Chapter 4 - Isolation and Preliminary Characterisation of cnox1Am and barhAm genes from *Acropora millepora* - no clear functional correlation with the Hox11-like genes or the Bar family

4.1 Introduction

The Bar and Hox11/Tlx families are defined by the presence of a threonine residue at homeodomain position 47, a feature not found in other homeodomain families. Although many members of the Hox11 family are as yet largely uncharacterised, most of the characterised members function in nervous system patterning. Prior to the start of this project an *A. millepora* genomic clone was isolated in our laboratory, and shown to contain a *Tlx/Hox11*-related gene, that was most similar to the predicted *Drosophila* gene *cg13424* and is hereafter known as *cnox1Am*. In addition, a preliminary EST analysis of *Acropora* (Kortschak *et al.*, 2003) led to the identification of a Bar-related gene, referred to here as *barhAm*. As members of these gene families function in nervous system patterning in higher animals, their temporal and spatial expression patterns were investigated.

4.1.1 cnox1Am and cg13424 – Related to the Hox11 and Mnx families

Preliminary BLASTx analysis indicated that *Drosophila* CG13424 was the closest sequence in the database to the partial sequence available for the *Acropora* Cnox1Am from the EST; the predicted homeodomains have 72%/80% identity/similarity. Comparison of the CG13424 protein with others clearly placed it into the Tlx/Hox11 group, with no obvious matches to other homeobox gene families, thus the implication was that Cnox1Am was likewise a Tlx/Hox11 family member. However, analysis of the full-length Cnox1Am protein indicates similarities not only to the Tlx/Hox11 group, but also to HB9/Mnr2 proteins from chicken and zebrafish, which define a subfamily of the Mnx class of homeodomain containing transcription factors. Additionally, several clear orthologs of CG13424 have subsequently been added to the database (see Fig 4.2).

So far only 3 members of the Mnr2 family have been described, chick Mnr2 and zebrafish Mnr-2a and –2b, and it appears this subfamily of Mnx genes has been lost in mammals. In the chick, studies on the effect of ectopic expression of Mnr2 have shown that it plays a major role in specifying various aspects of motor neuron identity and in interneuron fate decisions (Tanabe *et al.*, 1998). Related genes of the HB9 family have been identified not only in vertebrates (e.g., mouse, *Xenopus* and chicken), but also in invertebrates (e.g., sea urchin, amphioxus and *Drosophila*) (Saha *et al.*, 1997; Bellomonte *et al.*, 1998; Ferrier *et al.*, 2001; Broihier and Skeath, 2002; Odden *et al.*, 2002). In vertebrates, HB9 genes are involved in the first stages of pancreas morphogenesis and in the differentiation of β-cells. In invertebrates (which lack a pancreas) embryonic expression of *hb9* is in the primitive trunk endoderm (Bellomonte *et al.*, 1998; Ferrier *et al.*, 2001; Broihier and Skeath, 2002; Odden *et al.*, 2002). The relevance of this highly conserved expression remains to be investigated. In both vertebrates and invertebrates, *hb9* genes are essential for motoneuron consolidation (Arber *et al.*, 1999; Thaler *et al.*, 1999) and are required for neuronal migration and axonal outgrowth in mouse, chicken and *Drosophila* (Tanabe *et al.*, 1998; Arber *et al.*, 1999; Thaler *et al.*, 1999; Broihier and Skeath, 2002; Odden *et al.*, 2002; William *et al.*, 2003). The N-terminal domain of MNR2 is required both *in vivo* and as a transcriptional repressor in *in vitro* cell-based reporter assays (William *et al.*, 2003). In mouse, HB9 has been shown to repress its own expression (Arber *et al.*, 1999; Thaler *et al.*, 1999). The precise mechanism of MNR2- and HB9-mediated transcriptional repression remains unclear. MNR2, like many other HD proteins (Muhr *et al.*, 2001), possesses a well-conserved Eh1 motif that, in other contexts, can recruit Groucho class co-repressors (Smith and Jaynes, 1996). However, deletion of the Eh1 motif in MNR2 does not affect its role in motor neuron generation and when the HD of MNR2 is spliced to a powerful Groucho recruitment domain, induction of motoneurons *in vivo* is poor (William *et al.*, 2003). It therefore seems that interaction with Groucho co-repressors is not essential for the repressive function of MNR2 (and by implication HB9), and it may be that that MNR2 repressor activity requires the recruitment of Ctbp class co-repressors (William *et al.*, 2003). Evidence has suggested that in some *Drosophila* proteins, Eh1-recruitment and Ctbp-recruitment domains might cooperate to mediate repressor function (Hasson *et al.*, 2001; Barolo *et al.*, 2002).

Tlx /Hox11 genes are orphan homeobox genes that play critical roles in the regulation of early developmental processes in vertebrates. The name of this family of genes is confusing because they are not true Hox genes and are physically located outside the *Hox* gene clusters. In addition, the designation Tlx, (for T-cell leukaemia homeobox-containing gene) (Raju *et al.*, 1993) is not accurate as expression of these genes is not restricted to T-cells (Andermann and Weinberg, 2001), and the same name has subsequently been used to identify vertebrate homologs of the *Drosophila* zinc-finger *tailless* gene (Yu *et al.*, 1994). The core DNA motif recognised by HOX11 has been determined to be TAAGTG or TAATTG using *in vitro* selection techniques (Tang and Breitman, 1995) and direct interactions between the homeodomain of HOX11 and the regulatory regions of specific genes is most likely essential for its role as an oncogene (Owens *et al.*, 2003). Mouse HOX11 has both transcription activation and repression roles, and its repressive function is independent of DNA-binding; it is most likely mediated via protein-protein interactions with other transcription factors (Owens *et al.*, 2003).

In humans, the *HOX11* gene was originally identified due to its association with the breakpoint in specific chromosome translocations in patients with acute T-cell leukaemia, and has since been found to be essential for spleen development during embryogenesis (Dube *et al.*, 1991; Hatano M *et al.*, 1991; Kennedy *et al.*, 1991; Lu *et al.*, 1991; Heidari *et al.*, 2002). The mouse ortholog *Hox11*, is involved in splenogenesis, but is also expressed in the developing hindbrain, spinal cord and the neurons of the developing cranial sensory ganglia (Dear *et al.*, 1993; Raju *et al.*, 1993; Roberts *et al.*, 1995). The presumptive chicken ortholog exhibits a similar expression pattern (Logan *et al.*, 1998), while the potential Xenopus ortholog, *Xhox11* is expressed in regions similar to the mouse, but not the spleen (Patterson and Krieg, 1999). Zebrafish possesses three *Tlx/Hox11* genes; *Tlx-1, Tlx-3a and –3b*; *Tlx-3a* and *–3b* most likely represent the result of a duplication of an ancestral *Tlx*-like gene. Similar to the expression of other vertebrate *Tlx/Hox11* genes, *Tlx-1* is expressed early in regions of the brain and is also present in splenic primordial tissue, although it differs with that of its orthologs in vertebrates as it is not present in the cranial sensory ganglia or spinal cord (Langenau *et al.*, 2002). *Tlx-3a* and *Tlx-3b* appear to have complementary expression patterns in the developing nervous system – *Tlx-3a* is expressed in the

cranial ganglia, enteric neurons and some non-neural tissues while *Tlx-3b* is expressed in the dorsal root ganglia (Langenau *et al.*, 2002).

4.1.2 BarhAm – A cnidarian member of the Bar class of transcription factors

The Bar class of homeodomain-containing transcription factors have been described from a wide range of organisms. Two atypical characteristics define and distinguish the Bar class - the presence of a threonine residue at homeodomain position 47 (T47) rather than an isoleucine or valine residue, and the presence of a tyrosine residue (Y49) rather than the near-universal phenylalanine (F49) residue in helix 3 (Burglin, 1994). Note that in the Hox11/Tlx family, a threonine at position 47 is also present.

Bar class genes were first identified in *Drosophila*, after investigation of the locus involved in the Bar mutation in *Drosophila melanogaster* and the related *Om*(1D) mutation of *Drosophila ananassae*, which impedes ommatidium differentiation. In both cases, the loci responsible contained a novel homeobox gene of the *Bar* class, *BarH1* (Kojima *et al.*, 1991). Characterisation of the locus revealed a second *Bar* gene, *BarH2* and both *Drosophila Bar* genes have since been found to have multiple roles in eye development including the formation of pigment and cone cells (Higashijima, 1992; Higashijima *et al.*, 1992). In addition, the *BarH1* and *BarH2* regulate microchaetae formation in the anterior notum and are important in the development of distal leg segments (Sato *et al.*, 1999; Kojima, 2000).

In the mouse, four *Bar* class genes have been identified, *Barhl1, Barhl2, Barx1 and Barx2*. While the BARX proteins had previously been reported to be closely related to the *Drosophila* Bar proteins, it is now accepted that they belong to a different Bar subfamily and the BARHL1/2 proteins are orthologous to the BarH1/2 proteins of *Drosophila*. The *Barhl1* gene is restricted to the developing CNS, where it may play a role in cell fate determination (Bulfone *et al.*, 2000), and to the hair cells in the inner ear (Li *et al.*, 2002). Although expression data are not yet available for murine *Barhl2*, expression of the rat ortholog, *MBH1* is specific to the eye and CNS (Saito *et al.*, 1998). The *Barx1* and *Barx2* genes have overlapping expression patterns in the central nervous system, including in the telencephalon and spinal cord, while in other non-neural tissues

(for example the developing facial structures), their expression patterns are complementary (Tissier-Seta *et al.*, 1995; Jones *et al.*, 1997).

Two Bar class homeoproteins have also been isolated from the chick; Barx-1 and Barx1b (Barlow *et al.*, 1999; Nakamura *et al.*, 2001). The homeodomains of both chick Bar proteins are identical to each other, and to mouse Barx1 (Tissier-Seta *et al.*, 1995; Barlow *et al.*, 1999; Nakamura *et al.*, 2001), however their NH_2 and COOH termini show no conservation (Nakamura *et al.*, 2001). In both chick and mouse, *Barx-1* expression is present in the stomach, limbs and early facial structures (Tissier-Seta *et al.*, 1995; Barlow *et al.*, 1999). Chick *Barx1b* is also expressed in similar regions, but expression extends to the smooth muscle cells of the upper digestive system (Nakamura *et al.*, 2001).

Members of the *Bar* class of homeobox genes have also been found in *Xenopus* (*XBH1, XBH2, Xbr1a*) (Patterson *et al.*, 2000), sponge (*prox2*) (Seimiya *et al.*, 1994), and the medaka fish (*OlBar*) (Poggi *et al.*, 2002). Prior to this study, only one cnidarian member of the *Bar* class of genes had been isolated (*cnox3; Chlorohydra viridissima* (Schummer *et al.*, 1992)). However, other cnidarian genes have been given the name *cnox3* since, even though they are not related to the *Bar* class, such as in *Hydra vulgaris* (Gauchat *et al.*, 2000), *Podocoryne carnea* (unpublished; AC# BAC56129), and *Eleutheria dichotoma* (Kuhn *et al.*, 1996). Preliminary analysis implied that the *Chlorohydra Bar* ortholog *cnox3*, was more highly expressed in the head region than along the body axis, and on this basis a role during head regeneration has been suggested (Schummer *et al.*, 1992).

4.1.3 Statement of goals

Two partial sequences encoding cnidarian orthologs of the *Tlx/Hox11* and *Bar* homeobox genes were recently isolated from the coral *Acropora millepora*. The *cnox1Am* locus was identified in a genomic clone by Dr Julian Catmull, while a ~380bp *Bar*-like fragment was identified in a preliminary EST analysis of *A. millepora*. In view of the roles of related genes in higher animals, these coral genes were characterised at the level of expression patterns. In parallel, the *Drosophila* homolog of *cnox1Am*

(*cg13424*) was studied. Characterisation of the *Tlx/Hox11*-related genes *cnox1Am* and *cg13424* is described in sections 5.2.1 – 5.2.4 while characterisation of the Bar ortholog *barhAm* is presented in sections 5.2.6 – 5.2.9.

4.2 Results

4.2.1 Isolation and preliminary characterisation of *cnox1Am* and *cg13424*

4.2.1.1 The cnox1Am cDNA

The *cnox1Am* cDNA clone was isolated using a portion of the ORF generated by PCR from a genomic clone isolated by Dr Julian Catmull. The complete cDNA was obtained from a λZAP-II *Acropora millepora* cDNA library following a number of repeat screens of approximately 5 x 10^5 plaques each, from a number of different libraries. This ultimately led to the identification of a single cDNA clone from the post-settlement library of ~1kb comprising 30bp of 5' untranslated region, 666bp of open reading frame and 363bp of 3' untranslated region (Figure 4.1). Conceptual translation of the open reading frame of *cnox1Am* results in a putative protein, Cnox1Am of 222 amino acids. Translation initiation is predicted to begin at base pair 31 of the cDNA clone – while no in-frame upstream termination codons in the cDNA sequence, initiation was deduced to most likely to begin from this codon from comparison with a genomic clone containing *cnox1Am,* which contains the first of a series of in-frame stop codons at 42bp upstream of the putative translation start site. The homeodomain begins 89 amino acid residues from the N-terminus of the Cnox1-Am protein. A motif similar to the Eh1 domain found in other homeoproteins (Smith and Jaynes, 1996) is present 59 amino acid residues N-terminal of the HD (see Figs 4.1 and 3.8).

4.2.1.2 The Cnox1Am protein - Assignment to the Hox11/Tlx family using phylogenetic analysis

The homeodomain of Cnox1Am is located at approximately the midpoint of the protein at amino acid residues 90 – 149 and is related to the vertebrate Tlx/Hox11 and the Mnx

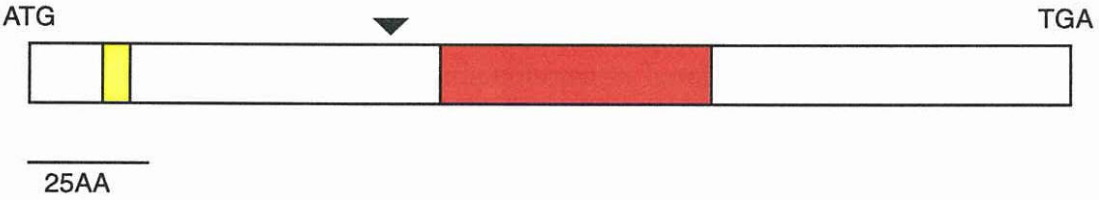
homeodomain families as well as to Drosophila CG13424 and its orthologs (see Fig 4.2). Relationships between these homeodomain families are complex and have not previously been investigated. Although the invertebrate CG13424-related sequences are clearly related to the Hox11-family, these latter are distinguished by the presence of four highly conserved motifs designated TH1 – TH4 (Cheng and Mak, 1993), which function in transcriptional activation (Zhang *et al.*, 1996). These regions are not present in Cnox1Am, however, Eh1 motifs (see Fig 3.8) are a common feature of Cnox1Am, the CG13424-like, Hox 11 and Mnx family proteins.

To better understand the evolutionary position of Cnox1Am and CG13424, maximum likelihood phylogenetic analyses were carried out on the homeodomains using MolPhy Version 2.3 (Adachi and Hasegawa, 1996). Phylogenetic analyses (Fig 4.2A) implied that Cnox1Am is likely to be orthologous to the CG13424-like proteins which have been identified from *Drosophila*, the honey bee (*Apis mellifera)* and the sea squirt (*Ciona intestinalis*). This group forms a clade that is closely related to, but distinct from, the Tlx/Hox11-like proteins. Although related, the Mnx family were clearly distinct from CG13424 and Tlx/Hox11-types (Fig 4.2A).

4.2.1.3 The cnox1Am and cg13424 genomic loci

Comparison of genomic sequence to the *cnox1Am* cDNA sequence revealed the presence of a single intron of 933bp located 29bp 5' of the homeobox (Fig 4.1). In comparison, the *Drosophila cg13424* locus contains two introns - one (891bp) located 17bp 5' of the homeobox, and another (156bp) between nucleotides 132 and 133 of the homeobox (corresponding to amino acid positions 44 and 45 of the homeodomain). Virtual northern analysis indicated the size of the *cnox1Am* transcript to be approximately 1kb (data not shown), which is similar to the size of the *cnox1Am* cDNA clone. This implies that the transcription start site must be proximal (within 100-200bp) – a size discrepancy that could be accounted for by the inaccuracy of agarose gels) to the first exon. TATA and CAAT boxes located at –68 and –161bp respectively fit the spacing requirements for a start site located at approximately –60bp.

A



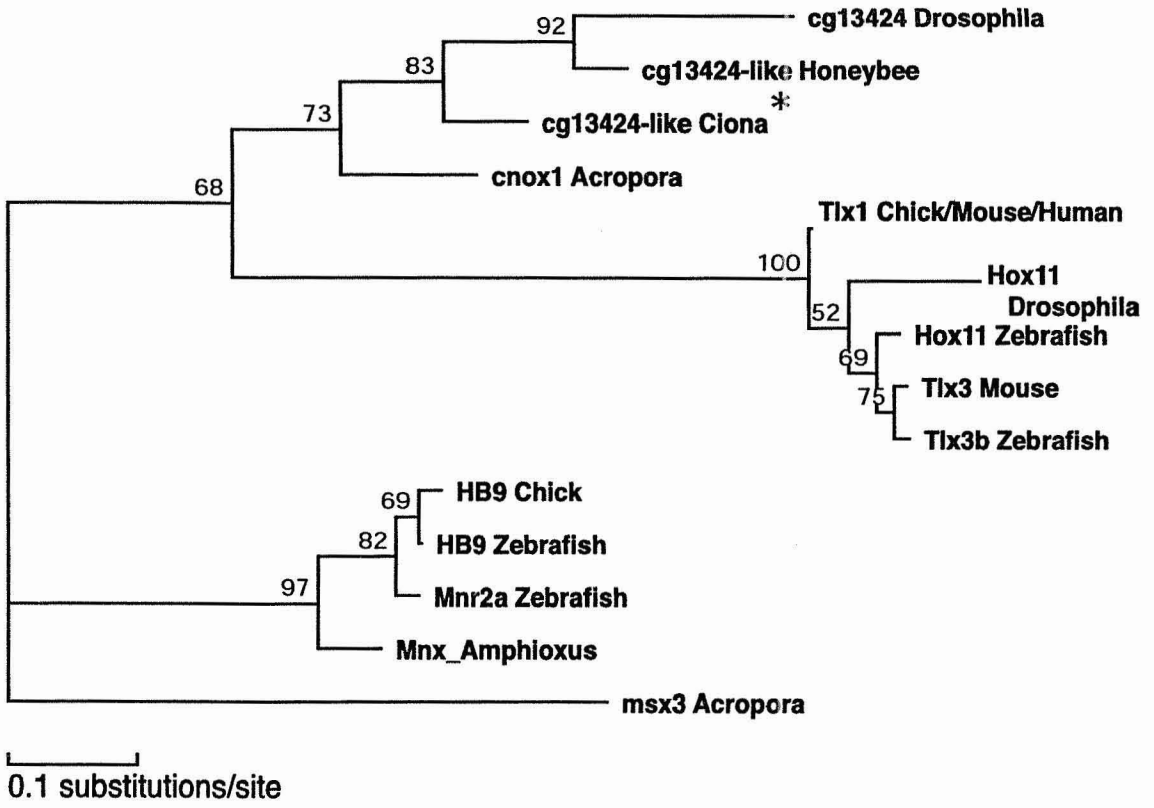
B

```
GCAGGAAGTTCGGGCACGAGGCTCGTACCGATGGAACATTGCCCCCCATCTCCAACA
                                M  E  H  C  P  P  S  P  T

ACAAACAGCAAAGTCCTCAAGTTCACCGTAGAACACATTCTAAGCAAAGAACCTTGT
 T  N  S  K  V  L  K  F  T  V  E  H  I  L  S  K  G  P  C

GGCGGGTCTATAAGCGGTAGTGATGAAAACTCCAGGAGGGATTGGCAATCATATCTT
 G  G  S  I  S  G  S  D  E  N  S  P  P  D  W  Q  S  Y  L

TCGGGAACTAACTTCAAAACACGCACAACTTCTAAAACTTTGGATCTCTCAAGAGAA
 S  G  T  N  F  K  T  R  T  T  S  K  T  L  D  L  S  R  E

ACATCTGACGGTGCAAACTCAGAGGGACGTGCCGATCCAGACTTCAGAAAACGATCA
 T  S  D  G  A  N  S  E  G  R  A  D  P  D  F  R  K  R  S

AGCCTAGATTTGAAGAAGCGACCTCGTACGGCCTTCACCAACGAACAAATCAAAGAA
 S  L  D  L  K  K  R  P  R  T  A  F  T  N  E  Q  I  K  E

CTGGAGAACGAATTTCACAAGAACAAGTACGTTAGCGTTGCGCGTCGAATTGAGCTT
 L  E  N  E  F  H  K  N  K  Y  V  S  V  A  R  R  I  E  L

TCCAAGTTGCTTAAGCTAACAGAAACCCAGATCAAAATTTGGTTTCAAAACAGACGC
 S  K  L  L  K  L  T  E  T  Q  I  K  I  W  F  Q  N  R  R

ACCAAATGGAAGCGACAATTGGCCGCGGAAATGGAATTCACGCTTGGAGCGCAAGGT
 T  K  W  K  R  Q  L  A  A  E  M  E  F  T  L  G  A  Q  G

TATCTGTTTCCCTCACATCCGAGCTGCCGAAGATTTCAGTATTCTCTCGCCAACTCT
 Y  L  F  P  S  H  P  S  C  R  R  F  Q  Y  S  L  A  N  S

AGTTCCCTGCCTGACCCCCGAATAATACAACCGGTAACTTCAAATTCAATGTTCTAC
 S  S  L  P  D  P  R  I  I  Q  P  V  T  S  N  S  M  F  Y

CGACCAACTTTTATACCTGCACCTGGCCAACCCAGTACAAGCCTCTACTCATCGATC
 R  P  T  F  I  P  A  P  G  Q  P  S  T  S  L  Y  S  S  I

GATGGATTTTGATGTCATCGTTTCCCAATTTGAAATTTTACAAGGTGAACCAGCAAGA
 D  G  F  *

AAACAAATTTACATAAGGCGTTTTTATAGTTATATGTTAGTGGACTTCATTTCGGACT
AACGAAGTTCAAAACTGTTTGGATGCAGGGTATTCAAACCTTTTCCATACAGGCGCTC
ACTTAAGTTTACAACAGTATCGTTCTTTTGTAAGAATTGCCTTATTTATAACGGCTTT
TCATCGAGGTATCCTAACAACAAATAAACTAGCTTGAATTAATCACATGAGCTTGTTA
AACACTTCATCTCGTAATTGTGCACTATGCTGTAAATATGCGGTAGGTAATCTTGTTG
TCTTTGTTAATAAAAGCGTGTGTTTCTAAAAAAAAAAAAAAAAA
```

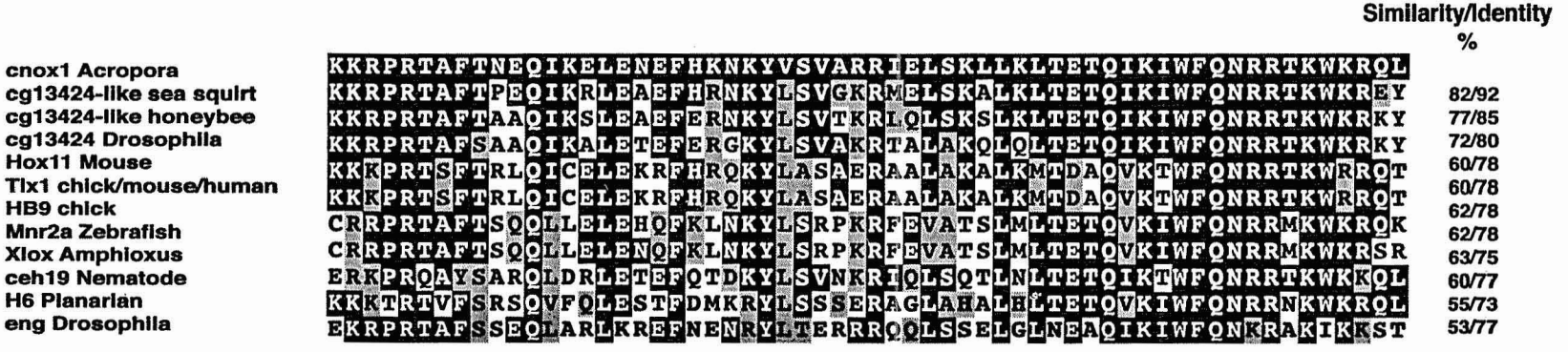
Figure 4.1: The *cnox1Am* cDNA. In both (A) and (B) red shading represents the homeodomain and yellow shading the Eh1 motf. (A) Schematic representation of the *cnox1-Am* ORF. The start (ATG) and stop (TGA) codons are shown at the top. (B) The nucleotide sequence of the *cnox1Am* cDNA and the predicted amino acid sequence of the protein.

Figure 4.2: Comparison of the *Acropora* Cnox1Am protein to other related proteins. (A) Cnox1Am and related homeodomains were analysed by Maximum-Likelihood phylogenetic analysis in MolPhy version 2.3 (Adachi and Hasegawa, 1996) using the Dayhoff model of protein evolution and local rearrangement of the NJ trees. Numbers against branches indicate the percentage of 1000 bootstrap replicates supporting topology. The asterisk indicates the Cnox1Am protein. The Msx3-Am homeodomain from *Acropora* served as an outgroup. (B) A Boxshade alignment of the homeodomain sequences used in the phylogenetic analyses. Identical residues are shaded black and conserved substitutions are shaded grey. The column to the right of the alignment indicates the overall identity and similarity of each protein with the Cnox1Am homeodomain. The species name and GenBank Accession number of each protein used are as follows: *Drosophila melanogaster* cg13424 (NP_611491), Hox11 (Z22959), engrailed (P02836); *Apis mellifera* cg13424-like (XP_393940); *Ciona intestinalis* (sea squirt) cg13424-like (AK174982); *Gallus gallus* Tlx1 (O93366), HB9 (AAC64925); *Mus musculus* (mouse) Tlx1 (P43345), Tlx3 (CAI25153), Hox11 (P43345); Human Tlx1 (NP_05512); *Danio rerio* (zebrafish) Hox11 (NP_705937), Tlx3b (NP_739572), HB9 (AAS07018), Mnr2a (AY445045); *Branchiostoma floridae* (amphioxus) Mnx (AAG33015); *Discocelis tigrina* (planarian) H6 (AJ300664); *Caenorhabditis elegans* (Nematode) ceh19 (NM_068928).

A


92
cg13424 Drosophila
83
cg13424-like Honeybee
*
73
cg13424-like Ciona
cnox1 Acropora
68
Tlx1 Chick/Mouse/Human
100
Hox11
Drosophila
52
Hox11 Zebrafish
69
Tlx3 Mouse
75
Tlx3b Zebrafish
69
HB9 Chick
82
HB9 Zebrafish
97
Mnr2a Zebrafish
Mnx_Amphioxus
msx3 Acropora
0.1 substitutions/site

B


Similarity/Identity
%
cnox1 Acropora
cg13424-like sea squirt
cg13424-like honeybee
cg13424 Drosophila
Hox11 Mouse
Tlx1 chick/mouse/human
HB9 chick
Mnr2a Zebrafish
Xlox Amphioxus
ceh19 Nematode
H6 Planarian
eng Drosophila
82/92
77/85
72/80
60/78
60/78
62/78
62/78
63/75
60/77
55/73
53/77

While upstream promoter sequences controlling gene transcription are generally located within approximately 100bp of the transcription start site (Lewin, 1997), without additional evidence from functional studies it is difficult to ascertain the regions responsible for the regulation of a gene. In closely related species, the regulation of a developmentally important gene is likely to have been conserved. In this way it is possible to identify putative regulatory regions by comparing the non-coding sequence upstream of genes, or their introns in related species and determining if any stretches of sequence are conserved. These methods have been remarkably effective in identifying mammalian regulatory sequences (Pennacchio and Rubin, 2001) and preliminary studies in *Drosophila* suggest their usefulness will extend to the study of insect regulatory sequences (Berman *et al.*, 2004; Sinha *et al.*, 2004). A number of transcription factor databases, such as MatInspector 2.2 and TFSEARCH are available to examine potential transcription factor binding sites for a given sequence, and programs such as BLASTalign allows for pairwise alignment of two nucleotide sequences to identify conserved segments. *D. melanogaster cg13424* is located only 4774bp downstream of another gene, *cg33008*. For this reason, 4700bp of the *Drosophila melanogaster cg13424* putative promoter sequence was compared with the equivalent region of *Drosophila pseudoobscura*. In addition, the first introns in the *cg13424* loci were compared between these species. BLASTalign analysis revealed a number of short stretches of DNA, ranging from 33 to 122bp with 80% - 95% similarity - eight in the 5' UTR and four in the first intron. TRANSFAC® 6.0 (Heinemeyer *et al.*, 1998) analyses of the arthropod database revealed that all but one of these stretches contained binding sites for a number of transcription factors, and binding sites for chorion factor 2 (CF2-II) heat shock factor (HSF), deformed (Dfd) and broad complex (BR-C) were conserved in equivalent positions between species (see Fig 4.3).

4.2.1.4 Spatial expression patterns of cnox1Am and cg13424

The spatial expression patterns of both *cnox1Am* and *Drosophila cg13424* were analysed by *in situ* hybridisation (*Acropora*, *Drosophila*), double *in situ* hybridisation (*Drosphila*) and *in situ* hybridisation/22c10 antibody staining (*Drosophila*).

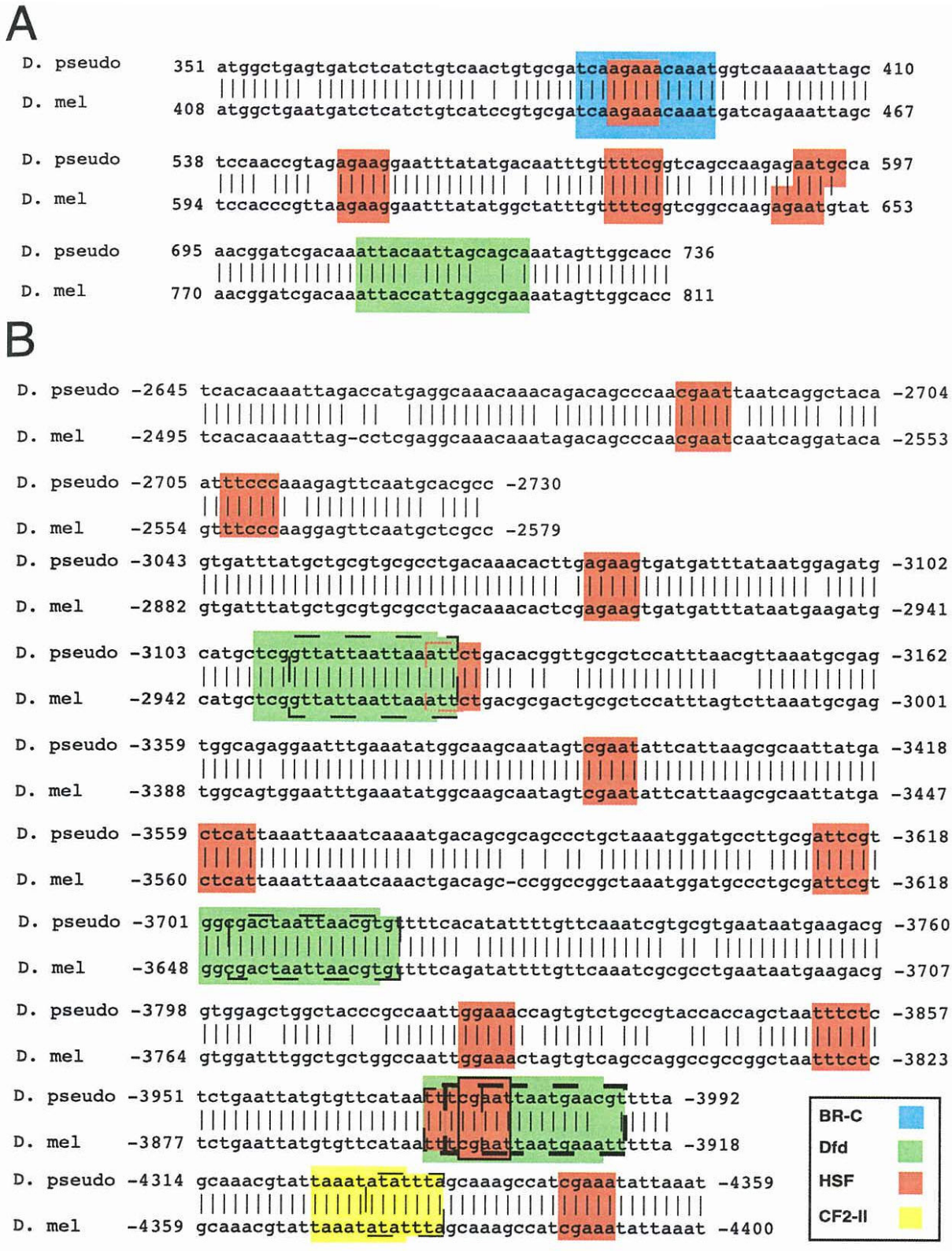


Figure 4.3: Transcription factor binding sites in the conserved between the first intron and putative promoter region of *D. melanogaster* and *D. pseudoobscura cg13424* genes. (A) Conserved transcription factor binding sites in the first intron of the *cg13424* gene of both species. The relevant species is shown at the left of the sequence, and numbers refer to the position from the first base of the intron. (B) Conserved transcription factor binding sites in the 4700bp putative promoter sequence of the *cg13424* gene from both species. Numbering refers to the base position relative to the start codon. Blue boxes represent Broad Complex (BR-C) binding sites, green Deformed (Dfd) binding sites, red Heat Shock Factor (HSF) binding sites and yellow Chorion factor (CF2-II) binding sites. Overlapping sites are shown by dotted lines.

A digoxigenin (DIG)-labelled riboprobe generated from linearised plasmid containing the complete *cnox1Am* was used in the *Acropora in situ* experiments. Only post-settlement stages of *Acropora* were used for *in situ* experiments, on the basis of the virtual northern results (data not shown). *In situ* hybridisation showed *cnox1Am* to be expressed in a spatially restricted pattern, although staining did not appear to be cell type specific. Occasionally, *Acropora* embryos appear mushroom shaped (see Fig 4.4A(i) – (iv)) although it is not known which stage of development this represents. It was initially thought this stage represented the intermediate progression from pre-settlement planulae to post-settlement embryos as the planula shorten and thicken along their oral-aboral axis. However, 'mushroom' embryos are not seen frequently enough in developing embryo collections to be known with certainty to represent an intermediate stage (personal observations). It is now thought that this may be an abnormality of certain embryos around the time of settlement. In mushroom-shaped embryos, *cnox1Am* expression is restricted to the ectoderm covering the entire oral surface, and is completely absent from the aboral 'stalk' region (Fig 4.4A(i) – (iv)). In the, flattened disc post-settlement stage *cnox1Am* continues to be expressed in the ectoderm of the entire oral surface but not in the basal surface ectoderm (Fig 4.4A(v)a – (v)b); this is most apparent in transverse sections (Fig 4.4B (i) – (iii)).

In situ hybridisation experiments were also conducted on *Drosophila* embryos using a *cg13424* riboprobe. The complete *cg13424* cDNA was obtained from ResGen™ (Invitrogen Corporation) via the Flybase website (http://flybase.bio.indiana.edu) and was sequenced prior to manipulation. A 1kb portion (base 21 to 1020) of the *cg13424* ORF was PCR amplified and subcloned into pGEM-T. The *cg13424*/pGEM-T clone was then linearised to act as template for production of DIG-labelled and Fluorescein(FL)-labelled riboprobes. *In situ* hybridisation was performed using the DIG-labelled *cg13424* riboprobe and NBT/BCIP as a substrate.

In order to facilitate interpretation of the CG13424 expression patterns, double *in situs* were also attempted using a combination of either DIG-labelled *cg13424* riboprobe with FL-labelled even skipped riboprobe or FL-labelled *cg13424* riboprobe with DIG-labelled *even skipped* or DIG-labelled *bagpipe* riboprobes, kindly provided by Dr David Hayward and Dr Masha Smallhorn respectively. Primary antibodies conjugated to different enzymes were also used; either anti-FL antibody conjugated to alkaline

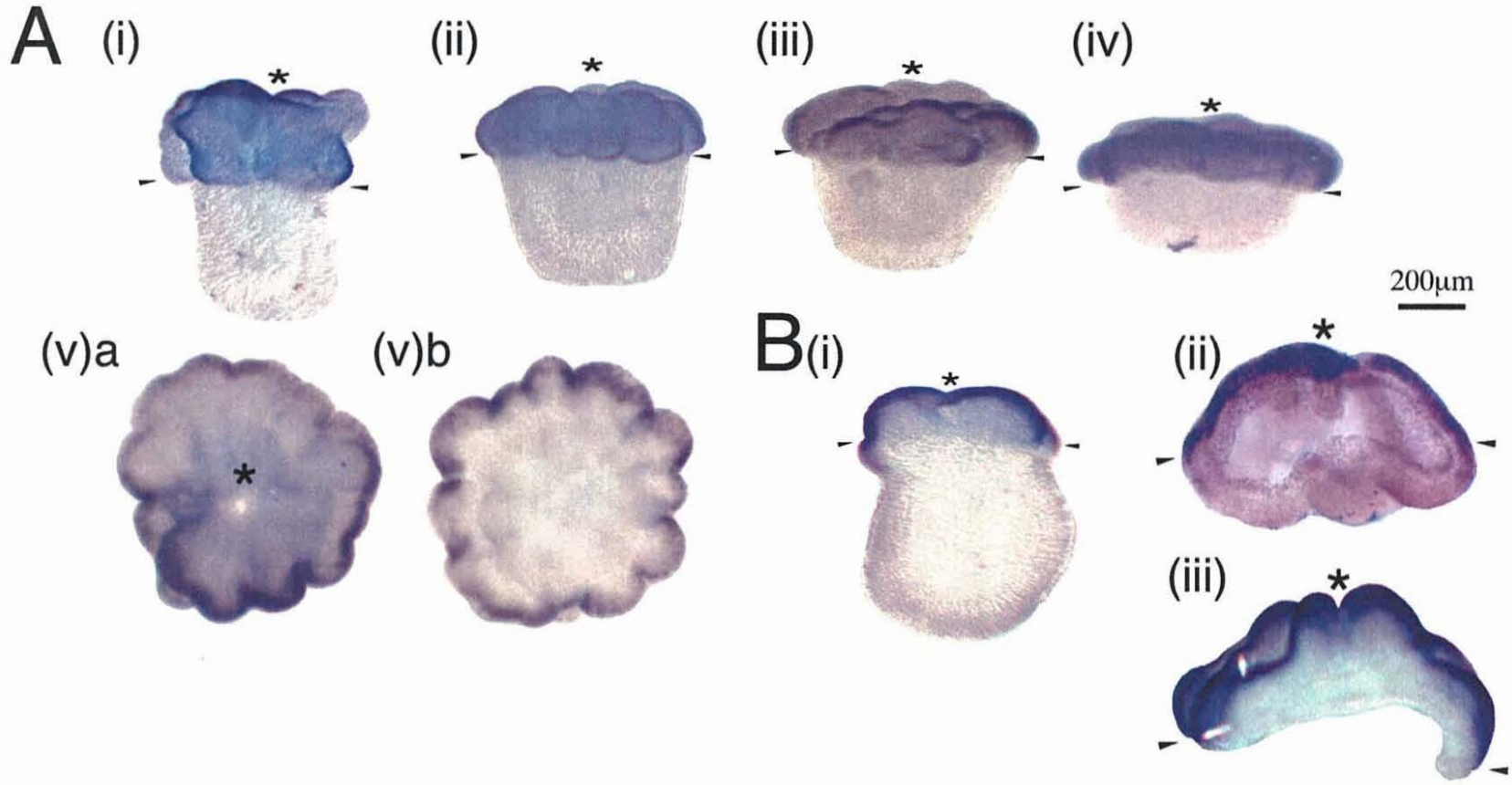


Figure 4.4: Spatial expression pattern of *cnox1Am* in post-settlement polyps. *In situ* hybridisation pattern observed in post-settlement polyps when hybridised with a DIG-labelled *cnox1Am* probe. Asterisks mark the oral pore and arrows mark the limits of *cnox1Am* expression. 4.4(A)(i) - (iv) show expression in the mushroom shaped polyps which may be an abnormality of development (see section 4.2.1.4) but which clearly show expression in the oral surface. 4.4(v)a and (v)b show expression in the oral side and aboral side respectively. B(i), (ii) and (iii) show expression in cross-sections which show the expression in the oral ectoderm, which is absent from the aboral ectoderm.

phosphatase (AP), or an anti-DIG antibody conjugated to AP or horseradish peroxidase (HRP). Substrates used to detect message were Vector Red (Vector Laboratories) or NBT/BCIP (Vector Laboratories) used in reactions catalysed by AP, and DAB used in reactions catalysed by HRP. In addition, direct visualisation of the fluorescein riboprobe after hybridisation was attempted, without the use of an antibody or colour substrate. Unfortunately, no combination was found which allowed for the visualisation of both *cg13424* with either *even skipped* or *bagpipe* (results not shown).

Further attempts to relate the expression of the *cg13424* transcript to known tissues involved *in situ*/antibody double staining experiments performed using DIG-labelled *cg13424* riboprobe and the 22c10 antibody, which stains neurons of the peripheral nervous system (Zipursky *et al.*, 1984). For these experiments, *Drosophila* embryos that had been fixed in paraformaldedyde for 12 – 15 min ('short fix') were used, as the longer fixation times (30-60 mins; 'long fix') generally used for *in situ* hybridisation lead to loss of the epitopes recognised by the 22c10 antibody. Successful *in situ*/antibody double staining was achieved following overnight hybridisation of DIG-labelled *cg13424* riboprobe followed by incubation of the embryos with both antiDIG-AP and 22c10 primary antibodies simultaneously (see section 2.2.2.2). Colours were developed using first the NBT/BCIP substrate to detect the *cg13424* signal, followed by addition of goat anti-mouse-HRP secondary antibody and finally detection of 22c10 protein with the use of DAB as a substrate.

The results of the *in situ* hybridisation and *in situ* hybridisation/22c10 Ab double staining experiments are shown as Figs 4.5 and 4.6. The *in situ* data imply that expression is initiated around stage 9 in the maxillary bud (Fig 4.5A). By stage 12, expression can be seen in the thoracic segments T1, T2 and T3 (Fig 4.5C); the expression domain continues to expand towards the posterior in stages 14 and 15 (Fig 4.5D-E), until in stage 16 all segments show staining in the body wall (Fig 4.5F). The double in situ/22c10 antibody results clearly shows that the cells expressing CG13424 lie deeper than the peripheral nervous system, which is in the surface ectoderm of the *Drosophila* body wall. Axons and sensory cells of the peripheral nervous system are clearly seen to pass over the top of the cg13424 stained cells (See Fig 4.6). The three obvious cg13424 stained cells (See Fig 4.6(A)) are located in a position where there are

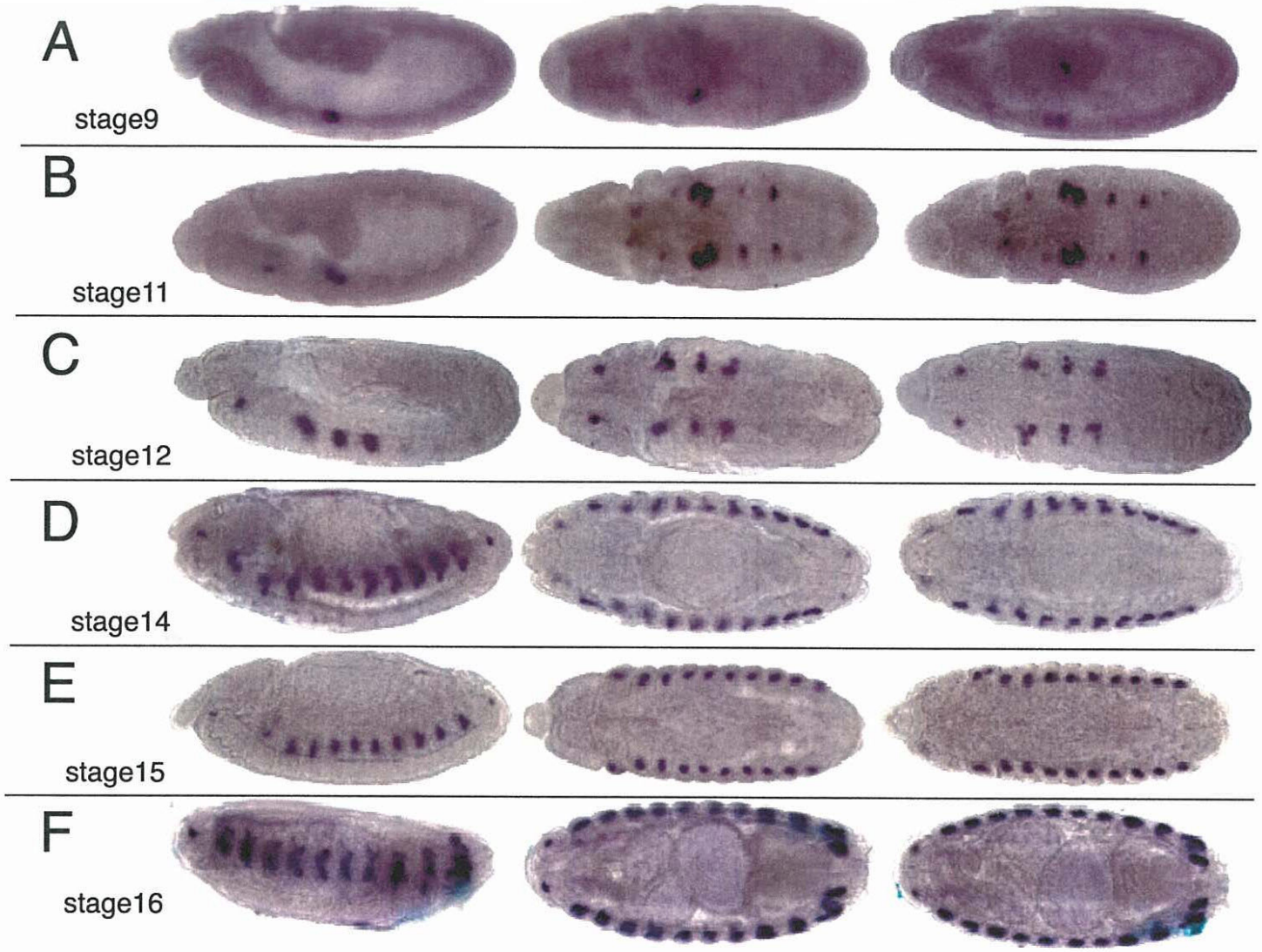


Figure 4.5: Spatial expression pattern of *cg13424* at various stages of *Drosophila* development. The stages are noted to the left of each series row. The first column of each shows a lateral view, the middle column a dorsal view and the far right column a ventral view. Expression initiates at stage 9 in the maxillary bud and extends to the thoracic segments (T1, T2, T3) by stage 12. In later stages (14 - 16), expression is seen in the lateral transverse muscles of the body wall.

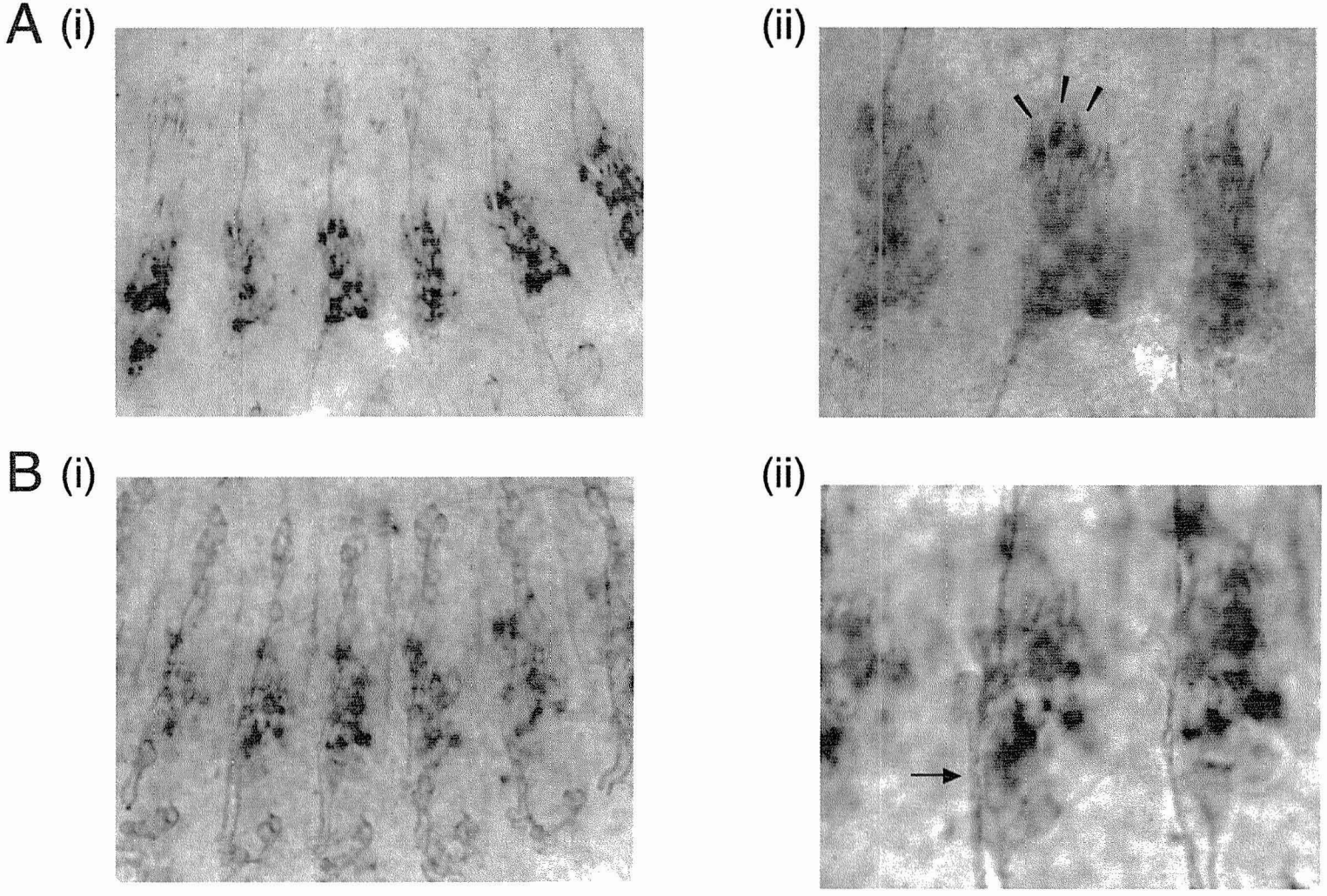


Figure 4.6: Coexpression of *cg13424* and 22c10 in the body wall of late stage *Drosophila* embryos. In each, *cg13424* cells are stained with NBT/BCIP (purple/blue), while 22c10 is stained with DAB (light brown). Fig 4.6A(ii) is an enlarged version of A(i) as B(ii) is an enlarged version of B(i). The three arrowheads in A(ii) show the three characteristically shaped *cg13424*-expressing cells. The arrow in B(ii) shows the axon of the 22c10 stained peripheral nerves passing over the *cg13424*-expressing cells.

no sense organs, i.e. in abdominal segments A1-7 between the 1cg5 and dbd neurons. The position and alignment of cells expressing *cg13424* are reminiscent of the lateral muscle group which consists of three muscles that run dorso-ventrally – the lateral transverse muscles (Landgraf *et al.*, 1997). The above results imply that *cg13424* is specifically expressed at the dorsal attachment point of these muscles.

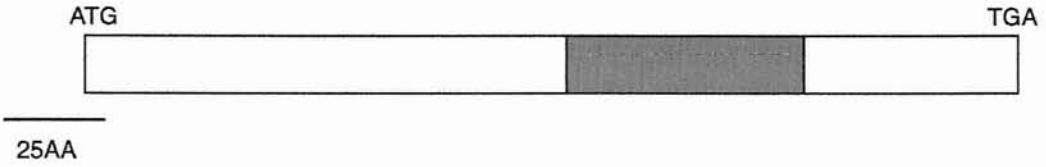
RNA interference (RNAi) was proposed in order to determine the knockout phenotype of *Drosophila cg13424* null mutants. Initiation of this work is presented as Appendix C, and is ongoing.

4.2.2 BarhAm – a second cnidarian member of the Bar class

4.2.2.1 The barhAm cDNA

Virtual northern analysis of *barhAm* indicated expression in post-settlement and adult stages, represented by a single band of ~1.5kb (data not shown). A screen of $5x10^5$ clones of a λZAP-II post settlement cDNA library using the complete *barhAm* EST as a probe yielded four positive clones, three of which covered different regions of the same gene, together spanning ~1.4kb and thus correlating with the results obtained from virtual northern blotting. The contig derived from the three cDNA sequences contained a putative ORF of 669bp with 159bp 5' UTR and 555bp of 3' UTR (see Fig 4.7). Conceptual translation of the ORF of *barhAm* gives a putative protein BarhAm of 223AA. In the absence of confirmatory evidence, the assignment of the start codon is tentative. The sequence of the only other known cnidarian Bar family member, *cnox3* from *C. viridissima* is incomplete, therefore it is not possible to make inferences about translation start sites etc by comparison with this gene. A number of unsuccessful attempts were made to isolate *barhAm* genomic clones, using the longest cDNA clones as probes. The reasons for failure are not clear; one possibility is that the coding sequence may be highly interrupted.

A



B

```
ACCCTCACTAAAGGGAACAAAAGCTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAA
CTAGTGGATCCCCCGGGCTGCAGGAATTCGGCACGAGGAGGTCGTGCTAGCTTCGGT
TTTCGCGTGAACTCAACCTCGCGTCAACGATATCGCCCAGTAGCCGACATGTATTAC
                                                  M  Y  Y

CAAGATTCGTTCCTGTTTTACCAAAAGCCTCAGTATATGCAGAATAACCTCTTTTCT
 Q  D  S  F  L  F  Y  Q  K  P  Q  Y  M  Q  N  N  L  F  S

CCGCCGACACATTTCCTCGACGAAAGCAGAATTTGCAACCCTTCACAGCCTTGTTTT
 P  P  T  H  F  L  D  E  S  R  I  C  N  P  S  Q  P  C  F

ATGACCAAACAAACCAGCGCCTTTACTCAACCCGAAGCGTTCAACTCAAACTTGCCA
 M  T  K  Q  T  S  A  F  T  Q  P  E  A  F  N  S  N  L  P

TCTCAATTCCCATTCAGTTACGGAACTGACCCGAGAGCCCTTCTTCCGTCTTCCTCA
 S  Q  F  P  F  S  Y  G  T  D  P  R  A  L  L  P  S  S  S

GGAAAACTCAGCATTCCTGAGTCCTCCGCAAGCTCAGAGATAACCATTCACCACCCT
 G  K  L  S  I  P  E  S  S  A  S  S  E  I  T  I  H  H  P

CTGCCAGTTTACTTCCGTGTGAAACCAGCTCTCCGAACACCCAGTGGAAAACGATGT
 L  P  V  Y  F  R  V  K  P  A  L  R  T  P  S  G  K  R  C

CGAAAATCTCGCACTGTCTTCACCGACTTACAACTAAGAGTTCTGGAGAAAAAGTTC
 R  K  S  R  T  V  F  T  D  L  Q  L  R  V  L  E  K  K  F

TCGGAACAACGATACCTAGACTCAACCAATCGCACGAGGCTGTCACAGATTCTAGGA
 S  E  Q  R  Y  L  D  S  T  N  R  T  R  L  S  Q  I  L  G

CTGAACGAAGCGCAAGTTAAAACCTGGTTTCAGAACAGGCGAATGAAGTGGAAAAGA
 L  N  E  A  Q  V  K  T  W  F  Q  N  R  R  M  K  W  K  R

AGAGAAGCAAAAACGGACAAACCCGACAGCTTCAGTATGGGAAAGAAAGGAGAGAAC
 R  E  A  K  T  D  K  P  D  S  F  S  M  G  K  K  G  E  N

CTAGCTCGAGATCAAGGAAATACAGACAAAGAAACAATGATAAACGGAGGAGAGAAG
 L  A  R  D  Q  G  N  T  D  K  E  T  M  I  N  G  G  E  K

TTTGGGAATACAACGCGTAGTAGTTATGGCCGAGACAATTGAAAGTATCAACGAAGA
 F  G  N  T  T  R  S  S  Y  G  R  D  N  .

AAGATTCAAAAACTGTTGTTACTTCATCATTGATTCAGAGCGCAGAAATGGATACTA
AGTAACTTAAGAAATTATTGTGCCAAAACAACGGAAGTTTACGTTAGTTGTTAAGAA
TCTGGAAAAGAAAAATTTCATTGCAGAAGAGATAAGAAATAATCAGGCTAGCTAATC
AGTAACCTTCCTTCCAGAGGAAAAAAGGCATCTCGTTTCGCCCCCTTTTTTGTCAGC
GGGTCTTGCCTTTCATCTCGAAAATGTAAATATACCAACGATACTAATCTCACTCGC
GAATCCCAGTTATAATAAGGTAAAAAATGAAAGAGAAAGCTAGTAATTTGTGTGAAC
AGAAATCTTTTTTCAAGTGCTTATATAACATGCCTTTTCTCAAGATGTATCTTTTTG
TCGTCAGGCATTTGAATCAGCTGCCTGCCTTGCGGTTTATATATACAATGCATTTAG
CAAAATTGTAAATATCAACTTTAGTATATCGTGATCTATTAAAATTTACAACTTCAA
AAACACGAAAAAAAAAAAAAAAAAA
```

Figure 4.7: The *barhAm* cDNA. In both (A) and (B) red shading represents the homeodomain. (A) Schematic representation of the *barh-Am* ORF. The start (ATG) and stop (TGA) codons are shown at the top. (B) The nucleotide sequence of the *barh-Am* cDNA and the predicted amino acid sequence of the protein.

4.2.2.2 The BarhAm protein

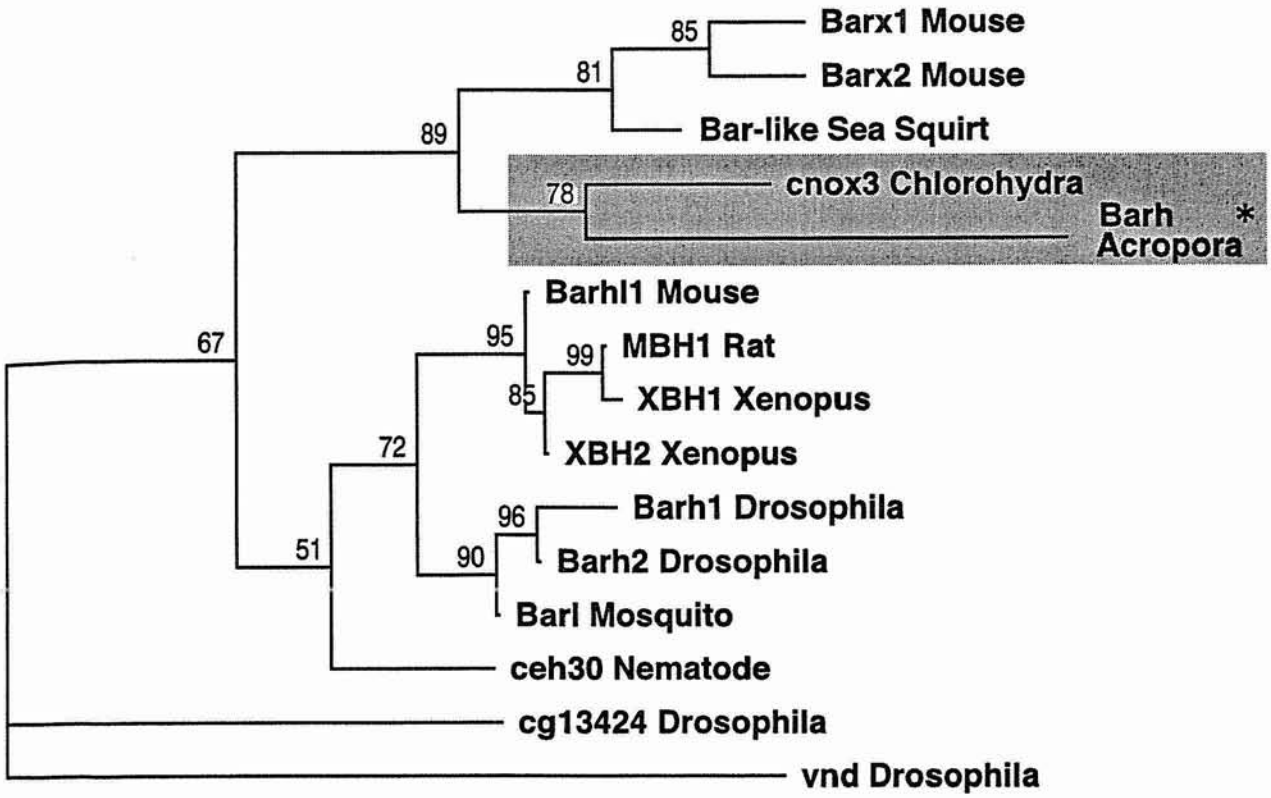
The *Acropora barhAm* gene encodes a protein containing a *Bar* class homeodomain in the C-terminal portion of the protein, at amino acid residues 117 – 176. This homeodomain is more similar to the vertebrate Barx genes than to BarH1/2 from *Drosophila* (see Fig 4.8). As might be expected, the BarhAm homeodomain is most similar to that in Cnox3 from *Chlorohydra* (62/77% identity/similarity). Although BarhAm clearly falls into the Bar family, the homeodomain does not have the characteristic tyrosine residue at position 49, shared by all other known Bar class homeoproteins except for sponge prox2 and *Xenopus* Xbr1a (see Fig 4.8B). Outside of the homeodomain BarhAm does not share any of the features that are present in some other members of the Bar class. For example BarH1 and BarH2 (*Drosophila*), Barhl1 (mouse), XBH1 and XBH2 (*Xenopus*), Bar (medaka) amd cBarx2b (chick) each contain single Eh1 motifs upstream of the homeodomain; MBH1 (rat) contains two such motifs. Mouse Barx1 and Barx2 both contain a basic region of 17 amino acids C-terminal of the homeodomain (Barx basic region/BBR) (Edelman *et al.*, 2000), and Barx2 also possesses a leucine zipper. polyalanine tract and an acidic domain (Jones *et al.*, 1997). Both rat MBH1 and *Drosophila* BarH1/2 proteins contain strings of alanine, glutamine and histidine residues (Higashijima, 1992; Saito *et al.*, 1998).

4.2.2.3 Evolutionary relationships amongst the Bar class of homeodomains

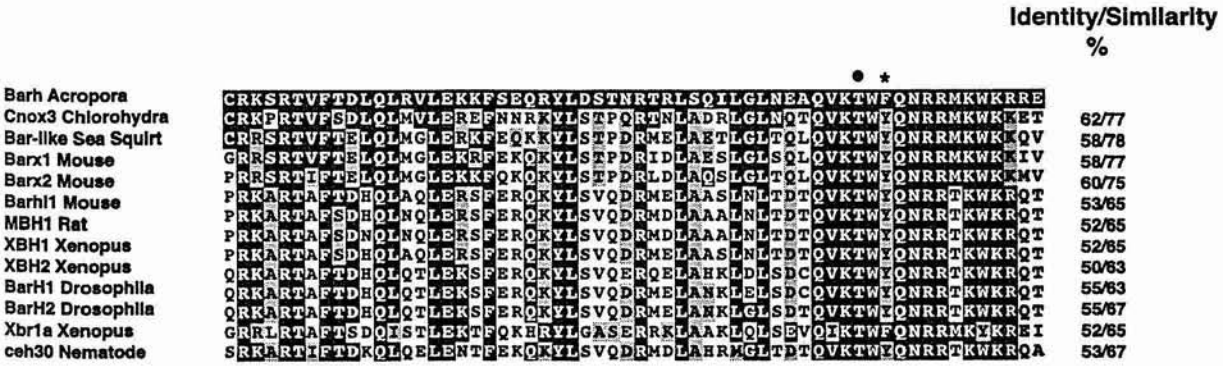
To better understand the evolutionary position of BarhAm, maximum likelihood phylogenetic analyses were carried out using MolPhy Version 2.3 (Adachi and Hasegawa, 1996) based on the homeodomain sequence. Phylogenetic analysis separates most members of the Bar class into two major clades – one containing proteins most similar to the Barx class, and the other containing proteins more similar to BarH from *Drosophila.* The two cnidarian sequences (BarhAm and Cnox3) form a well supported clade that is more closely related to the Barx than to the Barh clades. The length of the BarhAm branch suggests that the protein is more derived than its *Chlorohydra* counterpart, which is unusual; in general, *Acropora* proteins are less

Figure 4.8: Comparison of the *Acropora* BarhAm protein to other Bar-related proteins. (A) BarhAm and related homeodomains were analysed by Maximum-Likelihood phylogenetic analysis in MolPhy version 2.3 (Adachi and Hasegawa, 1996) using the Dayhoff model of protein evolution and local rearrangement of the NJ trees. Numbers against branches indicate the percentage of 1000 bootstrap replicates supporting topology. The asterisk indicates the BarhAm protein. The CG13424 (accession #NP_611491) and Vnd (CAA60619) homeodomains from *Drosophila* served as outgroups. (B) A Boxshade alignment of the homeodomain sequences used in the phylogenetic analyses. Identical residues are shaded black and conserved substitutions are shaded grey. The column to the right of the alignment indicates the overall identity and similarity of each protein with the BarhAm homeodomain. The black dot represents the distinguishing threonine residue at position 47 (T47) and the star the tyrosine residue (Y49) conserved in most Bar-type homeodomains. The species name and GenBank Accession number of each protein used are as follows: *Mus musculus* Barx1 (AAG18573), Barx2 (AAH12684), Barhl1 (NP_031552); *Chlorohydra viridissima* cnox3 (CAA45910); *Rattus norvegicus* MBH1 (O88181); *Xenopus laevis* XBH1 (AAG14450), XBH2 (AAG14451), Xbr1a (AAB03566); *Drosophila melanogaster* Barh1 (Q24255), Barh2 (Q24256); *Anopheles gambiae* Barl (XM_320094); *Caenorhabditis elegans* ceh30 (NP_508524).

A


85
Barx1 Mouse
81
Barx2 Mouse
89
Bar-like Sea Squirt
78
cnox3 Chlorohydra
Barh Acropora
*
Barhl1 Mouse
67
95
99
MBH1 Rat
85
XBH1 Xenopus
72
XBH2 Xenopus
96
Barh1 Drosophila
51
90
Barh2 Drosophila
Barl Mosquito
ceh30 Nematode
cg13424 Drosophila
vnd Drosophila
0.1 substitutions/site

B


Identity/Similarity %
Barh Acropora
Cnox3 Chlorohydra
Bar-like Sea Squirt
Barx1 Mouse
Barx2 Mouse
Barhl1 Mouse
MBH1 Rat
XBH1 Xenopus
XBH2 Xenopus
BarH1 Drosophila
BarH2 Drosophila
Xbr1a Xenopus
ceh30 Nematode
62/77
58/78
58/77
60/75
53/65
52/65
52/65
50/63
55/63
55/67
52/65
53/67

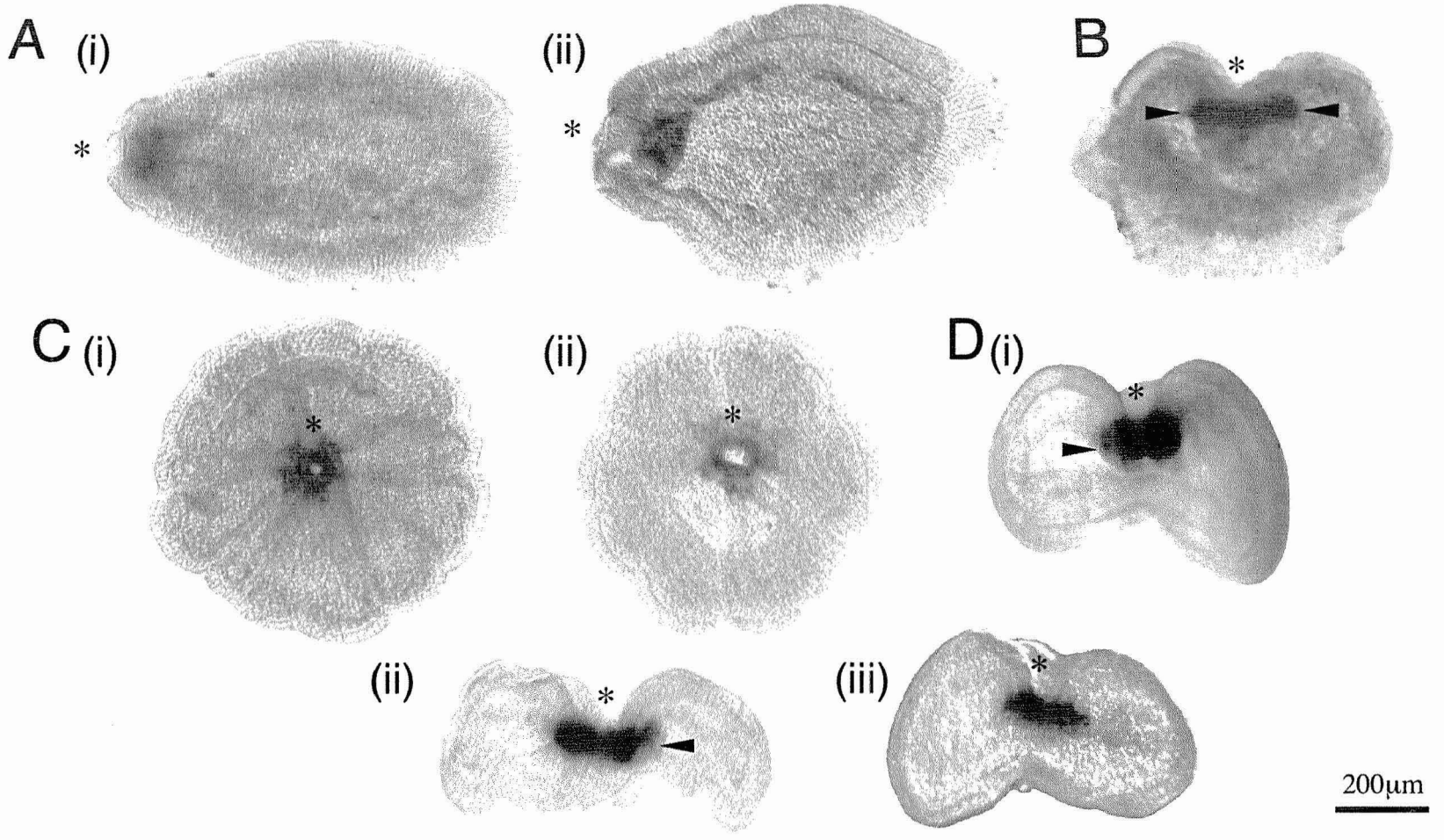


Figure 4.9: Spatial expression patterns of *barhAm* in pre-settlement and post-settlement stages. Asterisks mark the position of the oral pore in each plate. A(i) and (ii) show expression in a whole pre-settlement larvae and a cross-section of a similarly aged larvae respectively. B shows staining in a cross-section of a 'mushroom' stage polyp, while C-D show expression in post-settlement polyps. C(i) and (ii) are polyps of the same age showing expression exclusively around the oral pore, while D(i) - (iii) are cross-sections of polyps of the same age as those shown in C. Expression is restricted to the ectoderm around the oral pore, although 'bleeding' of stain can make the expression appear more extensive (as indicated by arrowheads).

derived than are their orthologs from hydrozoan cnidarians. Within the Barh clade, the arthropod and vertebrate sequences are clearly resolved (see Fig 4.8A).

4.2.2.4 Spatial expression pattern of barhAm

The spatial expression pattern of *barhAm* was analysed by *in situ* hybridisation. A digoxigenin (DIG)-labelled riboprobe was generated from linearised plasmid containing the complete *barhAm* cDNA. While virtual northern results indicated expression only post-settlement (data not shown), *in situ* hybridisations were conducted on both pre- and post-settlement material, as certain collections contained representatives of both stages. Surprisingly, expression in pre-settlement was seen, in apparent contradiction to the virtual northern results.

Expression in pre-settlement larvae is seen in the ectoderm immediately surrounding the oral pore (Fig 4.9A(i)-(ii)). In sections of 'mushroom' stage post-settlement material, expression is restricted to the ectoderm at the cup-like oral pore (Fig 4.9B), and this specific pattern persists as the coral settles and flattens, (Figs 4.9C (i)-(ii), D(i)-(iii)). Virtual northern analysis indicates that *barhAm* is also expression in the adult colony, albeit at lower levels; unfortunately, the calcified skeleton greatly complicates the application of *in situ* hybridisation methods on adult.

4.3 Discussion

4.3.1 Structural features of the Cnox1Am and BarhAm proteins

The Bar and Hox11/Tlx families are typical members of the Antennapedia superfamily in that they contain a glutamine residue at position 50 in the third helix of the homeodomain (Q50) which results in preferential binding at sites containing the core sequence CATTA (Scott *et al.*, 1989; Kappen *et al.*, 1993). However, these same families are atypical in that they are the only known homeodomains with a threonine residue at position 47 (T47) within helix3 – in other homeodomains, isoleucine or

valine is most commonly found in this position, but asparagine, leucine or histidine are also known (Laughon, 1991; Dear *et al.*, 1993). Position 47 is one of the four positions in helix 3 known to confer DNA binding specificity (Burglin, 1994), proven to be the case for the HOX11 HD *in vitro* (Dear *et al.*, 1993); most likely the specificity determined by the threonine in this position extends to roles *in vivo*. A further characteristic of Bar class homeodomains is the tyrosine residue at position 49 (Y49) replacing the near-universal phenylalanine residue. It has been suggested that together, these two positions (T47 and Y49) could play a role in influencing DNA-binding after a posttranslational modification, such as phosphorylation changes the protein conformation (Jones *et al.*, 1997). As the *Acropora* BarhAm homeodomain has F49 rather than Y49 residue, this raises the question of whether this is an ancestral or a derived feature. The fact that *Chlorohydra* Cnox3 HD has Y49 suggests that *Acropora* may contain a second Bar class protein with the F49/Y49 substitution. This is also indicated by the derived nature of BarhAm in comparison to Cnox3, shown by the length of the BarhAm branch in the maximum likelihood analysis (see Fig 4.8).

With the exception of the Eh1 motif (see sections 4.2.1.2 and 4.2.2.2), the Bar and Hox11 proteins exhibit no significant homology outside of the homeodomain. All Barx family members contain an Arg/Lys-rich region N-terminal of the HD, a feature shared by some members of the NK2 family. This region appears to be required for serum response factor (SRF) binding (Chen and Schwartz, 1995), and a direct interaction between SRF and Barx2b has been demonstrated (Herring *et al.*, 2001). The vertebrate members of this class contain additional motifs which may reflect the need for increased complexity of action of Bar proteins in vertebrates, relative to their activity in 'simpler' animals such as cnidarians and *Drosophila*.

Phylogenetic analyses of Cnox1Am and BarhAm places both homeodomains in specific classes – Cnox1Am as a member of the Hox11/Tlx class and BarhAm as a member of the Barx subfamily of the Bar class. It is interesting to note that there are no clear vertebrate orthologs of *Drosophila cg13424* in the database. The presence of an ascidian ortholog implies that loss of members of this gene family has occurred in vertebrate species. No cnidarian homologs of the BarH subfamily of the Bar class of genes have been isolated, although the possibility of cnidarians possessing this subfamily cannot be discounted. Vertebrates contain members of both subfamilies,

while *Drosophila* has only those of the BarH type. Clearly both the BarH and Barx subfamilies were present in the last common ancestor of higher metazoans, and that the latter has been lost in the arthropod lineage, but maintained in the vertebrate lineage. Alternately if cnidarians were conclusively shown to possess only Barx-type genes, it could be hypothesised that while the last common ancestor had both, it is possible that the Barx-type were ancestral, and that the BarH subclass originated after the cnidarian-bilaterian split.

4.3.2 Expression patterns of cnox1Am, cg13424 and barhAm

The *Hox11/Tlx* and *Bar* genes are both expressed in the nervous system in all organisms studied thus far, although each also have diverse roles in other tissues. On the basis of apparent conservation of function in the nervous system, these genes were characterised in *Acropora* in the expectation that they would also play roles in nervous system development or patterning in cnidarians. However, both *cnox1Am* and *barhAm* were more generally expressed in the ectoderm around the time of settlement. While expression in the nerve net cannot be discounted, it is apparent that expression of neither gene is limited to any one cell type; rather, they are both uniformly expressed in spatially restricted patterns. *BarhAm* expression is seen as a single band in the ectodermal layer surrounding the oral pore and is initiated at settlement. During this time, the nervous system of *Acropora* is completely remodelled (E Ball, pers comm.), as has also been described in hydroids (Martin, 2000). Settlement initially involves the receipt of settlement cues which induce the planula larvae to attach to the substratum at its aboral end and contract along the oral-aboral axis to form a flattened disc, giving the characteristic post settlement morphology. It is possible that *cnox1Am* might function in regulation of this process, as it is exclusively expressed in the ectoderm on the oral surface of the embryo, from approximately the time of settlement. The expression pattern of the Chlorohydra gene *cnox3* is similarly restricted to the head region, and it has been suggested that this gene plays a role in head regeneration (Schummer *et al.*, 1992). The role of *barhAm* in *Acropora* is unclear; it can be speculated that its restricted expression pattern might reflect a role in specifying a specific cell type(s) to ensure the correct functioning of the oral pore.

The expression pattern of *cg13424* was investigated in *Drosophila* embryos in an attempt to determine the function of this uncharacterised gene. Here, the expression pattern was very specific, in what are thought to be the lateral transverse muscles of the body wall, based also on the morphology of the *cg13424*-specific cell staining. The fly gene was not expressed in the nervous system. Thus, although clearly related in sequence, the fly and coral genes have completely different expression patterns, implying unrelated functions. However, there are many known examples of *Drosophila* genes having diverged function (eg (Strecker *et al.*, 1986)), and it would be of interest to determine the expression patterns of *cg13424*-related genes in a wider variety of animals. The phylogenetic analyses presented here indicate that *cg13424* and *cnox1Am* define a family of homeobox genes that are clearly related to, but distinct from the *Hox11/Tlx* type; from this preliminary characterisation their functions do not appear to be related.

4.3.3 Regulation of cnox1Am and cg13424

Recently, a second *Drosophila* genome (*Drosophila pseudoobscura*) was sequenced by the National Human Genome Research Institue (NHGRI) to aid in the identification of conserved regions, and guide the annotation of functional sequences in *D. melanogaster* (Berman *et al.*, 2004). A secondary use for the data was in identifying conserved regulatory regions of genes. Although conservation of DNA sequence cannot always accurately determine functional regulatory sequences, conservation of binding-site clustering between *D. melanogaster* and *D. pseudoobscura* has been used to discriminate functional binding-site clusters from those with no function, as demonstrated by several recent studies (Bergman *et al.*, 2002; Berman *et al.*, 2004; Grad *et al.*, 2004; Sinha *et al.*, 2004). For this reason, a comparison of the 5' upstream sequence and first intron of both *D. melanogaster* and *D. pseudoobscura cg13424* genes was compared. It was found from a search of the 4.7kb 5'UTR and first intron (see section 4.2.1.3), only four of the binding sites were shared between the two species; those for heat shock factor (HSF), deformed (Dfd), Chorion factor 2 (CF2-II) and broad (Br;BR-C) (Fig 4.3). For any of these proteins to play a role in the regulation of *cg13424*, it would be necessary for these genes to be expressed in analogous positions, i.e. the maxillary bud in early stages and in the lateral transverse muscles in later stages.

Based on published expression data alone, it appears that of these four transcription factors Dfd and CF2-II are possible regulators of *cg13424*.

The binding of HSF homotrimers to heat shock elements (HSEs) is cooperative, but requires close proximity of HSEs, and a minimal complete binding site for HSF is made up of three contiguous units (Fernandes *et al.*, 1995). Results of the TF search of the 4.7kb promoter region and first intron of *D. melanogaster* and *D. psuedoobscura cg13424* genes do not identify more than two contiguous HSEs indicating that none of the HSF binding sites identified can operate as a functional HSF promoter element. Br-C (Broad Complex) is essential for the progression of larave through metamorphosis, and those which lack BR-C die without pupriating (Stewart *et al.*, 1972; Kiss *et al.*, 1978). It is expressed in numerous tissues including the forming CNS, the developing eye and is required for the attachment of the thoracic muscles to the body wall (DiBello *et al.*, 1991; Emery *et al.*, 1994; Sandstrom *et al.*, 1997; Brennan *et al.*, 1998). However, while extensive, the expression of BR-C does not seem to coincide with CG13424 and can thus be discounted as a possible regulator. Dfd is a Hox gene in the antennapedia complex expressed in the eye-antennal and labial imaginal discs, and also in the mandibular and maxilliary segments of the head (see (Hughes and Kaufman, 2002); it is therefore possible Dfd plays a role in regulating *cg13424* in the maxillary bud, but regulation in the muscles of the body wall at later stages is not possible. CF2-II however is expressed in the developing muscles of the embryo where it first appears at stage 12; later it is expressed in all muscle lineages including skeletal, visceral and cardiac, and is in the nuclei of all skeletal muscles (Bagni *et al.*, 2002). Thus, the combinatorial actions of both Dfd and CF2-II might specify the region of *cg13424* expression in early and late embryonic stages respectively. Interestingly, in *Drosophila*, CF2-II expression is dependant on the highly conserved MADS-box gene MEF2; a MEF2 homolog has recently been isolated from the hydrozoan jellyfish *Podocoryne carnea* suggesting the possibility of a conserved muscle patterning cascade (Bagni *et al.*, 2002; Spring *et al.*, 2002).

In orthologous genes, intron positions are often conserved. However, *Drosophila cg13424* contains two introns, whereas *Acropora cnox1Am* contains only one. The first in both of these genes is located at seemingly equivalent positions, located 29bp 5' of the homeobox in *Acropora*, and 17bp 5' of the homeobox in *Drosophila* while the

second *Drosophila* intron is located between amino acids 44/45 of the homeodomain. A plausible evolutionary scenario is that the ancestral *cnox1/cg13424*-type gene contained an intron close to the 5' end of the homeobox, and that the additional intron that is present in the arthropod lineage has accumulated during evolution. Alternately, it may be that secondary loss is responsible for the reduced intron density in *Acropora*; further sampling of *cnox1/cg13424* genes from other organisms will help to resolve this issue.

4.4 Conclusions

In this Chapter, the isolation and expression patterns of two coral genes, a Barx-type gene (*barhAm*) and Hox11/Tlx-related gene (*cnox1Am*), and the preliminary characterisation of a *Drosophila* gene related to the latter are described. Like most members of the corresponding protein families, threonine residues are present at HD position 47 in both BarhAm and Cnox1Am. However in the BarhAm HD F49 is present, rather than Y49 as in most other Bar class proteins. There is no significant conservation outside of the homeobox in either BarhAm or Cnox1Am, with the exception of an Eh1 motif N-terminal of the homeodomain in Cnox1Am which is also present in all members of the Hox11/Tlx family.

Both *barhAm* and *cnox1Am* are expressed around the time of settlement; *cnox1Am* is only present in post-settlement and adult stages, while *barhAm* expression appears to initiate immediately prior to the settlement stage. Both genes are expressed in the ectoderm in spatially restricted patterns. While vertebrate homologs of both *barhAm* and *cnox1Am* are expressed in the nervous system during development, the expression of the *Acropora* genes is not cell type restricted. The *Drosophila* homolog of *cnox1Am*, *cg13424,* is expressed from approximately stage 9, and is restricted to the lateral transverse muscles of the body wall in later stages (14 – 16). Analysis of the promoter region and first intron revealed that the Dfd and CF2-II proteins are potential regulators of *cg13424*.

4.5. Future Directions

The work described here on *barhAm*, *cnox1Am* and *cg13424* suggests several lines of further investigation. While *Acropora* lacks obvious muscle, many cnidarians (eg Podocoyne) have well developed striated and smooth muscles that are derived from a mesoderm-like structure, the entocodon (Spring *et al.*, 2002). Hence it may be informative to characterise *cnox1Am*-like genes from cnidarians with well-defined muscles, as well as *cg13424* family members from other "higher" animals. In terms of understanding the function of the Drosophila gene *cg13424*, it is important that loss of function phenotypes are examined, and the most promising way of addressing this is via RNAi expression under control of the GAL4-UAS system. The identification of downstream targets of this range of genes can now also be attempted via the use of yeast one-hybrid technology.

Chapter 5 - DNA-Binding Characteristics of Cnidarian PaxC and PaxB Proteins In Vivo and In Vitro: No Simple Relationship With the Pax6 and Pax2/5/8 Classes

5.1. Introduction

5.1.1 Common Molecular Mechanisms of Eye Development

To a surprising extent, common molecular mechanisms appear to underlie the early morphogenesis of eyes across the animal kingdom. Classically, eyes were thought to have polyphyletic origins, and to have evolved at least 40 times independently in the various animal lineages - an idea that can be traced back to Darwin's *Origin of Species* (Darwin, 1859; Salvini-Plawen and Mayr, 1977). However, the discovery that the *Drosphila eyeless* (*ey*) gene is orthologous and functionally interchangeable with the mammalian and *Xenopus Pax6* genes (Quiring *et al.*, 1994; Halder *et al.*, 1995; Onuma *et al.*, 2002) led to recognition that aspects of the molecular basis of eye specificiation are similar despite major morphological differences between fly and vertebrate eyes, and suggested a monophyletic origin of eyes. Subsequently, conservation not only of *Pax6* function, but also of much of the genetic network of eye specification and patterning has been demonstrated between *Drosophila* and mammals (Wawersik and Maas, 2000; Kumar and Moses, 2001). In higher animals, clear Pax6 orthologs have been universally implicated in eye specification, and have been cloned from a diverse range of animals including the squid *Logilo* (Tomarev *et al.*, 1997) and ribbon worm *Lineus* (Loosli *et al.*, 1996).

Not only are the Pax6 proteins conserved across deep evolutionary time, so too are functional aspects of their regulatory elements. Autoregulation is a general feature of *Pax6*-like genes, and in the case of the *eyeless* gene of *Drosophila* this is mediated by the two intronic enhancer elements that are required for expression in the nervous system and eye. These enhancer elements share a significant degree of sequence identity with other *Drosophila* species and with *Anopheles gambiae* (Hauck *et al.*, 1999; Adachi *et al.*, 2003). Although lacking significant sequence identity in the non-coding

regions, these *Drosophila* enhancers can recapitulate the major features of mouse *Pax6* expression during development when placed in an appropriate reporter construct (Xu *et al.*, 1999). Regulation of mouse *Pax6* is complex; transcription can start from any one of three promoters under the control of six enhancer elements - four of these regulate different aspects of *Pax6* expression in the developing eye (Williams *et al.*, 1998; Kammandel *et al.*, 1999; Xu *et al.*, 1999; Griffin *et al.*, 2002). These enhancer elements are recognisable across a broad range of non-mammalian vertebrates including quail, pufferfish and *Xenopus*, and the pufferfish enhancer elements for example, can substitute for their mouse equivalents (Kammandel *et al.*, 1999). Identity between vertebrates and invertebrates in the non-coding regions of *Pax6* genes is limited to short motifs (Morgan, 2004) that closely match the *Pax6* consensus binding sites in all of the genes studied (Xenopus, Mouse, Zebrafish, Human, Ciona, *Drosophila* and *Caenorhabditis)* and these are often at equivalent positions relative to the first exon. This suggests that an ancestral control mechanism that has subsequently been 'fine-tuned' in different lineages (Morgan, 2004).

5.1.2 Pax Genes

The defining characteristic of Pax genes is they encode a 128 amino acid DNA binding motif, the paired-domain (PD). The term paired box (from which the name pax originates) was coined due to the fact that the motif was first recognised in the *Drosophila* segmentation gene *paired* (Bopp *et al.*, 1986; Frigerio *et al.*, 1986). In addition to the paired domain, many *Pax* genes also encode another DNA-binding domain, a paired-type homeodomain (HD), and/or a small octapeptide-motif which binds Grouch-class corepressors (Eberhard *et al.*, 2000). *Pax* genes have been cloned from many representatives of a diverse range of phyla, including the cnidarian *Acropora millepora* (Catmull *et al.*, 1998). In many cases, those *Pax* genes which contain a homeobox are expressed in the nervous system during development.

The paired domain is responsible for sequence specific DNA-binding via two helix-turn-helix subdomains, termed PAI and RED, separated by a short linker (reviewed in (Callaerts *et al.*, 1997)). The paired domain binds as a monomer to a non-palindromic recognition sequence in the two major grooves on the one side of the DNA helix

(Czerny *et al.*, 1993). The PAI subdomain recognises the more extensive 5' consensus half-site motif, whereas the RED subdomain interacts with the 3' consensus sequence. Specificity of binding is achieved through the co-ordinated actions of these two half sites (Czerny *et al.*, 1993) and is due to particular amino acids in the paired domains themselves. For example, positions 42, 44 and 47 in the N-terminal region of the paired domain, were shown to be responsible for differences in DNA binding specificity between Pax5 and Pax6 (Czerny and Busslinger, 1995). Mutation of these three specific residues in Pax6 to those of Pax5 resulted in a complete switch of the DNA-binding specificity from Pax 6 toPax5, proving that these three residues are the primary determinants of DNA binding specificity (Czerny and Busslinger, 1995).

The second DNA-binding domain, the paired-type homeodomain, is able to form homo- and hetero- dimers on DNA, and binding of one homeodomain protein has been shown to increase the affinity of a second by up to 300-fold (Wilson *et al.*, 1993). Most homeodomains recognise almost identical sites, and although distinct and independent from the paired domain, recent studies have indicated that the absence of the homeodomain greatly decreases the affinity of paired domain binding. This suggests that the paired domain and homeodomain must interact cooperatively to activate their targets (Jun and Desplan, 1996). Homeodomain proteins have been shown to cooperatively dimerise on palindromic binding sequences known as P_2 or P_3 sites on the basis of the number of nucleotides between the palindromic half sites [TAAT (N_2 or N_3) ATTA] (Wilson *et al.*, 1993; Jun and Desplan, 1996). The homeodomains encoded by different Pax gene classes have characteristic preferences for either the P_2 or P_3 sites.

The diversity of *Pax* gene function contrasts sharply with their conservation at the amino acid level and is puzzling in light of their *in vitro* DNA-binding specificity. The solution to how *Pax* genes achieve their many roles may lie in the ability of the corresponding proteins to use multiple combinations of their helix-turn-helix motifs in order to recognise more targets. In this way, *Pax* genes are able to achieve a wide variety of functions (Jun and Desplan, 1996). Members of the Pax2/5/8 group contain an 'incomplete' homeodomain, which is unable to bind to DNA, and yet has been shown to be conserved in animals from sea urchin to humans, suggesting a non-DNA-binding function for this protein; the partial homeodomain of Pax-5 has been shown to be able to bind TATA-binding protein (TBP) and Retinoblastoma protein (Rb)

(Eberhard and Busslinger, 1999). The C-terminal domain of several Pax proteins contains a proline/serine/threonine (PST)-rich region with cell-type independent transactivation activity, and an adjacent inhibitory sequence (Dorfler and Busslinger, 1996). The octapeptide motif mediates transcriptional repression by recruiting Groucho class co-repressors (Eberhard *et al.*, 2000).

5.1.3 Specification of Mammalian and Drosophila Eyes

The development of the *Drosophila* eye begins in the eye-antennal imaginal disc from a single layer epithelium. Initially, the eye primordium is subdivided into dorsal and ventral (D/V) compartments via wingless (wg) which activates the Notch signalling pathway and leads to elevated Notch levels on both sides of the compartment boundary and functions to control eye growth (Dominguez and de Celis, 1998; Cavodeassi *et al.*, 1999; Irvine and Rauskolb, 2001). The initiation of the D/V patterning mechanism also functions to set up a midline organiser and 'firing centre' from which the morphogenetic furrow (MF) forms, leading to the formation of photoreceptor cells (ommatidia) of the eye. The morphogenetic furrow is a coordinated wave of cell-shape regulation and patterning which moves from posterior to anterior across the eye field, initiated by a pulse of *Hedgehog* expression at the posterior margin, and which also requires the presence of decapentaplegic (dpp), and wg at the lateral margins which inhibits ectopic furrow initiation (Dominguez and Hafen, 1997; Borod and Heberlein, 1998; Baker, 2001). The ommatidia are placed in a precisely spaced array as the furrow moves with normal cell proliferation both anterior and posterior to the furrow dependent on the Egfr/Ras pathway that is activated by wg expression (Baker and Yu, 2001; Yang and Baker, 2003). Prior to the D/V and A/P patterning mechanisms however, is the expression of a network of eye specification genes at the posterior margin of the eye disc where differentiation will later initiate (Quiring *et al.*, 1994; Czerny *et al.*, 1999). This complex regulatory network consists of what are known as the eye specification genes, *eyeless (ey), twin of eyeless (toy), sine oculis (so), eyes absent (eya), dachshund (dac), eye gone (eyg)* and *optix* (see (Gehring and Ikeo, 1999; Wawersik and Maas, 2000; Kumar and Moses, 2001). These genes do not function in a simple linear hierarchy, but rather as a complex interwoven regulatory network (Fig 5.1); loss of function of any of these seven genes in the eye primordium results in a reduction or

deletion of the adult eye, whereas ectopic expression of any of these genes (with the exception of *so)* in the eye/antennal imaginal disc results in ectopic retinal development (Halder *et al.*, 1995; Shen and Mardon, 1997; Bonini and Fortini, 1999; Czerny *et al.*, 1999; Seimiya and Gehring, 2000). *Toy, ey* and *eyg* are expressed first, while expresson of *eya*, *so* and *dac* begins immediately prior to morphogenetic furrow formation, although these genes also have an earlier role independent of *toy* and *ey* in the embryonic development of the whole visual system. *Ey*, *eya*, *so* and *eyg* are responsible for growth of the early eye disc, *ey, eya, so, eyg* and *dac* are required for initiation of the morphogenetic furrow, *eya* and *so* for photoreceptor differentiation, and *ey* for rhodopsin gene expression during pupal eye development (see (Treisman, 1999)). What genes act upstream of the network of this eye specification cascade is unclear, although it has been theorised to be a combinatorial code of HOX genes (see (Treisman, 1999)).

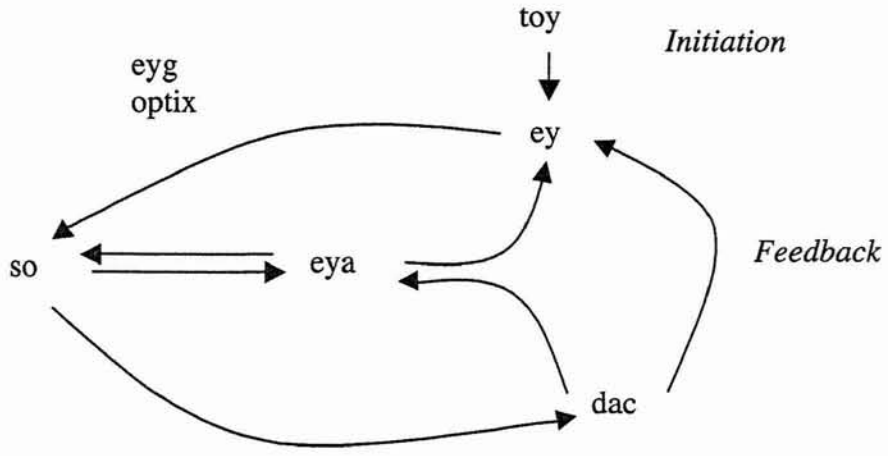


Fig 5.1: Representation of the complex regulatory network of the eye specification genes. Twin of eyeless (toy) initiates the cascade, with the remaining gene products involved in feedback regulation. (Adapted from Nordstrom, 2003).

The two pairs of Pax6-related genes, *eyeless/twin of eyeless* and *eyegone/twin of eyegone*, exist as paralogous pairs on *Drosophila* chromosomes 4 and 3 respectively (Flybase). *Eyeless* and its paralog *twin of eyeless* both encode canonical Pax6 products, and twin of eyeless functions upstream of eyeless by directly regulating a specific enhancer element, activating eyeless expression, which in turn activates expression of *so* and *eya* (Halder *et al.*, 1998; Czerny *et al.*, 1999; Hauck *et al.*, 1999; Niimi *et al.*, 1999). Twin of eyeless is more similar to vertebrate Pax6 proteins than eyeless with

regard to overall sequence conservation and DNA-binding function, (suggesting that *eyeless* is insect-specific), but the two genes share a similar expression pattern in the developing visual system (Czerny *et al.*, 1999). *Eyegone,* like its paralog *twin of eyegone,* encodes a Pax6-like protein with a truncated paired domain, lacking most of the N-terminus (Jun *et al.*, 1998; van Heyningen and Williamson, 2002; Jang *et al.*, 2003). The function of Toe is unclear although it has recently been reported that while Eyg is essential for eye growth, it seems dispensable for eye specification, thus there may be functional redundancy between Eyg and Toe with respect to the latter (Jang *et al.*, 2003; Dominguez *et al.*, 2004). In addition, it has been shown that Notch signalling in the eye acts through Eyg and, while overexpression of Eyg can fully rescue *Notch*$^{-}$ mutants, Toe overexpression results only in a partial rescue (Dominguez *et al.*, 2004). Eyg binds to a consensus site through the C-terminal (RED) subdomain of the paired domain which corresponds to the consensus binding site for a mammalian Pax6 isoform PAX6(5a) (Epstein *et al.*, 1994; Jun *et al.*, 1998). In human and mice, the PAX6(5a) isoform is generated by alternative splicing of the *PAX6* primary transcript, and overexpression of human PAX6(5a) induces strong overgrowth *in vivo* whereas the canonical PAX6 variant hardly effects growth showing that these two isoforms are required for different functions – PAX6 for differentiation of the eye, and PAX6(5a) for growth (Dominguez *et al.*, 2004). Like Eyg, PAX6(5a) recognises DNA exclusively through the RED region of its paired domain (Jun *et al.*, 1998). Therefore, the PAX6 and its fly counterparts, Ey and Toy, act principally to specify eye fate whereas the PAX6(5a) isoform and its functional conterpart in the fly Eyg (and perhaps also Toe) function in growth of the eye. PAX6(5a) can functional substitute for Eyg in *Drosophila*, as can PAX6 for ey (Dominguez *et al.*, 2004).

5.1.4 Cnidarian Pax Genes

The evolutionary origins of *Pax* genes are unclear; no *Pax* genes have yet been identified outside of the Metazoa, thus there are no obvious 'outgroups' for phylogenetic analyses. In addition, alternative splicing and multiple roles during development complicate the identification of ancestral functions. As the Cnidaria are the simplest animals at the tissue level of organisation, considerable attention has

focussed on defining their Pax gene complement and expression patterns in the hope that this might lead to insights into ancestral functions and *Pax* gene evolution.

The first *Pax* genes identified in Cnidarians were PaxA and PaxB from the sea nettle *Chrysaora quinquecirrha* (a scyphozoan) and *Hydra littoralis* (a hydrozoan; (Sun *et al.*, 1997). Phylogenetic analyses and preliminary DNA binding assays demonstrated that three of the cnidarian paired domains (sea nettle *PaxA* and *-B* and Hydra *PaxB*) bound to *Pax5/6* sites and implied that both the cnidarian *PaxA* and *PaxB* genes were more closely related to the *Pax2/5/8* and *Pax6* classes rather than to *Pax1/9* and *Pax3/7*. Thus, it was proposed that modern *Pax2/5/8* and *Pax6* genes evolved from an ancestral gene similar to cnidarian *PaxB* (Sun *et al.*, 1997). To date, PaxA and/or PaxB genes have been identified in a broad range of medusazoan cnidarians, including the hydrozoans *Podocoryne* (*PaxB*) (Groger *et al.*, 2000), *Polyorchis penicillatus* (*PaxB*) (Nordstrom, 2003) the scyphozoans *Aurelia aurita* (*PaxA*) (Nordstrom, 2003), *Chrysaora* (*PaxB*) and *Cladonema* (*PaxA/B*) (Sun *et al.*, 1997; Sun *et al.*, 2001) and the cubozoan *Chiropsalmus sp* (*PaxB*) (Nordstrom, 2003).

Our laboratory has identified four Pax genes from the anthozoan cnidarian *Acropora millepora*, *PaxAam (am = Acropora millepora), PaxBam, PaxCam* and *PaxDam*, with *PaxCam* and *PaxDam* apparently unique to *Acropora*. One of the two initial genes discovered, *PaxAam* is clearly homologous to the Hydra and sea nettle *PaxA* genes, whereas the other, *PaxCam* appears to be relatively distantly related to the cnidarian *PaxB* genes (Catmull *et al.*, 1998). The PaxAam and PaxCam paired domains most resemble those of the vertebrate Pax2/5/8 class, although the homeodomain of PaxCam is somewhat more closely related to the Pax6 type than are the cnidarian PaxB homeodomains, suggesting that PaxC rather than PaxB might correspond to a Pax-6 precursor. This hypothesis is supported by the fact that the homeodomain of Hydra PaxB has only 55% identity with mammalian Pax6 (compared to 70% identity of PaxCam with Pax6) and PaxB proteins typically contain an octapeptide motif, whereas true Pax6 proteins do not. The fact that the PaxAam and PaxCam proteins share several distinctive substitutions in their paired domains led to the suggestion of a common origin via a duplication event (Catmull *et al.*, 1998). Subsequently two more *Pax* genes were cloned from *Acropora*; *PaxBam* is orthologous to *PaxB* isolated in other Cnidarians, whereas *PaxDam* (like *PaxCam*) appears to be unique to *Acropora* (Miller,

2000). *PaxDam* clearly belongs to the *Pax3/7* class, while *PaxBam* appears to be an ancestral *Pax2/5/8* gene although it is shown in this work that the PaxBam protein can bind to Pax6 binding sites (Miller, 2000). Of the four *Acropora Pax* genes, three (*PaxBam, -Cam* and *-Dam*) encode complete homeodomains and *PaxBam* also contains an octapeptide motif (typical of the *Pax2/5/8* class). The PaxAam protein contains neither homeodomain or octapeptide, suggesting that *PaxAam* may be related to *Drosophila pox neuro* (Miller, 1999; Miller, 2000). The possession of common splice sites has frequently been used to support common ancestry, and the fact that several splice sites are shared between the Cnidarian and triploblastic metazoan *Pax* genes supports the monophyly of the *Pax* gene family (Miller, 2000).

5.1.5 Eyes in Cnidaria

Despite the absence of a central nervous system with which to process images, distinct eyes ranging in complexity from simple eye-spots to complex lens eyes are present in many representatives of three of the four cnidarian classes and photosensitivity is considered to be a general property of the phylum. The most sophisticated eyes are present in the most motile of cnidarians, the cubozoans (box jellyfish). The adult cubozoan eye may consist of up to 11 000 sensory cells and has an epidermal cornea, spherical lens and a retina with distinct sensory, pigmented nuclear layers (Brusca and Brusca, 1990). In contrast, the larvae of *Tripedalia cystophora* display photosensitivity in the complete absence of a nervous system; the ocelli which act as photoreceptors have no neural connections to any other cells, but each has a well-developed motor cilium (Nordstrom *et al.*, 2003b). Some cnidarians, including *Hydra,* lack any obvious photoreceptors but clearly react to light (Tardent and Frei, 1969) and generally cnidarian larvae respond to light (Svane and Dolmer, 1995). Members of the cnidarian class Anthozoa (of which *Acropora* is a member) lack the motile medusa stage that is characteristic of the other cnidarian classes and also lack eyes, however they do display photosensitive behaviour. For example, coral polyps are extended at night and retracted during the day, and coral larvae display a variety of phototactic behaviours (reviewed in (Harrison and Wallace, 1990). As previously mentioned *Acropora* remains the only cnidarian species in which PaxC and PaxD genes have been identified despite considerable efforts by several laboratories (eg (Sun *et al.*, 1997; Groger *et al.*, 2000;

Sun *et al.*, 2001; Nordstrom, 2003)) suggesting that either these genes have been lost in other lineages or evolved in the anthozoan lineage after the Anthozoa/Medusozoa split. One problem with the hypothesis that PaxC represents a Pax6 precursor is that this implies that a PaxC gene should be involved in specifying jellyfish eyes, although to date there is no evidence that this is the case.

5.1.6 Statement of Goals

To better understand the relationship of the cnidarian genes with the Pax6 and Pax2/5/8 classes, we studied the DNA-binding specificity of PaxCam and PaxBam Paired domains *in vivo* and *in vitro*. This work was done in collaboration with Dr Serge Plaza and Prof Walter Gehring of Biozentrum at the University of Basel, Switzerland. The results reported in this chapter were a necessary extension of my honours project, and have been published (Plaza *et al.*, 2003). In order to put the new data in context, some of my honours work has been included in this chapter. The expression of *PaxBam* in *Drosophila* was performed during my PhD enrolment, but the corresponding *PaxCam* work was conducted during my honours year (sections 4.2.1 and 4.2.5), as were the in vitro DNA binding assays (described in sections 4.2.2 – 4.2.4). The yeast one-hybrid experiments (section 4.2.6) were conducted during my PhD enrolment. The *Drosophila* work and aspects of the yeast one-hybrid analyses were performed by Dr Serge Plaza.

5.2 Results

The results indicate that both PaxBam and PaxCam proteins bind to EY targets *in vivo* and *in vitro*, and thus indicate that the relationship between these cnidarian proteins and the Pax6 and Pax2/5/8 classes of bilateral animals is unlikely to be simple. The literature suggests that Pax gene loss may be an ongoing process within the Cnidaria. We suggest that in non-anthozoan cnidarians, PaxB may have acquired the roles of PaxC or alternatively, that within the Anthozoa PaxC may have arisen from a PaxB-like ancestor to fulfil more restricted roles.

5.2.1 Expression of PaxCam in Drosophila imaginal discs results in a dominant negative-like phenotype

In order to test the ability of *PaxCam* to generate ectopic eyes, several independent *Drosophila* lines carrying the *PaxCam* cDNA transgene under the control of the yeast GAL4-UAS regulatory sequence were generated via P-element-mediated germ line transformation (Brand and Perrimon, 1993). The UAS sequence upstream of the cDNA results in the transcription of the cDNA when yeast GAL4 is expressed in *Drosophila* cells. A UAS-*Drosophila eyeless* cDNA line was used as a positive control for the effects of induced EY expression. The UAS-*PaxCam* fly lines were crossed with the dpp-GAL4 driver, permitting expression of the *PaxCam* cDNA in the wing, leg and eye/antenna discs. In the case of the UAS-*ey* lines, GAL4-directed expression of EY in any of these imaginal disc types resulted in ectopic eyes in the corresponding adult structure. GAL-4 driven expression of *PaxCam* however, did not only result in ectopic eye formation, but also appeared to interfere with the development of adult structures arising from the disc in which it was expressed (wing, leg and eye). The legs were malformed and truncated, the wings did not develop correctly and the eyes were reduced in size (Fig 5.3a).

5.2.2 Chimeric PaxCam constructs encoding the EY C-terminal domain result in eye formation in Drosophila

Comparing the PaxCam C-terminal sequence with a range of PAX6 proteins from other animals suggested that the *Acropora* protein may lack the C-terminal transactivation domain and that the phenotype seen with ectopic expression of full length PaxCam might be due to a dominant negative effect with respect to EY targets in vivo. To test this hypothesis, a series of UAS-constructs were generated via splicing by overlap extension (Clackson *et al.*, 1991) in which the C-terminal region of EY was transposed onto regions of PaxCam encoding the DNA-binding domains and vice versa, generating four constructs in total (see Fig 5.2). Crosses of fly lines carrying these UAS-domain swap constructs with the dpp-GAL4 driver line showed that UAS lines in which the EY C-terminal region was present were capable of triggering ectopic eye development in the wing disc. This effect was seen to be independent of the presence of the

homeodomain, i.e. with constructs in which the PaxCam PD and the EY C-terminal domain were present. The reciprocal constructs - that is EY constructs featuring the PaxCam C-terminal region - were incapable of inducing ectopic eyes. Eyes induced in response to expression of PaxCam/EY fusion proteins were always significantly smaller than those resulting from EY misexpression, but were otherwise morphologically normal. In scanning electron micrographs regular ommatidia and inter-ommatidial bristles are clearly visible (Fig 5.3f).

5.2.3 Ectopic sine oculis expression is induced by PaxCam/EY constructs

The *Drosophila* experiments described above implied that the PaxCam protein bound to EY targets *in vivo* and this was tested directly by examination of the effect of PaxCam on *sine oculis (so)* expression *in vivo*. *sine oculis* is one of the best characterised direct targets of EY; the EY protein activates expression of *so* by binding to an eye-specific enhancer called *so10* in the *so* gene (Niimi *et al.*, 1999). Figure 5.4 shows LacZ staining patterns in wing discs from fly lines in which expression of the enhancer trap *so-LacZ* (Cheyette *et al.*, 1994) was driven by various EY, PaxCam or PaxCam/EY chimeric constructs. Both of the constructs in which the EY C-terminal domain was present (Fig 5.4c-d) activated *so* expression, albeit at significantly lower levels that did EY (Fig 5.4b). Constructs consisting of the EY PD and the PaxCam HD and C-terminal region did not drive significant levels of so expression (data not shown).

5.2.4 In vitro binding properties of Pax-Cam and Pax-Bam PDs

The 128bp fragment in the *so* enhancer found to be a direct target of EY (*so10*) by band shift assays and DNA footprinting experiments (Niimi *et al.*, 1999; Punzo *et al.*, 2002) was used *in vitro* to investigate the DNA-binding specificity of PaxBam and PaxCam PDs. In addition, the ability of the PaxBam and PaxCam PDs to bind to consensus Pax2/5/8 and Pax6 binding sites were also examined. In each case, the recombinant PDs bound specifically and with high apparent affinity to the labelled oligonucleotides, and although the method does not permit quantitation of the interaction, no major differences were apparent between the PDs in affinity for the oligonucleotides (Fig 5.5).

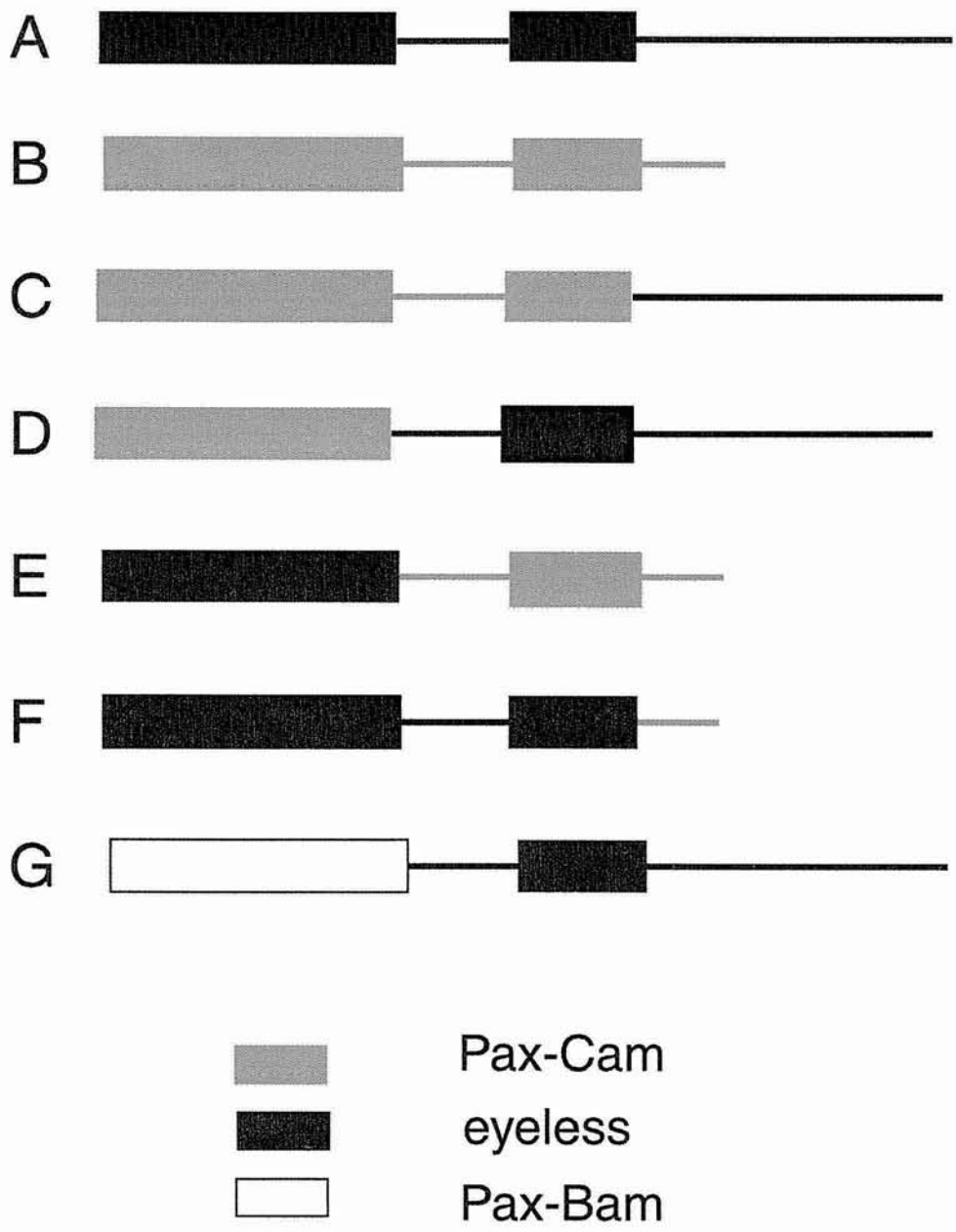


Fig. 5.2: Schematic representation of constructs expressed in Drosophila imaginal discs. (A) The EY protein, (B) Pax-Cam, (C-F) EY / Pax-Cam chimeric proteins, (G) Pax-Bam / EY chimera.

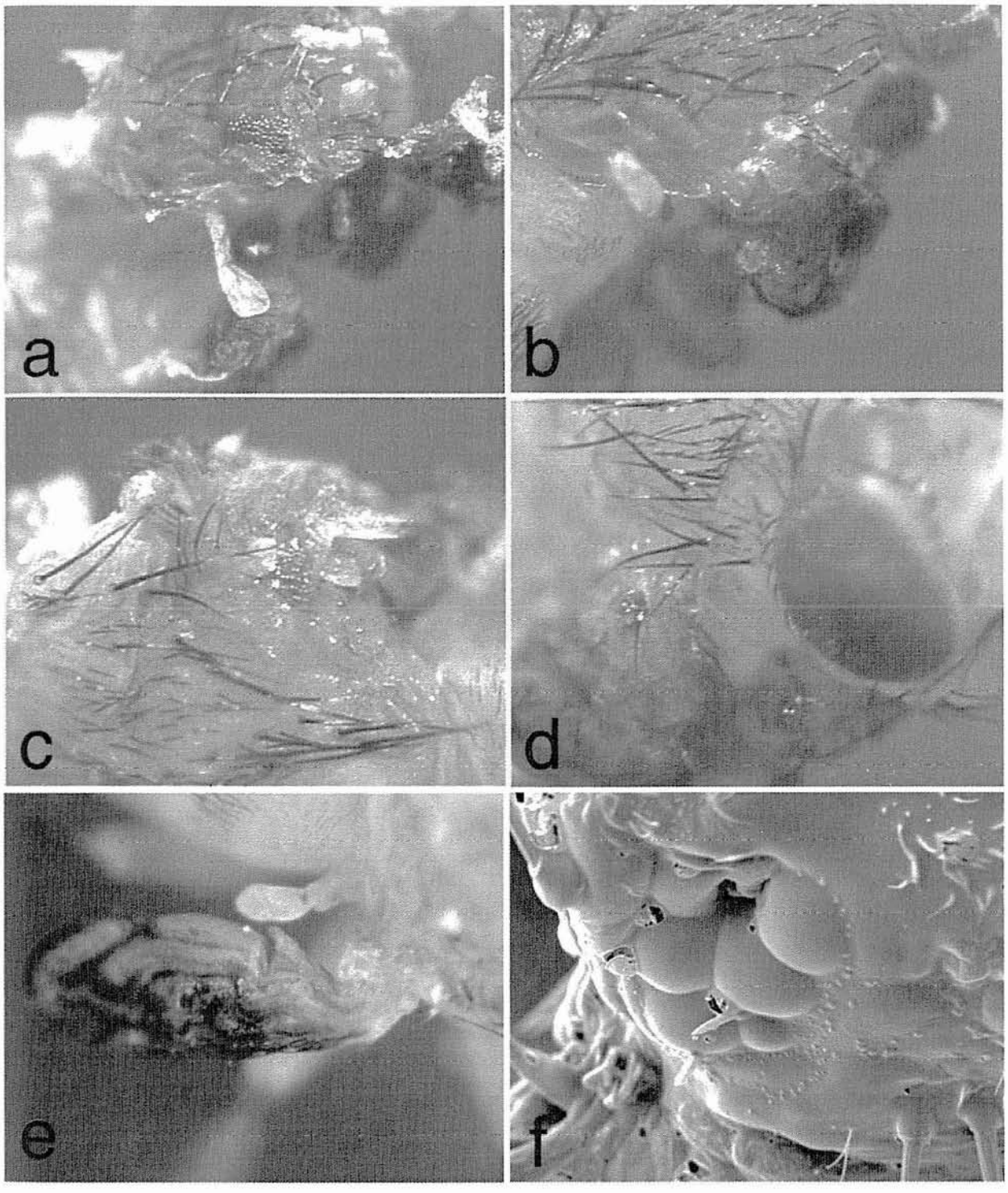


Fig. 5.3: Phenotypes resulting from expression of *Pax-Cam* constructs in *Drosophila* wing imaginal discs. The constructs shown in Fig. 5.2 were expressed under GAL4-UAS control in wing discs. Expression of the eyeless cDNA leads to the formation of ectopic eye tissue seen as the red-pigmented structure (a), whereas expression of the Pax-Cam cDNA does not result in eye formation (b). (c-d) Expression of chimeric constructs (c and d in Fig. x) encoding the C-terminal region of EY result in eyes that are smaller than those induced by EY. (e) Expression of Pax-Cam in the wing disc causes severe abnormalities. (f) At the SEM level, the morphology of the eyes induced by the Pax-Cam / EY constructs can be seen to include regular ommatidia and inter-ommatidial bristles.

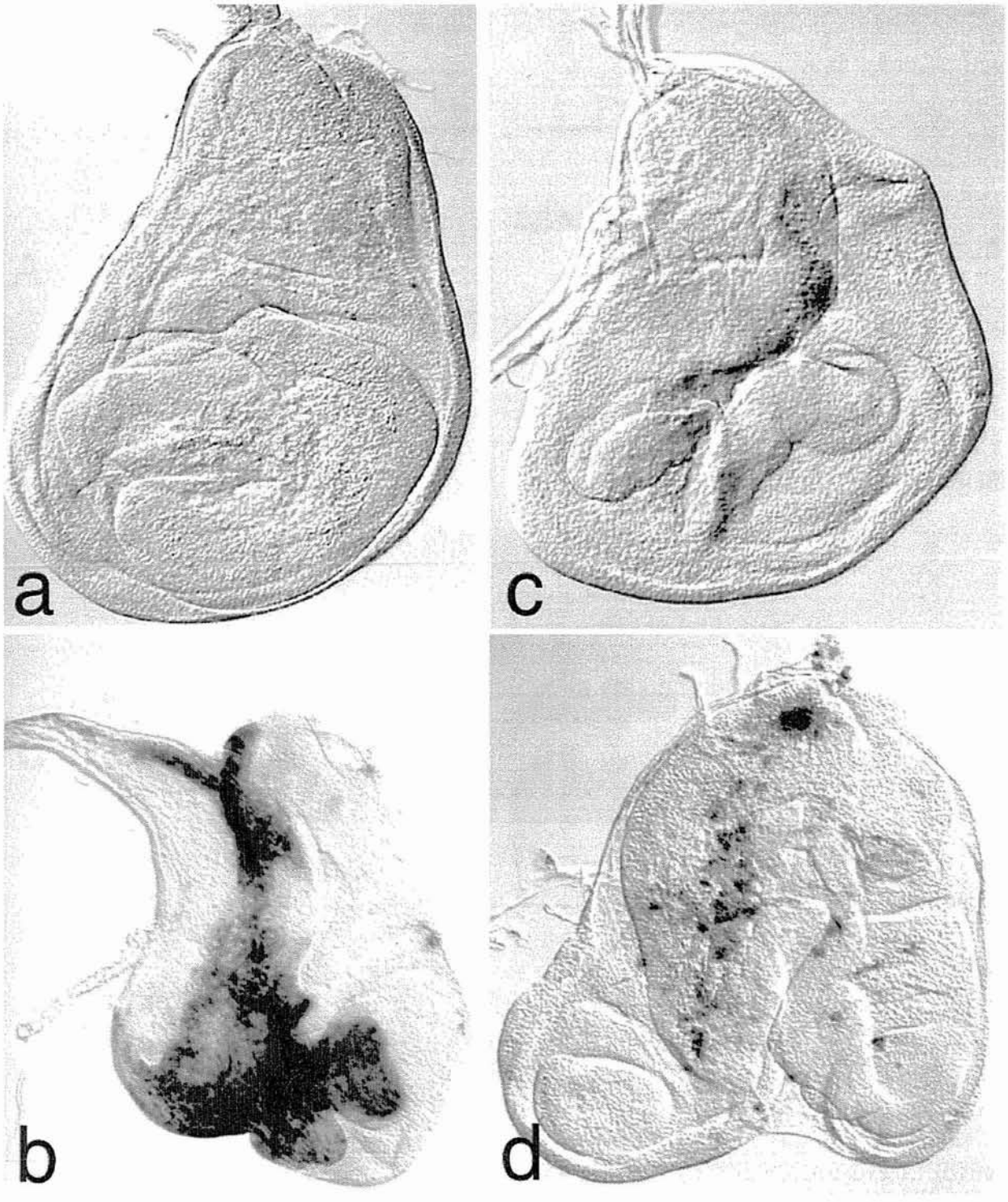


Fig. 5.4: b-galactosidase expression in sine oculis (so)-lacZ line driven by *dpp*GAL4-UAS *Pax-Cam* constructs. Wing discs are shown in which the constructs shown in Fig. 5.2 were expressed. (a) Pax-Cam does not induce significant so-lacZ expression, (b) positive control by misexpression of EY. (c-d) Both domain swap constructs encoding the C-terminus of EY (i.e. constructs c and d in Fig. 1) induced so-lacZ expression, whereas the reciprocalconstructs did not (data not shown).

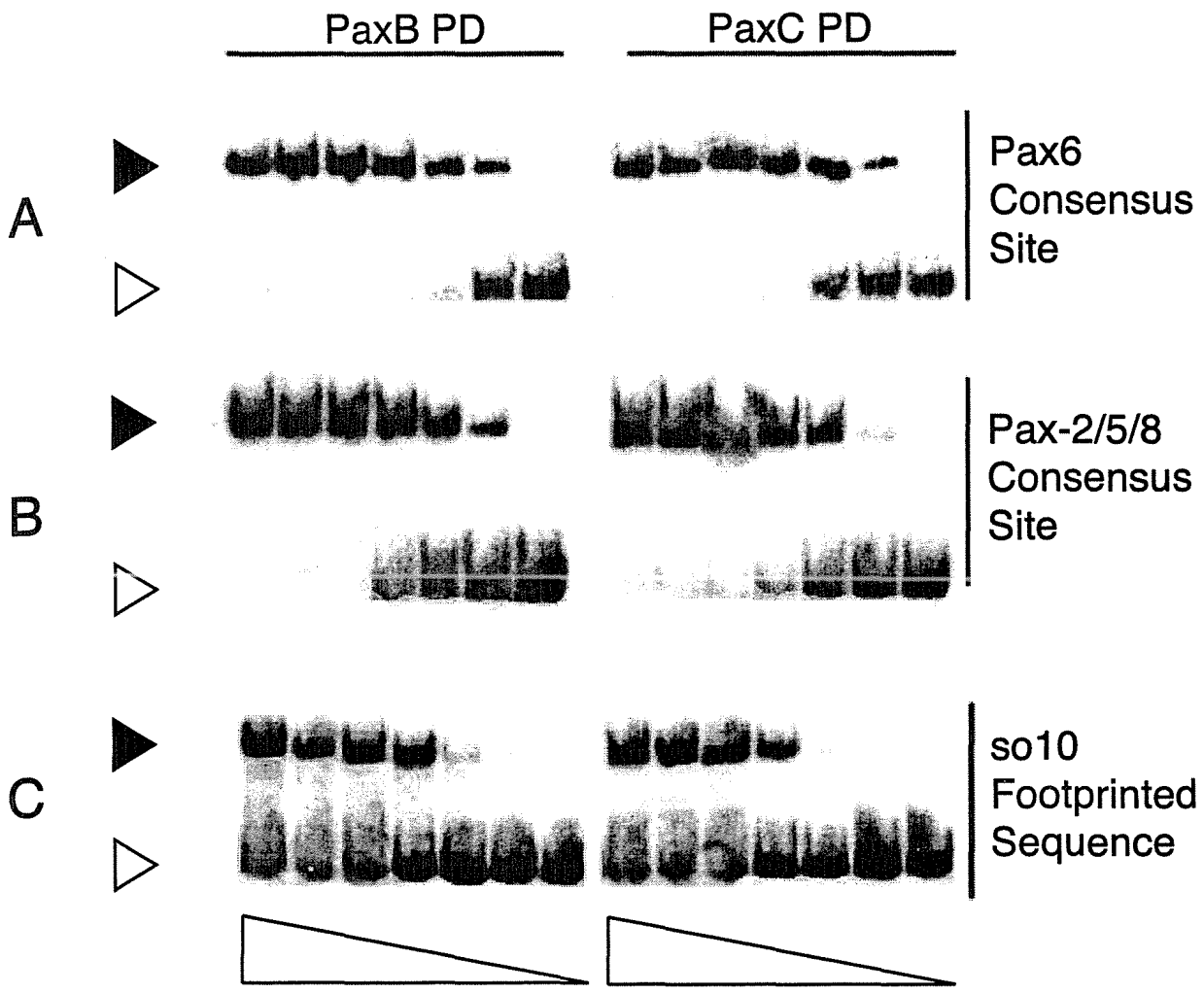


Fig. 5.5: DNA-binding assays using the Pax-Cam and Pax-Bam Paired domains. The binding of recombinant Pax-Cam (right panel) and Pax-Bam (left panel) PDs to (a) a consensus Pax-6 binding site (5'-AGGTTCACGCTTCAGTTAGTCAGC-3'), (b) a consensus Pax-2/5/8 binding site (5'-CTAGTCATGCATGAGTGTTCCAGC-3'), and (c) a known EY target (the so10 oligonucleotide 5'-GCAAACAAGTAAAAATTAATTCCCCCTCACTGGGCACAACT-3') was determined by electrophoretic mobility shift assays. In each case, the concentration of oligo-nucleotide was held constant, and the concentration of the PD varied from 2.6 x 10-5 M to 7.9 x 10-10 M (corresponding to 8-fold dilution between lanes); the right hand lane in each case is a negative control in which no protein was added. Open and closed triangles on the left of the figure indicate the positions corresponding to free and bound probe respectively.

5.2.5 Expression of a PaxBam/eyeless chimera in Drosophila

Because the PaxBam PD bound the same range of target sites *in vitro* as did that of PaxCam, the morphogenetic properties of a PaxBam/EY domain swap construct were examined in *Drosophila* imaginal discs. To avoid potentially complicating protein-protein interactions mediated by the Pax-Bam octapeptide motif, and to enable direct comparison with the Pax-Cam/EY phenotypes, the construct encoded only the N-terminal region and PD of PaxBam, and the region of EY C-terminal of the PD (Fig 5.2G). When expressed either in the leg or wing discs, the PaxBam/EY construct was able to induce ectopic eyes, albeit with lower efficiency than was the corresponding PaxCam/EY construct (Fig 5.6 – note the extremely weak wing disc phenotype). Thus the *in vivo* data are consistent with the *in vitro* DNA-binding experiments, indicating that the PDs of both PaxCam and PaxBam bind EY targets and, in the presence of the EY C-terminal domain, can initiate compound eye morphogenesis in the fly.

5.2.6 PaxCam binds to the sine oculis eye-specific enhancer region in a yeast one-hybrid system

The yeast one-hybrid system described by Mastick *et al* (1995) was used to examine the interaction of PaxCam constructs with a defined EY target. This method was employed as the *Drosophila* transgenic experiments implied that PaxCam/EY chimeras were capable of activating expression of EY targets, such as *sine oculis, in vivo*. To better understand this interaction, we examined the ability of the corresponding chimeras to bind to the *sine oculis* eye-specific enhancer region in a yeast one-hybrid assay. Yeast activator constructs expressing the PaxCam/EY chimeras corresponding to those used in *Drosophila* were cloned into the *BamHI* site of the activator plasmid pBM258T (Mastick *et al.*, 1995) and used in conjunction with the *so10-HIS3* reporter in pHR307a previously described (Niimi *et al.*, 1999). The insert fragment was generated by PCR using the domain swap constructs as templates with *BglII* restriction sites introduced at either end of the primers to facilitate cloning into pBM258T. Prior to expression, the PCR fragment was cloned first into pGEM-T to enable verification of the sequence,

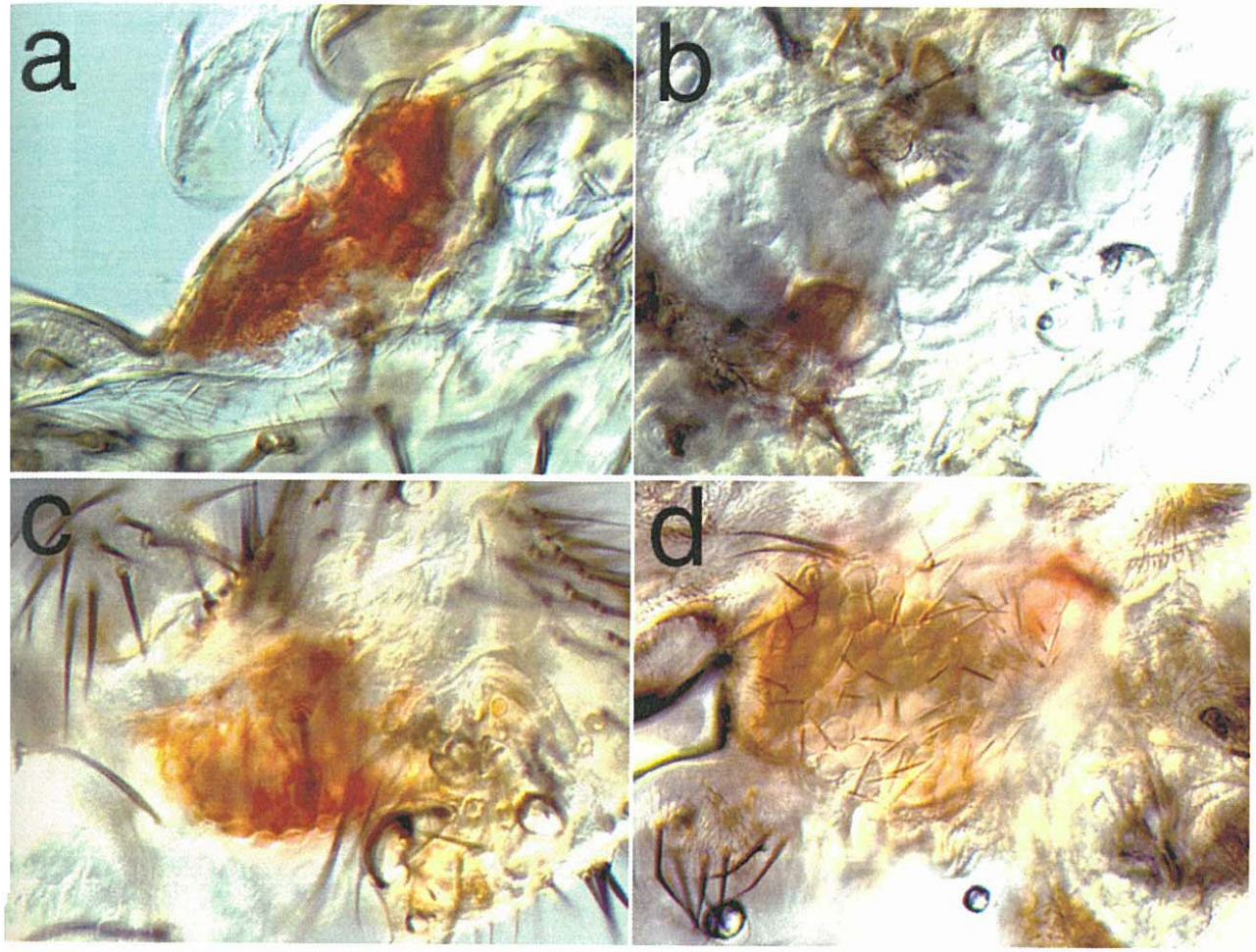


Fig. 5.6: Comparison of phenotypes resulting from expression of Pax-Bam / EY and Pax-Cam / EY constructs in leg and wing imaginal discs. The Pax-Bam / EY construct (shown schematically in Fig. 5.2G) was capable of inducing eye morphogenesis in the leg disc only (a); typical results of expressing this construct in the wing disc are shown as (b). The corresponding Pax-Cam / EY construct (shown schematically in Fig. 5.2D) displayed stronger morphogenetic properties in both the leg (c) and wing (d) discs.

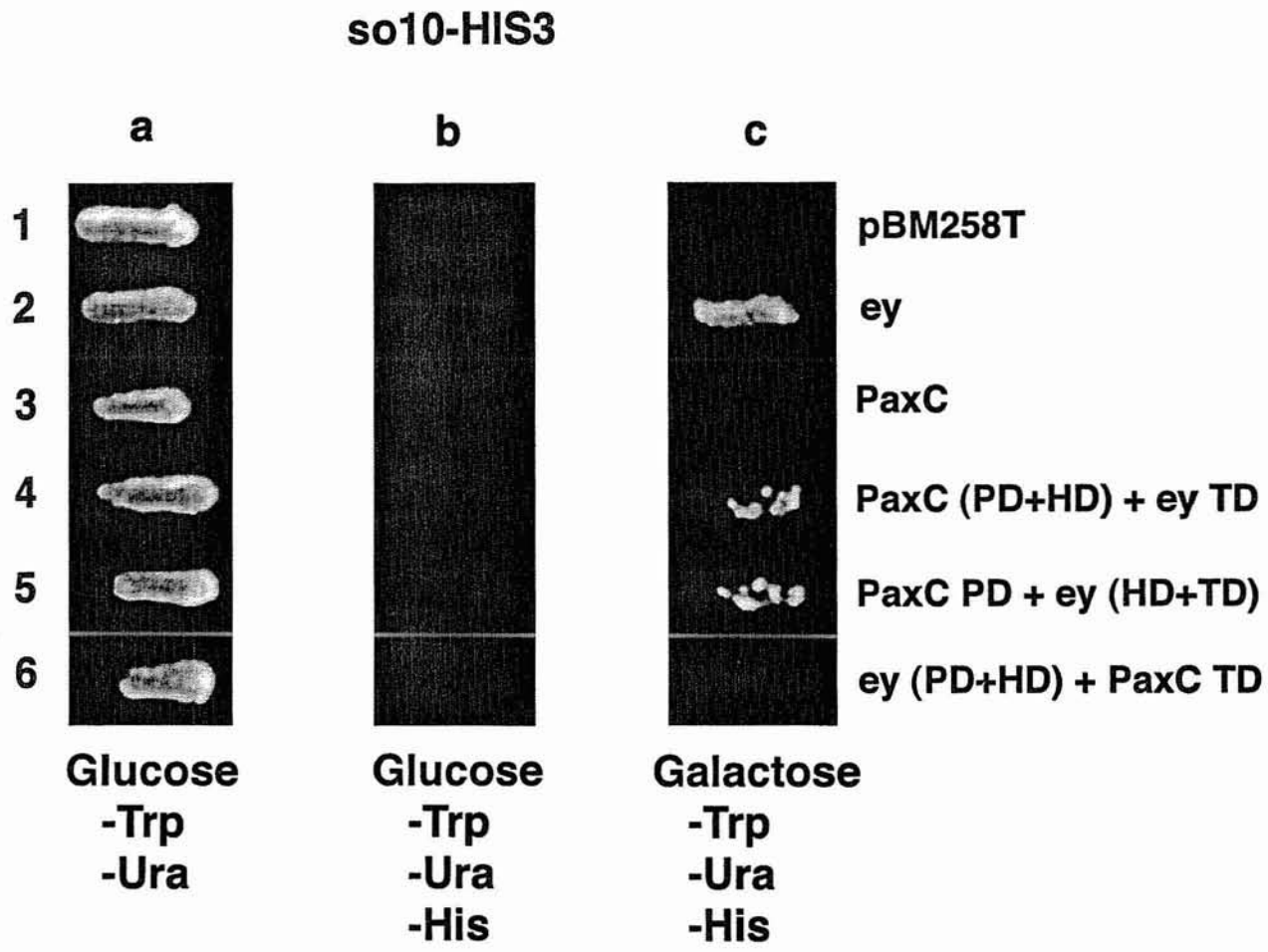


Fig. 5.7: Pax-Cam binds to the sine oculis eye-specific enhancer so10 in a yeast one-hybrid system. Expression of the proteins indicated on the right of the figure was driven by a galactose-inducible promoter that is strongly repressed by glucose. Each panel (a, b and c) represents the same colonies plated onto different media, the composition of which is indicated at the bottom of the figure. Panel a: growth control experiment. The presence of histidine (His) allows all of the colonies to grow; the medium lacks tryptophan (Trp) and uracil (Ura) to select for maintenance of the so10-His3 reporter and pBM258T activator plasmids respectively. Panel b: negative control experiment; no growth is observed on medium lacking histidine in the presence of glucose since activator proteins are not produced. Panel c: In the presence of galactose, yeast colonies are able to grow on media lacking histidine if the protein produced binds to the so10 target and activates transcription of the HIS3 reporter gene. Lane 1: pBM258T empty vector as negative control; lane 2: pBM258T Eyeless expressing vector as positive control. Lanes 3 to 6: various PaxC/Ey chimeras cloned into pBM258T as indicated to the right of the figure. Lanes 4, 5 and 6 correspond to constructs c, d and f respectively in Fig 1. Note that no growth is observed with all constructs on the empty His plasmid lacking the so10 sequence (not shown and Niimi *et al.*, 1999).

then excised using *BglII* and ligated into the *BamH1* site of pBM258T (n.b. *BamH1* and *BglII* have compatible ends). Cloning into pBM258T was complicated by the size of the vector (~14kb); once achieved, insert orientation was verified by PCR. Yeast cells (strain YM4271) transformed with the activator plasmid were selected on media lacking uracil, and these were then transformed with the appropriate reporter construct. The basis of this system is that double transformants are only capable of growth on galactose media (galactose drives expression of the activator construct) lacking histidine if the chimeric protein constructs are capable of binding the *so10* region and activating expression of the reporter gene (*HIS3*). Yeast transformation and expression were performed by Dr Serge Plaza in the Biozentrum (University of Basel). Results of these experiments are shown as Figure 5.7.

5.3 Discussion

Our initial goal was to test the hypothesis that *PaxCam* represents a precursor of the Pax6 class by examining the morphogenetic properties of the PaxCam protein expressed in *Drosophila* imaginal discs. Although PaxCam was unable to initiate eye morphogenesis in imaginal discs, this effect appears to result from the lack of a transactivation domain and hence an inability to activate transcription in *Drosophila*. Expression of chimeric PaxCam proteins containing the C-terminal region of EY in imaginal discs resulted in eyes that were morphologically normal, but smaller than those induced by EY misexpression. The PaxCam/EY chimeras conferred a number of phenotypic characteristics normally associated with Pax6 proteins, including activation of a *so-lacZ* construct *in vivo*. Similarly PaxCam/EY chimeras were able to activate transcription of a HIS3 reporter by binding to the *so10* fragment in a yeast one-hybrid system. The C-terminal region of the PaxCam protein is much shorter (only 81 AA residues C-terminal of the HD) than that in EY and PAX6 proteins in general (the EY C-terminal region is 387 AA residues; that of PAX6 is 152), and contains no obvious transcription activation domain (Czerny and Busslinger, 1995; Tang *et al.*, 1998). It is therefore likely that PaxCam functions primarily as a transcriptional repressor, as does the mammalian PAX6-related protein PAX4 (Smith *et al.*, 1999).

Whilst the experiments in which PaxCam/EY chimeras were expressed in *Drosophila* support the hypothesis that PaxCam is a precursor of the Pax6 class, the *in vitro* DNA-binding properties of a second *Acropora* Pax protein, PaxBam led us to question this assumption. The PaxCam and PaxBam PDs bound the same range of sequences *in vitro*, including a known EY target site – a footprinted sequence in the *so* eye-specific enhancer. The PDs of the cnidarian Pax proteins appear to have relatively low DNA-binding specificites; published data for the *Cladonema* and *Chrysaora* PaxB proteins (Sun *et al.*, 2001) are broadly consistent with the *Acropora* data. Although the specificity of the *Acropora* PaxAam PD has not yet been determined, binding to EY targets is not a universal property of cnidarian PDs, as the *Acropora* PaxDam PD does not bind to these same sites *in vitro* (Nordstrom *et al.*, 2003a). The DNA-binding behaviour of the PaxBam PD *in vitro* led us to examine the morphogenetic properties of a PaxBam/EY chimera in *Drosophila*. The PaxBam/EY chimera was able to induce ectopic eyes in the leg disc and to a limited extent in wing discs, but with lower efficiency than the corresponding PaxCam/EY construct. Phylogenetic analyses clearly show that cnidarian PaxB and PaxC both belong to the Pax supergroup which also includes the Pax6 and Pax2/5/8 classes (Balczarek *et al.*, 1997; Catmull *et al.*, 1998; Groger *et al.*, 2000; Miller, 2000). Although we have previously suggested otherwise, (Catmull *et al.*, 1998; Miller, 2000), the results presented here suggest that there is unlikely to be a simple correspondence between the cnidarian PaxB and PaxC genes and the Pax2/5/8 and Pax6 classes in higher animals.

Clearly, Pax proteins are an ancient class of transcription factors (Hoshiyama *et al.*, 1998) that are absent from fungi and plants (Galliot *et al.*, 1999) but diversified very early in animal evolution (Miller, 2000). Other groups investigating the evolution of the pax family have suggested an evolutionary scheme, based solely on comparisons between the HDs and PDs of Pax proteins from a variety of Metazoa. Their scheme, which accommodates both vertebrates and arthopods, is based on the assumption that the failure to isolate PaxC and PaxD sequences from other cnidarians is due only to incomplete libraries or the fact that primers used are not compatible with the target sequence. The scheme allows only for a minimum of domain acquisitions and losses and proposes that PaxA/PoxN and PaxC/Eyg form two subgroups separate from the other four groups of Pax genes proposed by Noll, 1993; (Pax1/9, Pax3/7, Pax4/6 and Pax2/5/8), and that the PaxA/PoxN and PaxC/Eyg groups have since been lost in

vertebrate evolution but retained in certain cnidarian and arthropod lineages (Nordstrom, 2003). The PaxA/PoxN genes were proposed to be founded by a PD-containing gene which lacked a HD. Upon acquisition of the HD the PaxC/Eyg-like group of genes were formed (note that Eyg has lost part of its PD). The acquisition of an octapeptide and subsequent duplication then gave rise to the PaxB and PaxD genes. Loss of the octapeptide or part of the HD of the PaxB-like gene gave rise to Eye/Toe/Pax6 genes, and the Pax2/5/8/sparking genes respectively, while the PaxD-like gene gave rise to the Pax1/9/Poxmeso group and the Pax3/7/paired/GooseberryN group (Nordstrom, 2003) (See Fig 5.8). If it is assumed that undersampling or other issues are not the cause of the lack of PaxC and PaxD genes found in other cnidarians other than *Acropora*, it can be supposed that there has been ongoing loss of Pax genes throughout the Cnidaria. Using this scenario, the PaxB and PaxC types are likely to post-date the Cnidaria/Bilateria split; PaxC either originated within the common cnidarian ancestor or within the Anthozoa after the Anthozoa/Medusozoa (Hydrozoa, Scyphozoa and Cubozoa) split. In either case, the specificity associated with true Pax6 genes presumably arose after the Cnidaria/bilateral Metazoa split.

From our data we cannot exclude the possibility that the specificity of PaxB and PaxC proteins is influenced by regions other than the PD; it is quite possible that the activity and specificity of the PD is influenced by the overall protein environment. Therefore the fact that the *in vitro* experiments described here were carried out with PDs alone, and the *in vivo* work was carried out on *Acroproa* PDs in the context of the EY protein, is one major limitation in interpreting the results. Although their DNA binding characteristics are similar, the two proteins are likely to have distinct roles – PaxCam presumably functions primarily as a repressor of transcription, whereas sequence comparisons imply that PaxB proteins may be able to act either as transactivators (via the C-terminal domain) or repressors (via the octapeptide) depending on context. The presence of complete HDs in PaxB proteins distinguishes these from the Pax2/5/8 class proper; presumably the full HD enables PaxB proteins to also act via their HD to regulate specific gene expression. In addition to common roles throughout the Cnidaria, the functional flexibility of PaxB proteins may have enabled them in some cnidarians to effectively fulfil the roles of PaxCam in *Acropora*. Either the roles of PaxCam may have been subsumed by PaxB in medusozoans (the non-anthozoan cnidarians), or in contrast to the Nordstrom evolutionary scheme, PaxCam may have derived from a

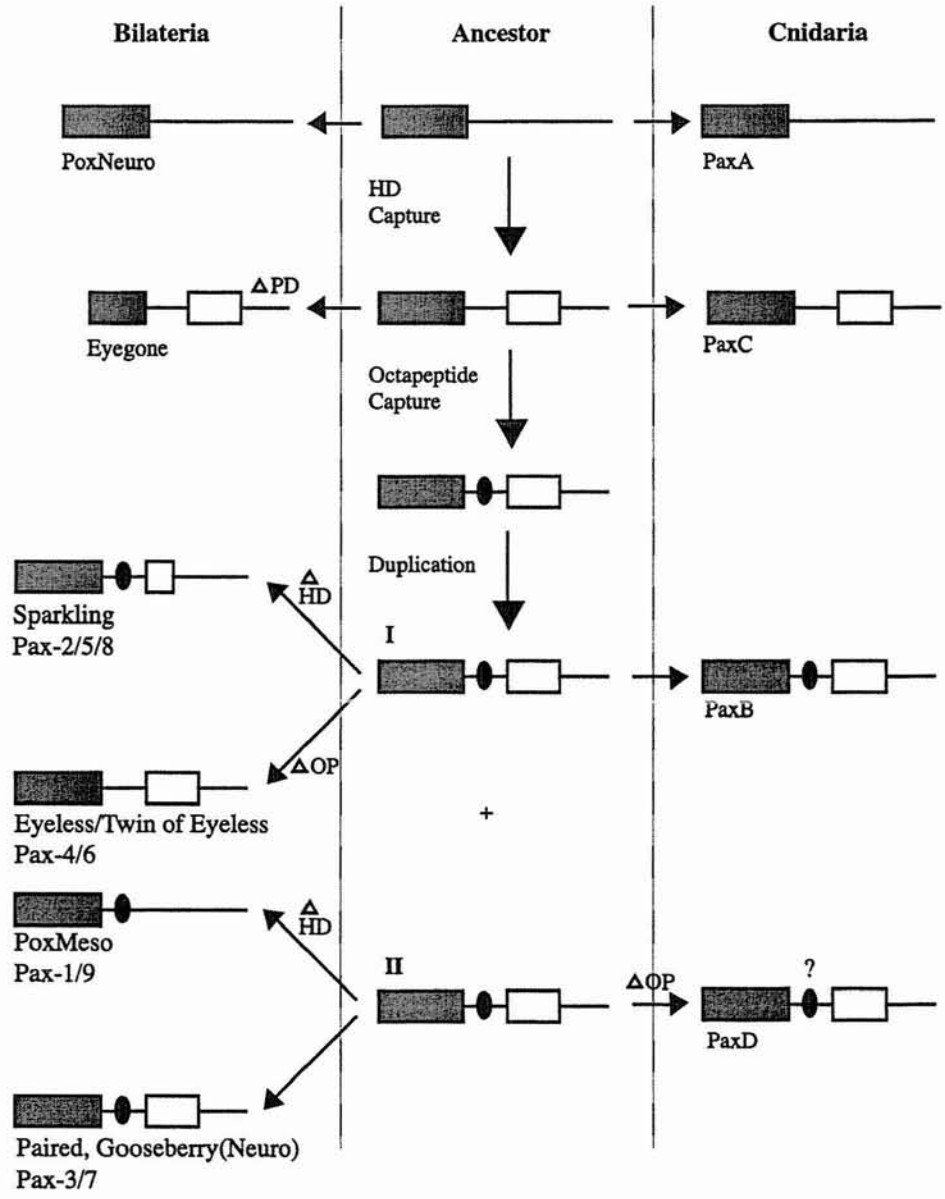


Figure 5.8: Suggested scenario of Pax evolution. The left column indicates genes in the Bilateria, the middle column the last common ancestor of bilaterians and cnidarians, and the right column the cnidarian Pax proteins. Paired domains are shown as grey shaded boxes, homeodomains by open boxes and octapeptides by black circles and the appropriate protein names shown. Complete or partial domain loss is shown by the delta symbol and the domain abbreviation. The question mark on cnidarian PaxD indicates the ambiguity of the octapeptide. Initially, an ancestral PD-containing gene founded the PaxA and PoxNeuro genes. Capture of the HD led to the development of the PaxC and loss of part of the paired domain led to *Drosophila* Eyg. After capture of the octapeptide and duplication in the ancestral organism, the cnidarian PaxB and PaxD genes were created. Further duplication events in bilaterians gave rise to Sparkling/Pax-2/5/8 and Eyeless/Pax-4/6 families and PoxMeso/Pax-1/9 and Paired, Gooseberry (Neuro)/Pax-3/7 families. Thus following this scenario, the last common ancestor of cnidarians and bilaterians had at least four Pax genes (adapted from Nordstrom *et al,* 2003).

PaxB-like precursor to fulfil more specific roles. One prediction of the above model is that we might expect the expression patterns of PaxB genes in non-anthozoan cindarians to correspond to the sum of the patterns of *PaxCam* and *PaxBam* in *Acropora*. Unfortunately, expression data are available only for *Acropora PaxCam* and *Podocoryne PaxBPc*. *PaxCam* has a very specific pattern of expression, in a subset of presumed neurons in the planula larva (Miller, 2000). At the same stage in *Podocoryne*, *PaxBPc* is expressed throughout the entire ectoderm (Groger *et al.*, 2000). In *Podocoryne* polyps, *PaxBPc* expression is restricted to ectodermal cells that are either interstitial cells or neurons (or both) and in medusae, the (endodermal) pattern of *PaxBPc* expression is again consistent with a role in nerve cell differentiation (Groger *et al.*, 2000). However, testing these ideas requires expression data for more Pax genes in a variety of cnidarians.