelled black box.

The 3 kb of promoter sequence proximal to the 5'- end of the *emx-Am* cDNA sequence was examined for potential transcription factor binding sites using MatInspector 2.2 (Quandt et al., 1995) and the TFSEARCH (Heinemeyer et al., 1998) transcription factor databases. For clarity, the first nucleotide of the first exon is labelled +1 and the first proximal nucleotide in the promoter sequence is labelled –1. In addition to known regulators of *Drosophila ems*, putative binding sites of other transcription factors active during early development were also identified using the above-mentioned databases. Binding sites for proteins linked to external stimuli such as cAMP were also identified.

Putative basal transcriptional regulator elements including TATA and CAAT boxes were identified using MatInspector 2.2 (Quandt et al., 1995). Northern blot results indicated the *emx-Am* transcript is approximately 1.3 kb and the *emx-Am* cDNA is 1262 bp, therefore the start site of *emx-Am* transcription should lie within 200 bp of the start of the cDNA. Given that functional TATA and CAAT boxes must lie within 50 bp and 100 bp of the initiation start site respectively (Lewin, 1997), those TATA and CAAT motifs lying within –1 to –300 were identified. The TATA and CAAT boxes located at –109 and –176 respectively (Figure 3.15: marked by an asterisk) fit the spacing requirements for a start site located between –80 and –60 (Lewin, 1997).

3.1.9. DNA-binding characteristics of Emx-Am

The regulatory activity of ems/Emx transcription factors is relatively unknown. Little is known regarding the binding specificity of the homeodomain, nor of its downstream targets. The yeast-one hybrid system and electrophoretic mobility shift assays (EMSA) were employed to begin a preliminary investigation of the binding properties of the Emx-Am homeodomain.

3.1.9.1. Yeast-one hybrid screen

The yeast-one hybrid system was used to examine the DNA-binding characteristics of the Emx-Am homeodomain *in vivo*. An *A. millepora* genomic reporter library was constructed using genomic DNA partially digested with *Sau*3AI so that at least half the DNA fragments were <1 kb. Fragments between 500 bp and 1.5 kb were selected by electrophoresis through a 1% agarose gel and purified. The DNA was then ethanol-precipitated and the purified fragments ligated into *BamH*I-digested, dephosphorylated

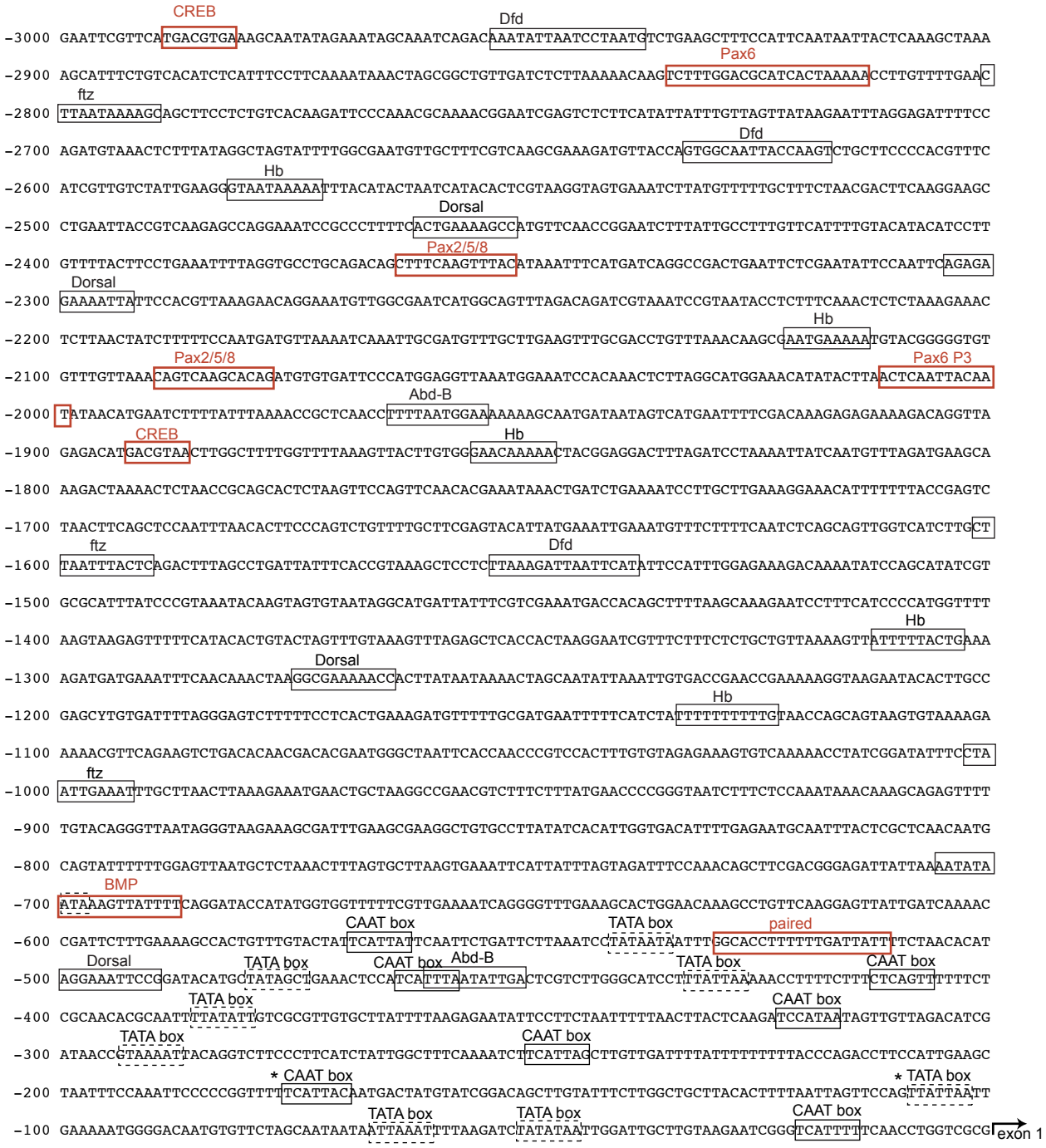


Figure 3.15: Basal promoter of *emx-Am*. Nucleotide sequence of the proximal 3 kb of genomic clones λEMX-1 and λEMX-2, upstream of exon 1. Putative binding sites of possible emx-Am regulators and basal transcription elements are boxed and labelled. Regulators corresponding to putative binding sites labelled in red have been identified in cnidarians, while those labelled in black are not known in cnidarians Potential regulators include paired-like proteins (paired, Pax6 P3, Pax2/5/8), homeoproteins and transcription factors (Dfd, Bcd, ftz, Hb, Dorsal, Abd-B, BMP) and proteins responding to external stimuli such as cAMP (CREB). Putative TATA and CAAT boxes identified within the proximal 400 bp are marked. All sites were identified using MatInspector 2.2 (Quandt et al., 1995) and the TRANSFAC 4.0 (Heinemeyer et al., 1998) binding site databases. Numbers to the left of the nucleotide sequence refer to the genomic clone with zero defined as the 5'-end of exon 1.

pBM2389. Aliquots of the ligation were electroporated into *E. coli* DH5 strain, generating six separate library pools (two of 5,000 clones and four of 10,000 clones) and a total library of 50,000 clones. Plasmid DNA for transformation of yeast was prepared from the pooled bacterial colonies. Best estimates of the size of the *A. millepora* genome are 1 - $4x10^8$ bp (D. Miller, pers. comm.). Based on this figure, the pooled *A. millepora* reporter library represents up to 48% of the coral genome. Figure 3.16A shows maps of both the activator and reporter plasmids of the yeast-one hybrid system employed in this screen. The location of the Emx-Am homeodomain sequence in the cloning junction of the activator plasmid is indicated in Figure 3.16B.

The Emx-Am homeodomain and surrounding region was cloned into the yeast activator plasmid pBM2463. A 528 bp DNA fragment encoding the homeodomain and 76 N-terminal and 40 C-terminal residues was amplified by PCR using primers (F 5'-CCCGCGGCCGCCCTAACTCTACAATTCAGCCT-3'; R 5'-CGCCTCGAGCA ATGTGGTCGCTCTCTTCCTT-3) containing *Not*I and *Xho*I adaptors respectively to facilitate directional cloning. The PCR product was cloned first into pGEM-T for sequencing then excised using *Not*I and *Xho*I and ligated into the corresponding sites of pBM2463 (Figure 3.16B). The Emx-Am fragment was cloned in-frame with an upstream lexA domain and a downstream GAL4 domain and then sequenced to ensure the insert was in-frame and in the correct orientation.

The selection of clones containing Emx-Am homeodomain binding-sites from the *A. millepora* reporter library was based on their ability to confer Emx-Am-dependent activation of the minimal promoter in yeast cells. Using the *Saccharomyces cerevisiae* strain YM4271, competent yeast cells were transformed with the lexA/emxAm/GAL4 activator plasmid. A single, positive colony containing the activator plasmid was then picked and prepared for transformation with the reporter library. The screening strategy illustrated in Figure 3.16 (C and D) was used to identify reporter plasmids carrying inserts that mediated activation of the *HIS3* reporter gene through the Emx-Am homeodomain.

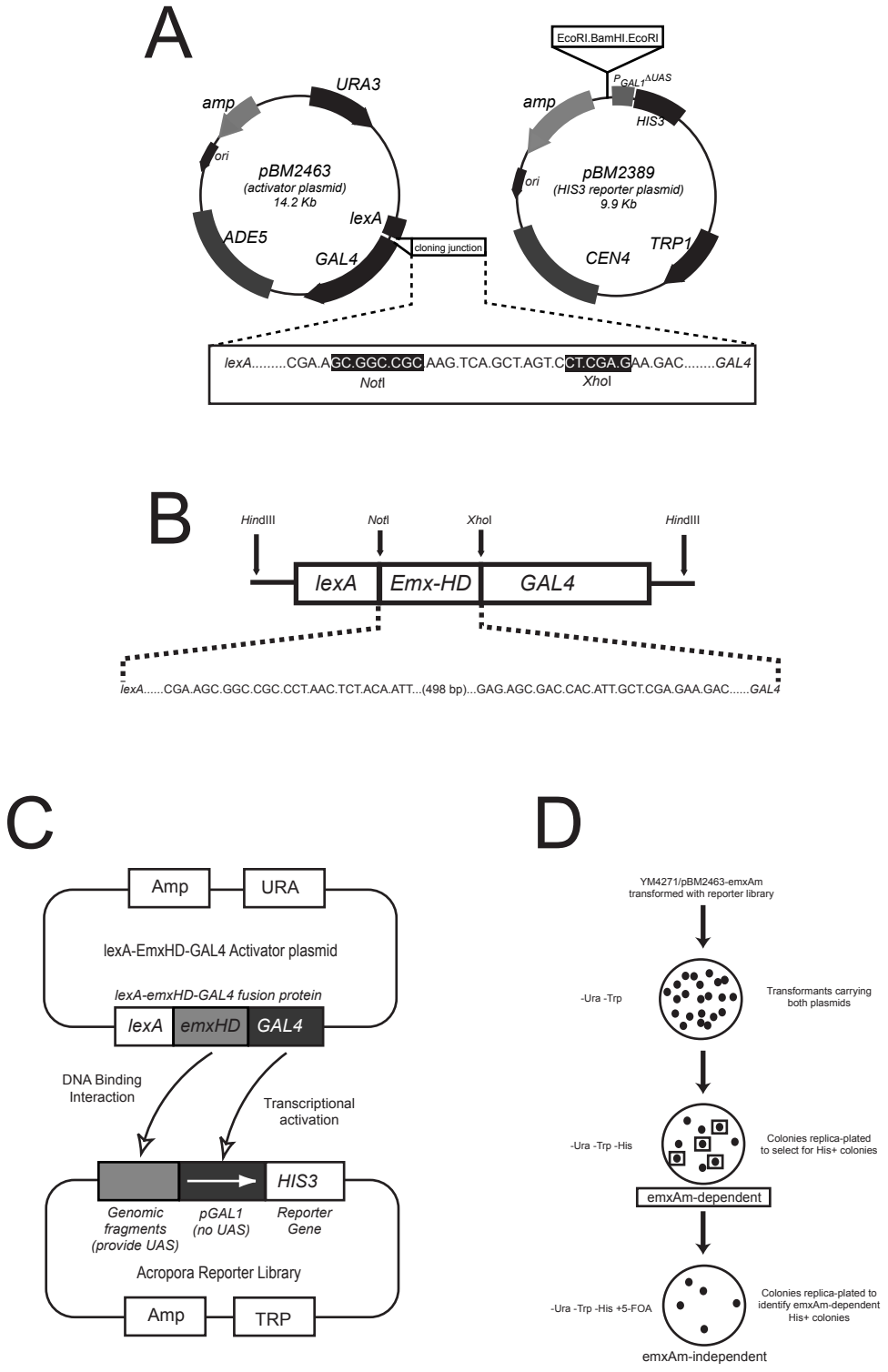


Figure 3.16: The yeast-one hybrid screen for potential targets of the Emx-Am homeodomain.
(A) Plasmid maps of activator plasmid pBM2463 illustrating the cloning sites for the Emx-Am homeodomain fragment into *Not*I and *Xho*I sites of the cloning junction. The *BamH*I site of the reporter plasmid pBM2389 was used to clone *Sau3*AI-digested fragment of *Acropora* genomic DNA. (B) Expanded view of the cloning junction of the pBM2463 activator plasmid and cloning of the Emx-Am homeodomain to form a fusion protein with lexA and GAL4 subdomains. (C) Schematic representation of a positive interaction between the Emx-Am homeodomain and a genomic fragment of the reporter library. Positive interactions lead to the transcriptional activation of the *HIS3* reporter gene. (D) Diagrammatic representation of the methods used to screen the pBM2389 *Acropora* reporter library with the pBM2463-Emx-AmHD activator plasmid and isolate activator-dependent positive interactions.

Four clones (7.2, 18.2, 24.2, 26.2) interacting with the Emx-Am homeodomain were identified from a screen of 13,000 plasmids, representing approximately 3.5-14% of the *Acropora* genome. Only a limited screen was attempted due to a lack of sequence data available for the coral genome. The DNA inserts in each of the positives from the yeast screen were completely sequenced and compared. Of the four positive clones, two were identical (clones 18.2 and 26.2) so only the three different clones were used to screen the *A. millepora* genomic library and identify the genomic regions surrounding these sites, which interacted with the Emx-Am homeodomain.

The positive reporter plasmids were isolated from the yeast cells and transformed into *E. coli* strain JM109. The inserts from each reporter plasmid were amplified by PCR and the fragments cloned into pGEM-T for sequencing. The nucleotide sequence of each clone (Figure 3.17 A-C) was determined and compared with the Genbank database. All open reading frames greater than 200 bp were found to have no significant identity with any previously identified proteins, based on comparisons with the BlastX database (Altschul et al., 1990). The nucleotide sequence of each fragment was compared with the BlastN database (Altschul et al., 1990) but did not reveal any putative promoter regions. However small regions (~50 bp) of repetitive sequence were identified in clone 7.2. While some homeoproteins can bind simple AT-rich sequences (Beachy et al., 1988; Ekker et al., 1991), which are present in some repetitive sequence elements, very little is known about the DNA-binding specificity of Emx proteins. Consequently, all three different DNA fragments isolated in the yeast-one hybrid screen were included in further analyses.

Using the insert fragment corresponding to each positive clone (7.2, 18.2, 24.2), digested from the plasmid using *Pst*I and *Nco*I, 50 000 plaques of the *A. millepora* genomic library were screened with each positive fragment in order to isolate clones containing the corresponding region. However, screening difficulties were experienced with positive clones 7.2 and 18.2; almost all plaques in the primary screen hybridised to the probe suggesting these fragments contained regions of repetitive DNA. A reverse southern blot was used to determine sections of the fragment lacking stretches of repetitive DNA. Each positive clone (in pGEM-T vector) was digested to its fragment lengths with *Hinc*II/*Hind*III and *Sau*3AI, resolved on a 0.8% agarose gel and blotted to nylon membrane. The membrane was allowed to hybridise with radioactively labelled

genomic DNA and fragments that hybridised strongly most likely contained repetitive regions of DNA. Those non-hybridising fragments that were unique to each insert were isolated and used as probes in subsequently screens of the genomic library.

Genomic library screening isolated pure populations of plaques hybridising to clones 7.2 and 24.2 respectively (screening with insert 18.2 was unsuccessful), but problems were experienced while attempting to purify lambda DNA of high quality for DNA sequencing from these phage. Only low quantities of lambda DNA could be purified but these were not sufficient for DNA sequencing protocols. Time constraints limited further experimentation after repeated attempts to purify quality DNA using a Qiagen Lambda Phage DNA kit were unsuccessful.

Data regarding ems/Emx binding sites are limited. A preliminary study identified potential ems binding sites within the *Drosophila* ems promoter with the consensus sequence AANTNTAATGACA (Taylor, 1998). This binding site closely resembles that of Antp-type homeodomain proteins and, based on this information, the positive yeast sequences were screened for potential binding sites using the TAAT core sequence. From the three positive yeast clones, a total of fifteen possible sites were identified that conformed to the TAAT core consensus sequence and a putative *Acropora* Emx-Am binding site was designed based on these data (Figure 3.17D).

Both the *Acropora* and *Drosophila* binding sites shared a conserved TAAT core sequence as well as an AAT trinucleotide 5'-, and a single adenine nucleotide 3'- of the central TAAT core sequence (Figure 3.17D; marked by an asterisk). The two binding sites differ at all other positions.

3.1.9.2. Electrophoretic Mobility Shift Assays (EMSA)

To further analyse the binding properties of the Emx-Am homeodomain, its *in vitro* binding properties were examined using EMSA. The recombinant protein used in these assays contained the Emx-Am homeodomain plus five residues N-terminal, and one residue C-terminal to it. This protein was generated using the Glutathione-S-Transferase (GST) Gene Fusion System (Pharmacia), and involved the cloning of the DNA region of interest into the pGEX-6P-2 vector (Pharmacia), enabling expression of the DNA-binding domain as a fusion protein with a GST moiety. This GST domain

A

```
TCTAGCTTTACCCTACAGTGTTTGCTAAGAGGTTGATCAACGAGTCTCTCCTTTTTAATAGTGTGGCATGTTCGTAGACA    80
GGAGACGTTGGGAAAGCTTTGCCATATTAACATTCCCAAAGGAAAAACGGTTAAACTTTAAAAAATTTAACTCACGATCA   160
ATTTTTTAACAAAATACTCCAACAGACTAACAGTTCAACAACTTGGCAAACACTATAAAGTAAAGCTCGATCAGGTACTA   240
TTCATGGGATAAATATGAGAAATAAATCACCATCCACTATTCAATCTTCGAACAACTGGACCCACTTCTAAGGCTCGGAG   320
CCACAGAGCAGCCATTTCCCGTTCCCCTCGAGAGGGTTGACAACAACCCGCAAGTGTTCTTAGGAACCCGGCACCAAGCA   400
GTCCAAGAAAGAAGGTGAGACAAAACCTTTGCTTCTTCCTGACTTCATAAGTCCTGCGAATTTCGATGGGTCTTAGGGAA   480
CATGAAGATCTCAGTTAGGGGTGTCTTCCGGCTCAAATTTGGAAACCTAACCTCATCTACTGAACAATTTCAATGTTGCT   560
GTAAACCTTCTAGAGCAGTTTTTAAATGCATGACTTGTTGCACTTCACAGATTGTGTTTGCTACGTTTATTGGCTGGCTA   640
AAAAATTTCTCGCCAGCTTTTCAACCAATGAGAAACTGAACCAAAACCAATAATTCCACCTTAAACTCGCGATTTTTCGC   720
TCTGAGCAAGTCACCGGATTGGTTCATCACGAATTTTAAATTGTATATCACGAAGCTTCAGTGCACATCGCACACAGTGT   800
TTGAAACCTTGAATTTCAATTGCCGATGTAAACACAATATAACACTTTTAACCATTTAAACGTTACTTTACATTTTTATG   880
CAATGAGACTACGAGTAAATTTATACTAAAACAATTAGACTACTCGCCCTCGTTTTCTACGAGCGATAGTCAACTCGGCT   960
GCGCCTCGTTGACTATCTGCTCGTAGAAAACTCGGGCTCGTAGTCTAATTGTTAATTATAACCCACTGGTTAATTTCTGT  1040
GGGCTCTCATTGGATAAAACTGATTCCGTGGTCTTCACCAAGTGTCTTCTGTTTCTTCTGTATACACTTCCTTGCAGGTC  1120
AAAACCCTAGTGCACGGTACTTTGCAGAAAAAGTTGCTTATCTCGAGGAGATCCGGGGAATTCCTGGCCCCACAAACCTT  1200
CAAAGGAACGAAATCACTAGTGCGGCCGCCTGCAGGTCGACCATATGG                                    1248
```

B

```
GGGCCTCCGACGTCGCATGCTCCCGGCCCCCATGGCCGCGGGATTTCGTTCCTTTGAAGGTTTGTGGGGCCAGGAATTCC     80
CCGGATCTGAAAGTATATAGAATGCGCGCGCGAGCTCTTTTGGGATTTTTGGCCGCTGCCCCGCCTAATTGCTCGAAGTT    160
AAAAAAAAAAAAAAAAACTCTTCTTAAATGTTTATTTTAACACGGTGAAAAATTGATCGCCAACTAAAAGTTATCGAGAAC    240
GAAAAAACGCAGTCTGTGGCGCAAGTTGTCGACAACTAAACCCCAAATTATTAGCTGTCGACGGCCTCATTTTTTTTGCG    320
TTTTTCTCGTGTGTGAAAACTTTTGTTTTAGTTGTCGACTACTTTTAGATCCTGTACTAAAATCTCTGCTCCGAGTATTT    400
CGAGTAATCTCGAATAGTAATTTTGGCAGGCTGACTTGTTACAATAGAAGATTGAAATATGCAAAATTGGGAAAAGGTCA    480
TCAGTGATGTCGGAAACAATCTTCCCACTGGCAAAGTACTCGGAGAACTTATTTGTGAAGATATTATAGATAAGTGTTGC    560
TGAATTATGAAAGAAATGTTATATGAAGTGCGGTGTTTGAAATCAAATGAAGATATGATCTTATTTGACTTTGTTGGTTT    640
TATATCCTTTAAATACCCATGATCTCCACGCAGTGTTTTTTTATAGTTGACAACTACTTGCTAGATTATAAAGACACAAC    720
TCAACAAGCTGCCAGCTGACATTAAGGAAAACCCAGAAGATTCAAAACAATAGTCAACAAATGGATTTCTTTTGCTCTCC    800
CAAAATTTGTTCTTTTTTCTCTAATGATCTTGTGACATAATCTTGATGTCGCATGAAAAGACAACTCTCTATAACCTGTA    880
AAAATTAAAACAAATGGTTTAGAAGTCTTTCAAAATTGACTCCAAAATTCTATGAAAACTCAAAGAGA                948
```

C

```
GTTTGTGGGGTCAGGAATACCCCGGATCTGAATGATTGCTTCCCCATGTTTCAAGTAAAATCGTTTCCACCAGGATATGT     80
                                                T
GTGATCGCATTTTACTCTGTTCAATTCCAAGGACCTTAACCTCAAATCAAGGTAATTTAAATTTCTCTTCCATATCATAT    160
                           T  A
CCTAATTAGCCTGACTGATTTCCTTAAATTAAAACGATCAGTAGTGTGACCTTTCGCTCACAGTCAGTAAAAATGACATG    240
ATGATACAAATGGTTGATAAGGCAAACACATCTCTTATTGGTCAAGATGGGGAAACTTTGACCTTCAAAGCTTTGCTTTA    320
TTCAAATGTTGCTGTTCATGTGCAAATCTGTAGACTGTTTTGCGTGGTGTATATCCCATTGCTCAGCCTGGTTCAGCTTG    400
GTTACCCCGCTATAACGATTTTTTTTGCTGTTAACCAGATAATTCGTTATAACGGGGTCTTCGCTATAACGAAGACCCCG    480
                                                          T
CTATAACGGACATTTTTTTCGGTACCGTGACACTTCGTTATAGCGGGGTTCCACTGTATTCTGTTTGTTATTTCTTTTTT    560
GCAGTACTAGCATTAATTAAACAGCTGTTTGTGCAAATGTCACCCTGGTAAAAGCTCTGAAAATGTAGAAAGGAAAAACA    640
GTTTACAGGTGTACAAATGAGACGATTTCTGTTTAACATAATTATATTAATACATCATTCTACTTACATTGTGACTGCAT    720
                      T                                                   T
TTTATCATTTAAGGAAGAAGCAAGCAAAAGCAAAAACGTTGAGACAGAAAGGACAGATGAATTAATTCTAGGACTTTACA    800
GACGACTGGACCGCATGATTGACATTCTGGATAGAATTGCAGGGCAAAATGA                                852
  A                          A          A
```

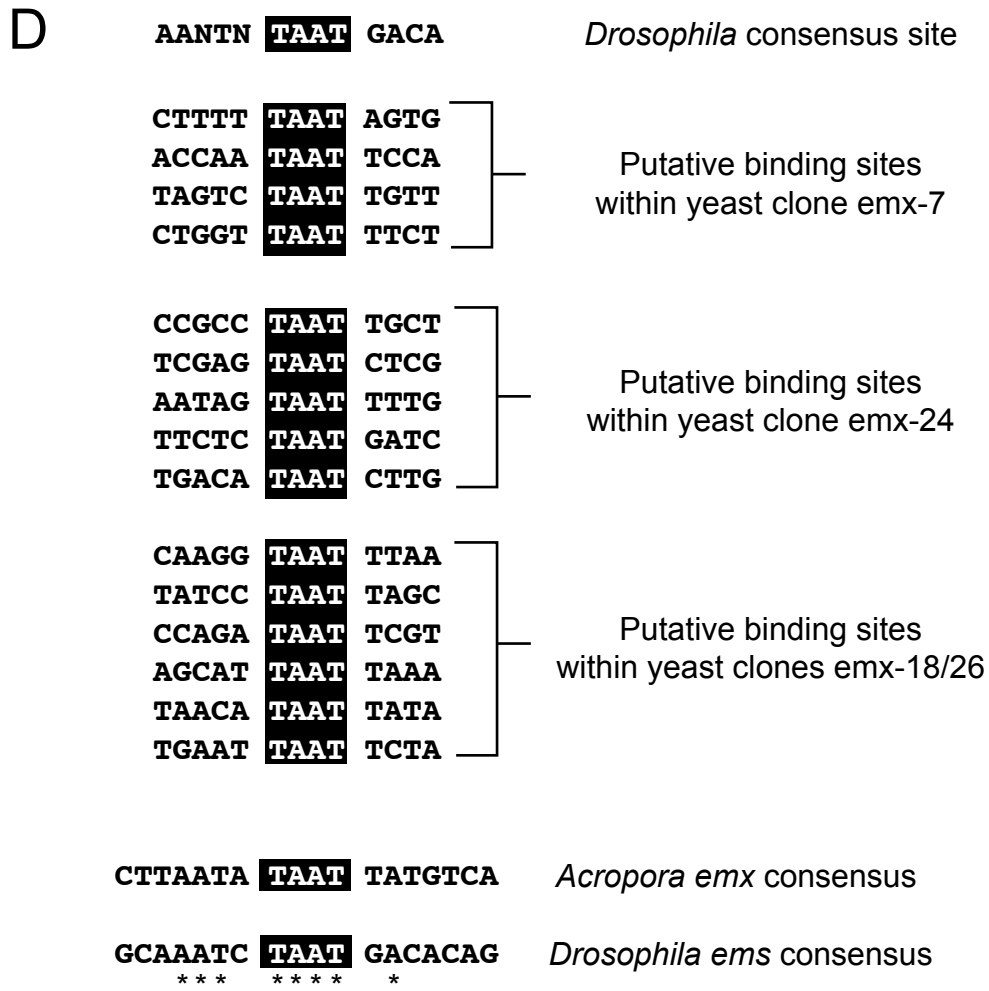


Figure 3.17: Positive clones from the yeast-one hybrid screen of the *Acropora* reporter library.
(A) Nucleotide sequence of positive clone 7.2. (B) Nucleotide sequence of positive clone 24.2. (C) Nucleotide sequence of positive clones 18.2 and 26.2. Nucleotide substitutions between the two clones are indicated beneath the sequence. Numbers to the right of each sequence refer to the nucleotide sequence of that clone. (D) Putative binding sites of the Emx-Am homeodomain identified within each positive clone compared with the consensus site identified for the *Drosophila* Ems homeodomain [Taylor, 1998 #17]. Asterisks indicate differences between the two consensus sequences. A consensus *Acropora* oligonucleotide based on these results was designed for use in EMSA experiments using a recombinant Emx-Am homeodomain protein.

facilitated the purification of this fusion protein through its strong affinity for glutathione-Sepharose beads.

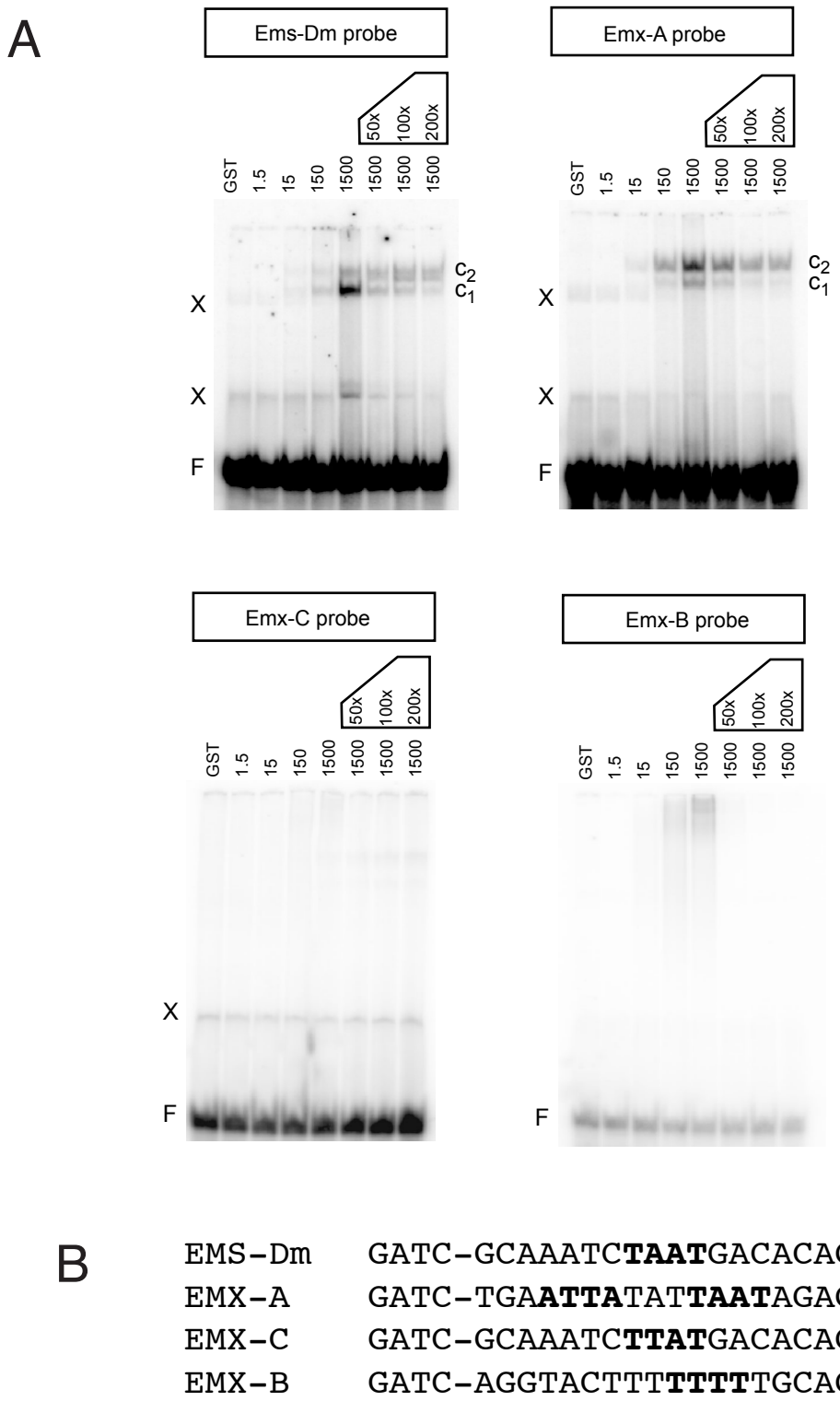
The DNA of interest was amplified by PCR using primers designed to anneal to regions either side of the *emx-Am* homeobox. *Bam*HI and *Eco*RI restriction sites were engineered into the primers (pGEX-emxHD-F: 5'- CGGGATCCGTGGCATGCCGA CGA-3'; PGEX-emxHD-R: 5'-CGGAATTCCTGCTTGTTGACGTTC-3') to facilitate the directional cloning of the 199 bp fragment into pGEM-T (Promega) for sequencing. Once the sequence of the fragment was confirmed, the insert was digested and cloned into the corresponding sites of pGEX-6P-2 expression vector. The nucleotide sequence of the pGEX-6P-2-emx construct was confirmed by sequencing, and the predicted size of the Emx-Am-GST recombinant protein was calculated to be approximately 36kDa (Figure 3.13C).

The recombinant Emx-Am-GST protein was expressed using the BL21 strain of *E. coli* in one litre cultures. The protein was extracted from the total cell lysate using glutathione beads and subjected to SDS-PAGE analysis. A 36 kDa protein was present at low concentrations and quantified at 3000ng per ml of induced bacterial culture. Figure 3.18 shows the results of EMSA experiments using the recombinant GST Emx-Am homeodomain protein and four different oligonucleotides with variations to the core TAAT consensus sequence. Oligonucleotides used in these EMSA experiments are shown in Figure 3.18, and include the oligonucleotide shown to bind a recombinant *Drosophila* Ems homeodomain protein (Ems-Dm) (Taylor, 1998), and a similar sequence identified in the yeast-one hybrid screen in *Acropora* (Emx-A). The Emx-C oligonucleotide was included to illustrate the effect of changing the TAAT consensus to TTAT, while the Emx-B oligonucleotide was included as a negative control.

While these binding assays determined if the recombinant Emx-Am homeodomain protein could bind to each oligonucleotide, the main focus was to determine if the binding affinity varied between the sites identified in *Drosophila* and *Acropora*. The *Drosophila* binding sites contained a single TAAT core motif, while the *Acropora* binding site included two TAAT core motifs on complimentary strands separated by three spacer nucleotides.

Complementary pairs of single-stranded oligonucleotides were designed to include a 5'-GATC extension in one oligonucleotide that would create a 5'-overhang when the oligonucleotides were annealed. The 5'- overhangs were end-filled using labelled nucleotides and Klenow enzyme to facilitate radioactive-labelling of each oligonucleotide. Varying amounts of recombinant Emx-Am homeodomain protein were incubated with 40 fmol of labelled oligonucleotides on ice for 30 minutes, before the mixture was subjected to polyacrylamide gel electrophoresis on a non-denaturing gel. Unlabelled oligonucleotides were used in some binding reactions to illustrate the specificity of the interaction between the GST-EmxAm fusion protein and the radioactively labelled probe. Increasing concentrations (50-200x) of unlabelled probe were added to the binding reaction to compete with the radioactively labelled probe.

With both the Ems-Dm and Emx-Am probe, the binding reaction was shown to be specific. The recombinant Emx-Am homeodomain protein bound strongly to both the Emx-Am and Ems-*Drosophila* sites, but in both cases two bands were observed that most likely correspond to a monomer (C_1) and dimer (C_2). The recombinant protein preferentially complexed as a dimer with the Emx-A probe, but mostly formed monomer complexes with the Ems-Dm probe. This was to be expected, given that two TAAT core motifs are present in the Emx-A probe in comparison to the single motif in the Ems-Dm oligonucleotide. No binding was observed with Emx-C oligonucleotide, suggesting that a complete TAAT core sequence is essential.



B

EMS-Dm	GATC-GCAAATC**TAAT**GACACAGA
EMX-A	GATC-TGA**ATTA**TAT**TAAT**AGAGA
EMX-C	GATC-GCAAATC**TTAT**GACACAGA
EMX-B	GATC-AGGTACTTT**TTTT**TGCAGG

Figure 3.18: Electrophoretic Mobility Shift Assays (EMSA) using the recombinant GST Emx-Am homeodomain fusion protein. (A) Phosphorimages of binding reactions run on 8% non-denaturing polyacrylamide gels. F = free oligonucleotide; X = non-specific interaction; C1 = oligonucleotide/protein monomer complex, C2 = oligonucleotide/protein dimer complex. Numbers above the gels indicate the amount of recombinant protein (in fmol) incubated with 40 fmol of each labelled oligonucleotide in the binding reaction. GST was used as a negative binding control using 1500 fmol of the GST moiety. The specificity of the binding reaction was analysed using 10x, 100x and 200x the amount of each unlabelled oligonucleotide to compete with the binding of the labelled oligonucleotide. (B) Oligonucleotides used in the EMSA experiments, with the potential homeodomain binding site highlighted in bold. EMX-B was

3.3. DISCUSSION

3.3.1. Conserved protein motifs

The hexapeptide motif has been shown to function in cooperative binding interactions of *Drosophila* and vertebrate Hox proteins with their co-factor proteins Extradenticle/ Pbx (Chan et al., 1994; van Dijk and Murre, 1994; van Dijk et al., 1995), which in turn interact with Homothorax/MEIS (Berthelsen et al., 1999; Kurant et al., 1998). The hexapeptide is generally viewed as necessary but not sufficient to promote co-operative interactions and increase the DNA-binding specificity of Hox proteins. The position of the hexapeptide relative to the homeodomain is also important, and changes to the length of this linker region can alter DNA-binding affinities (Neuteboom et al., 1995). The fact that the cnidarian Ems/Emx proteins appear to lack hexapeptide motifs may indicate the absence of co-factors known to interact with this motif in Hox class proteins, but this is unlikely given that plants and fungi have MEIS/TALE homeodomain proteins (Burglin, 1997). A more plausible evolutionary scenario would suggest the interaction between the homeodomain and its cofactors had not evolved before the Cnidaria/Eumetazoa split.

3.3.2. Phylogenetic relationships within the *ems/Emx* gene class

The recent identification of three *Emx* genes in the dogfish (*S. canicula*) has provided evidence for the presence of three distinct vertebrate *Emx* classes (Derobert et al., 2002), with good statistical support for each clade. However, from this analysis it is difficult to interpret the emergence of each class.

Of the vertebrate species examined, only *S. canicula* and *Takifugu rubripes* possess *Emx* genes of all three classes. However, given that zebrafish *Emx1* is clearly orthologous with the dogfish *ScEmx3* sequence, it is likely the three types are present throughout bony fish and elasmobranchs and that the third type is either present in zebrafish or has been secondarily lost. Tetrapod *Emx* genes fall into two distinct clades corresponding to mouse *Emx1* and *Emx2*. The single lamprey *Emx* gene is most closely related to the *Emx1* class, suggesting that this is the ancestral type within the chordate lineage, and that the duplication events within vertebrates post-date the split with the

Agnatha. The relationship of the single ascidian sequence (*Hremx*) to other types is unclear, as the sequence appears to be derived.

This analysis supports the hypothesis of independent duplication events in *Drosophila* and the vertebrate lineages, and probably also in amphioxus. Given the important roles of *ems*/*Emx* genes in *Drosophila* and vertebrate development, frequent duplication of this gene family may have accompanied the diversification of anterior structures in these metazoans. Often, duplication of developmentally important regulatory genes is followed by subfunctionalisation, whereby the functional complement of the ancestral gene is partitioned between the two daughter genes (Force et al., 1999; Lynch and Conery, 2000; Lynch and Force, 2000). Each gene may then retain aspects of the original gene function, but each may also acquire novel functions through mutational changes in the regulatory region of a gene. Such duplication events effectively provide novel genetic material with which to increase specialisation. The two mouse *Emx* genes are an example of such an event, each having evolved differential patterns of expression in the dorsal telencephalon, while Emx2 has acquired a secondary role in the urogenital system (Gulisano et al., 1996; Miyamoto et al., 1997; Yoshida et al., 1997).

3.3.3. Expression of *emx-Am* in developing coral larvae

Expression of *emx-Am* during early embryonic and planulae development is limited to a subset of transectodermal cells at the aboral end, which are presumed to be neurons. Cnidarian neurons were first reported in methylene blue stained tissue of *Hydra* (Hadzi, 1909; Schneider, 1890). In the literature, two types of neurons are described as (i) 'sensory cells' which are "…long, slender neurons orientated perpendicular to the mesoglea…..and projecting to the surface of one of the cell layers"; (ii) 'ganglion cells' which are "neurons with a more round perikaryon located in the basal parts of ecto- or endoderm" (Grimmelikhuijzen and Westfall, 1995). (Chia and Koss, 1979) also described two types of sensory cells present in the ectodermal layer of the sea anemone planula larva as: "Sensory cell A….is a spindle-shaped cell with a nucleus located in its middle region; Sensory cell B…is a flask-shaped cell with a long, narrow neck. The nucleus is at the basal end of the cell." At least two distinct morphologies of cells expressing *emx-Am* are observed in the ectoderm of the central and aboral regions. Both these cell types match well with the above descriptions of cnidarian neurons and fail to fit the description of other non-neural cell types reported in cnidarian ectoderm

(Fautin and Mariscal, 1991). The morphology of *emx-Am* expressing cells also closely resembles that of RFamide-expressing cells that are regarded as neurons. Based on this evidence, I propose that these *emx-Am* expressing cells are neurons.

As the planula continues to develop, it lengthens and becomes thinner, extended and spindle-like in appearance. At this stage, *emx-Am* expression is no longer observed in the ectoderm. This disappearance may be associated with degeneration of the nervous system in preparation for settlement. This phenomenon has been documented in hydroids (Martin, 2000), and is apparent in *Acropora* upon staining of the corresponding late larval stages with anti-RFamide (Ball, pers. comm).

Localised expression along the O/A axis in pre-settlement embryos may reflect a role for *emx-Am* in patterning the O/A axis. Bilaterian *ems*/*Emx* genes are involved in A/P patterning, but how this axis correlates with the cnidarian O/A axis is equivocal. Expression of *cnox2Am* is also spatially restricted along the O/A axis during early coral development (Hayward et al., 2001). The *cnox2Am* gene is orthologous with the *Gsx* class genes, which play central roles in specifying the D/V axis of the developing CNS in *Drosophila* and vertebrates. Restricted expression along the single cnidarian axis of coral orthologs of bilaterian axis-patterning genes suggests that no simple relationship exists between the major longitudinal axes of the Cnidaria and Bilateria. It is unlikely that *cnox2Am* and *emx-Am* interact, as their domains of expression overlap considerably, however, they could function in a combinatorial fashion in specifying neuron identity. As cnidarian orthologs are known of other components of both A/P and D/V pattern systems in bilaterians (including *Otx, msh* and *vnd*), it will be fascinating to examine the expression patterns of these during coral development.

After settlement, the morphologies of staining cells observed also reflect descriptions of "ganglion" and "sensory" cells given by (Grimmelikhuijzen and Westfall, 1995). However, the distribution of staining cells is far less clear than it is during pre-settlement stages. It may be informative to compare the expression pattern of *cnox2Am* in post-settlement polyps with that of *emx-Am*. To date it has not been possible to establish unequivocally a correlation between the distributions of *emx-Am* expressing cells and those staining with RFamide antibodies. It is possible that cells expressing *emx-Am* represent a subset of the nervous system in which neuropeptides other than

RFamide predominate. There are precedents for this type of expression (Higashijima et al., 1996; Udolph et al., 2001).

3.3.4. The genomic structure and its evolutionary implications

The *emx-Am* locus contains a single intron located 5'- of the homeobox (Figure 3.22); a characteristic shared by other *ems*/*Emx* genes. The exact position of this intron cannot be compared as sequence conservation in this region is low, but in orthologs that encode a hexapeptide motif, the intron lies between the homeodomain and the hexapeptide. The homeobox sequences of vertebrates, *C. elegans, Drosophila E5* and *Hydractinia* also contain a second intron, but its location differs; in vertebrates and *C. elegans*, this intron lies between codons 44/45 of the homeodomain, while in *Cn-ems* its position is between codons 46/47.

It is unclear whether these introns in vertebrates and *Hydractinia* are homologous. Introns are observed at position 46/47 in many Paired-type homeodomains and phylogenetic analyses of Antennapedia-related homeodomains place the Ems/Emx clade in the basal position (reviewed by Gauchat et al., 2000). Therefore, it is possible that an intron in position 46/47 was present in the ancestral gene of Paired and Antennapedia-type homeobox gene superclasses. A plausible evolutionary scenario is that the ancestral Emx gene contained an intron at position 46/47 (as observed in *Cn-ems*) and that slippage gave rise to the altered position of the intron in many of the other members of this gene family (Rogozin et al., 2000; Stoltzfus et al., 1997). Under this scenario, it is assumed that *emx-Am* has undergone secondary loss of the homeobox intron. Alternatively, the introns in *Hydractinia* and bilaterian *Emx* genes may have independent origins and thus be non-homologous.

3.3.5. Regulation of *emx-Am*

Regulators of *ems*/*Emx* transcription are slowly being elucidated in both *Drosophila* and vertebrates, but many regulatory pathways may affect *ems*/*Emx* transcription indirectly through mediators that have yet to be characterised. Only a few coral orthologs of developmentally significant bilaterian genes have been identified, so it is difficult to assess the likelihood of conserved genetic regulation between the coral, flies and vertebrates.

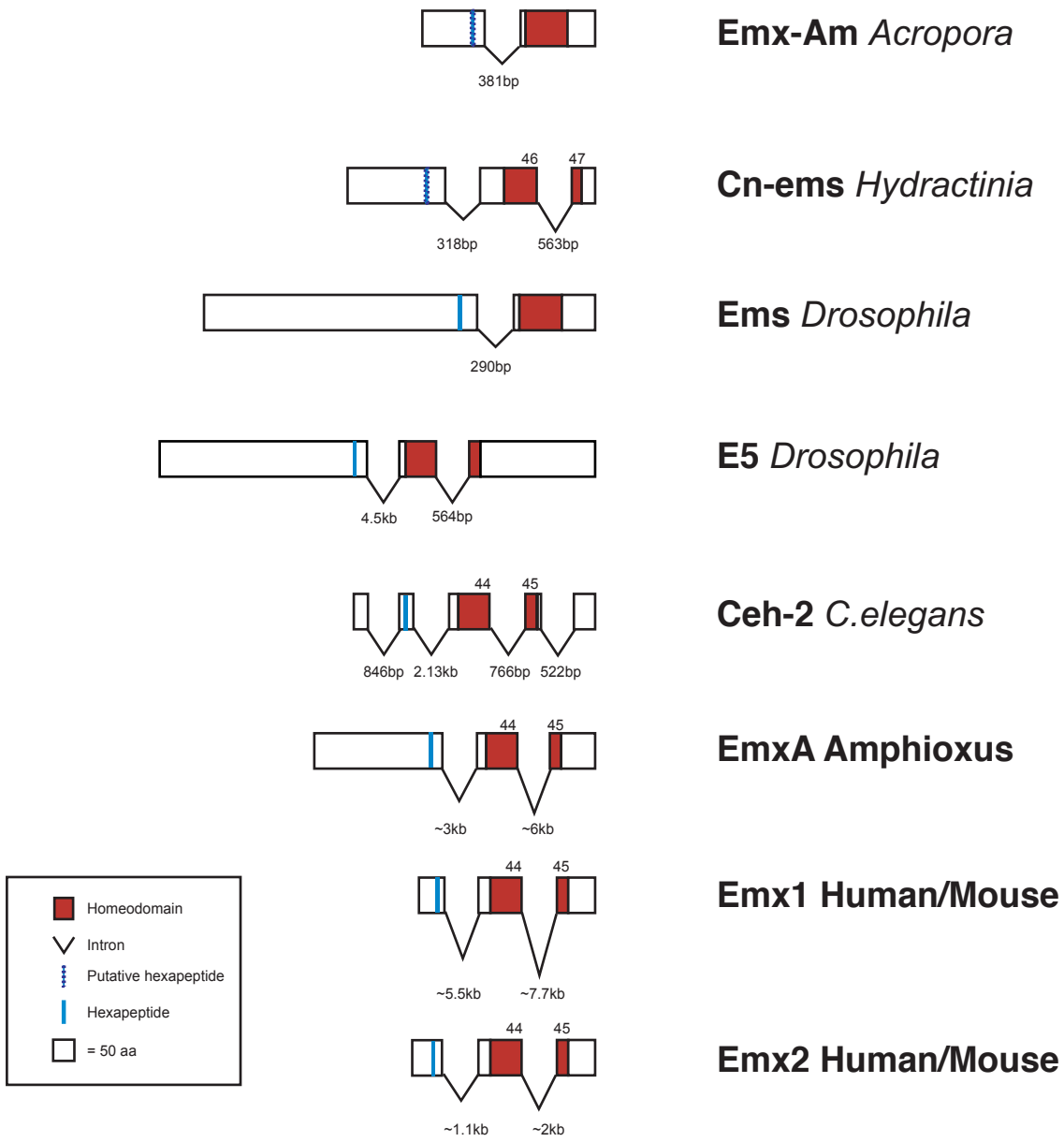


Figure 3.19: Comparative genomic structure of *ems*/*Emx* genes.
Schematic representation of the intron/exon structure of representative *ems*/*Emx* genes. Open boxes indicate the coding region of each gene. The homeobox of each gene is shown by a red box. The hexapeptide motif is shown by a blue line; putative hexapeptides are shown by a zig-zag blue line. Intron locations are indicated by open triangles with the size of each intron shown below the triangle. Numbers above each homeobox indicate the position where the encoded homeodomain is interrupted by the intron. Drawings are to scale (bar indicating 50 amino acids is given in the key).

Abdominal-B has a direct influence on *ems* expression in the fly, mediating its abdominal expression domain (Jones and McGinnis, 1993). Two putative Abd-B binding sites were identified in the *emx-Am* proximal promoter. While it is possible that this genetic interaction is conserved, an ortholog of *Abd-B* is yet to be identified in the coral. Genes corresponding to *Hox 10*/*Abd-B* have been reported in a number of cnidarian species (Finnerty and Martindale, 1997; Gauchat et al., 2000; Murtha et al., 1991; Schierwater et al., 1991; Yanze et al., 2001), and while it is unlikely these cnidarian genes are strict homologs of *Abd-B* Hox gene, it is possible they represent a precursor to both posterior Hox and Cdx genes.

In *Drosophila*, the Tailless (Tll) protein has been shown to directly inhibit *ems* expression in the head and developing brain (Hartmann et al., 2001). However, no Tll binding sites (AAGTCA) were identified in the *emx-Am* proximal promoter and expression studies of the coral *tll* homolog, *tlx-Am*, suggest the temporal patterns of expression do not significantly overlap as would be necessary for a regulatory relationship in the coral (discussed in Chapter 4). It is possible that this regulatory function of Tll is a derived characteristic in *Drosophila* and does not reflect its ancestral function, although some evidence suggests conserved upstream and downstream genetic programs are in place in flies and vertebrates (Yu et al., 1994). In contrast, it now appears unlikely that vertebrate Tlx proteins interact directly with *Emx* genes. Investigations of Tll/Tlx proteins in other model organisms will promote an understanding of the early functional role of this gene family.

Binding sites for other transcription factors active during early development were also identified and include those of a number of proteins such as hunchback, Deformed, Dorsal and Fushi tarazu (Ftz), which are involved in early segmentation and dorsal-ventral patterning in *Drosophila*. Given the origins of the oral-aboral axis and its equivocal relationship to the A/P and D/V axes of bilaterians, it seems likely that regulators of both pathways may have once functioned together in a single axis.

3.4. CONCLUSIONS

The *emx-Am* gene from *Acropora* is a clear member of the *ems*/*Emx* gene class. Phylogenetic analyses based on the homeodomain sequence indicate Emx-Am is more

closely related to vertebrate Emx sequences than to the *Hydra* Emx sequence, reflecting the basal position of the anthozoans within the Phylum Cnidaria. As such, the role of *emx-Am* may be informative with respect to the ancestral role of this gene family.
Sequence conservation outside the homeodomain of the Ems/Emx family is limited, with the exception of a hexapeptide motif N-terminal of the homeodomain. Unlike the triploblastic Emx proteins, cnidarian Emx proteins possess only a partial hexapeptide motif and share only two of the six conserved residues observed in other Emx proteins. Whether or not this hexapeptide sequence is of functional importance to Ems/Emx proteins is still to be determined.

Temporal and spatial patterns of *emx-Am* expression suggest this gene is involved in anterior patterning and nervous system specification of the developing coral larvae. In both pre- and post-settlement tissue, *emx-Am* transcripts are detected in cells that fit descriptions of cnidarian neural cells, although this could not be confirmed unambiguously. In post-settlement polyps, no clear distribution pattern is apparent, but in pre-settlement planulae, expression of *emx-Am* in these cells types is restricted along the O/A axis. No expression is detected at the oral end of planula larvae, but cells expressing *emx-Am* are abundant at the aboral end of the planulae, which is anterior with respect to direction of swimming. Although direct comparisons with triploblastic animals are difficult, this pattern of expression resembles the role of *ems/Emx* genes in anterior patterning in higher metazoans.

Data from the yeast-one hybrid and EMSA experiments suggest the Emx-Am homeodomain shares a similar DNA-binding affinity with the *Drosophila* ems homeodomain. While only preliminary, these data represent a first step in understanding the downstream regulatory activity of the Ems/Emx family.

3.5. FUTURE DIRECTIONS

Future studies of cnidarian *Emx* genes are open to several lines of investigation. In situ hybridisation studies suggest *emx-Am* functions in anterior patterning and nervous system specification in the coral, corresponding to the role of this gene family in more specialised metazoans. To unequivocally confirm this hypothesis, double-labelling experiments are needed. This requires double-antibody staining techniques using an

anti-RFamide and an anti-Emx-Am HD antibody to be optimised, or if an RFamide cDNA can be isolated, double in situ hybridisation experiments could be attempted.

Conservation of function between the mouse *Emx2* and *Drosophila ems* gene has been demonstrated through partial rescue experiments of the fly *ems* null mutant. A similar transgenic approach using the coral *emx-Am* gene to rescue the fly null phenotype has been initiated in collaboration with Prof. Heinrich Reichert at the University of Basel. In order to investigate the role of the hexapeptide in *Drosophila* ems, I have mutated the fly hexapeptide (YPWL to AAAL) and created a hexapeptide in the *Acropora* emx (YPCA to YPWL) at a similar location. Four pP(UAST) constructs containing: (i) fly ems, (ii) coral emx, (iii) fly ems with hexapeptide deleted and (iv) coral emx with hexapeptide, were cloned and have been sent to the University of Basel to begin fly rescue experiments. Results from these experiments will help to clarify the role of the hexapeptide and whether functional conservation of the *ems*/*Ems* gene family bridges the diploblast/triploblast division. Comparative expression studies of *ems*/*Emx* genes within the Cnidaria would be further supported by data from another anthozoan, such as the sea anemone *Nematostella*. Similar pattern of expression of *ems*/*Emx* genes during embryogenesis in another anthozoan would strength the hypothesis that this gene family has ancestral roles in nervous system development.

The downstream regulatory activity of Ems/Emx proteins is still in doubt. Given that the complete *Drosophila* genome sequence is now available, yeast-one hybrid experiments using the *Drosophila* Ems homeodomain to screen the fly yeast-one hybrid reporter library would quickly determine if this domain interacts with the regulatory regions of specific genes. Comparative studies in other triploblastic organisms would determine if conserved regulatory pathways are in place.